CROSS-DIALECTAL VARIABILITY IN PROPOSITIONAL ANAPHORA: A QUANTITATIVE AND PRAGMATIC STUDY OF NULL OBJECTS IN MEXICAN AND PENINSULAR SPANISH

DISSEPTION

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Graduate School of The Ohio State University

By

Maria Asela Reig, PhD

The Ohio State University

2008

Dissertation Committee:
Professor Scott A. Schwenter, Adviser
Professor Rebeka campos-Astorkiza
Professor Terrell Morgan
Professor Rena Torres-Cacoullos

Approved by

Adviser Graduate Program in Spanish and Portuguese
ABSTRACT

In this dissertation, I analyze the linguistic constraints that condition the variation in Spanish between the null pronoun and the clitic *lo* referring to a proposition. Previous literature on Spanish has analyzed null objects referring to first order entities, mostly in varieties in contact with other languages. This dissertation contributes to the literature on anaphora in Spanish by establishing and analyzing the existence of propositional null objects in two monolingual dialects, Mexican and Peninsular Spanish.

A variationist approach was used to discover the significant constraints on the variation of the null pronoun and the overt clitic *lo* in Mexican and Peninsular Spanish. Following the generalizations from the previous literature on two separate areas of study, anaphora resolution and null objects (Chapter 2), several internal factor groups were included in the coding scheme. In Chapter 3, I provide an explicit statement of the envelope of variation and I specify the coding scheme employed.

Chapter 4 offers the results of the multivariate analyses of Mexican and Peninsular Spanish. These results show that some of the linguistic constraints conditioning the variation are shared by both dialects (presence of a dative pronoun, type
of antecedent, sentence type), suggesting that the null pronoun has the same grammatical role in both dialects. Nevertheless, divergences in the conditioning of the null pronoun also emerge from the analysis and the overall rate of null objects in Mexican Spanish is significantly greater than in Peninsular Spanish.

Chapter 5 offers a discussion of the results and their theoretical implications for studies of anaphora. The accessibility of the referent hypothesis, operationalized with the factor groups referential distance, turn, and number of mentions, finds no support in the data, suggesting an important difference in the behavior of propositional versus noun-phrase anaphora. Instead, I provide qualitative evidence indicating that a crucial constraint on the variation is the degree of completeness of the referent: a complete proposition favors the overt clitic lo whereas null objects more likely refer to incomplete propositions. Finally, Chapter 6 explores the hypothesis of a change in progress and provides concluding remarks and suggestions for further research.
ACKNOWLEDGMENTS

This dissertation would not have been possible without the contribution of many people. I wish to express my deepest gratitude to my adviser, Scott Schwenter. His contributions to this dissertation are countless, from the first general idea to the final result. I am thankful for his generosity, dedication and sincere trust. His questions and comments were always challenging and encouraging, and I am glad I could benefit from many stimulating conversations with him (en mesas altas y bajas...). Outside academia, Araceli and you made us feel “at home” so many miles away. Muchísimas gracias.

I want to extend my thanks to the members of my committee. I was fortunate enough to have access to Terrell Morgan’s broad knowledge on dialectology and pronoun variation in Spanish. Talking about linguistic (and non linguistic) issues with Terrell’s great smile and warmth was always a pleasure.

My gratitude to Rena Torres-Cacoullos, for generously dedicating time and effort to this unknown student. Her expertise and insightful comments were a great contribution to this dissertation. I also thank Rebeca Campos-Astorkiza for generously accepting our invitation and for her dedication. Her remarks and questions helped improve previous versions of this dissertation.
The time spent in graduate school would not have been so pleasant without my friends and fellow graduate students, Edith Hernandez, Manuel Delicado, Carolina Castillo, Marissa Vargas, Magdalena Mejia and Gilberto Velazquez. I am grateful for their friendship. Un enorme placer haber compartido estos años.

A mi familia: mis padres, Asela Alamillo y Rafael Reig, y mis hermanos Juan, Elena y Jorge. Gracias por soportar la distancia todos estos años y por mostrarme siempre vuestro incondicional apoyo y confianza.

Por último, a mi compañero desde el principio de esta aventura, Carlos. Gracias por hacer que este viaje haya merecido tanto la pena. Me alegro mucho de haber recorrido esta primera etapa juntos. Que el viaje prosiga por mucho tiempo.
VITA

2000 ...............B.A. In Hispanic Philology Universidad Autónoma de Madrid
2006 .............. PhD Spanish Language and Linguistics Universidad Autónoma de Madrid

PUBLICATIONS


FIELDS OF STUDY

Major Field: Spanish and Portuguese

Studies in:

Pragmatics

Variationist linguistics
TABLE OF CONTENTS

ABSTRACT............................................................................................................................ii

ACKNOWLEDGMENTS..............................................................................................................iv

VITA........................................................................................................................................vi

LIST OF TABLES.......................................................................................................................x

LIST OF FIGURES....................................................................................................................xiii

CONVENTIONS........................................................................................................................xiv

CHAPTER 1: INTRODUCTION .................................................................................................1
  1.1. Phenomenon and goals.................................................................................................1
  1.2. The linguistic variants...............................................................................................6
    1.2.1. Propositional lo.................................................................................................6
    1.2.2. The null element ...............................................................................................9
  1.3. Anaphora and discourse referents. Terminology......................................................19
  1.4. Dissertation overview..............................................................................................20

CHAPTER 2: ANAPHORA RESOLUTION AND NULL OBJECTS ..................................22
  2.1. Studies on anaphora resolution ...............................................................................23
    2.1.1. The accessibility of NP antecedents...............................................................23
    2.1.2. The accessibility of clausally introduced antecedents.....................................25
  2.2. Null object with propositional antecedents in other languages............................34
  2.3. Null objects in Spanish ............................................................................................41
  2.4. Prototypical direct objects, the animacy scale and propositions.........................46
  2.5. Prototypical objects and the transitivity of the clause............................................51
    2.5.1. Complement taking predicates and transitivity..............................................57
5.2.1. Contexts of categorical null object ............................................................ 183
5.2.2. Incomplete content of the proposition .................................................... 189
5.2.3. Complete propositions ........................................................................... 202
5.3. Transitivity .................................................................................................. 206
   5.3.1. Transitivity and Completeness of the proposition .................................. 210
5.4. Conclusions ............................................................................................... 213

CHAPTER 6: CONCLUSION .................................................................................... 215

   6.1. The variationist analysis and the comparison of Mexican and Peninsular
   Spanish ........................................................................................................... 218
   6.2. Completeness of the proposition ............................................................... 222
   6.3. Accessibility of the referent and the null/lo variation ............................. 224
   6.4. Variation and linguistic change: socio-demographic information .......... 227
   6.5. Future research ...................................................................................... 233

REFERENCES ................................................................................................. 235
LIST OF TABLES

Table 2.1. Parameters of transitivity from Hopper and Thompson (1980) .......................... 53
Table 2.2. Individuation in Hopper and Thompson (1980) ........................................... 54
Table 4.1. Overall frequency of “lo” and null DO ......................................................... 109
Table 4.2. Factors contributing to the choice of the null DO in Peninsular Spanish ...... 112
Table 4.3. Internal factors contributing to the choice of the null DO in Mexican Spanish ........................................................................................................................................ 113
Table 4.4. Social factors contributing to the choice of the null DO in Mexican Spanish ........................................................................................................................................ 114
Table 4.5. Internal factors contributing to the choice of the null DO in Mexican and Peninsular Spanish ........................................................................................................................................ 116
Table 4.6. Distribution of null pronoun and clitic lo according to absence or presence of dative pronoun in Mexican Spanish .............................................................................................................. 118
Table 4.7. Distribution of null pronoun and clitic lo according to absence or presence of dative pronoun in Peninsular Spanish .............................................................................................................. 118
Table 4.8. Distribution of null pronoun and clitic lo by Cooccurrence of dative pronoun in Mexican Spanish ........................................................................................................................................ 119
Table 4.9. Distribution of null pronoun and clitic lo by Cooccurrence of dative pronoun in Peninsular Spanish

Table 4.10. Varbrul weights for the null pronoun by Cooccurrence of dative pronoun:
Mexico and Spain

Table 4.11. Rates of null and overt objects with monotransitive verbs: Mexico and Spain

Table 4.12. Rates of null pronoun and clitic lo with ditransitive verbs without dative pronoun: Mexico and Spain

Table 4.13. Distribution of null pronoun and overt lo according to type of antecedent in Mexico

Table 4.14. Distribution of null pronoun and overt lo according to type of antecedent in Spain

Table 4.15. Varbrul analysis of the Group Type of antecedent in Mexican Spanish

Table 4.16. Varbrul analysis of the Group Type of antecedent in Peninsular Spanish

Table 4.17. Distribution of null pronoun and clitic lo in the "no (lo) sé" construction by type of antecedent in Peninsular Spanish

Table 4.18. Distribution of null pronoun and clitic lo in the "no (lo) sé" construction by type of antecedent in Mexican Spanish

Table 4.19. Results of the Varbrul analysis of the group Polarity in Mexican Spanish

Table 4.20. Results of the Varbrul analysis of the group Polarity in Peninsular Spanish

Table 4.21. Internal factors contributing to the choice of the null DO in Mexican Spanish. Analysis without external factors.
Table 4.22. Distribution of the null pronoun and the clitic lo in negated sentences in Peninsular Spanish........................................................................................................139

Table 4.23. Distribution of the null pronoun and the clitic lo in negated sentences in Mexican Spanish........................................................................................................139

Table 4.24. Distribution of null objects and the effect of “no sé” .......................141

Table 4.25. Overall frequency of lo and null DO in data without "no (lo) sé" ..........141

Table 4.26. Factors contributing to the choice of null pronoun in the data without "no (lo) sé" in Peninsular Spanish .................................................................142

Table 4.27. Factors contributing to the choice of null pronoun in the data without "no (lo) sé" in Mexican Spanish .................................................................143

Table 4.28. Factors contributing to the choice of the null DO in Peninsular Spanish, excluding negation .................................................................144

Table 4.29. Internal factors contributing to the choice of the null DO in Mexican Spanish, excluding negation .................................................................145

Table 4.30. Completeness of the proposition and lack of knowledge.................147

Table 4.31. The expression "Quién (lo) sabe" in CREA........................................153

Table 4.32. Results of the Varbrul analysis for the group manner adverbial (ma) in Mexican Spanish for null object.................................................................155

Table 4.33. Crosstabulation of manner adverbial (ma) with Type of antecedent in Mexican Spanish.................................................................156

Table 4.34. Results of the factor group Adverb in Mexican Spanish.....................159
Table 4.35. Results of the Varbrul analysis of the group Sentence type in Mexican Spanish ................................................................. 161
Table 4.36. Distribution of null pronoun and overt lo by non declarative sentence types in Mexico ............................................................................................................................ 162
Table 4.37. Sentence type in the Multivariate analysis of Mexican Spanish data without “no (lo) sé” ......................................................... 163
Table 4.38. Sentence type in the Multivariate analysis of Peninsular Spanish data without “no (lo) sé” ......................................................... 163
Table 5.1. Distribution of null pronoun and overt clitic according to referential distance in Peninsular Spanish ................................................................. 169
Table 5.2. Distribution of null pronoun and overt clitic according to referential distance in Mexican Spanish ................................................................. 169
Table 5.3. Occurrences of DO pronouns and demonstratives in Madrid and Mexico Habla Culta ................................................................................................................................. 173
Table 5.4. Distribution of null pronoun and clitic lo by number of times the proposition was referred to in Mexican Spanish ................................................................................................................................. 174
Table 5.5. Distribution of null pronoun and clitic lo by number of times the proposition was referred to in Peninsular Spanish ................................................................................................................................. 175
Table 5.6. Distribution of null pronoun and clitic lo by Turn in Peninsular Spanish ................................................................................................................................. 176
Table 5.7. Distribution of null pronoun and clitic lo by Turn in Mexican Spanish ................................................................................................................................. 177
Table 5.8. Null objects by type of antecedent and dative pronoun (DP) in Mexican Spanish ................................................................................................................................. 204
Table 5.9. Null objects by manner adverbial (ma) and Type of antecedent in Mexican Spanish

Table 5.10. Individuation (Hopper and Thompson 1980)

Table 6.1. Internal factors contributing to the choice of the null DO in Mexican and Peninsular Spanish

Table 6.2. Null object and propositional lo in Mexican data from CORDE

Table 6.3. Null object and propositional lo in 17th century Spanish data
LIST OF FIGURES

Figure 4.1. Distribution of verbs in Peninsular and Mexican data.................................110

Figure 5.1. Percentages of null object by Referential distance (number of sentences). . 171

Figure 5.2. Completeness of the proposition .................................................................183
CONVENTIONS

ACC: accusative
BP: Brazilian Portuguese
DO: direct object
DAT: dative
EP: European Portuguese
FEM: feminine
MASC: masculine
NCA: null complement anaphora
NP: noun phrase
SG: singular
PL: plural
VP: verb phrase

Corpora
Mex Cult: Corpus de Habla Culta de Mexico
Mex Pop: Corpus de Habla Popular de Mexico
Mad Cult: Corpus de Habla Culta de Madrid

Corec: Corpus de Referencia de la Lengua Española Contemporánea

ALCORE: Alicante - Corpus Oral del Español

Monterrey: Habla de Monterrey

CREA: Corpus de Referencia del Español Actual

CORDE: Corpus Diacrónico del Español
CHAPTER 1

INTRODUCTION

1.1. Phenomenon and goals

Studies from a wide range of different linguistic perspectives have contributed to the understanding of the cross-linguistic phenomenon of anaphora, yet these approaches are usually kept separate from each other and the complex phenomenon of anaphoric reference still offers many unexplored areas of research.

From formal semantics and pragmatics, dynamic approaches to meaning have been interested in the interpretation of anaphora, focusing on the identification of absolute semantic constraints on what can be anaphoric to what, and on describing the nature of the meaning of the anaphors (Kamp and Reyle 1993; Beaver 2004; Roberts 2005). These studies, however, do not attempt to explain the discourse or pragmatic constraints on the distribution of anaphoric expressions, an issue that is central to a different trend in the literature, devoted to anaphora resolution. Studies such as Ariel (1990) or Gundel et al. (1993) concentrate on the relation between various anaphoric expressions and the discourse or cognitive status of their referents. Alas, most of these
proposals are based on English, pertain mostly to anaphoric elements referring to first order entities, and, for the most part, observe expressions in subject position.

As in English, referring expressions in Spanish have been the object of studies from diverse interests and perspectives: anaphoric expressions in subject position have been of interest to syntacticians because of the pro-drop nature of the Spanish language, but the distribution of noun phrases, personal pronouns and null subjects has also been thoroughly analyzed from a pragmatic and variationist point of view in different dialects (Cameron 1992, 1995; Otheguy 2004, Silva-Corvalán 1982, 2003; Hurtado 2005). Some phenomena related to pronouns in object position in Spanish, such as leísmo, laísmo, loísmo, clitic climbing and clitic doubling have been carefully studied in the last few years (Myhill 1988; Fernández-Ordóñez 1993; Colantoni 2002; Gutierrez-Rexach 2002), but little has been said about the distribution of available anaphoric expressions in object position and the differences between the constraints that condition the distribution of object and subject elements.

Spanish is traditionally described as a non object-drop language, one which encodes anaphoric direct objects by means of overt pronominal elements, such as the clitics lo(s)/la(s), or the demonstrative esto/e/a, eso/e/a and aquel/lla/llo. Null objects (a.k.a. “object drop” or “null direct object pronominalization”) in Spanish are traditionally conceived as restricted to non-referential and non-countable referents, i.e. mass nouns and bare plurals (Campos 1986; Clements 1994, 2006). Null objects in Spanish have lately awakened the interest of several scholars who have shown that definite null objects are a frequent feature in several contact Spanish varieties (e.g.,
Paraguay [Choi 1998, 2000] and Quito [Yépez 1986; Suñer and Yépez 1988], Basque Spanish [Landa 1995]), but the case of null direct objects with propositional antecedents, which refer anaphorically to the clausal complements of cognition and communication verbs, has not received the same attention, and remains, until now, mostly uninvestigated.

Anaphoric direct objects referring to a proposition are variably coded by either the clitic lo or a null pronoun in Spanish. Such variation, illustrated in (1.1) and (1.2), will be the focus of this dissertation.

(1.1) -Bueno, mi trabajo consiste en arreglar za... en... reparación de zapatos; ya le Ø dije anteriormente. Este... ps hago mucho trabajo en cuestión de... no únicamente de zapato: yo he trabajado en talleres, he trabajado en fábricas de zapato. (Mex Cult)
Well, my job consists of repairing... repairing shoes; I told you Ø before. I do a lot of work in... no only shoe; I have worked in workshops, I have worked in shoe plants’

(1.2) -Yo creo que a las cuatro, yo es que a las cuatro me es imposible venir. Ya se lo dije a él; que me era imposible. (Mad Cult)
I think that at four; for me, at four it is impossible to come. I already told him LO, that it was impossible for me

Although this phenomenon has not been studied until now (except in Reig and Schwenter 2007), and has only been noticed in some dialects of Latin American Spanish, variation between the two pronominal forms seems to exist in every dialect of Spanish. The distribution and productivity of the null pronoun, however, are at first glance different among dialects, and the similarities and differences between dialects have never been analyzed in the literature.

In this dissertation, I will compare the distribution of the propositional clitic lo and the null pronoun in two monolingual dialects of Spanish: Mexican and Peninsular Spanish. As noted in Schwenter (2006), while the form le dije [him- tell:PRET-3SG] is
perceived as more natural in Mexican Spanish (as well as other dialects, such as Peruvian and Puerto Rican), this form is only marginally accepted in Peninsular Spanish, where the alternative *se lo dije* [him- it- tell:PRET-3SG], with the explicit direct object pronoun *lo*, is preferred. This is true not only for ditransitive constructions, as Kany (1945) and others suggest, but also for monotransitive sentences like (1.3), where the null object is commonly accepted in Mexican Spanish but not usually found in Peninsular speech:

(1.3)  -Ahí es mi casa; y tu casa también, cuando vayas
  -No, pues yo no Ø sabía. Pero ahora ya Ø sé (Mex Pop)
  - My house is there; and your house too, when you go
  - No, *I didn't know Ø*. But now *I know Ø'*

Although the null pronoun seems to be more frequent in Mexican Spanish than in Peninsular Spanish, it is by no means the only form used in this dialect: it alternates with the propositional clitic *lo*. Similarly, although the clitic *lo* is the preferred form in Peninsular Spanish, the null object with a propositional antecedent is also found in natural speech:

(1.4)  A: Yo es que, como no sabía dónde estaba... tenía una idea, ¿no?, pero...
  B: Sí, pero no Ø sabías exactamente (Mad Cult)
  A: ‘Since I didn't know where it was... I had an idea, but...
  B: yes, but you didn't *know Ø* exactly.

The factors that condition the distribution of the two forms remain unexplained. The purpose of this dissertation is to explain, from a variationist perspective, the syntactic variation between the direct object pronoun *lo* and the null pronoun with propositional antecedents in two monolingual dialects of Spanish, Peninsular and Mexican Spanish.
I start from the assumption that there is no free variation in the distribution of the two pronominal forms, and the goal of my research is to explain the pattern of distribution of these variants, that is to say, provide an answer to the following questions:

- What is the envelope of variation of this phenomenon? i.e., in what linguistic contexts is this variation possible?
- What linguistic factors play a role in speaker choice between variants? Under what conditions may a direct object with a propositional antecedent be (preferably) null or overt in Mexican and Peninsular Spanish?
- Are the factors that condition this distribution similar or comparable in both dialects?
- Are the factors that constrain this variation similar or comparable to the internal factors proven to affect the distribution of anaphors with noun phrase antecedents?
- And, only partially, is the distribution of these two variants constrained by social factors? If this is the case, does this variation indicate any kind of linguistic change, or stable variation?

In this introductory chapter I will define the two variants under study: I will first clarify what clitic lo is considered in this dissertation (1.2.1.) and I will provide a brief characterization of the null object under consideration here in comparison with the phenomena of inherent objects, VP ellipsis, and Null Complement Anaphora (NCA) (1.2.2). This chapter ends with the dissertation overview (1.3)
1.2. The linguistic variants

Before moving on, I need to establish exactly which type of *lo* and which type of null object I am interested in for this dissertation. In the following sections I will characterize the pronoun *lo* and null object whose variation is studied here.

1.2.1. Propositional *lo*

One of the variants studied in this dissertation is the pronoun *lo* found in (1.5):

(1.5) -Yo creo que a las cuatro, yo es que a las cuatro me es imposible venir. Ya se lo dije a él; que me era imposible. (Mad Cult)

I think that at four; for me, at four it is impossible to come. *I already told him LO,*

that it was impossible for me

In this section, I will differentiate this object clitic *lo* from other so called "neuter *lo*" forms that won't be considered in this dissertation, and I will establish the terminology that will be used in the following chapters, namely "propositional *lo*”.

The notion of a "neuter *lo*” is commonly found in traditional Spanish grammars (Bello 1945: 119; Gili y Gaya 1964:237) to refer to the *lo* occurring in a range of functions and syntactic positions, exemplified in (1.6): DO pronoun (a), pronoun in cleft construction (b), preceding an adjective or adjective phrase (c) and as a predicate attribute (d):

(1.6) a. Van a vender su casa pero yo no *lo sabía.*

They are selling their house, but I didn't *know LO*  

b. *Lo* que yo quiero es paz en el mundo.

What I want is peace in the world
c. Lo triste de esta situación es…
    The sad thing about this situation is...

d. María es muy simpática pero Juan no lo es
    Maria es very nice but Juan isn't LO

As this shows, any *lo* different from the DO pronoun referring to a masculine NP has been traditionally referred to as "neuter *lo*" in the Spanish grammatical tradition, although this terminology has not been unanimously accepted.

Several scholars (Leonetti 1999; Bosque and Moreno 1990; Ojeda 1984; Otheguy 1978) consider that using the term "neuter" to characterize this pronoun (as well as the demonstratives *eso, esto, aquello*) is misleading and inappropriate, since there is no morphological neuter gender in Spanish and, therefore, there is no formal neuter agreement. According to these scholars, since neither nouns nor adjectives have neuter forms in Spanish, it is odd to propose a neuter pronoun or determiner. These "neuter pronouns", in fact, always show agreement with the masculine form of adjectives (Leonetti 1999):

\[(1.7)\]
\[a. \text{Lo que dijo fue muy acertado/*acertada} \]
\[\text{What he said was right-MASC/*right-FEM}\]
\[b. \text{Lo dijo y me pareció clarísimo/*clarísima} \]
\[\text{He said LO and it was very clear-MASC to me/*clear-FEM}\]

Among the scholars that coincide in rejecting that the feature that differentiates *lo* from other members of the paradigm is gender, the interesting issue is what feature captures the opposition between *lo* and the corresponding masculine and feminine pronouns. It should be noticed that this debate is mostly centered on the uses of *lo*
illustrated in sentences (1.6b) and (1.6c) above, and, therefore, on the opposition between 
lo and el/la, and not lo vs lo/la (masculine/feminine). In this regard, different proposals
have arose in which scholars try to identify a semantic feature that explains the
opposition lo vs el/la. Othegey (1978) argues that the semantic feature that differentiates
lo from el/la is discreteness: la and el mean discrete while lo conveys non-discrete, i.e.
unclear, diffuse and not well-delineated boundaries, which would explain the contrast
between Dame el/la bueno/a [Give me the good one (MASC/FEM)] and Dame lo bueno
[Give me what's good], or Dame el/la que te compré [Give me the one (MASC/FEM) I
bought you] and Dame lo que te compré [Give me what I bought you]. Ojeda (1984)
considers that the feature that best characterizes the forms traditionally defined as neuter
is [-count]: "neuter" pronouns are [-count] and, as such, need not provide a criterion of
individuation for the reference of the NPs they head. Evidence for this claim, according
to Ojeda, is the fact that these pronouns lack plurals (Ojeda 1984:172).

The uses of lo that have received more attention in the literature, represented in
sentences (1.6b), (1.6c), or (1.6d) above, are those uses of lo not considered in this
dissertation. The pronoun lo which is in variation with a null pronoun in object position,
and which will be the object of this dissertation, is the clitic pronoun exemplified in
(1.6a). Not much is said about this "neuter lo" in the literature, and the observations
devoted to its use are little more than "negative definitions", which describe its use by

---

1 These studies also debate whether this lo should be considered a pronoun (Bosque and Moreno; Garrido
1986) or an article  (Álvarez Martínez 1986; Plann 1980).

2 In fact, observations about the semantics of this non-masculine, non-feminine forms are also found in
grammarians who characterize the form lo as neuter: Gili y Gaya affirms that the "carácter colectivo y
de alusión indeterminada hace que usemos los pronombres neutros cuando no queremos determinar el
concepto a que nos referimos, bien por ignorancia, bien por deseo de no aclararlo demasiado, y a veces
por eufemismo" (Gili y Gaya 1964:178).
saying that this pronoun does not refer to nouns or NPs (Ojeda 1984). In this sense, Lenz (1925) and the Real Academia (1973) argue for the non-nominal nature of neuter anaphora and Bello (cited by Fernández Ramirez 1951) asserts that lo either reproduces a predicate or another neuter pronoun, or it refers to a fact or situation mentioned in the discourse by means of a sentence. More recently, Butt and Benjamin (2004:88) state that “lo as a neuter pronoun does not refer to any specific noun, but to an idea, clause or sentence, to some unspecified thing that has no gender, or to something not mentioned before.”

In this dissertation, I will avoid using the term "neuter lo", in agreement with the observation that such a terminology is inappropriate in light of the fact that there is no morphological neuter gender in Spanish. I will instead use the term "propositional lo", which allows me to distinguish the direct object pronoun lo studied in this dissertation (1.6a) from the so called "neuter lo" in different syntactic constructions, exemplified in (1.6b), (1.6c) or (1.6d).

1.2.2. The null element

I have mentioned, so far, some characteristics of the null element that alternates with the propositional lo: we are interested in a null object whose content is propositional and would be expressed, if it were explicitly stated in the discourse, by means of a sentence (as opposed to an NP). More needs to be said about this null object before I proceed with the analysis. In this section I wish to further characterize the null element under study and illustrate the differences between this null pronoun and other phenomena
previously described involving a phonologically zero object, such as non referential null objects or "inherent objects" (Haegeman 1987), VP ellipsis, and null complement anaphora (NCA). As part of this characterization, I am interested in distinguishing the null object under consideration in this dissertation (1.8) from other non-phonologically realized objects in Spanish that illustrate different phenomena (1.9, 1.10, 1.11, 1.12):

(1.8) Bueno sí, sí ha habido [accidentes], nada menos antier se ahogó un niño allá, allá, allá, ¿sí Ø sabías? (Monterrey)  
Well, yes, yes, there have been some [accidents], the day before yesterday a boy drowned there, there. Did you know Ø?

(1.9) Yo tengo un cuñado también ingeniero y, sin embargo, pues es un hombre bastante preparado, con bastantes inquietudes, un hombre que opina, un hombre que sabe Ø, pero, es un poco excepción a la regla. (Mad Cult)  
I have a brother in law who is also an engeneer and, nevertheless, he is a pretty well qualified man, with interests on the side, a man that has opinions, a man that knows Ø, but he is the exception to the rule

(1.10) - Marisa.  
- Dime Ø.  
- Y por fin el barco de Naxos, ¿a qué hora sale? (Corec)  
- Marissa  
- Tell Ø me  
- And, finally, the boat from Naxos, at what time does it leave?

(1.11) Luis gana mucho dinero y María también (from Brucart, ex. 56)  
Luis makes a lot of money, and so does Maria

(1.12) Luis fue al acto; María, en cambio, no pudo (from Brucart, ex. 88)  
Luis attended the event; Maria, on the other hand, could not do it.
1.2.2.1. Referential null object

Previous literature has differentiated two types of "null objects" ("pragmatically controlled zero anaphora"; "understood arguments" or "linguistically unrealized arguments"), that Fillmore (1986) named "indefinite null objects" and "definite null objects". Although much has been said about the inappropriateness of Fillmore's terminology (Groefsema 1995; Junker et al. 2005), the studies that followed maintain his intuition in distinguishing between what has later been termed referential and non-referential null objects (Cummins and Roberge 2005; Cornish 2005).

The concept of "non-referential null objects", or "indefinite null objects" in terms of Fillmore, also called "inherent objects" (Haegeman 1987) or "understood objects" (García and Portero 2002), aims to explain the lack of a phonologically overt direct object in cases where the content of the zero element is unknown or indifferent to the speaker, like in (1.13) and (1.14) (examples 3 and 4 from Fillmore 1986).

(1.13) We have already eaten
(1.14) I spent the afternoon baking

In these examples, no discourse referent is being introduced by the null anaphor that could be the object of subsequent predications. To decide whether a null object is "indefinite" in Fillmore's terms, the author proposes the following test: if the zero anaphor is a case of "indefinite null complements", it should not sound odd for a speaker to admit ignorance of the identity of the referent of the missing phrase (He is eating. I
wonder what he is eating), but this continuation of the discourse would sound odd when
the null element is, in Fillmore's distinction, a definite null complement (#They found
out. I wonder what they found out).

In cases like (1.13) and (1.14) the interpretation of the object is constrained by verb meaning, by means of some kind of selection restriction, and only some verb tenses, such as the generic present tense, the infinitive and imperative forms, allow this kind of non referential null object. It is a common observation that these objects do not contribute other semantic information besides the information that was already contributed by the verb itself, and this, according to Cummins and Roberge (2005), is why these constructions are described as focusing on the "action" of the verb. According to Cornish (2006), it is the event (token or type) denoted by the verbal predicate which is highlighted by the null complement realization, its participants being backgrounded thereby. The indeterminate or generic reading of the direct object leads the predicate to be interpreted as an activity, and not an accomplishment (García and Portero 2002; Cornish 2006).

Not everybody agrees with this analysis of sentences like (1.13) and (1.14): while for some authors the verb licenses a null object, which is syntactically present though phonetically null, other researchers argue for an intransitive use of a commonly transitive verb; Junker et al. even defend the view that verbs that allow uses like (1.13) and (1.14) (eat, bake, drink, etc.) are in fact intransitives marked for optional addition of an argument (Junker et al. 2005:9; Iten et al. 2005). The issue of what the best analysis of sentences like (1.13) and (1.14) is not the object of this dissertation. These sentences are only of interest here to illustrate that the null object observed in this dissertation is clearly
not a case of "inherent" or "non referential" null objects. Observe the differences between examples in (1.15) and (1.16) [which include (1.6) and (1.7) above] and (1.17) [which includes (1.5) above]: the description just outlined, although proposed to refer to first order entities, could possibly be applied to cases like (1.15) and possibly (1.16), but not to (1.17).

(1.15) a. Yo tengo un cuñado también ingeniero y, sin embargo, pues es un hombre bastante preparado, con bastantes inquietudes, un hombre que opina, un hombre que sabe, pero, es un poco excepción a la regla. (Mad Cult)  
I have a brother in law who is also an engeneer and, nevertheless, he is a pretty well qualified man, with interests on the side, a man that has opinions, a man that knows, but he is the exception to the rule

b. ... y es lo que luego yo les digo: que no me dejen la puerta abierta. Pero es gente que no entiende. Unos días traen llave, y otros días están toca y toca; que ninguno trae llave. (Mex Pop)  
... and that's what I then tell them: don't leave the door open. But it's people who don't understand. Some days they bring the key, and other days they are ringing and ringing, none of them brings the key.

(1.16) a. - Marisa.  
- Dime.  
- Y por fin el barco de Naxos, ¿a qué hora sale? (Corec)  
- Marissa  
- Tell me  
- And, finally, the boat from Naxos, at what time does it leave?

- Buenos días.  
- ¿Qué tal?  
- Aquí estoy, hija.  
- ¿Menos liada?  
- En la mesa de mi vecino porque en la mía - tienen mi teléfono cogido. Cuéntame.  
- Mira, que - yo quería decirte que pusieras ya un P.T.A - . (Corec)  
- Could I talk with Marica? This is Teresa, from Celaje, good morning  
- Good morning  
- How are you?  
- Here I am,
- not as busy?
- at my colleague's desk because at mine, they are using my phone. Tell me
- Look, I wanted to tell you to send a P.T.A.

(1.17) a. I: Bueno sí, si ha habido, nada menos antier se ahogó un niño allá, allá, allá, ¿sí sabías Ø? (Monterrey)
   Well, yes, there have been some [accidents], the day before yesterday a kid drowned there, there, you knew Ø?

   b. y le digo, no vayas a decir a tu papá que fuimos anca tu padrino, dice, no... dice mami, ¿si me pregunta?, ¿vedá que no le Ø voy a decir? (Monterrey)
   and I tell him, don't tell your dad that we went to your godfather's place, he says No, he says, Mom, if he asks me, I am not going to tell Ø him, right?

In the examples in (1.15) and (1.16), there is no entity in the discourse or discourse context that is being referred to by the object of the verb. Like in (1.13) or (1.14), it could be said that the meaning is focused on the "activity" and the content of the zero element is unknown or indifferent for the speaker, following Fillmore's characterization. Although there are differences between (1.15) and (1.16), the important point here is that these cases could be argued to be examples of an inherent object or non referential null object, and, crucially, that these are not examples of the null object studied in this dissertation. On the contrary, in examples (1.17a, b), the null object of the verb is anaphoric, i.e. refers to an element that can clearly be retrieved from the discourse or discourse context (antier se ahogó un niño allá [the day before yesterday a child drowned there] in (a) and fuimos anca tu padrino [we went to your godfather's place] in b), and the abstract entity referred to by means of the null element can be the object of subsequent references. These direct objects could also be coded in Spanish with the propositional lo and this is the type of null object studied in this dissertation.
1.2.2.2. Null pronoun, VP ellipsis and Null Complement Anaphora

Previous literature in Spanish has documented and studied mainly two cases of phonologically null elements referring to propositions or previously introduced VPs: VP ellipsis (1.11) and Null Complement Anaphora (NCA) (1.12), repeated here as (1.18) and (1.19).

(1.18) Luis gana mucho dinero y María también
Luis makes a lot of money, and so does Maria

(1.19) Luis fue al acto; María, en cambio, no pudo
Luis attended the event; Maria, on the other hand, couldn't (go)

Both NCA and the null pronoun studied in this dissertation have clear differences with VP ellipsis: In VP ellipsis, the head of VP is empty whereas in the constructions of NCA and the null pronoun studied here, the head of VP is full, occupied by a verb in finite form which selects a clause, which is elliptical, and whose content is present in the discourse context. Besides this, VP ellipsis is "regular", in the sense that any predicate can be elided in these sentences, while NCA (and, as we will see, the null object studied here) is restricted to certain predicates, while other verbs, that apparently have very similar characteristics and are semantically very close, do not allow this construction. Other difference is that VP ellipsis, unlike NCA and the null pronoun observed here, follows the "verbal identity requirement": the syntactic antecedent (when there is one) has to be structurally identical to the form that the anaphorized complement would have taken were it present (Hankamer and Sag 1976:413).
In turn, the null direct object observed in this dissertation and NCA (Hankamer and Sag 1976; Depiante 2001; Brucart 1999) also show clear differences. NCA, described as one type of elliptical construction by Brucart, and one type of VP anaphora by Huang (2000), refers to the fact that certain verbs in Spanish (and other languages) allow their infinitival complements to be null and, as noted in Depiante, these sentences do not allow extraction of an element that would have appeared inside the clausal complement (Depiante 2001). The verbs that accept the NCA construction in Spanish are modal and aspectual verbs (restructuring verbs for Depiante), such as *poder, deber, saber* (to be able to) or *empezar*; verbs or predicates that express predisposition, attitude or intention (*aprender a, dudar de, saber, estar acostumbrado, estar preparado...*) and other causative predicates that express permission or influence on someone else's attitude (*autorizar, ayudar, dejar, disuadir, enseñar*) (Brucart 1999). These verbs do not accept the subordinate clause that they select be pronominalized with the clitic *lo*, in many of these cases probably because the complement of the verb is necessarily introduced by a preposition. On the contrary, verbs that reject NCA usually allow their propositional object to be represented by the clitic pronoun *lo*:

(1.20) Juan necesitaba aprobar esa asignatura pero no Ø pudo/*lo pudo
Juan needed to pass that class, but he could not/*LO could not

(1.21) Juan necesitaba aprobar esa asignatura pero no *Ø consiguió/no lo consiguió
Juan needed to pass that class, but he didn't manage to
The null object observed in this dissertation does not fit the description of NCA in Brucart (1999) and Depiante (2001): first, this null object is not the complement of modal or aspectual verbs. Second, the null object studied here would not be explicitly coded as a non-finite sentence, but expressed with a finite clause in Spanish if it were present in the discourse, and the finite clause would not be introduced by a preposition. And, finally, the null object studied here is in clear variation with the pronominal form lo, and the phenomenon of null complement anaphora has been described as applying to verbs that do not allow the clitic pronoun as their complement.

The term Null Complement Anaphora, however, has been used in the literature on other languages with less restrictions: For Huang (2000), null complement anaphora refers to an elliptical construction in which a VP or IP complement of a verb is dropped, and he offers the following two examples of the phenomenon: They asked him to pay at least lip-service to the principle, but he refused/John told Mary that the beauty of Cambridge rivals the splendor of Venice, and Peter told Susan". The latter example would not be a case of NCA following the description of Brucart and Depiante.

In turn, Cyrino and Matos (2006) include under the category NCA constructions that are comparable to the null object studied in this dissertation: to argue that NCA in Portuguese should not be classified as a case of deep anaphora in Hankamer and Sag's (1976) classification, Cyrino and Matos point out (among numerous arguments to support their claim) that in Portuguese the null object is not in complementary distribution with the neuter pronoun (it/so, lo/eso, o/issso) and that both the null and the overt pronoun can sometimes alternate (examples 34 and 35 from Cyrino and Matos):
(1.22) a. Ainda que queiras __, não podes resolver esse problema. (EP)
   b. Ainda que o queiras, não podes resolver esse problema. (EP)
       Even if you want (it), you can't solve that problem

(1.23) a. Nós pedimos aos rapazes para nos visitarem, e todos se recusaram __. (BP, EP)
   b. Nós pedimos aos rapazes para nos visitarem, e todos se recusaram [a isso].
       (BP,EP)
       We asked the children to visit us, and they all refused (to that)

They argue that in Portuguese, NCA is legitimized by (semi)auxiliary verbs (modal or aspectual) that select non finite sentences as well as by main verbs that select finite sentences, and that the latter can also carry the overt neuter pronoun as a complement. For Cyrino and Matos, therefore, the null object that complements a non-auxiliar verb, and that would be expressed by a finite clause if it were present in the discourse (comparable, therefore, to the null object in Spanish studied in this dissertation), is included under the category of NCA, which led these scholars to affirm that in Brazilian Portuguese, NCA and null objects are very similar, only differentiated by the content of the object: the NCA denotes a situation (João me pediu para eu ir à festa, e eu aceitei [John asked me to go to the party, and I accepted]) and the null object refers to an entity (João me deu um convite, e eu aceitei [John gave me an invitation, and I accepted]).

Unlike Cyrino and Matos, I won't include the null objects with propositional antecedents studied in this dissertation under the phenomenon of NCA in Spanish because that would imply overlooking some differences that, in my opinion, are crucial in this study, namely the characteristics of the verbs that host the null complement, the
syntactic form that the clause would show if it were present in the discourse and, more important for a variationist study, the fact that the null pronoun observed in this dissertation alternates with the overt clitic pronoun *lo*.

To summarize, in this section I have characterized the null object that will be studied in the following chapters as an anaphoric pronoun, which is clearly different from other phonologically empty objects (or lack of object) that could be described as cases of inherent objects or non referential object, and I have argued that there are enough differences between the null object under consideration and the phenomenon described under the term NCA in Spanish as to distinguish both constructions.

### 1.3. Anaphora and discourse referents. Terminology

A clarification of the terminology that will be used in the remainder of the text and some assumptions about reference and anaphora in discourse are in order at this point.

In this dissertation I will be talking about the null and overt pronoun *lo* in variation as anaphors referring to propositions. An "anaphor" or "anaphoric element" is a linguistic element whose interpretation depends on another linguistic constituent. Since I exclude from my study cases of "forward anaphora" or cataphora, the linguistic constituent that provides the anaphor with semantic content is always previous in the discourse. I will use the traditional term "antecedent" to refer to this semantically-independent linguistic element.
I assume the idea of a discourse model that contains, among other things, discourse referents (Karttunen 1976), i.e. representations of entities associated with the meaning of noun-phrases and other linguistic expressions in the discourse-so-far, which can be referred to by subsequent anaphoric elements. A linguistic element, the antecedent, or a non linguistic element, introduces a discourse referent in the discourse model, and this discourse referent can be subsequently referred to by means of another anaphoric expression. According to the theories of dynamic interpretation, these discourse referents are managed dynamically across discourse, to reflect the way in which information changes.3

The kind of anaphora studied in this dissertation involves antecedents that are not NPs but large linguistic constituents, one or several sentences, and the discourse referents that these linguistic elements introduce are sometimes called "high-order entities" or "abstract entities", as opposed to individual or first-order entities, and this case of anaphora in which the anaphor refers to high order entities has previously been called "discourse deixis" (Lakoff 1974; Webber 1991).

1.4. Dissertation overview

The dissertation is organized as follows. In Chapter 2, I present a review of some relevant studies dealing with the distribution of anaphoric forms and the use of null objects in Spanish and other languages. The review of this literature will provide working hypotheses and research questions regarding the null/lo distribution. Chapter 3 will be

---

3 The issue of what exactly the discourse models of the different participants contain is out of the scope of this dissertation. For a discussion of this issue, see Roberts 2005b.
devoted to the methodology: I will define the envelope of variation of this phenomenon as well as the coding scheme used for the quantitative analysis.

In Chapter 4, I report and comment the results of the statistical analyses of Mexican and Peninsular Spanish and present some of the theoretical concepts that will be further developed in chapters 5 and 6. Chapter 5 is devoted to discuss the main theoretical implications of this analysis: I analyze the results of the analysis regarding the notion of accessibility, present the concept of completeness of the proposition as a crucial constraint of the null/lo variation and discuss the relation of this variation and the transitivity of the clause. Chapter 6 provides concluding remarks and suggestions for further research.
CHAPTER 2

ANAPHORA RESOLUTION AND NULL OBJECTS

No previous studies have addressed specifically the phenomenon of variation observed in this dissertation and, for this reason, we lack specific observations or predictions from previous literature regarding this phenomenon. There is, however, a related background on anaphora and the distribution of different anaphoric expressions that will be useful for this study. I will first review some general studies of anaphora resolution (section 2.1) and some attempts to apply the observations extracted from these studies to anaphora involving high order entities (2.1.2.). In section 2.2. and 2.3, I will turn to the literature that specifically deals with the use of null objects in Spanish and other languages, and in sections 2.4. and 2.5., the ideas of prototypical direct object and prototypical transitivity and their relation with null objects will be considered. The review of this literature will provide us with several specific research questions, presented in section 2.6.
2.1. Studies on anaphora resolution

2.1.1. The accessibility of NP antecedents

The distribution of anaphoric expressions in discourse has been addressed from several theories that share the departure point of considering the accessibility of the discourse referent as the key factor to account for the choice of a particular anaphoric form at a particular point in discourse (Ariel 1994, 1996; Givon 1983; Gundel et al. 1993). The distribution of the different anaphoric expressions, according to these theories, is conditioned by the cognitive status of the discourse reference (Ariel 1994, 1996; Gundel et al. 1993) or by topic accessibility in discourse (Givon 1983).

In Givon’s approach, the notion of topic continuity or accessibility essentially determines anaphoric encoding in discourse. Topics (i.e. what is being talked about) are more or less continuous in the discourse and the degree of topic continuity or accessibility will determine the anaphoric expression chosen by the speaker. The notion of topic continuity can be measured in the discourse by three factors: referential distance (measurement of the gap between the previous occurrence of a referent/topic and its current occurrence in a clause), potential interference (measurement of the disruptive effect which other referents within the immediately preceding register may have on topic availability) and persistence (measurement of the number of clauses to the right in which the topic/participant continues an uninterrupted presence as a semantic argument of the clause) (Givon 1983).
Gundel et al. address the notion of accessibility of the discourse referent proposing a “givenness hierarchy”, a scale of six cognitive statuses, relevant to the form of referring expressions in natural language discourse.

In focus > activated > familiar > uniquely identifiable > referential > type identifiable.

In this proposal, each status on the hierarchy is a necessary and sufficient condition for the appropriate use of a different form or forms. The statuses signaled by the different forms are not mutually exclusive, but implicationally related, such that each status entails (and is therefore included by) all lower statuses, but not vice versa. In this scale, therefore, the statuses are ordered from most restrictive (in focus) to least restrictive (type identifiable) (Gundel et al. 1993).

Ariel (1990:70) proposes the following accessibility marking scale, which precedes from low accessibility markers to high accessibility markers:

Full name + modifier > full name > long definite description > short definite description > last name > first name > distal demonstrative + modifier > proximate demonstrative + modifier > distal demonstrative + NP > proximate demonstrative + NP > distal demonstrative (-NP) > Proximate demonstrative (-NP) > stressed pronoun + gesture > stressed pronoun > unstressed pronoun > cliticized pronoun > verbal person inflections > zero.

In order to choose among the available anaphoric expressions, the speaker must take into account factors related to the antecedent’s salience (the mental representations of speaker and addressee, and entities encoded as topics or subjects are more salient than the representations of entities not present, non-topics or non-subjects) and the nature of
the relation between the antecedent (clause) and the potential anaphoric expression (clause) (short distance usually implies high accessibility, and embedding creates a closer relation between the clauses containing the antecedent and the potential anaphoric expression than conjoining does).

These proposals, despite their differences, agree on the generalization that the more accessible or salient the antecedent, the more likely speakers are to use the minimal anaphoric form available (zero, when possible) and, conversely, the less accessible the antecedent, the more likely speakers are to use the longer or more complex anaphoric form. Since one of the most uncontroversial ways of making an NP referent highly accessible is by means of introducing it into the discourse in a prominent syntactic position, a personal pronoun (and null pronouns in the languages that accept them) is the anaphoric form most likely used to refer to a NP that has been linguistically introduced in a recent sentence and is not in competition with other potential antecedents. Interestingly enough, when the antecedent is not a NP, the same pattern of distribution doesn't seem to apply in English, as will be explained in the next section.

2.1.2. The accessibility of clausally introduced antecedents

It has been shown in the literature that reference to clausally introduced entities, unlike anaphoric reference to NP antecedents, is more often established in English by means of a demonstrative pronoun (that), and less frequently with the personal pronoun “it” (Asher 1993; Eckert and Strube 2000; Webber 1991) or, in other words, that entities introduced by a clause are accessible to immediate subsequent reference with
demonstrative pronouns but less accessible to reference with the personal pronoun *it* (Gundel, Hegarty and Borthen 2003). This is not only true in examples like (2.1) and (2.2), cases in which the use of the personal pronoun would have two possible referents (one first order entity and one clausally introduced entity), and the pronoun takes preferably the NP antecedent while the demonstrative is used to refer to the clausally introduced entity. Examples like (2.3) (example 2 in Hegarty 2003) show that even if there is no NP-referent potentially competing with the abstract entity, the demonstrative *that* is used rather than the personal pronoun (Gundel, Hegarty and Borthen 2003).

(2.1)  [Jane bought a new bike]$_i$
       a. It’s great
       b. that’s great

(2.2)  a. There was a snake, on my desk. It$_i$ scared me
       b. [There was a snake on my desk]. That$_j$ scared me, and it$_i$ scared my office mate too.

(2.3)  a. For more sophisticated mathematical ideas, it is necessary to study the cognitive mechanisms that characterize mathematical concepts. Lakoff and Nuñez argue that these are the same ones that characterize ordinary ideas. *This* leads to the authors’ main thesis, that mathematics develops by means of metaphors

       b. # It leads to the authors main thesis.

Both demonstratives and personal pronouns, however, can be used to refer to both kinds of antecedents, as shown by contexts that force either an individual (NP-antecedent) (2.4) or a discourse deictic interpretation (2.5) (Eckert and Strube 2001, examples 19 and 20):
At first sight, the use of demonstratives rather than personal pronouns referring to propositions contradicts the clear predictions made by the anaphora-resolution theories mentioned above: Why would a demonstrative be preferred to refer to an entity that is introduced in the immediately preceding discourse, very often with no competing antecedent, and that should be therefore very accessible to the speaker and interlocutor?

In order to explain these facts, it is suggested that abstract objects are not introduced into the discourse model by the constituents that describe them but rather by virtue of anaphoric reference. Discourse-deictic reference is said to involve referent coercion (Dahl and Hellman 1995) or ostension (Webber 1991), a mechanism by which “the anaphor itself is used to create a new referent in the discourse model” (Eckert and Strube 2001) and, therefore, the referents of discourse-deictic anaphors do not exist in the discourse-model unless anaphorically referred to.

Some authors (Borthen, Fretheim and Gundel 1997; Gundel, Hegarty and Borthen 2003; Hegarty 2003) maintain this idea applying the scale proposed by Gundel et al. (1993) and argue that the abstract entity, when first mentioned in the discourse, has an “activated” status, but not the “in focus” status that NP-referents can achieve by means of the first mention in the discourse. This would explain that only the demonstrative pronoun, and not the personal pronoun, is available for reference. In order for that abstract entity to be “in focus”, and therefore be referred to with a personal pronoun, it...
has to be made more accessible somehow. The most typical way would be by means of a subsequent anaphoric mention in the discourse. Once the abstract entity has been referred to twice, it achieves the status “in focus” and the subsequent mention(s) can be made with the personal pronoun “it” (examples 32 from Eckert and Strube 2000 and 3 from Borthen et al. 1997):

(2.6) - and we make it so easy for them [to stay there with welfare that they can get by just signing some papers],
  - granted, they can do that, very easily. It’s easy to do, but look where it puts them

(2.7) [There was a snake on my desk]. That scared me, and it scared my office mate too.

The reason why the first mention of an abstract entity would not be at the focus of attention but merely activated (in Gundel et al.’s (1993) terms) is not clearly explained in this literature, as Hegarty (2003) points out. Borthen, Fretheim and Gundel (1997) suggest that it has to do with the fact that these entities are typically introduced by a whole sentence and, therefore, this first mention usually does not correspond to expressions that have prominent syntactic functions in the sentence (subject or object). This explanation, not developed by Borthen, Fretheim and Gundel (1997), seems to be supported by the fact that, as Gundel, Hegarty and Borthen (2003) show, abstract entities introduced by a subordinate sentence in subject position are accessible for mention with it (example 39 from Gundel, Hegarty and Borthen 2003):

(2.8) A: What surprised you?
    B: For Max to get drunk at his fundraiser would surprise me. And it would really hurt his image too.
Gundel, Hegarty and Borthen (2003) and Hegarty (2003) point out that, although mention by a clause typically does not bring an entity into focus of attention, there are various semantic and pragmatic factors that can affect the accessibility of abstract discourse entities and, consequently, the anaphors available for subsequent reference. These factors include covert arguments, non-linguistic moves such as an inquisitive look and presuppositional contexts (Gundel, Hegarty and Borthen 2003; Hegarty 2003).

Entities introduced in presuppositional contexts, such as complements of factive verbs (43, from Gundel et al. 2003, example 23), or within interrogative complements (27, from Hegarty 2001, example 12) can be immediately referred to by means of the personal pronoun “it”, unlike entities introduced by clausal complements of bridge verbs.

(2.9)  A. Alex verified that the company destroyed the file.
      B. That’s false; the file was submitted to the judge
      B’. It’s false; the file was submitted to the judge

(2.10) A: Alex wonders whether the company destroyed the file
       B: It’s not likely. The file contained no incriminating information
       B’: That’s not likely. The file contained no incriminating information
       [it/that: that the company destroyed the file]

This fact is explained because the propositions introduced in presuppositional contexts are “automatically accorded a high cognitive status, as familiar or activated” (Hegarty 2003), or capable of being accommodated as familiar (Gundel, Hegarty and Borthen 2003) and therefore its mention in the speaker’s utterance suffices to render it in focus (Hegarty 2003; Gundel, Hegarty and Borthen 2003).
The proposals that I have just reviewed are developed to provide an explanation for the fact that discourse-deixis, or reference to clausally introduced entities is done with a demonstrative pronoun rather than with a personal pronoun in English. These studies, however, don’t distinguish between the uses of the two anaphoric forms in subject and object position. Although they don’t refer to the syntactic position of the anaphor explicitly, the examples provided indicate that these studies are focusing on the use of the anaphor in subject position: in Hegarty (2001), only one of twenty two examples include the anaphor in object position; Hegarty (2003) includes five examples with the anaphor in object position and twenty one with the anaphor in subject position; In Borthen, Fretheim and Gundel (1997), one out of thirteen examples include the pronoun in object position; Eckert and Strube (2000): four pronouns appear in object position and fifteen in subject position, and Gundel, Hegarty and Borthen (2003) offer two examples with the anaphor in object position and twenty two with the pronoun in subject position. Moreover, looking only at the examples that they provide, one would at least question whether the same pattern is found in object position, since, for instance, in Hegarty (2003), three of the examples he includes with the anaphor in object position (out of five) accept the personal pronoun “it”. In Eckert and Strube (2001), two out of four examples of object anaphoric constructions accept “it” and in Gundel, Hegarty and Borthen (2003), the two examples that show the anaphor in object position allow the personal pronoun.

Although it is argued that “similar restrictions can be found in other languages” (Gundel, Hegarty and Borthen 2003:281), the same pattern does not seem to apply in Spanish: in subject position, where the demonstrative pronouns eso/esto compete
with the null pronoun (and very rarely with the null subject personal pronoun *ello*), the demonstrative does not seem to be the form most frequently used. Example (2.11), from Borthen Fretheim and Gundel (1997) (example 12), and its translation into Spanish (2.12) show this contrast between Spanish and English.

(2.11) A1: Why didn´t you come to the rehearsal yesterday?  
B1: I though I told you. I had to help Peter move.  
A2: Ah, that´s true  
A2’: ??Ah, it´s true

(2.12) A1: ¿Por qué no viniste al ensayo ayer?  
B1: Pensé que te lo había dicho: tenía que ayudar a Pedro a mudarse.  
A2: Ah, ??eso es verdad  
A2’: Ah, Ø es verdad

The expression of subject pronouns is beyond the limits of my study, but the generalization observed in English seems even less accurate in Spanish if we consider the anaphors in object position: the clitic *lo* and the null pronoun are intuitively more frequent than the demonstrative pronouns *eso/esto/aquello* as direct objects. Quantitative data will be provided in section 5.1.2 to test this intuition.

Some studies have noticed that the meaning of discourse units other than noun phrases do not have the same lifetime for anaphora remention via a pronoun (Byron 2004: 19): discourse referents evoked by these higher-level entities would only be available for pronominal mention in the discourse segment immediately following the source of the abstract entity, and would not allow long-distance pronominal reference (Schiffman 1985, Dahl and Hellman 1995). In the following example, the demonstrative pronoun *that* cannot refer to the information about house X in line b (from Webber 1990,
modified by Byron 2004):

(2.13) A. There are two houses you might be interested in:
    B. House X is in Palo Alto. It has three bedrooms and two baths. The owner is
    asking $425K.
    C. House Y is in Portola Valley. It has three bedrooms, four baths and a pool. The
    owner is asking $600K.
    D. I heard all this from a real-estate friend of mine.
    E. *But that’s all I know about House X.

The following constructed example from Spanish, however, shows that reference
can be made to a proposition introduced in the discourse several sentences before and that
the abstract entity is still active in the discourse model:

(2.14) A: No he podido trabajar nada en todo el fin de semana. Primero el sábado Juan
    se cayó por las escaleras y se rompió la pierna, así que tuvimos que ir al hospital y
    allí pasamos toda la tarde. Después, el domingo fui a visitar a mi madre que se
    puso enferma y pasé con ella todo el día.
    B: ¡Pobre Juan!, no (lo) sabía.
    B’: Ah. Pues el sábado por la noche hablé con Juan y no me (lo) dijo.
    ‘A: I could not work anything during the weekend. On Saturday, Juan fell on the
    stairs and broke his leg, so we had to go to the hospital and we spent there the
    whole afternoon. On Sunday I went to visit my mother, who got sick, and I spent
    the whole day with her.
    B: Poor Juan! I didn know (LO)
    B’: Mm.. I talked to Juan Saturday evening and he didn't tell me. (LO)

If, as this example suggests, reference can be made to a proposition introduced
several sentences before, the distance between the pronoun and its antecedent could affect
the anaphor choice. Asher (1993), in this respect, proposes that the distance between the
abstract antecedent and the pronoun can vary and points out that this distance can be used
to explain the choice between the personal and the demonstrative pronouns: the
demonstrative would be used when the abstract entity referred to is more than one
sentence away, and “it” would be used to refer to entities introduced in the previous sentence. Byron’s analysis of the Boston University Radio Corpus and the Trains93 corpus of task-oriented spoken dialogs, however, does not uphold Asher's generalization, since in her data demonstrative pronouns are commonly used to refer to entities introduced in the previous sentence (Byron 2004). The issue of whether distance between the antecedent and the anaphor plays a role in the distribution of the two variants studied here will be addressed in the following chapters.

To summarize, with respect to the literature concerning the distribution of anaphoric forms, or anaphora resolution, anaphors referring to first order entities have received much more attention than anaphors whose referent is an abstract element, such as a proposition. While we know a lot about the different linguistic concepts that play a role in the distribution of null objects, personal pronouns, demonstratives and other anaphoric forms when the referent is a NP-introduced entity, the choice of anaphoric forms referring to abstract entities is not satisfactorily explained yet, and the few studies that have been devoted to this issue suggest that the explanation provided for first order anaphora can not be directly applied to abstract anaphora. One of the research questions that will be addressed in the following pages is whether the notion of accessibility and the parameters of accessibility considered to explain the distribution of anaphoric forms referring to NPs are of use in explaining the variation under study in this dissertation.

A different line of research that can shed some light on the distribution of the null pronoun and the overt clitic lo comes from description of the use of null pronouns in different Spanish dialects and other languages. I will turn to this literature in the next
2.2. Null object with propositional antecedents in other languages.

The use of null pronouns to refer to propositions has not been studied in Spanish in spite of the fact that propositions seem to be more likely referred to with a null pronoun than first order entities in other languages.

Meyer Lübke (1923) mentions the use of the null pronoun that refers to an idea already introduced in the discourse in Italian, Romanian, Portuguese, Patois, and French, specially when there is a co-occurring dative pronoun and with verbs meaning, and states that this use was not unusual in Latin (Meyer Lübke 1923:417). More recently, several studies have pointed out that, although French is not characterized as an "object drop" language, the use of null objects is not impossible and is, in fact, common in spoken French (Fónagy 1985; Larjavaara 2000; Lambrecht and Lemoine 1996). These studies report the use of a null pronoun (objets latents, "null instantiation", or "zero anaphora") that is clearly anaphoric and whose referent has to be specific and identifiable in the context in order for the hearer to understand the utterance (Larjavaara 2000). These null pronouns can refer to first order entities (2.15 and 2.16) or to propositions (2.17 and 2.18) in French:

(2.15) (Larjavaara, ex. 3.9)
"Tu as lu les pages?" Hondo lui caressa le nez (...). Il avait lu Ø.
Do you have the pages? Hondo caressed his nose ... . He had read Ø.

(2.16) (Larjavaara, ex. 3.29)
"Pourquoi avoir choisi cette époque?" "Parce que j'adore Ø.
Why have you chosen that period? Because I love Ø.
(2.17) (Fónagy, 10.3):
[La femme de la victime au commissaire qui l'interroge sur sa liaison avec le secrétaire de son mari]: - Mon mari savait Ø [ = que nous avons eu une liaison].
[The victim's wife to the police captain who questions her about her affair with her husband's secretary] - My husband knew Ø [that we had an affair].

(2.18) (Fónagy, 12.2):
- J'ai eu un accident
  - Eva m'a raconté Ø [ = que tu as eu un accident]
  I have had an accident - Eva told me Ø [that you have had an accident].

Lambrecht and Lemoine (1996) point out that the referent of the null pronoun must not only be discourse-active, but it also needs to have the pragmatic status of a ratified topic: topical elements whose occurrence in the proposition is taken to be predictable at the time of utterance. The null pronoun is, according to Noailly (1997), more frequent in commands or interrogative sentences and also with verbs in the second person, present tense.

Fónagy (1985) affirms that the transitive verbs that most commonly get these kind of null objects are verba sentiendi, dicendi et cogitandi, and he speaks of well established structures of object lost with dire [to say], savoir [to know], and trouver [to find]: dieu sait! [God knows], qui sait! [who knows], je ne sais/j'sais pas [I don't know], si je savais [if I knew], c'est comme vous dites [it's like you say]. In fact, null objects with propositional antecedents seem to be the most accepted in standard French (Larjavaara 2000:83; Noailly 1996, 1997). In this regard, however, it is important to note that, although most of these authors point out the higher frequency of the null object with propositional antecedents, they usually include under this category null objects of (semi)modal verbs, that are necessarily coded as zeros in French and never (or very
rarely) as the neuter pronoun, together with the cases that I am interested in, i.e. cases where the null pronoun alternates with an overt pronoun (like examples (2.17) and (2.18) above). Noailly, for example, points out that we have clear cases of zero anaphora with semi-modal verbs, opinion verbs, and verbs marking the development of the process, and she is including under the category of zero anaphora verbs that select both non finite and finite clauses, without distinguishing between them. Similarly, Larjavaara (2000) affirms that standard French only admits propositional "latent objects", referring to cases in which the referent of the DO pronoun would be a subordinate finite clause as well as an infinitive or participle clause, and points out that in many cases of propositional object, it would not be possible or natural to have an overt pronoun instead of the null object. As we saw in 1.2.2., I consider it necessary to distinguish between NCA, where the overt pronoun is impossible, and the null pronoun that alternates with lo.

To address the question of why the null pronoun would be more likely used to refer to propositions, Noailly (1996, 1997) proposes an iconic explanation, arguing that propositional complements are cognitively perceived as the more abstract complements and the null pronoun materializes its abstract and "hardly existing" character. A different explanation is offered by Larjavaara (2000), who proposes that an account of null objects in French would have to take into consideration two forces: the animacy of the referent and the desire of no ambiguity. Regarding the animacy of the referent, this author considers it necessary to distinguish between four types of referents: 1. proposition - 2. non human, not (or little) concrete - 3. non human, concrete - 4. human, since this four-part distinction is important in French in as much as it concerns their likelihood to be
represented as a null object. This hierarchy acts together with another force, the "desire of no ambiguity". According to this scholar, since the null pronoun is a phonologically empty category, it doesn't have any disambiguating power in itself and, therefore, its use is more conditioned by the desire to avoid ambiguities than other pronouns. When a verb accepts any kind of object (the object is not restricted to one kind of object from the animacy hierarchy: not only human referents, for example), the null pronoun can not be used to distinguish between different types of object, i.e. verbs that participate in (many) different constructions are the ones with variable object pronouns. The null, therefore, should be accepted with all the verbs of that category. On the other hand, when the verb accepts objects of different kinds, the animacy hierarchy arises and only referents that are further from the "human" end of the scale are easily represented with the null pronoun (Larjavaara 2000).

There is another interesting difference regarding the behavior of the French null pronouns with propositional antecedents and null pronouns referring to first order entities. The phonological zero representation of a direct object had been traditionally described in French as a case of "haplologie": the lost of a direct object with a dative pronoun due to phonological reasons, i.e. the desire to avoid the cacophony that would result from the juxtaposition of two third person personal pronouns: le lui > Ø lui [ACC MASC SG; DAT > Ø DAT SG] (Grevisse 1993:970; Lambrecht and Lemoine 1996:303). Some authors oppose this explanation, pointing out that it would only explain the lost of the direct object when it is masculine and singular (le), but sequences such as la lui [ACC FEM SG; DAT SG] or les lui [ACC PL; DAT SG] could not be explained in the same
way (Armary 1997:380). What is interesting for this dissertation is that, as Armary (1997:379) and Yaguello (1998a:270) point out, the dative pronoun with a null object has to be a third person pronoun when the null refers to concrete entities, but propositional null direct objects can be used with a dative pronoun in first and second person (*il m'a dit; il m'a montré [he told me; he showed me] vs. *il m'a donné and *il m'a accordé [he gave me]). The distribution of null objects, therefore, seems to be less restricted with propositional antecedents and this kind of null pronoun can clearly not be explained as a case of haplopoigie, i.e. due to phonological reasons.

The loss of phonologically overt direct objects with propositional antecedents has been studied in Brazilian Portuguese (BP). Cyrino names the zero phonological realization of the direct object of verbs that allow the neuter clitic o in Brazilian Portuguese “sentential ellipsis”. Within these verbs, she includes querer [to want], tentar [to try] and saber, dizer, crer [to know, to say, to believe], noting some differences in the behavior of these three kinds of verbs (Cyrino 1997:200), like the fact that “propositional verbs” like saber, dizer and crer, accept both the clitic o and the zero as direct objects.

A very interesting piece of information comes from Cyrino’s diachronic analysis of the object system in Brazilian Portuguese: her study shows that the objects of propositional verbs were crucial in the development of the null object system that BP has today, since this was the context in which null objects first became more frequent: in the 18th century, 46.3% of direct objects with sentential antecedents were realized as zeros, while only 7.5% of the direct objects with a specific NP and 6.1% of the direct objects with a non-specific NP were coded as zeros. In the 19th century, direct objects of
propositional verbs like *saber* or *dizer* were realized as zeros 87.1% of the time and only in 12.9% of the cases was the neuter clitic *o* used (Cyrino 1997:250). Until 19\textsuperscript{th} century, therefore, the speaker could choose between using the neuter clitic or the null form for propositional direct objects: both structures were grammatical and Cyrino argues that the factor that made speakers in 18\textsuperscript{th} and 19\textsuperscript{th} centuries choose the null realization more often than the neuter clitic *o* was phonological: there was a phonological change in Brazilian Portuguese by which clitics are always phonologically enclitic. The emergence of this phonological constraint explains, according to Cyrino, why speakers increasingly chose the null realization of the direct object, since it became impossible to start a sentence with a clitic (Cyrino 1997:255).

Iliescu (1988) notes in Romanian the potential lack of an overt anaphoric direct object when the neuter pronoun would be expected to refer to an idea, a fact or an event, and she observes that the occurrences of the null pronoun with this kind of referent are as frequent as the use of the overt form. Regarding the distribution of these anaphors, Iliescu points out that, according to her observations, the use of the overt pronoun is more frequent with transitive verbs that, in general, are more rarely used as "absolute verbs" (*to repeat, to confess, to pretend*); and verbs that can be used as transitive and as "absolute verbs" are more likely to appear with a null pronoun than with an overt anaphor (*to know, to learn, to imagine, to understand*). Iliescu also points out that there is a phono-stylistic factor, namely the parallel syntactic repetition: if there is a clitic or a neuter pronoun in the first sentence, the second sentence is more likely to be constructed in a similar way, and suggests that the notion of "topic/theme" may also play a role in the
distribution of the null anaphor and the neuter pronoun: the null would be more likely when the topic/theme of the sentence is not the direct object but something else (the opposition between the verbs in her example (14) or the origin of the knowledge in her example (15)). Finally, the lack of an overt direct object with this kind of reference is, according to Iliescu, more frequent in negative than in affirmative sentences.

Outside of the Romance languages, Meyerhoff studies the occurrence of null objects in Bislama. In her study, only inanimate entities favor the use of a null object (while animate referents disfavor it) and, specifically, the referent type “event or proposition” has a very high favoring effect on the null object (.935 on the Goldvarb results) (Meyerhoff 2002:333).4

Fillmore (1986) observes the use of null objects in English and presents the null anaphor as being lexically determined: some verbs, with some of their meanings, accept null objects whereas similar verbs or the same verb with different meanings require an overt object:

(2.19) a. I forgot that she had fixed it: I forgot"
    b. I forgot my keys- *I forgot

(2.20) a. He noticed that she was blind: He noticed
    b. He noticed the mouse - *He noticed

---

4 These quantitative results, however, should be taken cautiously because Meyerhoff explains that “since clausal complements were excluded from the corpus, it is not clear whether these anaphors should have been included as well” (Meyerhoff 2002: 333).
Cornish (2006) points out that the verbs that allow a null pronoun in Fillmore's list frequently have a propositional complement:

(2.21) John saw the "No Entry" sign. But Bill didn't see it/*Ø

(2.22) A: You'll have to wait till Monday, sir. The Council offices are closed today.
     B: I see Ø/*it

The possibility of using a null pronoun, according to Cornish, is not completely lexically determined but has to do with how predictable the object is: when the verb is understood in its "cognitive sense", the kind of denotation that can be associated with its complement (proposition or event) is determined by the host predicate sense and, if the verb is very frequent, the null pronoun can be used. When the verb has a more concrete sense, the selection restriction that the verb with this meaning transfers to its complement is too general to determine one type of denotation, and therefore the potential referent of the complement is semantically too unpredictable to be coded as a null pronoun.

2.3. Null objects in Spanish

Null direct objects in Spanish have been traditionally considered to be only permissible when the noun referred to by the direct object is non-referential and non-countable, i.e. mass nouns (2.23a) and bare plurals (2.23b) (Campos 1986; Clements 1994, 2006):

(2.23) a. ¿Compraste café?
   -Sí, compré Ø.
   -Did you buy any coffee?
   -Yes, I bought some
b. Quería comprar libros pero no encontraba Ø.
    I wanted to buy (some) books but I didn't find (any)

c. *Quería comprar el libro pero no encontraba Ø.
    I wanted to buy the book but I could not find Ø

The phenomenon of null direct objects is less restricted in Spanish contact
dialects (Basque Spanish [Landa 1995; Eguía 2002]; Quiteño [(Yépez 1986; Suñer and
Yépez 1988] and Paraguayan [Choi 1998, 2000; Palacios 1998]), in which null objects
with definite and specific antecedents are commonly found, and in both Quiteño and
Paraguayan Spanish null DOs are now a typical feature of the speech of monolinguals as
well.

(2.24) (from Yépez): - Aún no le Ø he franqueado (referring to los documentos)
    I haven't mailed him Ø [the documents] yet"

(2.25) (from Palacios, ex 1) - ¿Sabes si Pedro trajo los libros?
    - Sí, Ø trajo.
    Do you know if Pedro brought the books?
    Yes, he brought Ø

Until now, Masullo (2003) has been the only study of null DOs (what he calls
"definite object drop") in a non-contact dialect of Spanish, the Spanish of River Plate,
where null objects are possible if the referent is recoverable from the immediate context
of utterance (examples (2) and (4) in Masullo 2003):

(2.26) ¿Le Ø retiro, señor?
    Shall I take Ø for you? (e.g. a tray)
Although it has not been studied yet, the null pronoun is also found in monolingual varieties of Spanish when the referent of the direct object pronoun is a proposition instead of an entity (Schwenter 2006). This case of variation between the null pronoun and the clitic *lo* has not received much attention in the literature and the only previous studies that deal with this exact topic are Reig and Schwenter (2007) and Reig (2007), both preliminary studies to this dissertation.

In spite of not being the topic of research in previous literature, the phenomenon of null objects with propositional antecedents has been attested in some descriptions of different Spanish varieties. The first mention of this phenomenon comes from Kany (1945: 146), who affirms that *lo* with cognition and communication verbs is frequently omitted in American Spanish, almost always when there is an explicit indirect object pronoun, due to the tendency to avoid double clitics. Solé and Solé (1977: 41) point out that the direct object clitic is often omitted in Spanish with verbs like *decir* (to say), *preguntar* (to ask) and *pedir* (to ask for) if an indirect object pronoun is present, as in *Pregúntale* (Ask him/her) or *Le diré* (I will tell him/her).

 Null pronouns with propositional antecedents can be found in examples included in studies that observe the use of null objects referring to definite noun phrase, mainly as the result of language contact, as mentioned before. The phenomenon was noticed in the Spanish of Ecuador by Toscano (1953), who describes it as more frequent than the null object with inanimate NP antecedents:

(2.27) Dame Ø que te Ø lavo
     Give Ø to me, so I can wash Ø
“En la sierra no suelen usarse los dobles pronombres se lo, me lo, te lo, etc. En vez de se lo dije, lo común es decir le dije. Y a menudo se omite también el pronombre complemento directo de cosa”.

‘In the mountains, double pronouns se lo [DAT:3 ACC:NEUTER/MASC], me lo [DAT:1 ACC:NEUTER/MASC], te lo [DAT:2 ACC:NEUTER/MASC], etc. are not usually used. And the direct object pronoun of thing is also often omitted’

Yépez (1986) and Súñer y Yépez (1988), observing the use of null objects in Quiteño, include under the inanimate antecedents both concrete entities and propositions, and offer some examples of null pronouns where the propositional clitic lo could be expected. Likewise, Palacios Alcaine (2000) observes the elision of inanimate object pronouns in Paraguayan Spanish and presents several examples in which the null pronoun has a propositional antecedent, although she does not distinguish between inanimate propositional and inanimate non propositional referents. Palacios explains that the null pronoun with an inanimate antecedent is found when the elided information can be recovered from the linguistic or extralinguistic context, yet in all her examples the propositional antecedent is explicitly mentioned in the previous discourse. Her examples show cases of null pronouns with ditransitive as well as monotransitive verbs:

(2.28) -Pedro llegó ayer a la mañana a Asunción
     -Yo Ø sé
     Pedro arrived yesterday morning in Asunción
     I Ø know

(2.29) -La piedra cayó por mi pie
     -Eso no es verdad
     -¿Quién Ø dijo?
     The stone fell on my foot
     That’s not true
     Who Ø said?
Landa and Franco (1992) observe that referential null objects are possible in Basque Spanish when they have a [-human] antecedent and find that having a clause as the antecedent is one of the factors that favor the null pronoun (Tengo que comprar un taladro pero no hay prisa, si no puedo hacer hoy, Ø hago mañana [I have to buy a drill but there is no rush, if I can't do it today, I do Ø tomorrow]). Another factor that favors the null object in their study is the presence of a dative pronoun, but the null pronoun is also possible with monotransitive verbs and, in this dialect, the null object is not limited to cognition and communication verbs.

As these brief references show, studies which have observed the use of null objects in Spanish dialects, analyzed mainly as the result of language contact, offer examples of null objects referring to propositions and, at least in some of them, it is pointed out that the null pronoun is more likely to occur as the complement of some cognition and communication verbs. This subtype of direct objects, however, is always included under the category of "inanimate" and no distinction is made between inanimate entities introduced by NPs in the discourse and propositions.
2.4. Prototypical direct objects, the animacy scale and propositions

As the review of the literature dealing with null objects in Spanish shows, null objects referring to first order entities have received the attention of several scholars whereas null objects (or, for that matter, objects in general) referring to propositions have either been excluded from the analysis or collapsed with other inanimate entities, without demonstrating that the distribution and behavior of the overt and null pronouns are, in fact, equivalent for both kinds of referents. Although it could be the case that concrete and abstract entities in anaphora behave similarly in the Spanish varieties studied so far, the opposite hypothesis, i.e. that there would be differences in the distribution of null and overt pronouns referring to abstract and concrete entities, would be, at least, worth exploring, especially provided that, as shown in section 2.2, propositional antecedents and NP antecedents seem to behave differently regarding the possibility of being referred to by a null pronoun in other languages.

The facts described in the literature reviewed in section 2.3 show that the Spanish dialects that allow null objects with definite antecedents either limit this option to inanimate antecedents or allow it for all kind of antecedents; none of them, however, has been described as accepting a null pronoun for animate direct objects and not for inanimate entities. As Schwenter points out, “those direct objects coded as null objects are those which display the prototypical properties of direct objects crosslinguistically” (Schwenter 2006: 23).
It has, in fact, been assumed in linguistic theory that noun phrases with certain properties correlate with specific syntactic positions: subject NPs are typically conceived as animate, definite and specific, and object NPs are typically inanimate, indefinite and non-specific (Croft 1988; Comrie 1989). This idea is exemplified by what has been called "Comrie's generalisation" (Næss 2007; de Swart 2003): "the most natural kind of transitive construction is one where the A is high in animacy and definiteness, and the P is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction" (Comrie 1989:128). This distribution of NPs with certain properties in the sentence is believed to reflect a typical pattern of human experience, where humans or animate entities generally act on non-human or inanimate entities. When the object is higher in animacy definiteness or/and specification than the subject, this is frequently reflected in the linguistic structure in order to help with clause processing and comprehension (Næss 2007). This generalization has its roots in the observation of two crosslinguistic phenomena: direct-inverse systems, and differential case marking, and it assumes an iconic relation between form and meaning: the formal opposition between a morphological marked form versus the absence of such marking is taken to reflect marked objects versus unmarked objects and, as a consequence, definiteness and animacy are considered marked properties for objects (Aissen 2000: 3).

The descriptions of the prototypical properties of the direct objects include features such as animacy, definiteness (Company 2002; Comrie 1979, 1989), specificity, topicality (Comrie 1989) and egophoricity (Dahl 2000) but all of these studies only refer to noun phrase direct objects, and propositional direct objects are either not considered at
all in the discussion or integrated under “inanimate”, together with inanimate noun-phrases. I would like to entertain the idea of adding the referent type “proposition” to the animacy scale, which will then be as follows:

+    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +    +7

If the idea of differentiating these two kinds of inanimate objects in the animacy scale is correct, and following Schwenter's observation that in these contact Spanish dialects the null pronoun is used to refer to prototypical direct objects, Spanish dialects that allow null objects for inanimate things would also necessarily show null objects with propositional antecedents, although we would expect to find varieties in which the null anaphor with propositional antecedents is accepted but is the only case in which the null object is allowed (besides with non-referential and non-countable antecedents). There seems to be evidence that supports this hypothesis. Some of the first descriptions of the dialects that allow null objects with definite inanimate antecedents in Spanish give the impression that objects with propositional antecedents were more frequently coded as null pronouns before those with NP antecedents (Toscano 1953, for Ecuador), as was clearly the case for Brazilian Portuguese (Cyrino 1997). Moreover, as seen in section 2.3, according to the limited data that we have, Spanish dialects that allow null objects for definite inanimate-things also accept the zero realization for propositional antecedents. Mexican and Peninsular Spanish would be cases of varieties in which the null pronoun is only accepted in the first level of the scale, direct objects with propositional antecedents
(as we will see, with vastly different frequencies).

Differentiating inanimate first order entities from abstract entities on the animacy scale also seems to be of use for other pronoun-related phenomena which are often mentioned in relation with the use of null direct objects in dialects of Spanish, such as leísmo, clitic doubling (Colantoni 2002; Yépez 1986) and differential object marking (Schwenter 2006). In this respect, the phenomenon called leísmo, i.e. the use of the dative etymological pronoun (le) instead of the etymological accusative clitic lo (or la), would affect the scale starting from the opposite end: the dialects that demonstrate leísmo would use the non-etymological pronoun le for direct objects most likely when the direct object is animate, less likely when it is inanimate–thing and only rarely with propositional antecedents, and the prediction is that only dialects that show leísmo for all the other cases (animate and inanimate-first order) would allow leísmo with propositional antecedents. Certainly, this hypothesis would need to be tested but it seems to be supported by the data described so far: Yépez, describing the pronominal system of the Spanish spoken in Ecuador (in which definite noun phrase antecedents license null pronouns) states that since le was essentially perceived as animate, it could not take the place of neuter lo to refer to a clause (Yépez 1986:42), and the null pronoun is used in these cases (Yépez, example 29: y así Ø hicimos [and so we did] (referring to how something was organized); example 31: Qué quiere decir eso no Ø sé [what that means I don’t know]). Fernández-Ordóñez writes “En el español de Paraguay, en la región guaranítica argentina, y quizá en Bolivia, se emplea la forma le como único clítico de tercera persona, con independencia de la posición sintáctica ocupada por el ente referido
y sin consideración de su género, número, e incluso, con referente neutro” (emphasis added)\(^5\), suggesting that this use of leísmo is the least frequent or expected (Fernández-Ordóñez 1999).

The distribution of null pronouns also provides parallels with Differential Object Marking (Schwenter 2006). Aissen (2003) uses the concept “prominence” to capture the characteristics shown by more typical objects or more typical subjects in languages. The feature “prominence”, for Aissen, includes two dimensions: animacy and definiteness (Aissen 2003:436). Each of these dimensions is conceptualized as a scale, represented here:

Animacy: Human > Animate > Inanimate  
Definiteness: Pronoun> Name> Definite> Indefinite specific> Non-Specific.

I argue that the scale that I proposed above, which distinguishes two categories within the inanimate entities, things and propositions, could also be of use in the description of this phenomenon: propositional direct objects, I believe, never exhibit direct object marking in Spanish. Although inanimate-thing direct objects may receive the Spanish object marking \(a\) when the animacy of the subject and the object is equal (2.31a)\(^6\), this is never the case for propositional direct objects, even when both subjects and objects display the same semantic features (2.31b, 2.31c).

---

\(^5\) ‘In Paraguayan Spanish, in the Guarani region of Argentina and maybe in Bolivia, the form \(le\) is used as the only third person clitic, independently of the syntactic position occupied by the referent and independently of its gender, number, and even with neuter referent’

\(^6\) For quantitative evidence about differential object marking "\(a\)" with inanimate DOs in Spanish, see Tippets and Schwenter (2007).
(2.31)  a. Una camioneta golpeó a un taxi luego embistió a un tercer vehículo (El Salvador Hoy, 06/10/2000)
A van hit a taxi, then hit a third vehicle’

b. Que estés embarazada indica (*a) que necesitamos contratar a otra persona
That you are pregnant indicates (*A) that we need to hire someone else’

c. Que sus estudiantes le nominen para el premio muestra (*a) que es muy buen profesor
That his students nominate him for the prize shows (*A) that he is a very good teacher

If one accepts the animacy scale just proposed, that would imply that propositional direct objects are the most prototypical direct objects in language. Would we want to defend that this is the case? Before addressing this question, it is pertinent to turn to the notion of transitivity in order to discuss how the previously described notion of prototypical objects fits with the characteristics attributed to the direct object of a prototypical clause in transitivity theories.

2.5. Prototypical objects and the transitivity of the clause

In parallel to the functional typology line of research, functional approaches to the notion of transitivity were developed which, as regards the characteristics attributed to the object of a prototypical transitive clause, are in contradiction to the description of a prototypical direct object presented in the previous section.

The traditional concept of transitivity is "the property of sentences which express the idea that something passes from one participant to another one" (Lazard 2002); In this sense, verbs that can or must take an object are defined as grammatically transitive. The concept of transitivity that I will refer to here, developed in functional grammar, is
understood as a global property of a clause and is conceived of as a gradient feature, which can be measured as more or less, in relation to a prototypical transitive clause that describes an activity that is carried over or transferred from an agent to a patient (Hopper and Thompson 1980).

In this approach, the concept of transitivity is conceived of as a prototype category. In the prototype theory used in cognitive linguistics, categories are defined in terms of similarity to a central exemplar and, in this sense, the notion of transitivity, like any other category, allows for degrees of membership: clauses can be considered more or less transitive, i.e. a more or less good member of the category (Næss 2007). The semantic prototype of transitivity is therefore defined as a set of semantic properties which characterize the construction known as a transitive clause across languages; clauses that, semantically, differ in one or several of these properties from the transitive prototype are also expected to differ from it formally i.e. to show grammatical differences with the grammatical structure that, in each particular language, encodes the semantically prototypical transitive clause.

The different approaches to the notion of transitivity share, with small differences, the essential features of the prototypical transitive clause: the event denoted is a concrete, dynamic action, that is presented as real and concluded, and which involves a volitional, controlling, acting agent participant, who is responsible for the event, and a non-volitional, non-controlling patient, affected by the action (i.e. registers changes of state associated with the event) (Lakoff 1977; Givon 1985a, 1990; Kittilä 2002a; Lazard 2003). The most influential list of parameters that define the transitive clause is the one
provided by Hopper and Thompson (1980:252):

<table>
<thead>
<tr>
<th></th>
<th>high</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Participants</td>
<td>2 or more participants, A and O</td>
<td>1 participant</td>
</tr>
<tr>
<td>B. Kinesis</td>
<td>action</td>
<td>non-action</td>
</tr>
<tr>
<td>C. Aspect</td>
<td>telic</td>
<td>atelic</td>
</tr>
<tr>
<td>D. Punctuality</td>
<td>punctual</td>
<td>non-punctual</td>
</tr>
<tr>
<td>E. Volitionality</td>
<td>volitional</td>
<td>non-volitional</td>
</tr>
<tr>
<td>F. Affirmation</td>
<td>affirmative</td>
<td>negative</td>
</tr>
<tr>
<td>G. Mode</td>
<td>reals</td>
<td>irrealis</td>
</tr>
<tr>
<td>H. Agency</td>
<td>A high in potency</td>
<td>A low in potency</td>
</tr>
<tr>
<td>I. Affectedness of O</td>
<td>O totally affected</td>
<td>O non-affected</td>
</tr>
<tr>
<td>J. Individuation of O</td>
<td>O highly individuated</td>
<td>O non-individuated</td>
</tr>
</tbody>
</table>

Table 2.1. Parameters of transitivity from Hopper and Thompson (1980)

These properties define the semantically prototypical transitive clause that, in each language, is encoded with a particular clause structure, which Lazard names the "major biactant construction". In any language, when a clause differs in one or more elements from the formally prototypical transitive clause (the major biactant construction for Lazard), this clause should not be prototypically transitive semantically, i.e. it is expected to differ from the semantically prototypical transitive clause in one or several of the semantic properties that define transitivity (Lazard 2002).

Going back to the issue that led me to introduce the idea of transitivity, namely the features typically associated with the object position, the object of a prototypical transitive clause in these approaches is defined as being "highly individuated", as the last row in Table 2.1 shows. The notion of individuation includes several properties that
determine the extent to which a participant is perceived as an independent entity in the context (Næss 2007). Hopper and Thompson include the following features defining individuation:

<table>
<thead>
<tr>
<th>INDIVIDUATED</th>
<th>NON INDIVIDUATED</th>
</tr>
</thead>
<tbody>
<tr>
<td>proper</td>
<td>common</td>
</tr>
<tr>
<td>human, animate</td>
<td>inanimate</td>
</tr>
<tr>
<td>concrete</td>
<td>abstract</td>
</tr>
<tr>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td>count</td>
<td>mass</td>
</tr>
<tr>
<td>referential, definite</td>
<td>non referential</td>
</tr>
</tbody>
</table>

Table 2.2. Individuation in Hopper and Thompson (1980)

As Næss points out, it appears that there is a direct contradiction between the transitivity approaches, such as Hopper and Thompson's, and the functional-typological analysis of the typical direct object presented above: Comrie's generalization (Næss 2007, de Swart 2003) and similar approaches attribute to the natural, prototypical direct object the characteristics of a low individuated object, whereas, in the transitivity model, prototypical transitive clauses have highly individuated object arguments (Næss 2007:18). Or, in Næss' words, "clauses which are labeled as less natural or more marked according to one approach are categorized as prototypical according to the other" (Næss 2007:20).

This scholar, however, maintains that it is only an "apparent contradiction", one that stems from a failure to keep apart two distinct levels of analysis and fundamentally different concepts, namely prototypicality or "transitive semantics" (which has to do with
the general properties of transitive objects irrespective of the formal system these objects are encoded in) and what Næss names “markedness or case-marking phenomena”, which "concerns the way in which objects which do or do not show these properties are formally encoded within a specific kind of structural system, namely accusative case-marking systems" (Næss 2007:20).

Næss considers that the difference between these two approaches can be clarified considering the principle of maximal distinction, a property of prototype categories by which a prototype category is defined in relation to opposing categories in the same domain. Each of these analyses takes different entities as their standard comparison against which the clauses with indefinite/inanimate object participants are measured (Næss 2007). In the analysis underlying Comrie's generalization, the properties considered "marked features" are precisely the properties which distinguish transitive from intransitive clauses: the presence of a second independent argument, clearly distinct from the general background or from the other argument of the clause. Næss affirms that, in this approach, two-participant clauses are taken to be "more natural" the closer they are, formally and semantically, to one argument clauses and, consequently, the underlying, implicit standard of comparison for scholars who consider typical objects to be less individuated is an intransitive clause.

7 “Markedness” is used here to refer to the overt morphological sign of the object (differential object marking) or the verb (direct-inverse systems) in some languages when the object has certain characteristics, as opposed to the lack of this morphological mark with other objects. The “more marked construction” is the construction with overt case-marking as opposed to zero (Næss 2007:19).
On the other hand, a prototype approach to transitivity necessarily defines the transitive clause in clear opposition to the intransitive. From a transitive semantics point of view, a definition of the prototypical transitive clause will have to be the construction most formally and semantically distinct from the intransitive, and two participant constructions will be categorized as more or less transitive on the basis of their resemblance to this prototype. The natural contrast between a highly individuated subject and a less individuated object participant that is crucial in Comrie's generalization is not relevant to determine the transitivity prototype (Næss, 2007).

For the purposes of this dissertation, it is of interest to observe how propositional objects and transitive clauses with propositional objects can be related to these accounts of prototypical direct objects and prototypical transitivity. These explanations focus on NP objects and, to the best of my knowledge, do not take propositional direct objects into consideration. If, as suggested in section 2.4, propositions are differentiated from first order inanimate entities on the animacy scale, the immediate question that arises is whether we would want to say that propositional direct objects are actually the most prototypical objects cross-linguistically. Based on the remarks made by Næss and referred to above, the answer to this question would depend on what we mean by "prototypical object". If we consider that the prototypicality of objects has to do with the difference between object and subject, then we could comfortably say that propositional DOs are clearly more different (i.e. different in more features) from a subject than inanimate-things, provided that the subject of cognition and communication verbs is necessarily a
human entity.\textsuperscript{8} In this sense, propositional direct objects would be more typical or less marked for the functional typology theories and, following Næss' observations, this would not imply that propositional objects correspond to the most transitive clauses.

Since the literature devoted to transitivity does not refer to clauses with these kinds of objects and the parameter of individuation is explicitly constructed taking into consideration only NPs, it is not clear where these clauses would fall in these theories. Nevertheless, if we could include under individuation the three-term animacy scale proposed, propositional objects would predictably be less individuated than inanimate-thing objects, and, therefore, a clause with this object would be less transitive than a clause with a more individuated object.\textsuperscript{9} Regarding the individuation of propositional objects, I will posit in chapter 5 that propositional objects, themselves, can be more or less individuated, taking into consideration a new parameter: the completeness of the proposition.

\subsection{2.5.1. Complement taking predicates and transitivity}

Although the theoretical proposals regarding the notion of transitivity do not mention clauses with propositional direct objects, the issue of how to characterize these kinds of clauses using the parameters of transitivity arises in studies that apply the notion

\textsuperscript{8} Propositional objects are the most different from the subject of the sentence, but they don't seem to be the "prototypical patient": these objects, using Dowty (1991) characteristics for a patient, do not suffer a change of state derived from the action denoted by the verb, they are not casually affected by another participant, and there is no movement of the subject that causes this effect on the object in these sentences.

\textsuperscript{9} Hopper and Thompson (1980) include the feature concrete/abstract under the properties that distinguish an “individuated” noun from a “non-individuated” one. If [abstract] is one of the features that characterize non individuated objects, we could argue that propositional direct objects, which are also abstract entities, would be more prototypical direct objects than concrete noun phrases. With these features, however, abstract noun phrases and propositions remain at the same level of the scale.
of transitivity to the analysis of naturally occurring language. Thompson and Hopper (2001), observing transitivity in spontaneous conversation, make the methodological decision of coding clauses traditionally analyzed as a main verb that takes a clausal complement as intransitive sentences, i.e. sentences with one participant. A sentence like *I was wondering why I hadn’t heard from him*, for example, is coded in their analysis as two different clauses: a first clause, *I was wondering*, with one participant (the subject *I*), and a second clause, in this case interrogative, *Why I hadn’t heard from him*. The complement clause, therefore, is not counted in this study as a participant of the main clause of which it is part.

This decision is based on the analysis of the expressions like *I wonder* as markers of epistemicity and evidentiality, and not as main clauses with complement taking predicates, following Thompson (2002) and Thompson and Mulac (1991). Thompson (2002) argues that what is traditionally described as the main clause or the head of complementation can be better understood as an epistemic, evidential or evaluative formulaic fragment expressing the speaker’s stance toward the content of the (complement) clause. To support this claim, she argues that there is no unitary category of “complement”; that there is little argumentation regarding the objecthood of complements (since these “main verbs” don’t usually take NP complements and other “tests”, such as passivization, do not work); and that it is inadequate to consider these “complements” as subordinate, since, in her opinion, “what conversationists are engaged in doing with their utterance involves the complement at least as much as the complement-taking predicate)” (Thompson 2002:131). The conversation data, according
to her analysis, shows that what speakers are engaged in doing with their talk crucially involves the complement, and not the “main clause”, which is there, in most cases, only providing the epistemic, evidential or evaluative frame for the clause they occur with. Evidence for this analysis, in Thompson’s opinion, comes from the fact that both fragments can occur independently in the discourse, and that the most frequent of these epistemic phrases tend to be reduced to formulas, usually with first person singular subjects, either affirmative or negative, and with no complementizer or other elements (I know, you know, I knew).

In my opinion, there are several issues that can be raised against this proposal. Even though cognition verbs expressing epistemic or evaluative values are most frequently used in discourse with first person singular subjects, and these constructions tend to grammaticalize into fixed discourse formulas, an analysis of these sentences as two independent clauses is not completely satisfactory. I believe that one important issue for objecting to this analysis comes from one linguistic construction that these scholars do not mention, namely sentences in which the “completive clause” is pronominalized. The fact that the second (or complement) clause can be pronominalized, and that this is done in many languages with the accusative pronoun, is one argument for the objecthood of these complements not mentioned in these studies. It is, in fact, not clear how these sentences are coded in Thompson and Hopper (2001): their observations regarding

---

10 Thompson's examples of traditionally called complement clause appearing independently in discourse are limited to cases in which the clause clarifies the content of a previously mentioned pronoun (Thompson 2002:145):
K: I just know what you told me
A: well, I can't remember what the deal was
K: that you were driving through Denver

A: that's interesting, I mean th- that you should pair the word aesthetics, ... with advertising.
“object deletion” and the examples provided in this section (that's the best time to find out; you can tell) would indicate that when the object is, in my terminology, a null object (phonetically null anaphor which refers to an abstract entity identifiable by the interlocutors, and that can be used for later re-mention), the sentence would be coded as a one participant clause. It is not clear, however, how clauses like I know that or Forget it are coded (although the last sentence in example 20, don't you remember that, seems to indicate that they would be coded as two participant clauses), and how transitivity is quantified in these cases, since clauses with propositional complements would be very low in transitivity if they were counted as objects (Thompson and Hopper 2001:31) and, at the same time, pronominal objects are said to be high in individuation (Thompson and Hopper 2001:36). How to measure the transitivity of the types of sentences observed in this dissertation is, therefore, unclear to me, and the idea of considering a sentence like No sabía a one participant sentence (with very low transitivity) and the alternative sentence with the propositional lo variant, No lo sabía, a two participant sentence (high in transitivity due to the pronominal object), when in both cases what the speaker did not know is exactly the same abstract idea, is, in my opinion, not satisfactory.

Another observation regarding Thompson’s and Hopper’s and Thompson’s analyses of these sentences has to do with their argument that the complement clauses are not really semantically subordinate to the “main clause” because the information in the complement clause reflects what the participants of the conversation are engaged in doing with their utterance, whereas the main clause does not constitute the “speaker’s interactional agenda”. I believe that, although the analysis of the main clause as a
fragment that merely provides the speaker’s stance towards the declarative or interrogative clause that is at issue in the conversation could be valid for many of the cognition verbs (*think, know, guess*), it cannot be so easily applied to the analysis of communication verbs such as *tell, say, thank, explain* or *ask*, all of them taking a clause and included in the data analyzed in this dissertation.\(^{11}\)

For the reasons just exposed, i.e. the fact that in the sentences that I am analyzing there the clausal complement is, in fact, pronominalized, and because it is not clear that the analysis of the main verb as an epistemic or evidential frame that is not at issue in the conversation is adequate for at least some of the communication verbs included in my analysis, I will not follow Thompson and Hopper in analyzing these clauses as one participant clauses: they will be considered two participant clauses whose verb takes a propositional complement; the issue of how to determine the individuation of the abstract complement, not addressed in Thompson and Hopper (2001) because there is no such complement-type in their analysis, is consequently still open.

Besides the individuation of the object, most of the other parameters that affect the transitivity of the clause can vary in the sentences with propositional objects studied in this dissertation (except for parameters A, B and E in Hopper and Thompson's list), which makes it interesting to explore whether the distribution of null and overt objects could somehow be related to the notion of transitivity. It is worth noticing at this point that some of the syntactic contexts that, according to studies of null objects in Spanish

---

\(^{11}\) The verb *to tell* is included in some examples, mainly in the construction *I can't tell*, analyzed as an evidential marker. *To say* is only included in Thompson (2002) in one of the examples that illustrate exceptional cases where the conversational action is carried by the complement-taking predicate phrase (ex 49).
and other languages, constrain the use of null pronouns (not necessarily with propositional referents) are in fact features that, according to the transitive approaches, affect the degree of transitivity of a sentence. It has not been proposed so far, as far as I know, that null objects in Spanish dialects or in other languages are related to the transitivity of the sentence but the fact that negative sentences (Iliescu 1988), commands and interrogative sentences (Noailly 1997) and perfective or telic vs. non-perfective or atelic clauses (García and Portero 2002) are mentioned as contexts favoring null objects in different language varieties and that a sentence with a null object is, in its external form, more similar to an intransitive clause makes the relation between transitivity and null objects an appealing one.

2.6. Further research questions

Several specific research questions arise from the review of the literature in this introductory chapter and will be addressed in the remaining of the dissertation:

- Is the presence of a dative pronoun necessary for the null pronoun to occur, or does it highly favor the null pronoun, as suggested in previous literature?
- The construction of NCA is only possible with certain verbs, whereas other verbs, apparently similar ones, require an overt direct object. Is the null pronoun that alternates with the clitic lo also restricted to a certain number of verbs and prohibited with other verbs of the same semantic class?
- Is the notion of accessibility a key notion to explain the distribution between the null pronoun and the propositional *lo*? If "accessibility" is a central idea here, how can it be defined?

- Can the data from Spanish support the idea that abstract objects are not introduced into the discourse model by the constituents that describe them but by virtue of anaphoric reference, as has been argued for English abstract anaphora? I.e. is the pronoun (null or clitic) preferably avoided in favor of the demonstrative pronoun when the abstract entity has only been introduced by a sentence in the discourse? or, in other words, do the examples of *lo* and the null pronoun analyzed in this study follow a previous anaphoric mention of the same proposition?

- Are the discourse referents evoked by the null pronoun and the propositional clitic *lo* necessarily available in the immediately preceding discourse segment, as some scholars have argued for other languages? If, as suggested by Asher, this is not necessarily the case, does the distance between the antecedent and the anaphor affect the use of the null or the clitic *lo* in Spanish?

- Are there syntactic contexts such as commands or interrogative sentences (Noailly 1997) or negative sentences (Iliescu 1988) that favor the null pronoun? If yes, is there any correlation between the distribution of the variants and the transitivity of the sentence?
2.7. Conclusions

In this chapter, studies belonging to two areas of research that, unfortunately, hardly ever find each other, have been reviewed. One of these areas of study corresponds to the studies on anaphora resolution and the distribution of anaphoric forms, mainly devoted to anaphora to first order entities but including some attempts to explain the distribution of anaphoric pronouns referring to abstract antecedents. The second line of research observes the use of null objects in varieties of Spanish and other languages that are traditionally described as non object-drop languages.

Although neither of these two areas of research addresses specifically propositional anaphora in Spanish, the observations in these studies will be the foundations of my coding system and the variationist analysis. The review of this literature provided several hypotheses regarding the variation under study in this dissertation. A clear prediction emerges from the literature on anaphora resolution: the shorter anaphor, in this case the null pronoun, should be preferred for more accessible entities and the more complex anaphoric expression should correspond to less accessible entities. This prediction will be tested in the following pages. Similarly, several of the research questions stated in the previous section are based on the observations on the use of null objects in other Spanish varieties an languages. In order to address these questions, the quantitative methodology characteristic of a variationist study will be combined with qualitative observations. The methodology used in this dissertation is described in the next chapter.
CHAPTER 3

METHODOLOGY

3.1. Variationist linguistics and morphosyntactic variation

I will address the distribution of the null pronoun and the propositional clitic *lo* from a variationist perspective (e.g. Labov 1969; Milroy and Gordon 2003; Sankoff 1988). The main characteristic of this linguistic approach is that, unlike other linguistic approaches such as generative linguistics, it accepts variation as a significant linguistic fact (Labov *to appear*). Indeed, language is considered to be inherently variable and the assumption is that there is a patterned component to it, beyond what Chomsky (1965:3) characterized as "such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors" (Paolillo 2002:2).

Variationist studies\(^\text{12}\) consider that linguistic features (morphosyntactic, phonological, lexical or discursive) may vary in a systematic way and that, whenever there is a choice among two or more linguistic alternatives and the choice may be

\(^\text{12}\) The name "variationist analysis" comes from the notion of variable rule (Labov 1969), a notation for variable rules which adapted that of generative phonology by annotating rules with probabilistic weights. The concept of a "variable rule", however, is not crucial for today's variationist linguistics and, in fact, most variationist analysis do not postulate the existence of variable rules (Paolillo 2002).
influenced by several external or internal factors, the distribution of the alternatives (or variants) can be quantitatively modeled using statistical techniques (Young and Bayley 1996; Sankoff 1988).

The goal of variationist analyses is not to discover individual occurrences or overall rates of occurrence, but patterns of variability in the data investigated (Tagliamonte 2001:731). A quantitative analysis allows the investigator to model these grammatical tendencies and regularities in the data when several factors are hypothesized or proved to influence a linguistic feature.

In the history of variationist studies, there has been some controversy as to whether morphosyntactic variation can be studied in the same way as phonological variation, since it is not clear for all scholars whether the analytical device proposed by Labov, the concept of “linguistic variable”, can be expanded to investigate morphosyntactic variation. A linguistic variable is realized in at least two variants that are linguistically equivalent ways of saying the same thing, but socially and stylistically different.

The concept of syntactic variable was soon used to analyze the complementizer *que* deletion in Montreal French (Sankoff 1973), the passive and active constructions (Weiner and Labov 1983) or the French auxiliaries *avoir/être* (Sankoff and Thibault 1977). However, the disagreement on the possibility or the convenience of extending this concept from phonology to syntax soon arose, the main concerns being whether the syntactic variants would be, in fact, “linguistically equivalent”, i.e. “mean the same”, and whether the morphosyntactic variation is correlated with external (social and stylistic)
As for the second issue, the question of whether the choice between two syntactic constructions, such as the active or passive sentences, is influenced in some way by external factors, it is pointed out that if syntactic variation is observed to show no social or stylistic meaning, using the concept of linguistic variable may be inappropriate (Lavandera 1978). Probably the main question raised, however, is whether the different variants of a pretended syntactic variable “say the same thing”, a requirement that Labov had first pointed out to describe the variants of a phonological variable. Lavandera (1978) considers that the variants of a syntactic variable, unlike phonological variants, carry cognitive meaning, and the question emerges of whether the concept of variable can be extended to syntactic variation assuming that it is defined by “sameness of meaning” in phonological variation and not in syntactic variation. Lavandera proposed to substitute the condition that referential meaning must be the same for all syntactic alternates with a condition of functional comparability (Lavandera 1978:181).

Romaine (1984), as well as Winford (1984) in a more detailed taxonomie, differentiates among “pure syntactic” and “morpho-syntactic or morpho-lexical” variables: pure syntactic variation involves the alternation of whole constructions or arrangements of items (Winford 1984:281) and, although according to Romaine these variables are not socially or stylistically constrained, Winford mentions cases of syntactic variation in Creole languages conditioned by social factors (Winford 1984:282). Morpho-syntactic variables, on the other hand, are conditioned by both linguistic and social/stylistic factors and often deal with the presence or absence of some linguistic item
Even though the use of the concept “linguistic variable” to define this kind of variation is controversial, I agree with Winford in considering that “linguistic variables (and variable rules) must be seen as the purely heuristic devices that they are, not as theories of language” and that “linguistic theory has profited from the analysis of variable ways of saying the same thing, and that quantitative studies provide powerful methods of proof” (Winford 1984:285). I will use this methodological device as the starting point of my analysis, assuming that the phenomenon that I am studying is a linguistic variable (morpho-syntactic variable in both Romaine’s and Winford’s typologies), with two variants that are equivalent in terms of function and cognitive content: they are equivalent in referential meaning, but I expect to find differences in their linguistic-discursive behavior as well as correlations with external information (although the latter is limited in my study).

From a variationist point of view, therefore, the two representations of the anaphoric propositional direct object that I am considering, the clitic lo and the null pronoun, are distinct concrete realizations of a linguistic variable, the anaphoric direct object with a propositional antecedent. I start from the assumption that the two variants are not in free variation, neither are they categorically conditioned by environmental-linguistic factors. I assume that the distribution of these forms is conditioned by linguistic (and possibly extralinguistic) factors that are manifested as quantitative relations. I expect several factors to play a role in the distribution of the two pronominal forms, as in the case of anaphors with NP antecedents, since, as the “principle of multiple causes”
explains, it is unlikely that any single contextual factor can explain the variability observed in natural language data (Bayley 2002). As it is usually done in morphosyntactic variation studies, I will use variationist analytical techniques to uncover the differences in use (regarding internal factors, in this study) that exist between the two constructions, i.e. the bases of that variation.

It is necessary to point out that this will not be a sociolinguistic study. I do believe that synchronic variation is systematic and socially symbolic, constrained by both linguistic and social factors, but, due to the fact that only limited socio-demographic information is available in the corpora that I am using, I will focus here on the linguistic constraints and the results of the analysis of some external factors will be presented in section 6.4., mainly as the basis for future research.

3.2. Corpus

The linguistic material analyzed in this dissertation consists of naturally occurring data extracted from six Spanish corpora: Habla Culta de México (Mex Cult), Habla Popular de México (Mex Pop), Habla de Monterrey (Monterrey), Habla Culta de Madrid (Mad Cult), Corpus de Referencia de la Lengua Española Contemporánea: Corpus Oral Peninsular, genero conversacional (COREC), and Alicante-Corpus Oral del Español (ALCORE). All of them are corpora of spoken language, non-task oriented, and are intended to be representative of natural and unconstrained speech.
The Habla Culta and Habla Popular de Mexico and Habla Culta de Madrid corpora belong to the corpora collected in the Habla Culta Project, which consists of recorded interview data gathered from cities in different Spanish speaking countries. The speakers were selected considering factors such as family, education, occupation and cultural experiences. All of the informants interviewed for the Mexico and the Madrid corpora were born or had lived in Mexico city or Madrid, respectively, for at least three quarters of their life and their parents were always Spanish speakers, preferably born and educated in the same city. The interviews in Mexico were conducted in the late 1960’s and early 1970’s, and the material corresponding to educated speech was published in 1971 (Lope Blanch 1971) and the interviews to non educated speakers were published in 1976 (Lope Blanch 1976). The interviews for the Madrid corpus were recorded in the 1970’s and published in 1981 (Esgueva and Cantarero 1981). The Habla Culta de México consists of approximately 167,000 words, Habla Popular de México has approximately 172,000 words, and Habla Culta de Madrid, over 140,000 words.

The COREC corpus (Corpus de Referencia de la Lengua Española Contemporánea: Corpus Oral Peninsular) was collected in 1991 and 1992 by the Department of General Linguistics of the Universidad Autónoma de Madrid (Marcos Marín 1992). Although the corpus includes samples of different language uses (administrative, scientific, political, journalist discourse, advertisements, etc.), I only used the data from the "conversation" and "debate" sections (approximately 404,000 words). These materials, unlike the Habla Culta and Habla Popular corpora, are not limited to interview data, but include other kinds of oral interactions such as TV talk shows,
secretly recorded spontaneous conversations or phone conversations.

The Habla de Monterrey corpus consists of six hundred interviews of about one hour each, which follow the Lavobian sociolinguistic interview and cover topics such as "celebrations and food", "work" and "the crisis in Mexico" (Rodríguez Alfano 2004). This corpus was collected with a sociolinguistic interest and includes valuable sociodemographic information of the speakers. This project started in 1985 and the interviews were collected through 19 years by researchers from the Universidad Autónoma de Nuevo León. I analyzed 120 interviews from this corpus.\textsuperscript{13}

Finally, ALCORE (Alicante-Corpus Oral del Español), collected by the Universidad de Alicante, includes sociolinguistic interviews and group conversations of 76 informants, all of them inhabitants of Alicante but not necessarily born there (Azorín 2002). The corpus offers the transcription of 800 minutes of conversation and includes detailed socio demographic information. I included the Alcore corpus in my analysis to extend the number of tokens from Spain and make it comparable to the data from Mexico, and for this reason I only used the first three sections of this corpus, Huertas-San Juan, Tómbola-San Agustín and Pla-Garbinet, which add up to over 260,000 words.

3.3. Defining the envelope of variation

Previous to the quantitative analysis, some crucial decisions have to be made regarding the sites where the two variants of the linguistic variable are possible, what is usually named "defining the envelope of variation" or "circumscribing the variable context".

\textsuperscript{13} Information about this project is available in the website http://www.filosofia.uanl.mx.
In this case, the envelope of variation was determined inductively: I first searched for the overt clitic pronoun *lo*, and took the occurrences of the verbs that appeared with the propositional clitic in the corpora. Then, I searched for those verbs and noted whether *lo* was present or not in all occurrences of the same verb with anaphoric reference to a previously mentioned proposition. This method ensures that all the verbs included in the analysis allow both variants as its direct objects, and allows for a more replicable procedure.

With this methodology, the analysis includes the DOs of a limited group of cognition and communication verbs which take sentential complements: *decir* (to say), *entender, comprender* (to understand), *saber* (to know), *contar* (to tell), *explicar* (to explain), *imaginar* (to imagine), *preguntar* (to ask), *agradecer* (to thank) and, only in Mexico, *platicar* (to tell).

Once the decision of defining the envelope of variation inductively was taken, it was necessary to exclude the contexts where variation cannot occur. I will provide here an account of contexts that fall outside of the envelope of variation and were excluded from the analysis: those contexts where variation is not possible (one of the variants is either necessarily present or absent), as well as other contexts where both anaphors seem to be acceptable but which, for reasons to be explained, are considered distinct from the context of variation defined here.
a. Cataphoric uses of the pronouns

Both the null pronoun and the overt *lo* appear in the corpus in cataphoric uses, i.e. referring to clauses that are explicitly introduced later in the discourse:

(3.1) Dice: "Le digo a tu padre que te cuelgue del pescuezo y que te... que te mate; que te mate tu padre, que tanto que le has costado, que ni... te mate otro". Pos ai me tiene usté con tantos miedos. Pos fue y se lo dijo mi tía... a él: que decía que no me quería dar mi mamá, y que mejor me llevara y que me llevara. (Mex Pop) She says: I'll tell your father to hang you by your neck, and to, to kill you, your father should kill you, after all the trouble you've caused him, no one else should do it". And there I am with so much fear. So she went and she told him LO, my aunt, told him: that I said that my mather didn't want to give me, and that it was better that he just took me.

(3.2) La doctora M. es una doctora americana. Ustedes saben: en Estados Unidos siempre nos ganan en eso. siempre tienen oportunidad de hacer todo tipo de experimentos. (Mex Cult) Dr. M is an american doctor. You know: in the US they always beat us on that, they always have the opportunity of doing every kind of experiments

The uses of the null pronoun and the propositional clitic in cataphoric reference were excluded from the analysis because anaphora to a previously referred entity and cataphora have previously been shown to be different phenomenon that behave differently in many respects.

b. Topicalization/left dislocation

Examples like (3.3), (3.4) and (3.5), instances of left dislocation (3.3) and topicalization constructions (3.4, 3.5) were excluded from the analysis:

(3.3) Mira, quería que subiese Lolita a dormir y en cuanto la dijimos que tenía que ser en su cama las dos dijo que a lo mejor, que a lo mejor se daba la vuelta y la daba con el brazo un escayolazo. Eso me lo dijo mamá. (COREC)
Look, she wanted Lolita to come to sleep [at hour home], and as soon as we told her that it had to be both of them in her bed, she said that she might, sí might turn around while sleeping and hit her with her plastered arm. That mum told LO me.

(3.4) "Que sueñes con los angelitos, mi hijito", y punto. No me voy a poner a atacar, porque muchas amistades de nosotros, que son así... también así laicos: "No, que eso no digas." (Mex Cult)
"Dream with the little angels, my son", and that's it. I am not going to attack, because many of our friends, who are also lay: "No, that, don't say"

(3.5) -Porque... pues sí: Ustedes hacen su trabajo. No tienen porque decirles nada. ¿O sí?
- Pero eso no comprenden, no entienden ellas. (Mex Pop)
- Because... yes: you do your job. They don't have to tell you anything. Do they?
- But that they don't understand, they don't understand

As we can see in these examples, both the duplication with an overt resumptive pronoun or the lack of overt pronoun are found in Mexican Spanish (not in Peninsular Spanish). Although examples like (4) and (5), where no resumptive pronoun is used in the clause to refer to the NP in TOP position, are very interesting, these examples were excluded from the analysis because they constitute a different syntactic construction than the one studied in this dissertation.

c. Impersonal and passive "se" sentences

The occurrences of the mentioned cognition and communication verbs in "se sentences" (impersonal or passive sentences) were excluded because the clitic pronoun can never be used in this context:

(3.6) Pues... este... ahora... francamente... al menos... yo, en lo personal -digo yo- estoy muy agradecida con estas señoras que... pues teniendo sus ocupaciones, sus hijos y todo, vienen a tomarse la molestia de... de venirnos... pues, ahora sí que a abrirnos los ojos, ¿verdad? Y, digo yo... pero pues se les agradece, porque... (Mex Pop)

74
Well, now, frankly, at least, I, personally, I am very thankful to those ladies who, having their occupations, their children, and everything, take the trouble to come and open our eyes, right? And, I mean... but, se-IMP thank them, because...

(3.7) Las calles estaban llenas. Y ¿por qué? P's no, no se sabe. (Mex Pop)
The streets were crowded. And why? no, se-IMP don't know.

\[ \text{d. Null Complement Anaphora with saber} \]

As explained in section 1.2.2.2., Null Complement Anaphora can occur with the verb saber, when it means "be able to":

(3.8) Pues no, ¡je, je! No me gusta, en absoluto, cocinar; y además, es que no sé. (Mad Cult)
Well, no, jeje, I don't like cooking, at all, and besides, I don't know [how to do it]

In sentences like (3.8), saber would take an infinitival clause as a complement and the null pronoun found here is an example of Null Complement Anaphora, discussed before. The propositional clitic lo is impossible in this context and, therefore, these cases were excluded from the analysis.

\[ \text{e. Idiomatic expressions} \]

Also excluded from the analysis are fixed or idiomatic expressions in which the object is necessarily present or absent. The decision of whether one expression should be excluded from the envelope of variation was based on whether that particular expression, that could potentially be considered an idiom or, sometimes, a discourse marker, showed variation in the corpora used for this dissertation. For example, although it should not be

---

14 The categorical use of one of the variants in these expressions in the corpora was the criterion chosen to exclude them from the envelope of variation. These expressions have certain characteristics that led me to consider them fixed or idiomatic expressions: their high frequency (specially in the speech of some
impossible to find, in language, cases of ¡no me lo digas!, the fact that the construction ¡no me digas! is perceived as an idiomatic expression, used in very specific pragmatic contexts, and, crucially, the fact that the expression with the overt pronoun ¡no me lo digas! was never documented in any of the six corpora analyzed led me to exclude it as a site of potential variation. With this criterion, the following expressions were excluded:

- ¿comprende(s)?
- ¿Cómo diré/ cómo diría?
- no me diga(s)
- ya (te/le) digo
- por decirlo así
- quién lo iba a decir
- ¿entiendes?
- ¿sabes?
- quién sabe
- digo (discourse marker, Mexico)\(^\text{15}\)
- no sé

\(^\text{15}\) De luz gastamos como setenta y nueve y centavos; de luz, ¿entiende? Digo... de... de lujos no tenemos casi nada, pero lo que tenemos es... por lo... por lo que ha salido de aquí. De ahí nos mantenemos... -digo- de ahí se mantiene toda la familia, en una palabra. Así es que... digo... Ps en este caso, asegún se vea la persona, así se le cobra. (Mex Pop)

‘On electricity we spend around seventy nine cents, on electricity, you understand? DIGO... luxuries we don't have any, but what we have is... because of... what came out of here. We get by on that ... DIGO... the whole family gets by on that. So... DIGO..., in that case, depending on how the person looks, we charge him.’
It should be clarified that *no sé* was only excluded in those uses where there was not an available clearly identifiable referent of the direct object of *saber*, and all these cases were generally uses of *no sé* that seem to interrupt the discourse, frequently preceding a reformulation. The following example is a good illustration of the uses of *no sé* that were excluded from the analysis.

(3.9) Luego, por otra parte, una cosa muy interesante para mí, desde mi punto de vista al menos, es el intercambio; prque... eso es... *no sé*, entonces sí que es vivir plenamente en un ambiente... de otro país. Por tanto, la misma familia donde se esté pues ya... hablando otro idioma. Y el... mismo, *no sé*, el mismo instinto de... de conservación pues hace... hace intentar un poco más seriamente hablar el idioma. (Mad Cult)
Then, on the other hand, an interesting thing for me, at least from my point of view, is the exchange; because... that's... *I don't know*, then it is really living in an environment... from another country. therefore, the same family where one is speaks another language. And the same... *I don't know*, the same instinct of conservation makes one try more seriously to speak the language.

I also excluded the expression *no contarlo*, meaning "not to last, to die":

(3.10) Abuelita está muy mal, y ahora voy a ir a verla, porque le ha dicho el médico que, *que no lo va a contar*. Tenía ochenta y cinco años. (Mad Cult)
Granma is very sick, and now I am going to visit her, because the doctor has told her that, that she is not going to last (*lit. she is not going to tell LO*). The was eighty five.

(3.11) -Claro, pero es que eso es, es un problema, ¿no?, y yo de verdad que... si sigo así mucho tiempo...
-No, claro, si puedes hacerlo, mejor.
-... *no lo cuento*, ¿eh? (Mad Cult)
Of course, but that is, that is a problem, isn't it? and I , really, if I continue like this for a long time..
- No, of course, if you can do it, much better.
-... I won't last (*lit. I don't tell LO*).
f. No anaphoric direct objects

As mentioned in section 1.2.2.1, some of the verbs that allow the null pronoun and the clitic lo in direct object position can also be used with non anaphoric, non referential direct objects or, for some scholars, intransitive uses of the verb:

(3.12) -Y luego, entran uno... este... estudiantes... unos estudiantes, y este... practicantes...
- Sí, practicantes.
- Entonces, este... como van... como son practicantes, pos no saben. (Mex Pop)
- And then, some students came in, students, and... assistants
- yes, assistants
- and, since they go, since they are assistants, they don't know

(3.13) Y... este... y es lo que luego yo les digo: que no me dejen la puerta abierta. Pero es gente que no entiende. Unos días traen llave, y otros días están toca y toca; que ninguno traen llave. (Mex Pop)
... and that's what I then tell them: don't leave the door open. But it's people who don't understand. Some days they bring the key, and other days they are ringing and ringing, none of them brings the key.

(3.14) [Javier] me dijo: "Mira, ¿ese parto quieres? O. Key, a mí también me parece que es el mejor; ve." No se opuso ¿yes? Pero él no conocía el sistema... Lo ha conocido... a través de lo que yo le he podido dar a conocer. Pero si tú no tienes ese interés, y no te empeñas, y no les dices, y no les explicas... pues ellos nunca llegan a comprender. Y además el apoyo del marido es muy importante. (Mex Cult)
[Javier] told me "Look, do you want that labor? OK, I also think that that's the best option. Go". He didn't object. But he didn't know the system... He has known it from what I could tell him, but if you don't have that interest, and you don't insist, and you don't tell them, and you don't explain to them... they never get to understand. And the husband's support is very important.

These uses of decir, saber, explicar, or entender do not accept in Spanish the clitic lo and were systematically excluded from the analysis.
Antecedent not identifiable

The same decision was made regarding cases where the referent of the direct object (either lo or null) was not clearly identifiable in the discourse. By way of illustration, consider the following examples:

(3.15) -Y al final me dicen:... iba yo, claro, co... ¿quien iba a ir?, iba yo por los resultados y me dicen: <<No, tienen que venir tus padres>>. Dije: << ¿Mis padres? Vamos, ¡si no lo saben! (Mad Cult)
And, at the end they say... I was going, of course, who else would go?, I was going to get the results and they say "No, your parents have to come". I said "My parents? Come on, if they don't know LO!

(3.16) Y el otro día voy con mi padre director, y antes de leerla [la carta] yo, la lee él. Dice: "Pues esto es una decl... se declara, es una declaración de amor; mira." Ahí la tengo guadada todavía la carta. "¡Ay, Padre!" --"¿Te simpatiza?"-- "Sí, Padre, mucho; mucho me simpatiza." --"pues vamos a ver, vamos a ver... Yo quisiera conocerlo, antes de decirte Ø." Pues entonces, que organizamos que... una entronización al Sagrado Corazón, precisamente, y convidamos al padre J. (MexCult)
And the following day I go with my ---, and before I read it [the letter], he reads it. He says "This is a... a declaration of love, look". I still keep it, the letter. "A, father". "Do you like him?" Yes, Father, a lot. I like him a lot". "Then let's see, let's see. I would like to meet him, before I tell you Ø." And so we organized a -------, and we invited father J.

Although in some cases it could be inferred from the discourse what the content of the direct object is (for instance, in example 3.15, based on the previous discourse, we can infer that what the parents did not know is that the girl went to have this test done), more than one option is often available: in (3.16), for example, the content of the direct object of decir could be qué me parece [what I think about it], que le digas que sí a su proposición [to say yes to his proposal], or other similar interpretations. In order to avoid second-guessing and because several of the factors that I included in the quantitative
analysis (referential distance, type of antecedent, etc.) could not be tested in cases where the antecedent is not clearly identifiable, I excluded such contexts as sites of potential variation.

Examples of anaphoric direct objects whose content was not clearly identifiable in the discourse were very frequent in the COREC due to the characteristics of this corpus: many of its conversations had more than two speakers, and some of them were secretly recorded. Both characteristics caused sometimes bad quality recordings and parts of the conversation are often missing in the transcription. It was not unusual to find anaphoric direct objects whose antecedent was probably in the missing portion of the discourse and which had to be excluded from the analysis.

h. No previously mentioned interrogative referent

In some of these cases, the reason to exclude the example from the envelope of variation is not only the fact that the antecedent is not clearly identifiable in the discourse: as examples (3.17), (3.18) and (3.19) show, when the referent of the anaphoric direct object is not present in the discourse but is inferable from the discourse or discourse context, and its interpretation is, necessarily, an interrogative sentence, the only anaphoric form available for the DO is the null pronoun, and the clitic lo is impossible:

(3.17) - Y le dimos el cuadro -
- Allí cenamos todos.
- y le encantó.
- ¿Le encantó?
- ¡Og!
- ¿Sí?
- Le encantó el cuadro.
- Están entusiasmados con el cuadro.
- Le habíamos regalado un edredón
- Sí.
- De la lista de bodas, ¿no?
- Sí.
- Le pregunté: "¿Qué quieres?", y tal, ¿no? que ya tenían lista de boda, ... Pues me dijo que tenía una, una lista de bodas y entonces, yo me fui al día siguiente a la lista de bodas y le regalamos una cosa de veinte mil pelos y un cuadro. Y, y el de, y luego, "No; una sorpresa", decía ella, pero yo no sabía, porque yo decía: "Si la regalo un cuadro a lo mejor pues no, la apetece más otra cosa que le haga falta, quizá (Corec)
- And we gave her the painting (...) and she loved it
- she loved it?
- ooooh
- did she?
- she loved the painting.
- They are delighted with the painting
- We had bought her a duvet
- yes
- from the weding list, right?
- yes
- I asked her: "What do you want?", and so, right?, that they already had a weding list, she told me that she had a, a weding list and so I went the next day to the weding list, and we bought her something for twelve thousand pesetas and a painting. And, then, "No, a surprise", she said, But I didn't know Ø, because I thought "If I give her a painting, maybe she prefers something else, that she needs, maybe"

(3.18) Mira, no me han dicho la, la edad de todos. Me han dicho que están entre esa edad. Yo no descarto para nada que ese hombre con 27 años pueda estar in/metido en eso. No, no descarto porque no he preguntado, y yo no he hecho ningún de ese, de, de esto, entonces no puedo afirmar mientras que yo no haga un estudio. (Corec)
Look, I haven't been told everybody's age. I have been told that they are around that age. I don't rule out that that man, who is 27, could be invoved in that. No, I don't rule that out because I haven't asked Ø, and I haven't done any , any... of that, therefore I can not affirm, while I don't do a study.

(3.19) -El Instituto Kipling. Es colegio laico relativamente, porque les dan religión, a los niños que quieren llevarla, les dan religión. El director es católico y todo eso, pero es mixta la escuela, y está muy bien preparada; los preparan muy bien.
-Pero ---digamos- la materia de religión es optativa.
-Optativa, sí, sí.
-Si quieren pueden llevarla.
The Kipling Institute. It is a relatively lay school, because they teach religion to the kids who want to take it, they teach them religion. The principal is Catholic, and all that, but it is a mixed school, and it's very well qualified, they prepare them very well.

- But, let's say, the subject "religion" is optional
- optional, yes, yes
- if they want, they can take it
- it's up to them (lit. they will know Ø, they will know Ø.)

In example (3.17), the content of the direct object of yo no sabía is not explicitly mentioned in the discourse. To interpret this sentence, one must infer, from the discourse, a plausible interpretation and this interpretation is necessarily an interrogative sentence:

yo no sabía qué regalarle/qué hacer/si regalarle un cuadro [I didn't know what to offer her/what to do/if I should give her a painting]. It seems that when the referent of the direct object is not mentioned in the discourse and is necessarily interpreted as an interrogative sentence, the clitic lo is not possible and the null anaphor is obligatory. The same is true for examples (3.18) and (3.19): in both of them, there is not a plausible antecedent for the direct object of preguntar and saber respectively, and in both cases, to interpret the sentence one must infer an interrogative clause as the complement of the verb: yo no he preguntado si un hombre con 27 años puede estar metido en eso/no he preguntado qué edad tienen los que están metidos en eso, [I haven't asked whether a 27 year-old man can be involved in that/I haven't asked what age the people involved are] etc. and ellos sabrán si quieren estudiar religión/ellos sabrán qué hacen [lit. They will know if they want to study religion/they will know what they should do]. Again, in both examples, where the content of the direct object is not explicitly mentioned in the discourse and has to interpreted as an interrogative sentence, the clitic lo is impossible.
Examples like (3.17), (3.18) and (3.19) were excluded from the quantitative analysis but we will come back to them in section 5.2., where they will be useful for the qualitative analysis of the null pronoun and the propositional _lo_.

*i. Duplicated DO*

Outside the envelope of variation were examples of duplication of the direct object: sentences like (3.20), (3.21), (3.22) and (3.23), where the propositional _lo_ and direct object clause both co-occur in the discourse:

(3.20) Entonces, como no se trata de una aislada sino de dos, y la del Seminario por un periodo de tiempo que fue tres años, entonces creo que... puedo decirlo, sin temor a equivocarme, que no es tan duro como se pone (Mad Cult)
So, since it was not one retreat but two, and the one in the Seminar for a period of time that was three years, then I think that, I can tell _LO_, without fear to be wrong, that it is not as hard as people say.

(3.21) - Es que, termina en agosto o así. Es que ahora va a venir Tere. Teresa va a venir en junio.
- Sí.
- Y va a estar aquí, en casa.
- ¿ un pase un mes y otro pase otro mes?
- _Lo vamos a preguntar, si se puede_. Si se puede, pues en junio pondremos a Teresa (Corec)
  It [the swimming pool pass] ends in August or so. Now Tere is coming. Teresa is coming in June.
  - Yes
  - And she's going to be here, at home.
  - one pass for a month and another one for another month?
  - _We are going to ask, if that is possible_. If that's possible, then in June we will put Teresa

(3.22) -Aunque no pongas ni media... gorda más. Te aprueba.
- ¿Quién te lo ha dicho eso? (Mad Cult)
- Even though you don't make any effort. She passes you.
- Who has told you that?
(3.23) instalamos los aparatos porque para mí el objetivo principal es que la gente
duerma tranquila; y me los llevo todos al "Banana" y al "Cielo de San Pedro", no
sé si lo sabes que lo inauguro ya. (Corec)
And we install the systems because for my the main goal is that people sleep well,
and I take all of them to the "Banana" and "San Pedro's heaven", I don't know if you
know LO that I open it already.

These examples were obviously excluded because it is impossible to affirm that
there is variation between the clitic lo and the null pronoun: if the clitic pronoun were
not expressed in these sentences, we would not postulate the existence of a null object but a
sentence with the direct object clause explicit and no anaphoric DO element used.

j. Truncated sentences characteristic of oral discourse

The last group of examples excluded from the analysis consists of disfluencies
characteristic of oral discourse, like false starts, hesitations or reformulations that result in
truncated sentences and very often make the discourse difficult to interpret. The following
examples serve as illustration:

(3.24) -Y... usted ¿qué... qué piensa... digamos... Si usted dijera...
tuviera facilidad para
dar alguna solución, a usted ¿qué... qué pensaría? (Mex Pop)
And you, what... what do you think... let's say... if you said... could give a
solution, to you, what, what would you think?

(3.25)-Cada cabeza es un mundo, señorita.
- Sí, eso sí. Él puede decir, por ejemplo... Yo le voy a decir... Bueno, cualquiera
de estos chamacos, voy a decir... "Bueno, pues yo puedo agarrar una carrera
grande o larga", (Mex Pop)
- To each his own, madam
- Yes, that's true. He can decide, for example... I will tell him... well, any of those
kids, I will say... "Well, I can take the longer career"

(3.26) ¡Ah!, porque lloraba... y decía... Cuando ella entró... le decíamos... Ella andaba
pa'lló y pa'cá...(Mex Pop)
Ah!, because she was crying..., she was saying..., when she came in..., we told her... she was here and there...

These and other examples of syntactically incomplete clauses, which make it difficult to define whether a null pronoun was expressed and what the potential antecedent could be were systematically excluded from the analysis.

3.4. Coding

A total of 1324 tokens of the dependent variable were collected from the corpora (669 from Mexican Spanish and 655 from Peninsular Spanish). Each of these tokens was coded for the value of the linguistic variable (i.e. the presence or absence of the DO clitic pronoun lo) as well as for 12 linguistic independent variables. In order to include the idea of the accessibility of the antecedent as previously considered for NP anaphoric reference, the factor groups Referential distance, Turn and Number of times the referent was mentioned were included. Based on intuitions or observations expressed in previous studies dealing with null objects in different languages (reviewed in chapter 2) I included the internal factors Prior form and grammatical role of the antecedent, Verb tense and Verb person, Polarity, Sentence type and Presence of a dative pronoun. Finally, based on my own observations of the data, I decided to include Kind of antecedent, Presence of the adverb ya and Presence of a manner adverbial. These factors and the hypotheses linked to each of them will be explained in this section.
As mentioned before, I will focus in this dissertation on the analysis of the linguistic constraints but I included in my coding schema three external factor groups: sex, age and, in Mexico, education, whose results will be presented in Chapter 6. I also coded for individual speaker and corpus. As I will further explain in 5, not all the internal factor groups were included in the final statistical analysis due to interactions between them.

3.4.1. Type of antecedent

The semantic features of NP referents that have been proved to affect the choice of an anaphoric form in the discourse, namely animacy, specificity and definiteness, cannot easily be extrapolated to the analysis of propositional anaphora. The factor group "type of antecedent" was included in the analysis as a way to capture a semantic feature that differentiates entities potentially referred to by the propositional lo and the null pronoun under consideration in this dissertation.

The idea that different types of antecedents could have different effects in the use of the two variants is also sustained by the intuition of Mexican speakers and my own observations of the corpora that there is a difference between the behavior of pronouns that have an interrogative (direct or indirect) clause as the antecedent and pronouns that have a declarative sentence as the antecedent. The factors included under this factor group were:
Declarative main sentence

(3.27) Luego los dos andaban con la misma chamaca; o sea que los dos andaban con la... el charro y el militar. Entonces, este... el padre Ø supo (Mex Pop)
Then both were dating the same girl; I mean that both were dating the... the charro and the soldier. Then, the father knew Ø

Complement clause (with "que")

(3.28) “¿Qué no le había dicho que quería que viniera un maestro?” "No me Ø había dicho, Joaquín.” (Mex Cult)
“Hadn't he tell you that he wanted a teacher to come?” “He had not told me Ø, Joaquin’

Main wh- interrogative sentence

(3.29) "Madres ¿dónde está mi hija?” "No lo sabemos.” "¿Se habrá muerto, madre?”"No lo sabemos”. (Mex Cult)
“Women, where is my daughter?” “We don't know LO.” “Could she have died, woman?” “We don't know LO”

Main yes/no interrogative sentence

(3.30) "Madres ¿dónde está mi hija?” "No lo sabemos.” "¿Se habrá muerto, madre?” "No lo sabemos”. (Mex Cult)
“Mothers, where is my daughter?” “We don't know LO.” “Could she have died, mother?” “We don't know LO”

Sustantivized relative sentence (lo que...)

(3.31) -Pero que nosotros tampoco les vamos a dar cien días. Vamos a decir lo que nos parezca desde hoy.
-But we are not going to give them one hundred days. We are going to say what we think starting today
- we are already saying LO
Indirect wh-question

(3.32) Inf. -Pero yo no sabía dónde vivías. Ahora ya Ø sé.
   Enc. - ¡Ah, sí Ø sabías! (Mex Pop)
   ‘Inf. -But I didn't know where you lived. Now I know Ø.
   Enc. - Yes, you knew Ø!’

Indirect si interrogative

(3.33) - ¿Sabes si hay algún servicio por aquí?
   - No, pero se lo preguntamos. (Mad Cult)
   DO you know if there is a restroom here?
   I don't know, but we can ask him LO

3.4.2. Prior form and grammatical role of the antecedent

Meyerhoff (2002), analyzing the null direct object is Bislama, considers the prior form (overt or null) and the role (subject, object) of the antecedent as it was mentioned in the last occurrence and concludes that when the antecedent was an object, the pronoun was more likely to be null. She also notes that if the antecedent of a direct object is null, there is a strong probability that the object under investigation is also realized as null (Meyerhoff 2002:331).

Iliescu (1988), based on her native speaker intuitions but without a quantitative analysis, observes a similar tendency in Romanian: it is frequent to find a parallel syntactic repetition, in such a way that if the clitic or the propositional pronoun are used in a sentence, the following sentence is very likely to contain the same anaphoric pronoun just used before.
Landa and Franco’s (1992) analysis of the object anaphor in the Spanish spoken in the Basque Country suggests that the form of the antecedent (whether it is coded as a clause or as a null pronoun) affects the form of the direct object pronoun (overt or null). In their study, they first show that the null pronoun is favored when the antecedent is a clause or verbal phrase. However, when this kind of antecedent is combined with antecedents coded as pronouns (todo, eso) to measure the effect of the gender of the antecedent (clauses and neuter pronouns are included under “neuter”), the results change and the null direct object is no longer favored. This suggests that different forms of the antecedent (clause vs. pronominal) may favor different codifications of the direct object.

In order to see whether the prior form and grammatical role of the antecedent affect the use of the propositional lo and the null pronoun, I coded for the following factors under this factor group. In each of the examples, the target propositional DO pronoun is underlined, and the last mention of its antecedent is marked in bold and italic:

**Overt subject demonstrative**

(3.34) A - ¿Por qué ha subido tanto la vida? ¿Por qué se imagina usted?
   B - Bueno, ps... si yo estuviera preparado, yo se lo... resolvía. Pero ps, como no estoy preparado, ora... ps eso... ps es cosa del gobierno, para mí, ¿no? ¿Que cómo?
   [Risas]... Si yo Ø supiera, bu yo le diría, ¿no? (Mex Pop)
   A- Why has life risen so much? Why do you imagine?
   B- Well... if I were qualified, I would solve that. But, since I am not qualified... that... is the government's business, in my opinion, right? How? [laughs] If I knew Ø, I would tell you

**Overt subject clause**

(3.35) Resulta que la señora esta vive aquí. Yo no lo sabía. (Alcore)
   The thing is that this woman lives here. I didn't know LO
Overt object pronoun

(3.36) que decían el gerente o, o los abajito del gerente, el jefe de personal o los ingenieros decían: yo quisiera que todos los trabajadores de Troqueles fueran como X, y pos si to’os fueran como X, dijo, Troqueles estuviera allá... bien alto, y eso me lo dijo una vez el gerente... general, y... ingenieros más abajito del gerente me lo dieron muchas ve’s a mí (Monterrey) that the manager, or the people below the manager, the personnel manager, or the engineers, would say: I would like ll the employees in Troqueles to be like X; if they all were like X, he said, Troqueles would be there, very high, and that a general manager told me LO once, and those below the manager told me LO many times

Overt object demonstrative

(3.37) -¡Ah! pero me van a quitar el colegio; no me diga eso porque me enfado. -No lo sé pero ese edificio parece que está ahí un poco condenado. (Mad Cult) - Ah!, but they are not taking away the school; don't tell me that because I get mad - I don't know LO but this building seems to be there, forced to dissapear

Overt object clause

(3.38) Bueno, a ver, Doris, si quieres tanto a Joaquín y él te quiere y yo veo que hablan y que tienen... este... pues un intercambio en sus gustos, en sus ideas, cual debe ser, si a mí no me dices qué te pasa y qué tienes, ¿por qué no le Ø dices a Joaquín, si le tienes más confianza?" (Mex Cult) Let's see, Doris, if you love Joaquin so much, and he loves you, and I see that you talk and that you have... an exchange of your tastes, your ideas, the way it should be, if you don't tell me what happens to you and what you have, why don't you tell Ø Joaquin, if you are on closer terms with him?

Null subject

(3.39) E: ¿Y por qué cree usted, así..., que se haya suscitado esto alza de precios y... nuestras deficiencias en el, en el dinero que no alcanzamos y que nunca vamos a completar? I: Pues Ø es bien complejo, realmente no Ø entiendo pero m... tiene mucho que ver el... lo que producimos, lo que exportamos, lo que importamos (Monterrey) E: And why do you think that this rise of prices has taken place, and our shortcomings financially, that we don't get by, and that we're never going to save
up enough?
I: That's very complex, I really don't understand Ø but, what we produce, what we export, what we import has a lot to do.

**Null object**

(3.40) A- ¿Por qué ha subido tanto la vida? ¿Por qué se imagina usted?
B- Inf. -Bueno, ps... si yo estuviera preparado, yo se lo... resolvía. Pero ps, como no estoy preparado, ora... ps eso... ps es cosa del gobierno, para mí, ¿no? ¿Que cómo? [Risas]... Si yo supiera , bu (pues) yo le Ø diría, ¿no? (Mex Pop)
A- Why has life risen so much? Why do you imagine?
B- Well... if I were qualified, I would solve that. But, since I am not qualified... that... is the government's business, in my opinion, right? How? [laughs] If I knew Ø, I would tell you Ø.

**Main/independent clause**

(3.41) A- No hacías, no hacías más que haberte marchado, cuando llama don X.
B- *Llamé el otro día allí.*
A- Ya nos Ø ha dicho, sí, sí. (Mad Cult)
A- You had just left when X called
B- I called the other day there
A- He has told Ø us, yes, yes

Finally, although it is grammatically possible, there are no examples in the corpora of an antecedent coded as an overt subject pronoun (el/lo).

### 3.4.3. Referential distance

One of the measures used in previous studies to account for the accessibility of the antecedent is "referential distance", proposed by Givon (1983) to measure topic continuity. According to Givon, the more continuous a topic, the more accessible the antecedent and, therefore, the less complex the anaphoric expression.
In addition, part of the literature reviewed in section 2.1.2. on abstract object anaphora share the idea that the antecedent of an anaphor referring to a clausally introduced entity has to be found in very recent discourse, most likely in the immediately preceding sentence (Schiffrin 1985; Dahl and Hellman 1995).

I include in the analysis the referential distance between the direct object pronoun and its antecedent with a double purpose: on the one hand, with this measure we will verify how accurate the idea that the antecedent of an abstract anaphor has to be found in the immediately previous discourse is, according to our corpora. On the other hand, if the occurrence of the null pronoun and the propositional clitic in any or both of the Spanish dialects studied here has to do with the accessibility of the antecedent, in a way comparable to the distribution of anaphors referring to first order entities, the shorter referential distance would favor the use of the null pronoun while the overt form of the pronoun would be expected when the last mention of the antecedent is more distant.

This measure was calculated counting the distance from the target anaphoric expression (here the propositional clitic or the null pronoun) to the most recent prior mention of the same referent, counting the number of clauses and establishing an upper limit of 10 clauses, a distance quite larger than what is expected to be found in propositional anaphora according to previous literature.
3.4.4. Turn

The idea that structural units of discourse could be a factor that constrains the occurrence of different anaphoric forms has been entertained by different scholars. Fox (1993) argues that the pattern of anaphora must be understood in terms of discourse units, and turns as well as the relation between turns (adjacency pairs) are perceived by speakers as marks of the sequence being open or closed, and therefore may have an effect on the choice of the anaphoric form. Schiffman (1985:163), nevertheless, examines the effect of speaker alternation in the distribution of the pronoun “it” and the demonstrative “that” with nominal reference and finds that it has no significant main effect on pronoun choice.

In Brazilian Portuguese, Schwenter and Silva (2003:124) note a pattern in the distribution of null and overt direct objects that has to do with turn change between speakers: lexical NPs were used anaphorically to refer to a referent that had only been mentioned previously by an interlocutor, not by the speaker. For Spanish, Clements (2006) suggests (based on a small sample of data from CREA) that the definite pronominalization of indefinite direct objects happens within the same speaker’s utterances in one turn in discourse, while null pronouns of such direct objects occur in utterances of different speakers.

I coded the tokens extracted as same or different turn, reflecting in which turn was the last mention to the abstract entity made.

(3.42) A- Pues le ha puesto unas inyecciones y le ha sacado radiografías, ¿verdad?
B- Sí.
A- Y no sé cuántas cosas más.
B- ¿De calcio?
A- Sí. Está muy mal el calcio. Es decir, está descalcificada.
B- Sí, ya me Ø imagino.
A- So they have given him some injections and taken X-rays, right?
B- Yes
A- And I don't know how many other things
B- of calcium?
A- Yes, the calcium is very bad. I mean, she's calcium deficient
B-, Yes, I imagine Ø

(3.43) Bueno sí, sí ha habido [accidentes], nada menos antier se ahogó un niño allá, allá, allá, ¿sí Ø sabías?
Well, yes, there have been some [accidents], the day before yesterday a kid drowned there, there, you knew Ø?

Cases like (3.42) would be coded as "different turn" because the antecedent, Está descalcificada, is mentioned in the turn of speaker A and the direct object pronoun occurs in the following turn, by speaker B. Example (3.43) illustrates a discourse in which the pronoun and its antecedents are in the same turn.

3.4.5. Number of times the proposition is referred to in the discourse

As mentioned above, it has been argued in previous studies that a sentence just introduced in the discourse (and, therefore, only referred to once) only has the status “activated”, and cannot be referred to with a pronominal form (or a null pronoun), since such a form requires that the antecedent have the status “in focus” in Gundel et al’s (1993) scale (Borthen, Fretheim and Gundel 2003; Gundel, Hegarty and Borthen 2003; Hegarty 2003). According to these scholars, for a higher-order entity to be “in focus” it has to be referred to at least twice in the discourse. In order to incorporate and test this intuition in my analysis, I included this factor group, coding for the number of times that
the proposition is referred to in the previous discourse. I established an arbitrary limit of a maximum of 10 times the proposition is referred to in the discourse and any number greater than this, would be coded as 10. The coding of this factor group is exemplified with (3.44) and (3.45)

(3.44) A- *El lunes ya clase normal, ¿no?, según lo que decían...*
B- Bueno, *eso* dicen, pero yo no lo sé seguro; (Mad Cult)
A- On Monday [we have] normal class, right? according to what they said
B- Well, so they say, but I don't know LO for sure;

(3.45) A- *Yo llevo una vida limpia de pensamiento, de cuerpo, de todo. ¿Por qué? Para merecer el respeto y cariño de mis hijos.*
B- Absoluta... bueno, sí lo entienden.
A- Sí, sí lo entienden.
B- - Digo, sí lo entienden; (Mex Cult)
A- I lead a clean life, spiritually, physically, in every sense. Why? To earn the respect and affection of my children.
B: Absolut... well, if they LO understand
A: Yes, yes, they LO understand
B: I mean, yes, they LO understand;

In (3.44), the target pronoun, *lo* in *no lo sé*, refers to the sentence "el lunes ya [hay] clase normal", and this proposition is referred to twice in the discourse: by the whole sentence, "el lunes ya [hay] clase normal" and by the demonstrative pronoun *eso*. In example (3.45), the target pronoun is the clitic *lo* in the last turn (*Sí, sí lo entienden*), and its referent is mentioned three times in the previous discourse: first by means of a complete sentence, and the second and third times by means of a propositional *lo* in two consecutive speech turns, marked in bold and italic in the example.
3.4.6. Verb tense/aspect

Observing the use of null objects (referring to first order entities) in the Spanish of River Plate, Masullo (2003) points out that the implicit object has to be anchored to speech time, which explains why null objects can only appear with punctual aspect (as opposed to progressive or habitual) and present or immediate past tenses in that dialect. In other studies, like Palacios Alcaine (2000) on null objects in Paraguayan Spanish, the differences in tense and aspect are said to be irrelevant. These observations led me to include this factor group in the quantitative analysis in order to test what effect, if any, will the aspect and tense of the verb have on the variation between the propositional clitic *lo* and the zero form. The following factors were included in this factor group:

- present indicative
- imperfect indicative
- preterite
- future (perifrastic and morphological)
- conditional
- present perfect
- past perfect (pluscuamperfecto)
- future perfect
- present subjunctive
- imperfect subjunctive
- present perfect subjunctive
- past perfect (pluscuamperfecto) subjunctive
- progressive periphrasis
- infinitive (antes de decirte, para qué decir...)
- imperative
3.4.7. Verb person

Noailly (1997), observing zero anaphora in French, posits that the object of verbs conjugated in second person (present tense) are more likely coded as zeros than the object of verbs with different grammatical person. In turn, in Reig and Schwenter (2007), first and second verb person were found to favor the null pronoun, while third person disfavored it. This factor group was included again with the hypothesis that there could be a difference between the behavior of objects of verbs conjugated in the first and second person and other grammatical persons, due to the role of egophoricity in discourse salience: entities that are related to the speaker and the hearer are usually more salient than entities not connected with these participants (Dahl 2000). Since salience in the context of utterance has been pointed out before as one of the pragmatic features that explain the use of null objects in other languages and other Spanish dialects (Masullo 2003; Meyerhoff 2002), the favoring effect of the first and second grammatical person may be interpreted as indicating that speakers conceive abstract entities invoked by the exchange participants as more salient (and, therefore, more accurately coded with the null pronoun) than entities not related to them.

In this factor group, for instance, example (3.46) would be coded as third person singular, and (3.47) as first person plural:

(3.46) Quieres decir que un momento antes de largarse de farra no le ha dicho: “Te repudio”. Si se lo ha dicho, ya vale (Alcore)
You mean that a moment before going to party he hasn't said to her: "I repudiate you". If he has told LO to her, that's enough.
(3.47) No, no sabemos qué tiempo... qué... qué cantidad seá la que deban. Si por eso le
digo a usted: no Ø sabemos. (Mex Pop)
No, we don't know how long... how, how much they own. Tha's why I told you:
we don't know Ø

3.4.8. Polarity

Iliescu (1988), observing the use of a null pronoun referring to abstract entities in
Romanian, argues that the null pronoun is more likely to occur in negative sentences. The
polarity of the host sentence was included in the analysis to test whether the occurrence
of the anaphor in the scope of a negation affects the anaphoric form preferred by the
speaker. Besides the accessibility of the antecedent, tested with other factors, the
anaphoric form choice could be constrained by the commitment of the speaker to the
existence (or even the truth) of the direct object: affirmative sentences would express the
commitment of the speaker to the existence of the propositional direct object and
negative predicates may have the effect of denying the existence of that referent or
manifesting the speaker’s uncertainty about its status. The occurrences that were coded
for negative polarity included these negated with no (not), as well as other negators
nunca (never), ni (nor) and jamás (never ever). The next sentence contains three instances
of propositional null object in a negative sentence, two of them negated with no and one
with nunca:

(3.48) -¿Cuánto gana más o menos, más del mínimo?
- Pos no Ø sé, fíjese que no, no Ø sé, porque nunca me, nunca me Ø quiere decir,
piensa que uno le va’ quitar todo el dinero ¿verdá? (Monterrey)
How much does he make, more than the minimum [wage]?
- I don't know Ø, I don't know Ø, beucase he never, he never wants to tell Ø me,
he thinks that I am going to take everything from him, right?
3.4.9. Sentence type

This factor group aims to test whether including the anaphoric propositional direct object in a declarative sentence, an interrogative sentence, an exclamatory sentence or a command has any effect on the choice of the anaphor used. In French, according to Noailly (1997), null objects (not only propositional null objects) are more frequent in commands or interrogative sentences. This factor group was selected in the Madrid data in Reig and Schwenter (2007) and declarative and non-declarative sentences actually showed different behaviors: the former slightly disfavored the null pronoun while the latter favored it.

Example (3.49) illustrates a declarative sentence, (3.50) exemplifies a sentence coded as interrogative and (3.51) illustrates a command.

(3.49) E: Es obrero, pero, ¿más o menos sabes qué hace?
    I: No nunca le Ø he preguntado
    E: He is a worker, but, more or less do you know what he does?
    I: No, I have never asked Ø him.

(3.50) si a mí no me dices qué te pasa y qué tienes, ¿por qué no le Ø dices a Joaquín, si le tienes más confianza?" (Mex Cult)
    if you don't tell me what happens to you and what you have, why don't you tell Ø Joaquin, if you are in closer terms with him?

(3.51) Digo: "Tú ya puedes estar atento cuando lo explique don Luis". Digo: "Y si no lo has entendido, lo dices. Don Luis, yo no lo he entendido" digo: "que él te lo va a volver a explicar. Pero díselo, no te quedes así, (Mad Cult)
    I say: "You have to pay attention when Mr Luis explains it". I say: "And if you haven't understood, you say so. Don't Luis, I haven't understood", I say, "because he is going to explain it again. But tell him LO, don't just stay like that.
3.4.10. Presence of a manner adverbial

This factor group may be a surprising choice to hypothesize as a conditioning factor in this variation. Although, to the best of my knowledge, it has never been mentioned in previous studies, in the process of coding the examples, it was notorious that the clitic lo was frequently used in Mexican Spanish when a manner adverbial was also found in the same sentence, and this observation led me to include this factor group in the analysis. Examples (3.52) and (3.53) illustrate the presence of a manner adverbial in a sentence:

(3.52) La última vez que fui, llorando, llorando, me dijo que ya se habían ido las madres de ahí, de su colonia: llorando me lo dijo, (Mex Cult)
   The last time I went, crying, crying she told me that the sisters had left the neighborhood, crying she told me LO

(3.53) -porque así le puse: "Hermosa Luna". Porque hay muchas canciones relacionadas a la luna, ¿no? Tú lo sabes perfectamente- (Mex Pop)
   because I named it "Beautiful moon". Because there are a lot of songs related to the moon, you know LO perfectly well

In these examples, llorando modifies lo dijo in (3.52), and perfectamente modifies lo sabes in (3.53). The presence of the manner adverbial does not always require an overt pronoun, but based on the results from a preliminary analysis (Reig and Schwenter 2007), it is expected to favor the propositional clitic.
3.4.11. Co-presence of the adverb *ya*

As with the presence of a manner adverbial, the inclusion of this factor group, the co-occurrence of the adverb *ya*, was exclusively based on my own observation that this adverb was very often found in the sentences under study and was spontaneously produced by speakers of both dialects in some of the examples constructed for the analysis. Moreover, in the studies that examine Spanish dialects with null objects referring to definite noun phrases, many of the examples mentioned contain the adverb *ya* (Toscano 1953 ¿Compraste el libro? –*ya* Ø compré [Did you buy the book? YA I Ø bought]; Palacios Alcaine 2000 ¿Adónde nos llevarán? -No Ø sé; dijo sin mirarlo. *Ya* te Ø dije [Where will they take us? I don't know Ø, he said without looking at him. I already YA told Ø you]; Masullo 2003 A: ¡Mozo! No me trajo el periódico. B: Le dije que *ya* le Ø traigo [A: You didn't bring me the newspaper. B: I told you that I YA bring Ø to you]; Echenique 1986 ¿Has cogido el jersey? –Sí, *ya* Ø he cogido [Did you take the sweater? Yes, I YA took Ø]). None of these authors, however, noticed the very frequent co-occurrence of this adverb. Example (3.54) illustrates a propositional direct object co-occurring with the adverb *ya* in the corpora:

(3.54) ¿Qué te han traído allí los Reyes? ¿*Ya* lo sabes?

What have the Wise Men brought to you there? Do you already know Ø?

This factor group was selected as significant in Reig and Schwenter (2007), with the effect of favoring the use of the null pronoun in both Peninsular and Mexican Spanish.
3.4.12. Presence or absence of a dative pronoun

Several of the scholars that analyze null objects in different dialects of Spanish and also in French consider that the absence of the overt clitic direct object is exclusive or more frequent when it co-occurs with a dative pronoun (Kany 1945; Landa 1995; Solé and Solé 1977; Suñer and Yépez 1988; Grevisse 1993; Lambrecht and Lemoine 1996). As noted before, the co-occurrence of a dative pronoun is clearly not necessary for the null pronoun in Mexican or Peninsular Spanish, since 4 out of 10 verbs found with null objects in the corpora are intransitive. Besides this fact, I wanted to reflect in my coding the fact that a sentence can lack a dative pronoun because the verb is a monotransitive verb and the dative pronoun is, thus, impossible, or because a potentially ditransitive verb does not actually present a dative pronoun in that particular sentence. To account for this very important linguistic difference, I coded not only for the presence or absence of a dative pronoun, but also for the monotransitivity or ditransitivity of the verb, and included both linguistic features in one factor group. Coded as monotransitive verbs were entender, comprender, saber and imaginar, and coded as ditransitives were decir, contar, explicar, preguntar, agradecer and platicar. I also differentiated between instances where only the dative pronoun was present and sentences where the dative pronoun was duplicated (although both factors were later collapsed in the analysis). The resulting factors are:

- presence of a dative pronoun (with a ditransitive verb)

(3.55) Si no le pareció al papá, que me lo diga después
If the father doesn't agree, he can tell me LO later
- presence of a duplicated dative (with a ditransitive verb)

(3.56) mi hermana para enterarse fue hace tiempo y me dijo que para una boda en junio, que como en junio es una época que hay bastantes bodas, que en vez de ir dos días antes o una semana antes, que es lo normal, que fuéramos antes, o sea a primeros de junio o finales de mayo; entonces yo cogí y se lo dije a Andrés. (Corec)

My sister went to find out some time ago and she told me that for a wedding in June, since June is a time where there are several weddings, instead of going two days before or one week before, which is normal, we should go before, at the beginning of June or at the end of May; so I told LO Andres

- lack of a dative pronoun with a ditransitive verb

(3.57) le ‘ije mira, para que se den cuenta que tú no trabajas/no hay necesidad que yo lo diga. (Monterrey)

I told him, look for them to realize that you don't work it is not necessary that I say LO, the boss knows it already.

- lack of a dative pronoun because of the monotransitivity of the verb

(3.58) cuando te casas y tienes hijos tu vida profesional es un se, es, pasa totalmente a segundo plano. Y ya está, eso ya lo sabemos. Porque no hay nada más importante que un hijo,

when you get married and have children your professional life is a, is, takes a back seat. And that's it, we already know LO. Because there is nothing more important than a child.

Note that presence/absence of a dative pronoun and mono/ditransitivity of the verb could not be included as two different factor groups because they were not orthogonal: all the cases of monotransitive verbs were also cases of "lack of dative pronoun".

Finally, I also coded for corpus and individual speaker, and for three external factors: age, sex and education (only in Mexico). I decided to allow for as many distinctions as possible in the coding system, which provides maximal flexibility in re-
coding the factor groups for different criteria in the quantitative analysis (Tagliamonte 2006: 116).

3.5. Statistical Analysis

Once the envelope of variation is delimited and the factor groups and factors within them are established, the coded data were analyzed using the Windows application of Goldvarb X (Sankoff, Tagliamonte, and Smith 2005).

Varbrul is an analytical tool that calculates logistic regressions, designed to handle the kind of data obtained in studies of variation\textsuperscript{16} (Bayley 2002:24). These data generally consist on badly distributed cells of language in use and many crosscutting factors that play a role in the distribution of the data. Varbrul allows researchers to consider simultaneously the statistical effect of numerous factors (independent variables) on a speaker's choice to use one linguistic form (variant) where other(s) are available. The null hypothesis in a variationist study is that none of the factors considered has any systematic effect on the choice of the dependent variable and that any differences in the choice of the speakers among the various contexts is to be attributed to statistical fluctuation (Tagliamonte 2006; Sankoff 1988 for more details). The multivariate analysis reveals whether a particular factor group has a statistically significant effect on the variation, as well as how much of an effect it had relative to other factor groups and, finally, within each selected factor group, whether the particular factors (linguistic contexts) favored or

\textsuperscript{16} Paolillo (2002) points out that the fact that Varbrul was created when statistical packages typically did not include logistic regression routines, together with the way it reports the results (transforming components of the model back into proportion-like probabilities), which is more user-friendly for linguists, and the fact that the program is free, explain the tradition of using Varbrul in variationist linguistics.
disfavored the choice of one of the variants analyzed.

3.6. Combining quantitative and qualitative analyses

This dissertation combines two linguistic disciplines that are not frequently found together: variationist linguistics and pragmatics. Both approaches can be seen as representing two different conceptions of what linguistic data should consist of. For variationist linguistics, the object of study is necessarily naturally occurring data, extracted from a corpus or elicited by means of sociolinguistic interviews or questionnaires. This discipline is among the linguistic subfields that share the idea that linguistic theories should be based on spontaneously-produced linguistic material.

The methodological approach usually linked to pragmatic studies is quite the opposite: Pragmatics, although described in general as the use of language in context, mostly utilizes constructed examples. It shares with other sub-fields such as syntax and semantics the assumption that native speakers (and, therefore, the linguist himself) have a sufficient knowledge of their language as to make the appropriate judgments about the availability of a specific construction in a language or the interpretation that a specific utterance can have. The use of constructed examples and speaker intuitions allows the pragmaticist to control over the linguistic and contextual factors that govern the interpretations available for a linguistic item and, in that way, tease apart subtle details that are difficult or impossible to control in naturally occurring data.

\[\text{17 A good example of a variationist pragmatics study is Torres Cacoulllos and Schwenter (forthcoming)\}]

105
In this dissertation, I will combine the quantitative analysis, characteristic of a variationist linguistics, with more qualitative observations and I will complement the corpus based data with constructed examples. The main part of this dissertation consists of a variationist study, and, as such, it relies on the quantitative analysis of real occurring speech. In variationist linguistics, the statistical analysis provides the linguist with quantitative data that must be, later, linguistically interpreted. Although real occurring speech is the main source of data in this dissertation, the observation of corpus-based data presents some limitations: a specific construction can be missing in the corpus, or its occurrence may be too low to obtain statistically significant generalizations. For these reasons, I will complement the naturally occurring data with constructed examples, which will allow me to observe the subtleties not found in the corpus and to control for details in the discourse and context of utterance that are out of the linguist control when working with corpus sources.

3.7. Conclusions

In this chapter, I explained the methodology that will be used in this study, mainly a variationist linguistic investigation in which qualitative observations will be used to complement the data not found (or not with enough frequency) in the corpora.

The envelope of variation between the null object and the propositional lo was defined inductively, and it comprises the referential anaphoric objects whose referent is a proposition or clausally introduced entity.
Excluded were objects whose referent was not clearly identified in the discourse, syntactic constructions such as cataphora and topicalization or left dislocation, impersonal or passive *se* sentences and null complement anaphora, as well as idiomatic expressions. Lastly, the coding scheme employed in the dissertation was explained and exemplified in the last part of the chapter.
CHAPTER 4

COMPARISON OF LINGUISTIC CONSTRAINTS: SPAIN AND MÉXICO

In this chapter, I will present the results of the quantitative analysis of the coded tokens. In section 4.1. I will offer the general results of the multivariate analysis of the Peninsular and Mexican data. The following sections will be devoted to closely observe the results of the different factor groups selected, starting with the factor groups selected in both dialects (sections 4.2, 4.3, and 4.4) and continuing with the factor groups selected only in Mexican Spanish (section 4.5). Further discussion of these results and their theoretical implications will be developed in chapter 5.

4.1. Results of the statistical analyses

Presented in Table 4.1 is the overall distribution of the variants in Mexican and Peninsular Spanish:
Anaphoric direct objects referring to propositions are realized as the "canonical" clitic lo 70% of the time in Spain but only 17% of the time in Mexico, while the null variant is used 30% of the time in Spain and 70% of the time in Mexican Spanish. The results corroborate the hypothesis previously stated above, that the Mexican dialect would show a significantly higher rate of null DOs than Peninsular Spanish: as expected, there is a highly significant difference between the dialects and, in fact, the rate of null DOs in Mexican Spanish is even higher than expected. The fact that the phenomenon of null DOs with propositional antecedents in Mexican, Peninsular or other varieties of Spanish is not mentioned in previous dialect studies, with the sole exception of Kany (1945), who nevertheless doesn't distinguish between different kinds of DO referents, makes the occurrence of null objects in Peninsular Spanish and, especially, the high overall frequency of null DOs in Mexican Spanish rather surprising.

One interesting question to consider is how much of the null/lo variation is lexically constrained, and how much of it is grammatically productive. This variation seems to be lexically constrained in the sense that only a small number of verbs admit the null pronoun, according to the corpora analyzed, and not all the verbs are equally represented in the data.

<table>
<thead>
<tr>
<th></th>
<th>Overt lo</th>
<th>Null DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spain (656)</td>
<td>70% (459)</td>
<td>30% (197)</td>
</tr>
<tr>
<td>Mexico (669)</td>
<td>17% (113)</td>
<td>83% (556)</td>
</tr>
</tbody>
</table>

Chi-square = 380.359  \( p < .0001 \)  \( df = 1 \)

Table 4.1. Overall frequency of “lo” and null DO
Figure 4.1. Distribution of verbs in Peninsular and Mexican data

Figure 4.1. shows the remarkable skewing of the distribution of verbs in the corpora: two verbs, *saber* and *decir*, are much more frequent than the remaining verbs and together account for 85% of the data in Spain and 86% in Mexico (Spain: *saber* 51%; *decir* 34%. Mexico: *saber* 45%; *decir* 41%) and the remaining 15% in Mexico and 14% in Spain is distributed over seven or eight verbs. A Varbrul analysis was run with factor group Verb that included as its factors the verb *saber*, the verb *decir* and other verbs, and in neither of the two dialects was this factor group selected as significant. The frequency of the verbs and the limited number of verbs identified to accept both variants suggests that the null/lo variation is not completely productive in either dialect, since it seems to be the case that some cognition and communication verbs do not accept both variants (for example, *rogar*, *asumir*, *sugerir*, or *decidir*).18

18Interestingly, the same is true in the case of Null Complement Anaphora: some verbs accept this null anaphor and verbs that are semantically close do not (Brucart 1999).
In order to compare the similarities and differences of the variation between the propositional pronoun and the null object in Mexican and Peninsular Spanish, two different multivariate analyses, one for each of the dialects, were conducted. Although, as described in chapter 3, the coding scheme first designed was very meticulous and included many internal constraints, several factor groups were later excluded from subsequent analyses due to crossovers in the data (explained in the corresponding sections), and groups were recoded in order to capture more general tendencies in the data. I will explain these recodes in the sections devoted to each factor group.

Tables 4.2, 4.3 and 4.4 show the results of the multivariate analyses of the data from Mexico and Spain. In Table 4.2 and similar tables in this chapter, each factor group is presented in the leftmost column (in caps), and the individual factor values for each group are presented immediately below. The factor groups are presented in decreasing order of strength (as indicated by the Range between the highest and lowest factor values for each group). Also included are the percentages of null DOs for each factor value, the total number of tokens per factor value, and the percentage of the data represented by each value. Statistically insignificant factor groups are not included in the tables.
<table>
<thead>
<tr>
<th>TYPE OF DATIVE PRONOUN</th>
<th>PROBABILITY</th>
<th>% Ø</th>
<th>TOTAL N</th>
<th>% OF DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monotransitivo</td>
<td>.61</td>
<td>44%</td>
<td>368</td>
<td>56%</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.54</td>
<td>19%</td>
<td>166</td>
<td>25%</td>
</tr>
<tr>
<td>Ditransitive w/o dp</td>
<td>.17</td>
<td>4%</td>
<td>122</td>
<td>19%</td>
</tr>
</tbody>
</table>

Range 44

<table>
<thead>
<tr>
<th>TYPE OF ANTECEDENT</th>
<th>PROBABILITY</th>
<th>% Ø</th>
<th>TOTAL N</th>
<th>% OF DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogative</td>
<td>.66</td>
<td>54%</td>
<td>220</td>
<td>33.5%</td>
</tr>
<tr>
<td>Declarative</td>
<td>.41</td>
<td>18%</td>
<td>436</td>
<td>66.5%</td>
</tr>
</tbody>
</table>

Range 25

<table>
<thead>
<tr>
<th>POLARITY</th>
<th>PROBABILITY</th>
<th>% Ø</th>
<th>TOTAL N</th>
<th>% OF DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>.64</td>
<td>50%</td>
<td>285</td>
<td>43%</td>
</tr>
<tr>
<td>Affirmative</td>
<td>.39</td>
<td>14%</td>
<td>371</td>
<td>57%</td>
</tr>
</tbody>
</table>

Range 25

Total N = 656, Input: 0.235 (30% null), Log likelihood = -320.430, p = 0.00

Factor groups not selected: Referential distance, Verb tense, Verb person, Sentence type, Manner adverbial, Age and Sex.

Table 4.2. Factors contributing to the choice of the null DO in Peninsular Spanish
<table>
<thead>
<tr>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATIVE PRONOUN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotransitive</td>
<td>.55</td>
<td>89%</td>
<td>338</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.54</td>
<td>84%</td>
<td>268</td>
</tr>
<tr>
<td>Ditransitive w/o dp</td>
<td>.15</td>
<td>48%</td>
<td>63</td>
</tr>
<tr>
<td><strong>Range 40</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MANNER ADVERBIAL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>.53</td>
<td>86%</td>
<td>623</td>
</tr>
<tr>
<td>Present</td>
<td>.13</td>
<td>37%</td>
<td>46</td>
</tr>
<tr>
<td><strong>Range 40</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TYPE OF ANTECEDENT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrogative</td>
<td>.67</td>
<td>95%</td>
<td>368</td>
</tr>
<tr>
<td>Declarative</td>
<td>.30</td>
<td>68%</td>
<td>301</td>
</tr>
<tr>
<td><strong>Range 37</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SENTENCE TYPE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.72</td>
<td>91%</td>
<td>81</td>
</tr>
<tr>
<td>Declarative</td>
<td>.47</td>
<td>82%</td>
<td>588</td>
</tr>
<tr>
<td><strong>Range 25</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>POLARITY</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>.59</td>
<td>93%</td>
<td>357</td>
</tr>
<tr>
<td>Affirmative</td>
<td>.40</td>
<td>72%</td>
<td>312</td>
</tr>
<tr>
<td><strong>Range 19</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total N = 669, Input: 0.91 (83.1 % null), Log Likelihood = -205.545, p = .078
Factor Groups not selected: Referential distance, Verb tense, Verb person.

**Table 4.3. Internal factors contributing to the choice of the null DO in Mexican Spanish**
<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>.61</td>
<td>88%</td>
<td>391</td>
<td>60%</td>
</tr>
<tr>
<td>Men</td>
<td>.34</td>
<td>75%</td>
<td>265</td>
<td>40%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 27</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-34</td>
<td>.62</td>
<td>87%</td>
<td>303</td>
<td>46%</td>
</tr>
<tr>
<td>55-</td>
<td>.45</td>
<td>82%</td>
<td>123</td>
<td>19%</td>
</tr>
<tr>
<td>35-54</td>
<td>.37</td>
<td>78%</td>
<td>229</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 25</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUCATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>.58</td>
<td>87%</td>
<td>369</td>
<td>55%</td>
</tr>
<tr>
<td>High</td>
<td>.40</td>
<td>78%</td>
<td>299</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 18</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4. Social factors contributing to the choice of the null DO in Mexican Spanish

Although the internal and external factor groups selected as significant in Mexico are presented in different tables\(^{19}\), the analyses whose results are presented in Tables 4.2, 4.3 and 4.4, included both internal and external factor groups, in both dialects.\(^{20}\) I present here the linguistic and social factor groups in two different tables for the sake of simplicity in my exposition, since, as previously explained, I will focus my analysis and discussion on the linguistic data and won't refer to the external data until section 6.4.

\(^{19}\) *Peninsular Spanish: non of the social FGs were selected as significant. The weights are: Sex: women: .49; men: .50, and Age: -34: .51; 35-54: .50; 55+: .45

\(^{20}\) It is noted in the literature that an analysis including both social and linguistic factors will usually differ little from one in which both internal and external variables are considered separately (Poplack 1997: 304). In this case, when social and linguistic factors were analyzed in two different runs, the results differ in the following way: in Mexico, the factor group Polarity was no longer selected and significant and Verb person was now selected; in Spain, the results of the linguistic factors are almost identical but Age is also selected as significant. External and internal factor groups, in this case, are not completely orthogonal.
Much has to be said about these results and they will be analyzed in detail in the following pages. To offer a first evaluation of the data just presented, it should be noted that the main difference between both analyses, besides the overall frequency of null objects in both dialects, is the fact that more factor groups were selected as significant in the analysis of Mexican Spanish than in Peninsular Spanish: only three linguistic factor groups were selected in the Spain data and five internal and three external factor groups are selected in the analysis of the Mexican data.

In spite of these differences, the results of the multivariate analyses show very interesting similarities between both dialects. Taking into consideration only the linguistic constraints, the statistical analyses show similarities that are easily appreciated in Table 4.5.\footnote{Not included in the table and not selected as significant in the analyses were the following factor groups: Referential distance: Spain: 1 clause: .51, more than 1 clause: .43; Mexico: 1 clause: .49; more than 1 clause: .53. Verb tense: Spain: present: .54; subjunctive: .50; preterite: .44; imperfect: .35; future/conditional: .21; Mexico: present: .45; subjunctive: .47; preterite: .61; imperfect: .72; future/conditional: .47. Verb person: Spain: 1\textsuperscript{st} person: .52; 2\textsuperscript{nd} person: .54; other: .43; Mexico: 1\textsuperscript{st} person: .55; 2\textsuperscript{nd}: .49; other: .36.}
<table>
<thead>
<tr>
<th></th>
<th><strong>MEXICO</strong></th>
<th><strong>SPAIN</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATIVE PRONOUN</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotransitivo</td>
<td>.55</td>
<td>.61</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.54</td>
<td>.54</td>
</tr>
<tr>
<td>Ditransitive w/o dp</td>
<td>.15</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td><strong>Range 40</strong></td>
<td><strong>Range 44</strong></td>
</tr>
<tr>
<td><strong>MANNER ADVERBIAL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>.53</td>
<td>[.51]</td>
</tr>
<tr>
<td>Present</td>
<td>.13</td>
<td>[.38]</td>
</tr>
<tr>
<td></td>
<td><strong>Range 40</strong></td>
<td></td>
</tr>
<tr>
<td><strong>TYPE OF ANTECEDENT</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.67</td>
<td>.66</td>
</tr>
<tr>
<td>Declarative</td>
<td>.30</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td><strong>Range 37</strong></td>
<td><strong>Range 25</strong></td>
</tr>
<tr>
<td><strong>SENTENCE TYPE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.72</td>
<td>[.55]</td>
</tr>
<tr>
<td>Declarative</td>
<td>.47</td>
<td>[.49]</td>
</tr>
<tr>
<td></td>
<td><strong>Range 25</strong></td>
<td></td>
</tr>
<tr>
<td><strong>POLARITY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>.59</td>
<td>.64</td>
</tr>
<tr>
<td>Affirmative</td>
<td>.40</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td><strong>Range 19</strong></td>
<td><strong>Range 25</strong></td>
</tr>
</tbody>
</table>

Table 4.5. Internal factors contributing to the choice of the null DO in Mexican and Peninsular Spanish

This table shows that the variation under study responds to similar patterns in both dialects: three of the factor groups selected as significant at the .05 level coincide, which means that the factor groups that are playing a role in the distribution of the two variants in Peninsular Spanish also have an effect in the variation in Mexico. Moreover,
the relative strength of these factor groups is also similar: the factor group Dative pronoun is most significant (shows the largest range), Sentence type follows and Polarity is the least significant (smallest range). And, finally, the table shows that the constraint hierarchy within each of these factor groups is also parallel in both dialects, that is to say, the linguistic factors included in each of the groups have the same effect on the choice of the null object and these factors are ordered in the same way.\footnote{Since the overall frequency of the variants is so different in Mexican and Peninsular Spanish, I conducted a Varbrul one step analysis, which analyzes the data according to the combined effect of the corrected mean and the factor weight (Poplack and Tagliamonte 1996). This analysis permits to compare the probability that the null pronoun will occur in a given context while controlling for its frequency of occurrence in that context. The results, however, were not very informative: in Peninsular Spanish, the combined Input and Weight for the null object is .35 in the most favorable contexts (interrogative antecedents), and in the Mexican data, the combined Input and Weight was always well above .50, and in all except two factors, above .80.}

I will first address the results of the factor groups that are selected as significant in both dialects (sections 4.2, 4.3 and 4.4) and later the factor groups that are only selected in Mexican Spanish (section 4.5).

4.2. Co-occurrence with a dative pronoun

As noted in section 3.4.12., several of the scholars who refer to the use of null objects (not necessarily with propositional antecedents) in different dialects of Spanish consider that the absence of the overt direct object pronoun is more frequent when there is a dative pronoun in the sentence (Kany 1945; Landa 1995; Solé and Solé 1977; Suñer and Yépez 1988). Kany suggests that null objects in American dialects of Spanish occur with dative pronouns because of a tendency to avoid double clitics.

To directly test Kany and other's intuitions, let's observe in first place whether the null pronoun is more frequent when there is a co-occurring dative pronoun:
The data presented in Tables 4.6 and 4.7 already show that the effect of the dative pronoun does not have the importance that Kany and others attributed to it: in Mexican Spanish, the null object is much more frequent than the *lo* pronoun with or without a dative pronoun, and the difference between the rates of null pronoun objects with and without a dative pronoun is very small (84% and 82% respectively); thus, the tendency to avoid double clitics does not explain the distribution of the variants. In fact, based on these numbers, in Peninsular Spanish the effect would be the contrary: the rate of null pronouns is higher when there is not a dative pronoun (39%) than when there is one (19%).

The distribution of the data in this analysis, conducted to directly test the intuition expressed in the previous literature, is, however, misleading and not very explanatory: from the observation of the linguistic data, it is obvious that the lack of a dative pronoun

<table>
<thead>
<tr>
<th></th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>dative pronoun</td>
<td>84% (225)</td>
<td>16% (43)</td>
</tr>
<tr>
<td>no dative pronoun</td>
<td>82% (331)</td>
<td>18% (70)</td>
</tr>
</tbody>
</table>

chi-square = 0.228  (No significant: .633)

**Table 4.6. Distribution of null pronoun and clitic *lo* according to absence or presence of dative pronoun in Mexican Spanish**

<table>
<thead>
<tr>
<th></th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>dative pronoun</td>
<td>19% (32)</td>
<td>81% (134)</td>
</tr>
<tr>
<td>no dative pronoun</td>
<td>39% (165)</td>
<td>69% (325)</td>
</tr>
</tbody>
</table>

chi-square = 12.23  p < .001

**Table 4.7. Distribution of null pronoun and clitic *lo* according to absence or presence of dative pronoun in Peninsular Spanish**
in a sentence can respond to two different situations: in some of the cases under "no dative pronoun" in Tables 4.6. and 4.7, the verb is a monotransitive verb and a dative pronoun is, thus, impossible, whereas in other cases we have ditransitive verbs that could, potentially, have taken a dative pronoun but that, in a given case, don't have it.

To account for this very important linguistic difference, I coded not only for the presence or absence of a dative pronoun, but also for the monotransitivity or ditransitivity of the verb, and included both linguistic features in one factor group. With three factors under this factor group, monotransitive verb (and, therefore, no dative pronoun), ditransitive verb with a dative pronoun, and ditransitive verb without a dative pronoun, the distribution of the two variants is as follows:

<table>
<thead>
<tr>
<th></th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ditransitive with dative pronoun</td>
<td>84% (225)</td>
<td>16% (43)</td>
</tr>
<tr>
<td>Ditransitive without dative pronoun</td>
<td>48% (30)</td>
<td>52% (33)</td>
</tr>
<tr>
<td>Monotransitive (no dative)</td>
<td>89% (301)</td>
<td>11% (37)</td>
</tr>
<tr>
<td>Total (669)</td>
<td>(556)</td>
<td>(113)</td>
</tr>
</tbody>
</table>

Table 4.8. Distribution of null pronoun and clitic lo by Cooccurrence of dative pronoun in Mexican Spanish

<table>
<thead>
<tr>
<th></th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ditransitive with dative pronoun</td>
<td>19% (32)</td>
<td>81% (134)</td>
</tr>
<tr>
<td>Ditransitive without dative pronoun</td>
<td>4% (5)</td>
<td>96% (117)</td>
</tr>
<tr>
<td>Monotransitive (no dative)</td>
<td>44% (160)</td>
<td>56% (208)</td>
</tr>
<tr>
<td>Total (656)</td>
<td>(197)</td>
<td>(458)</td>
</tr>
</tbody>
</table>

Table 4.9. Distribution of null pronoun and clitic lo by Cooccurrence of dative pronoun in Peninsular Spanish

---

23 Monotransitivity/ditransitivity of the verb and co-occurrence of a dative pronoun could not be included as two different factor groups because they were not orthogonal: all the cases of monotransitive verbs were also cases of lack of dative pronoun.
Observing the frequencies represented in these tables, it becomes clear that the distinction between lack of dative pronoun with monotransitive and ditransitive verbs was a necessary one: the percentages of null pronouns with these two kind of verbs are very different and collapsing them in the analysis would have been inappropriate. The new results show that, in both dialects, the rate of null pronouns with ditransitive verbs without a dative pronoun (48% in Mexico and 4% in Spain) is considerably smaller than the rate of null pronouns in the other two cases (monotransitive verbs and ditransitive verbs with a dative pronoun). The distribution of the variants in the Peninsular corpora also reflects a clear difference between the monotransitive verbs and ditransitives with a dative pronoun: null objects are more frequent with monotransitive verbs (44% of the objects are null) than with ditransitive verbs with a co-occurring pronoun (19% of the objects are null). The difference between these two factor groups in the Mexican analysis is very small in terms of percentages: 89% of the objects are null with monotransitive verbs and 84% of the objects are null when a dative pronoun co-occurs in the sentence. Let's turn to the results of the multivariate analysis for this factor group, repeated here:

<table>
<thead>
<tr>
<th></th>
<th><strong>Mexico</strong></th>
<th><strong>Spain</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATIVE PRONOUN (dp)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotransitivo</td>
<td>.55</td>
<td>.61</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.54</td>
<td>.54</td>
</tr>
<tr>
<td>Ditransitive w/o dp</td>
<td>.15</td>
<td>.17</td>
</tr>
<tr>
<td><strong>Range 40</strong></td>
<td><strong>Range 44</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.10. Varbrul weights for the null pronoun by Cooccurrence of dative pronoun: Mexico and Spain
The multivariate analysis shows that, contrary to Kany's observation, the presence of a dative pronoun has almost no effect (.54) on the choice of the null pronoun in both dialects. What is very interesting is that the lack of a dative pronoun when it could have been present, i.e. with a ditransitive verb like *decir, preguntar, agradecer*, etc. highly disfavors the null pronoun according to the multivariate analysis (.15 in Mexico and .17 in Spain).

If one observed the distribution of the two variants with ditransitive verbs in the first two rows of Tables 4.8 and 4.9 above, one could easily make sense of Kany's and other scholars' perception that the null pronoun is more frequent when there is a dative pronoun than when there is not one. The percentages of null pronouns show exactly this distribution of the variants: with ditransitive verbs, the null object occurs 48% of the time when there is no cooccurring dative pronoun and 84% when there is a dative pronoun in Mexico, and the percentages in Spain show the same tendency (4% without a dative pronoun and 19% with it). Nevertheless, the statistical analysis reveals that it is not the case that the presence of a dative pronoun favors the null (by a tendency to avoid two pronouns together), but rather that the lack of a dative pronoun with verbs that accept two objects disfavors the use of the null. The tendency that needs to be explained, therefore, is not the use of the null when a dative pronoun co-occurs but rather the preference for the overt clitic *lo* when no dative pronoun is used and, therefore, a ditransitive verb would have no overt pronominal objects if the null were used. Examples (4.1) and (4.2) illustrate this tendency:
In examples (4.1) and (4.2) there is no dative pronoun depending on the verb decir, and the overt lo is preferred over the null object in this context.

The results of the multivariate analysis tell us something else: not only is the null pronoun not limited to ditransitive verbs with a dative pronoun, but, in both dialects, monotransitive verbs also favor the null (with a weight slightly higher than the weight of ditransitives with dative pronoun: .55 for Mexico and .61 for Spain). The following are examples of monotransitive verbs with null pronouns:

(4.3) - Yo me comprometo a dar a Cicerón, a César por supuesto, claro, a Cicerón...V... pues...V... Tito Livio,
- Sí, a Tito Livio ya y...
- Y ya... con...V...
- ¡Cuidado!
- Salustio con muchas más reservas...
- ¡Cuidado!
- ...y poco más, vamos.
- Sí, sí, sí yo Ø comprendo, sí. (Mad Cult)
- I undertake to teach Cicero, Cesar, of course, Ciceron, mm, Tito Livio
- Yes, Tito Livio...
- And with...
- Caution!
- Salustio, with reservations
- Caution
- and, not much more
- yes, yes, I understand Ø, yes

(4.4) Luego le decía a mi hermana: "Oyes, oyes, Agripina -dice- que... este... ¿Por dónde salió el sol, por dónde salió el sol?", dice. -"Pos quien sabe, mamá. Solamente Juana sabe Ø". (Mex Pop)

Then she would tell my sister: "listen, listen, Agripina - she said-, where did the sun rise, where did the sun rise?, she said. - Who knows, mom. Only Juana knows.

Both *comprender* and *saber* are monotransitive verbs that occur with a null pronoun in the corpora.

It is possible to go further in this analysis wondering whether the favoring effect of the monotransitive verbs in these results is the effect of the high frequency of *saber* in the corpora.

<table>
<thead>
<tr>
<th></th>
<th>SPAIN</th>
<th></th>
<th>MEXICO</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NULL</td>
<td>LO</td>
<td>NULL</td>
<td>LO</td>
</tr>
<tr>
<td>saber</td>
<td>46% (154)</td>
<td>54% (184)</td>
<td>92% (274)</td>
<td>8% (25)</td>
</tr>
<tr>
<td>imaginar</td>
<td>33% (3)</td>
<td>67% (6)</td>
<td>95% (18)</td>
<td>5% (1)</td>
</tr>
<tr>
<td>comprender</td>
<td>20% (1)</td>
<td>80% (4)</td>
<td>75% (3)</td>
<td>25% (1)</td>
</tr>
<tr>
<td>entender</td>
<td>14% (2)</td>
<td>86% (12)</td>
<td>36% (5)</td>
<td>64% (9)</td>
</tr>
</tbody>
</table>

Table 4.11. Rates of null and overt objects with monotransitive verbs: Mexico and Spain

Table 4.11 shows that, in both dialects, both *saber* and *imaginarc* have rates of null objects higher than average in each dialect, whereas the results for *entender* in both dialects and *comprender* in Spain would contradict this pattern. Unfortunately, the number of tokens of *imaginarc, comprender* and *entender* (28 in Spain and 36 in Mexico)
is much smaller than the tokens of saber (338 in Spain and 299 in Mexico), and the conclusions that can be drawn from the behavior of imaginar, comprender and entender are limited.

Some relatively frequent collocations can be found in the data of monotransitive verbs. The most relevant skewing effect, that will be analyzed with detail in the following sections, is the high percentage of null objects with no sé in both dialects, which explains the high 46% of null objects with saber in Spain. The expression ya lo sé is repeated in Spain 23 times, whereas ya sé occurs one in this dialect. In Mexico, ya sé is found 6 times and there are no cases of ya lo sé. The number of tokens in imaginar, comprender and entender is low, but it is worth mentioning that out of the 12 cases of entender with an overt pronoun in Spain, 8 (66%) of them correspond to the construction no lo entiendo, negative and first person singular present indicative. In the case of imaginar, a close look at the few tokens of this verb in the Peninsular data shows that 4/6 uses of imaginar with lo correspond to the sentence me lo imagino, and the remaining two uses of imaginar with lo, the verb is also conjugated in the first person singular, whereas the three cases of imaginar with a null pronoun are not first person singular. In Mexico, out of 18 tokens of imaginar with a null pronoun, 13 correspond to (ya) me imagino, and the only occurrence of imaginar with lo is not in first person singular. These numbers are very small to draw conclusions but it could be worth exploring if the effect of these frequent collocations persists in a larger data set.
With the same rationale, one could ask whether the strong disfavoring effect of ditransitive verbs without dative pronouns is due to the very frequent *decir*. Let's observe the distribution of null and overt pronouns with ditransitive verbs when they lack a dative pronoun:

<table>
<thead>
<tr>
<th></th>
<th>SPAIN</th>
<th></th>
<th>MEXICO</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NULL</td>
<td>LO</td>
<td>NULL</td>
<td>LO</td>
</tr>
<tr>
<td>decir</td>
<td>4% (5)</td>
<td>96% (107)</td>
<td>45% (26)</td>
<td>55% (32)</td>
</tr>
<tr>
<td>contar</td>
<td>0</td>
<td>100% (4)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>explicar</td>
<td>0</td>
<td>100% (6)</td>
<td>66% (2)</td>
<td>33% (1)</td>
</tr>
<tr>
<td>preguntar</td>
<td>-</td>
<td>-</td>
<td>100% (2)</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 4.12. Rates of null pronoun and clitic lo with ditransitive verbs without dative pronoun: Mexico and Spain.**

Table 4.12 shows the distribution of null and overt direct object with transitive verbs only when there is not a dative pronoun in the sentence. For that reason, not all the transitive verbs are included in the table (*agradecer* and *platicar* have no occurrences without a dative pronoun) and the empty cells correspond to verbs that are not found in the corpora without a dative pronoun: the corpora from Mexico do not include cases of *contar* without a dative pronoun and *preguntar*, in the Peninsular corpora, is always found with a dative pronoun. The numbers in this table show that the majority of tokens of transitive verbs without a dative pronoun correspond to the verb *decir* in both dialects (112 in Spain and 58 in Mexico). Unfortunately, as in Table 4.11, we have very few data from other ditransitive verbs with a dative pronoun to be able to test whether the same pattern, i.e. the preference for *lo* over the null pronoun when there is no dative pronoun, would occur in all ditransitive verbs. What Table 4.12 shows is that in the Peninsular
corpora, the use of *lo* with ditransitive verbs when there is no dative pronoun in the sentence is almost categorical, with only 5 uses of null pronouns, all of them with *decir*. The following are some of these 5 cases of null pronoun:

(4.5) -anda, que dije ayer que no iba - que no iba Luis Reverter a - a vicepresidencia.
- ¿Ø Dijiste? Ah, no me acordaba. (Corec)
- I said yesterday that Luis Reverter was not going to Vice presidency
- Did you say Ø? I didn't remember

(4.6) -¿Y tú crees que en Cataluña por ejemplo tienen el mismo... se sienten iguales los de Gerona y los de Barcelona, o los de Barcelona y los de Tarragona?
-Es que tú eres de Alicante y eres española. Por lo tanto, tú no te llamas madrileña porque Madrid sea la capital.
- No, pero yo Ø digo por lo que estás diciendo de la rivalidad. (Corec)
-And do you believe that in Catalonia, for example, they have the same... people from Gerona and from Barcelona feel that they are the same, or people from Barcelona and from Tarragona
-you are from Alicante and you are Spanish. Therefore, you don't say you are from Madrid just because Madrid is the capital.
- No, but I say Ø because of what you are saying about rivalry

To summarize, the quantitative analysis contradicts the observation that null pronouns are more frequent with ditransitive verbs than with monotransitive verbs and, although the frequency of null pronouns is, as observed in previous studies, higher with a co-occurring dative pronoun than without it (when the verb is ditransitive), the statistical analysis reveals that the occurrence of null objects with a dative pronoun is not much higher than average in each dialect and that, interestingly, the tendency to be explained is that the lack of a dative pronoun with ditransitive verbs disfavors the null pronoun.

It looks like, regarding their phonetically overt arguments, the verbs with propositional objects observed here reduce their valency by one: when the verb is monotransitive, the only pronoun available to be "zero" is the DO; when the verb is
ditransitive, only one of the two objects can be phonetically "zero": the DO is realized as  
zero as long as there is another object explicit in the sentence, i.e. a dative pronoun. Yet,  
if there is not a second object in the sentence (the dative pronoun), the verb seems to  
require one explicit object, to maintain at least one of its two objects explicit in the  
sentence, reducing by one its expressed arguments, but not by two. The propositional  
direct object is coded then as an overt pronoun and the null object is highly disfavored in  
this case.24

4.3. Type of antecedent: declarative vs. interrogative sentence

In both Spanish dialects, the factor group Type of antecedent was selected as  
significant in the multivariate analysis. In this analysis, the tokens were recoded as either  
interrogative (originally coded as wh- or yes/no question, and subordinate or main clause)  
or declarative sentences (originally coded as main clauses, que subordinate clauses and lo  
que subordinate clauses). Tables 4.13 and 4.14 show the distribution of the variants by  
type of antecedent:

24 This distribution of the null pronoun cannot be explained employing the opposition "primary vs  
secondary objects" (Dryer 1986): A primary object is an indirect object in a ditranitive clause or a direct  
object in a monotransitive clause, while a secondary object is a direct object in a transitive clause. If the  
object coded as null were the primary object, that would explain that monotransitive verbs accept the  
DO-null but it would predict that in ditransitive verbs, the dative pronoun would be more likely coded  
as zero, and this is not the case: when the verb semantically carries a dative object, this one is usually  
explicit and the DO (secondary object) is coded as zero in our data. In other words, if this were a  
primary-secondary object system, in which the primary object were more likely coded as zero, the  
prediction would be that, together with "Tú sabes", we would usually find "Tu lo dices" meaning "tú se  
lo dices", with a null pronoun referring to the dative object, but this is not the case, since what we  
usually find is "tú le dices", with the null object referring to the propositional direct object.
As these tables indicate, in both dialects the percentage of null objects with interrogative antecedents is higher than the percentage of null objects with declarative antecedents. In Mexican Spanish, in fact, the rate of null objects with this kind of antecedent is close to being categorical (95%). These tables also show that, although in Mexico the null object is more frequent than the \textit{lo} for both kinds of antecedents, in Spain, surprisingly, the rate of null objects is higher than the rate of overt pronouns when the antecedent is an interrogative sentence. To truly determine the effect of the kind of antecedent on the variation and to be able to compare this effect in both dialects, the results of the statistical analysis have to be taken into account. The Varbrul analysis reveals that despite the very different percentages, the type of antecedent has a very similar effect in both dialects: when the antecedent is an interrogative sentence, the null object is favored in Mexican and Peninsular Spanish (.67 and .66 respectively), and in both of them, a declarative antecedent disfavors the null pronoun (.30 in Mexico and .41
Interrogative sentences account for around 20% of the antecedents of most verbs, with the exception of *preguntar* (90% in Spain and 87% in Mexico) and *explicar* (50% in Madrid and Mexico). Both of these verbs occur in few tokens in the corpora and neither *preguntar* nor *explicar* occur with a rate of null pronouns higher than average, and the effect of the type of antecedent cannot be explained by the effect of these two lexical units. The only clear skewing lexical effect in the data is due to the expression *no sé*, that will be commented below.

I will offer an explanation for these quantitative data taking into consideration the characteristics of the proposition that is introduced into the discourse by a declarative versus an interrogative sentence, that is to say, the characteristics of the referent of the direct object pronoun. I will only sketch here this explanation and it will be developed

---

25 An example of *preguntar* with a declarative antecedent is the following (from Corec):
- Pues usted ha sido un desorejado toda su vida, ¿eh? En todos los sentidos. En todos los sentidos. Y si no, que se lo pregunten a Erique Cresson
and supported with qualitative data in the discussion in chapter 5.

A declarative sentence (with some exceptions that will be considered later) typically introduces in the discourse a complete, saturated proposition. An interrogative sentence, on the other hand, can be described as an incomplete sentence (Peterson 1997, 39): the denotation of a question has been described as an incomplete or open proposition or as a set of propositions, corresponding to the set of its possible (true) answers (Hamblin 1973; Karttunen 1977).

The antecedent of the propositional direct object being a declarative or an interrogative sentence, thus, translates into an important difference regarding the referent of the DO pronoun. I will frame this semantic difference in terms of "completeness of the proposition": pronouns that have an interrogative antecedent in the discourse refer to an incomplete, non saturated proposition: the semantic content of the proposition referred to by the DO pronoun is not complete because part of that content remains unknown, namely the part of the content corresponding to the wh- word in wh- questions or whether the proposition is true or false in yes/no questions (Escandell 1999). In contrast, when the antecedent of a pronoun is a declarative sentence, the content of the proposition referred to is typically of greater completeness (although, as we will see below, this is not always the case). Going back to the results of the quantitative analysis, in both dialects the null pronoun is favored when the antecedent is an interrogative sentence, that is to say, when the referent of that pronoun is an incomplete proposition. On the other hand, declarative antecedents in both dialects disfavor the null pronoun, which can be explained by saying that when the referent of the DO pronoun is a complete, saturated sentence, the null is
Disfavored and lo is favored. This last statement is only partially true because a declarative antecedent does not necessarily mean a complete, saturated proposition for the DO pronoun, but I will keep this simplification now and this issue will be addressed below. The following examples illustrate the tendency found in the corpora:

(4.7)  A: Es que yo quiero que venga un maestro  
B: ¿Y hasta ahorita me lo dices? (Mex Cult)  
A: I want a teacher to come  
B: And you tell me LO now?

(4.8)  - ¿Cuál es el [nombre] del calendario? ¿No me Ø dices? (Mex Pop)  
- What is the name from the calendar (saint's name)? You don't tell me Ø?

In examples (4.7) and (4.8) the antecedent of the DO pronoun is a main sentence. In (7), it is a declarative sentence (yo quiero que venga un maestro), which introduces a full proposition that is referred to in the following sentence by the propositional lo. In (4.8), the antecedent is a wh- question (¿Cuál es el nombre del calendario?), which introduces an incomplete, not saturated proposition inasmuch as we don't have the part of the proposition corresponding to the wh- word in the question, cuál. The null pronoun is preferred in this case. The same pattern occurs in the following examples, which include subordinate clauses as antecedents of the DO pronoun:

(4.9)  A: Yo es que, como no sabia dónde estaba... tenía una idea, ¿no?, pero...  
B: Sí, pero no Ø sabías exactamente (Mad Cult)  
A: Since I didn't know where it was... I had an idea, but...  
B: yes, but you didn't know Ø exactly.

(4.10) -Yo sabía que el cadáver aparecería. Yo lo sabía, porque... (Corec)  
I knew that the body would turn up. I knew LO, because...
In (4.9), the direct object pronoun of *saber* (*no sabías*) refers to an entity introduced by means of an indirect interrogative sentence (*dónde estaba*); this interrogative clause does not introduce a complete proposition, since the content corresponding to the wh-word *dónde* is not available in the discourse. I argue that, because of the incompleteness of this proposition, the pronoun chosen to refer to it is a null pronoun. On the contrary, in (4.10) a completive declarative sentence (*que* sentence) is used in the first part of the discourse which introduces a proposition referred to by the DO of *saber* in the second sentence, and, because the declarative antecedent introduces in the discourse model a full, complete proposition, the reference with *lo* is favored.

As briefly mentioned before, having a declarative antecedent typically equals having a complete, full proposition as the referent of the DO pronoun, but this is not necessarily the case. It is possible for a propositional DO pronoun to refer to an incomplete proposition that would be expressed with an interrogative sentence if it were overtly expressed in the discourse. This is the case with the direct object of the expression *no (lo) sé*: a negative sentence with the first person singular, present indicative of *saber* (to know). The construction *no sé* accepts indirect interrogative clauses as its complement but it can not take a declarative clause as its complement, as shown in (4.11)

(4.11) a. No sé cuándo vendrá María
    I don't know when Maria will come

    b. No sé si vendrá María
    I don't know if Maria will come

    c. *No sé que María vendrá el lunes
    *I don't know that Maria will come
Because of this semantic-pragmatic restriction (it is impossible to say that speakers do not know something that they are asserting), when the direct object of no sé is a pronoun, null or overt, it has to be necessarily interpreted as an interrogative clause, or, based on what I exposed previously, as an incomplete proposition, even when the sentence licensing the direct object in the discourse is a declarative sentence:

(4.12) - [la isla] Será como un cuarto de España.
- No Ø sé, de - No, no, no. Porque - de una costa a otra - o sea, es que es casi un rectángulo, pero de una costa a otra hay 90 kilómetros. (Corec)
- [the island] is probably like a quarter of Spain
- I don't know, no, no, no, because from cost to cost, I mean, it's almost a rectangle, but from cost to cost there are 90 kilometers.

(4.13) - No ésa [tienda] no es de muñecas
- No, ésa lleva quince días.
- No esa no es que - esa grande de - muñecas.
- Sí, de muñecas, sí, en la misma acera, en la misma acera.
- Sí, pero ésa ya debe llevar varios años, eh. Porque -
- Y la de "Braun" es la que es una zapatería.
- No lo sé exactamente, (Corec)
- No, that [store] is not a doll's store
- No, that one has been there for fifteen days
- No, that one is not..., that big doll store
- yes, doll store, yes, in the same side of the street
- yes, but that one must have been there for several years, because...
- and the "Braun" store is the one that is now a shoe store.
- I don't know LO exactly

In (4.12) and (4.13), the sentence that serves as the antecedent (or licensor) of the direct object of no sé is a declarative sentence (será como un cuarto de España in 4.12 and la de Braun es la que es una zapatería in 4.13), yet the referent of that direct object (null in 4.12 and lo in 4.13) is not a complete, saturated proposition but an incomplete proposition that would be coded as an indirect interrogative sentence if it were explicit in
the discourse: (4.12) could be paraphrased as *No sé que la isla es como un cuarto de España* (*I don't know that the island is about a quarter of Spain*), and not as *No sé que la isla es como un cuarto de España* (*I don't know that the island is about a quarter of Spain*), and (4.13) is necessarily interpreted as *No sé que la de Braun es la que es una zapatería* (*I don't know that the Braun one is the one that is a shoe store*) and not as *No sé que la de Braun es la que es una zapatería* (*I don't know that the Braun one is the one that is a shoe store*).

Since this is the case, and if the effect of the factor group type of antecedent is correctly interpreted as the effect of a semantic difference in the referent of the direct object, i.e. a complete vs. an incomplete proposition, we would expect that the direct object of *no sé* when the antecedent is a declarative sentence, would also be preferably coded as a null pronoun. Tables 4.17 and 4.18 show the distribution of the null pronoun and the clitic *lo* with the tokens of *no (lo) sé* (195 in Spain and 156 in México) according to type of antecedent:

<table>
<thead>
<tr>
<th>Type of Antecedent</th>
<th>Null (%)</th>
<th>Lo (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogative antecedent</td>
<td>70% (99)</td>
<td>30% (42)</td>
</tr>
<tr>
<td>Declarative antecedent</td>
<td>57% (31)</td>
<td>43% (23)</td>
</tr>
<tr>
<td>Total</td>
<td>67% (130)</td>
<td>33% (65)</td>
</tr>
</tbody>
</table>

chi-square = 2.881  p = .089 (non significant)  df = 1

*Table 4.17. Distribution of null pronoun and clitic lo in the "no (lo) sé" construction by type of antecedent in Peninsular Spanish*
Table 4.18. Distribution of null pronoun and clitic lo in the "no (lo) sé" construction by type of antecedent in Mexican Spanish

<table>
<thead>
<tr>
<th>Type of Antecedent</th>
<th>NULL (%)</th>
<th>LO (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogative</td>
<td>98% (143)</td>
<td>2% (3)</td>
</tr>
<tr>
<td>Declarative</td>
<td>100% (10)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Total</td>
<td>98% (153)</td>
<td>2% (3)</td>
</tr>
</tbody>
</table>

chi-square = 0.21   p = .64  (non significant)   df = 1

In the data from Spain, 72% of the no (lo) sé constructions have interrogative antecedents and 28% have a declarative antecedent. The percentage of no (lo) sé constructions with declarative antecedents is considerably smaller in the Mexican data, where only 6% of these constructions (N = 10) have declarative antecedents. What is of interest for our discussion is whether the rate of null objects in no (lo) sé constructions with declarative antecedents is higher than the rate of null objects with declarative antecedents in the whole corpus, which should be the case provided that the referent of the direct object of no (lo) sé is always an incomplete proposition. The data from Spain show that, although the null pronoun is more often used with interrogative antecedents, when the antecedent is a declarative sentence, the rate of null pronouns (57%) is higher than the average use of null pronouns with declarative antecedents in the Peninsular corpora (18%, as shown in Table 4.16), as expected. In the Mexican Spanish data, we only have ten occurrences of no (lo) sé with declarative antecedents and in all of them the null pronoun is the variant used and the only three cases of no lo sé in the Mexican data have interrogative antecedents.26

---

26 - Pepe, el Tecnológico me imagino que es una institución tan fundamentada, tan estricta, que no, ¿no te has hecho sentir restringido en tu actividad como profesor?
Interestingly, when the tokens of *no (lo) s sé* were excluded from the analysis, the factor group kind of antecedent was not selected as significant in Spain, but it was still significant in Mexico (Tables 4.24 and 4.25). This can be explained because, in the corpus, many of the interrogative antecedents are licensing the DO of a *no (lo) s sé* construction, and when those tokens are excluded, the number of interrogative antecedents in the Peninsular data is reduced to 79. Although the factor group is not selected as significant in this analysis of the Peninsular data without *no (lo) s sé*, the weights of the factors "declarative antecedent" and "interrogative antecedent" show the same tendency (interrogative antecedent: .61; declarative antecedent: .48). In chapter 5 I will come back to the issue of the type of antecedent and its relation to completeness of the proposition to provide qualitative evidence of the importance of this linguistic factor in the distribution of the null pronoun and the propositional *lo*.

### 4.4. Polarity of the sentence

In the multivariate analyses whose results are presented in Tables 4.2 and 4.3, the factor groups polarity of the host sentence is selected as significant in both dialects. I repeat these results here:

- *No lo sé, sí, fijate que yo nuca m’he sentido ninguna, ningún tipo de restricción, ningún tipo de, de presión por parte de las autoridades del Tecnológico (Mex Cult)*  
  Pepe, the Technological [Institute], I imagine that it's such a based institution, so strict that, haven't they made you feel restricted in your activity as a professor?
- *I don't know LO, yes, look, I have never felt any, any kind of restriction, any kind of pressure by the Technological Institute authorities*

- *yo creo que son crisis de valores, qué hacer no lo sé, pero yo siento que de alguna manera tenemos que volver un poco, quizás a, a los tiempos de nuestros abuelos*  
  I think that these are value crisis, what to do I don't know LO, but I feel that we somehow need to go back a little, maybe to our granparents' time.
This factor group was the last one selected in both dialects with the same effect: negative sentences favor the null pronoun and affirmative sentences disfavor it.

Before trying to offer a linguistic interpretation of these quantitative data, it is worth observing whether the effect of the polarity of the sentence is consistent in the data or whether it is due to a particular construction: as Table 4.21 shows, when we run the Varbrul analysis without the social factors, that is to say, including only the internal factors (the same internal factors that were included in the first analysis, and with the same "recode"), the results for Mexican Spanish are somehow different: the factor group "polarity" is no longer selected as significant. Instead, verb person is selected:
<table>
<thead>
<tr>
<th>TYPE OF ANTECEDENT</th>
<th>PROBABILITY</th>
<th>% Ø</th>
<th>TOTAL N</th>
<th>% OF DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogative</td>
<td>.69</td>
<td>95%</td>
<td>368</td>
<td>55%</td>
</tr>
<tr>
<td>Declarative</td>
<td>.27</td>
<td>68%</td>
<td>301</td>
<td>45%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 42</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MANNER ADVERBIAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>.53</td>
<td>86%</td>
<td>623</td>
<td>93%</td>
</tr>
<tr>
<td>Present</td>
<td>.12</td>
<td>37%</td>
<td>46</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 41</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATIVE PRONOUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotransitivo</td>
<td>.55</td>
<td>89%</td>
<td>338</td>
<td>51%</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.53</td>
<td>84%</td>
<td>268</td>
<td>40%</td>
</tr>
<tr>
<td>Ditransitive w/o dp</td>
<td>.17</td>
<td>48%</td>
<td>63</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 38</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SENTENCE TYPE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.75</td>
<td>91%</td>
<td>81</td>
<td>12%</td>
</tr>
<tr>
<td>Declarative</td>
<td>.46</td>
<td>82%</td>
<td>588</td>
<td>88%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 29</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VERB PERSON</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First singular</td>
<td>.55</td>
<td>88%</td>
<td>418</td>
<td>64%</td>
</tr>
<tr>
<td>Second singular</td>
<td>.46</td>
<td>84%</td>
<td>136</td>
<td>21%</td>
</tr>
<tr>
<td>Other</td>
<td>.39</td>
<td>73%</td>
<td>98</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td><strong>Range 16</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Total N = 669, Input: 0.89 (83.1 % null), Log Likelihood = -221.170, p = 0.00*

Table 4.21. Internal factors contributing to the choice of the null DO in Mexican Spanish. Analysis without external factors.
Besides the fact that in Mexico the effect of polarity does not seem to be consistent, it is worth considering whether the effect of the negation is real or whether it is the result of one single construction repeated in the corpora, namely no sé. Tables 4.22 and 4.23 show the distribution of the null pronoun and the clitic lo in relation to negation and the different verbs in both dialects:

<table>
<thead>
<tr>
<th></th>
<th>NO + SABER</th>
<th>NO + OTHER VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no (lo) sé</td>
<td>other forms</td>
</tr>
<tr>
<td>NULL (144) (50.5%)</td>
<td>67% (130)</td>
<td>21% (10)</td>
</tr>
<tr>
<td>LO (141) (49.5%)</td>
<td>33% (65)</td>
<td>79% (38)</td>
</tr>
</tbody>
</table>

Table 4.22. Distribution of the null pronoun and the clitic lo in negated sentences in Peninsular Spanish

<table>
<thead>
<tr>
<th></th>
<th>NO + SABER</th>
<th>NO + OTHER VERBS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no (lo) sé</td>
<td>other forms</td>
</tr>
<tr>
<td>NULL (331) (93%)</td>
<td>98% (153)</td>
<td>89% (67)</td>
</tr>
<tr>
<td>LO (26) (7%)</td>
<td>2% (3)</td>
<td>11% (8)</td>
</tr>
</tbody>
</table>

Table 4.23. Distribution of the null pronoun and the clitic lo in negated sentences in Mexican Spanish

As these tables reveal, in Spain the rate of null objects in negated sentences (50.5%) is slightly higher than the rate of clitic lo (49.5%), but when we decompose these rates taking into consideration the verb, 140 (97%) of these null pronouns occur with no saber, and 130 specifically with no (lo) sé. In this dialect, out of 195 occurrences of the construction no (lo) sé, 130 (67%) have a null object and 65 (33%) include the clitic lo. For other verbs, the pattern is very different: out of 42 occurrences of negated verbs other
than saber with an anaphoric propositional DO, only 4 (9.5%) of these DO are null objects. These numbers suggest a strong effect of the construction no sé in Spain: most of the null objects in negative sentences occur in the no sé construction and, therefore, the selection of some factor groups as statistically significant for Spain could be a consequence of a skewing effect in terms of this one construction.

For Mexican Spanish, as in the overall distribution of the two variants, the null pronoun is more frequent than the clitic lo (93% of the pronouns in negated sentences are null), and this is the case for the verb saber as well as for other verbs. Like in Peninsular Spanish, the rate of null objects is higher in the no sé construction than in other forms of no saber, and higher there than in "no + other verbs", but the percentages are closer in Mexican Spanish and, therefore, it is not as clear that there is a skewing effect of the no (lo) sé construction in this dialect. Besides the no (lo) sé construction, it is worth noting that in the Mexican data, out of the 111 tokens of null pronouns in negated sentences with verbs other than decir, 52 are accounted for by the construction no saber decirte/le (present and conditional, first person singular), and 7 with no poder decirte/le, collocations that are not found in the data with propositional lo.

Table 4.24 shows that 66% of the null objects in Peninsular Spanish occur in the no sé construction, and the remaining 34% (N=67) are used in constructions other than no sé in this dialect. In Mexican Spanish, the construction no sé accounts for 28% of the null objects in the data set, and 72% of the null objects are found in other constructions.
Table 4.24. Distribution of null objects and the effect of “no sé”

<table>
<thead>
<tr>
<th></th>
<th>NO SÉ</th>
<th>OTHER CONSTRUCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexican Spanish</td>
<td>28% (153)</td>
<td>72% (403)</td>
</tr>
<tr>
<td>Peninsular Spanish</td>
<td>66% (130)</td>
<td>34% (67)</td>
</tr>
</tbody>
</table>

To fully test whether the polarity of the sentence is a consistent conditioning factor group in both dialects or whether the results in Table 4.20 are the effect of the no sé construction, I conducted the multivariate analysis excluding all the tokens of no sé and no lo sé from the data. The overall distribution of the two variants in the data, excluding the tokens of no sé and no lo sé is the following:

Table 4.25. Overall frequency of lo and null DO in data without "no (lo) sé".

<table>
<thead>
<tr>
<th></th>
<th>OVERT LO</th>
<th>NULL DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spain (461)</td>
<td>85.5% (394)</td>
<td>14.5% (67)</td>
</tr>
<tr>
<td>Mexico (513)</td>
<td>21% (110)</td>
<td>79% (403)</td>
</tr>
</tbody>
</table>

I conducted a multivariate analysis of the 461 tokens of Peninsular Spanish and the 513 tokens of Mexican Spanish, independently. In the Peninsular Spanish analysis, due to interactions, the external factor groups and the internal factor groups that had never been selected as significant in previous runs were excluded. The data from Mexico was analyzed using the same internal and external factor groups that were included in the original analysis (Tables 4.3 and 4.4), except for Verb tense, removed from the analysis due to interactions. Tables 4.26 and 4.27 show the results of the analyses of Mexican and Peninsular Spanish without the no (lo) sé tokens:

27 These factor groups were Referential distance, Verb tense, Verb person, Manner adverbial, Age and Sex.
<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATIVE PRONOUN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotransitive</td>
<td>.60</td>
<td>17</td>
<td>173</td>
<td>37.5</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.61</td>
<td>19</td>
<td>166</td>
<td>36</td>
</tr>
<tr>
<td>Ditransitive w/o io</td>
<td>.23</td>
<td>4</td>
<td>122</td>
<td>26.5</td>
</tr>
<tr>
<td><strong>Range 37</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SENTENCE TYPE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.70</td>
<td>13</td>
<td>403</td>
<td>87</td>
</tr>
<tr>
<td>Declarative</td>
<td>.47</td>
<td>28</td>
<td>58</td>
<td>13</td>
</tr>
<tr>
<td><strong>Range 23</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total N = 461, Log likelihood = -178.343, input = 0.12 (14.5%)  p  = 0.008

Factor groups not selected: Type of antecedent and Polarity

Table 4.26. Factors contributing to the choice of null pronoun in the data without "no (lo) sé" in Peninsular Spanish

---

28 Also selected as significant, but not included in this table are the three social factor groups: Sex (women: .60; men: .35), Age (-34: .61; 55-: .46; 35-54: .36) and Education (low: .57; high: .41). I did not exclude the external factors from this analysis because, as previously mentioned, when these were excluded in the first, general analysis, polarity was no longer selected as significant, and I wanted to know whether the effect of polarity changed due to the exclusion of the tokens no (lo) sé, and not due to the exclusion of the social factor groups.
As these tables show, the factor group polarity of the host sentence is no longer selected as significant in either of the two dialects, showing that the significance of polarity was due to the effect of the tokens *no (lo) sé* now excluded from the data. Since the effect of polarity in the results of the Varbrul analysis was only apparent, a new multivariate analysis was conducted excluding this factor group. The results, as shown in

---

29 The weights of the factors under the non-selected FG Polarity of the host sentence were: Spain: positive sentence: 0.50; negative sentence: 0.49. Mexico: positive sentences: 0.47; negative sentences: 0.54.
Tables 4.28 and 4.29, are very similar to the results obtained before: the same two internal factor groups were selected as significant in Peninsular Spanish, and the same factor groups plus the factor group verb person were selected as significant in Mexican Spanish.

<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATIVE PRONOUN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotransitivo</td>
<td>.67</td>
<td>44%</td>
<td>368</td>
<td>56%</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.44</td>
<td>19%</td>
<td>166</td>
<td>25%</td>
</tr>
<tr>
<td>Ditransitive w/o io</td>
<td>.14</td>
<td>4%</td>
<td>122</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>Range 53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYPE OF ANTECEDENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrogative</td>
<td>.71</td>
<td>54%</td>
<td>220</td>
<td>33.5%</td>
</tr>
<tr>
<td>Declarative</td>
<td>.38</td>
<td>18%</td>
<td>436</td>
<td>66.5%</td>
</tr>
<tr>
<td></td>
<td>Range 33</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total N = 656, Input: 0.239 (30% null), Log likelihood = -328.754 p = 0.000

Factor groups not selected: Referential distance, Verb tense, Verb person, Sentence type, manner adverbial, Age and Sex.

Table 4.28. Factors contributing to the choice of the null DO in Peninsular Spanish, excluding negation
<table>
<thead>
<tr>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DATIVE PRONOUN</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotransitive</td>
<td>.57</td>
<td>89%</td>
<td>338</td>
</tr>
<tr>
<td>Pronoun</td>
<td>.51</td>
<td>84%</td>
<td>268</td>
</tr>
<tr>
<td>Ditransitive w/o io</td>
<td>.15</td>
<td>48%</td>
<td>63</td>
</tr>
<tr>
<td><strong>MANNER ADVERBIAL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>.54</td>
<td>86%</td>
<td>623</td>
</tr>
<tr>
<td>Present</td>
<td>.12</td>
<td>37%</td>
<td>46</td>
</tr>
<tr>
<td><strong>TYPE OF ANTECEDENT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interrogative</td>
<td>.68</td>
<td>95%</td>
<td>368</td>
</tr>
<tr>
<td>Declarative</td>
<td>.29</td>
<td>68%</td>
<td>301</td>
</tr>
<tr>
<td><strong>SENTENCE TYPE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.73</td>
<td>91%</td>
<td>81</td>
</tr>
<tr>
<td>Declarative</td>
<td>.47</td>
<td>82%</td>
<td>588</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td>42</td>
<td></td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td>39</td>
<td></td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

*Total N = 669, Input: 0.91 (83.1 % null), Log Likelihood = -205.575, p = 0.044*

Factor Groups not selected: Referential distance, Verb tense, Verb person.

**Table 4.29. Internal factors contributing to the choice of the null DO in Mexican Spanish, excluding negation**

To sum up, the data just presented show that the effect of polarity originally found in the first analysis (mainly, the disfavoring effect of negative sentences on the null pronoun) mostly depends on the tokens of *no (lo) sé* and, when these tokens are removed from the data, the factor group polarity is no longer selected as significant. One can wonder, at this point, why this is the case. Is it only the effect of a fixed expression *no sé*? Is there something else, besides the existence of a fixed expression, that could be used to explain this effect? what does *no sé* have that "no + other verbs, other tenses" do not
have, and that favors the null pronoun? As explained in section 4.3., the referent of the DO of no (lo) sé is necessarily interpreted as an interrogative sentence, therefore as a "non complete proposition". Besides the arguable existence of a certain fixation of the expression no sé, the high frequency of null pronouns with no sé can be explained due to the necessary incompleteness of the referent of its DO and the lack of knowledge.

4.4.1. Lack of knowledge

In relation to the high rate of null pronouns with the construction no sé and to the effect of the incompleteness of the proposition referred to by the DO, I would like to entertain the idea that the null pronoun is somehow linked to "lack of knowledge": when the speaker does not have knowledge of the complete content of the proposition (and, therefore, cannot include it in the discourse and the complete proposition cannot be the referent of a pronoun used in the discourse), the null pronoun is favored.

As shown in Tables 4.17 and 4.18 above, representing the distribution of the two variants in the no (lo) sé construction according to type of antecedent, the rate of null objects in this construction with interrogative antecedents is very high (70% of nulls in Spain and 98% in Mexico). The fact that the context in which we find more null objects is the context of a negated sentence in which the direct object pronoun refers to an interrogative clause suggest that the idea of "lack of knowledge" can be a useful concept to explain the use of the null pronoun. The hypothesis would be that when the speaker does not know the content of the proposition, she would prefer to use the null pronoun rather than the overt lo. Note that the idea of lack of knowledge is different from the
completeness of the proposition, and both relate in the following manner:

<table>
<thead>
<tr>
<th>KNOWLEDGE</th>
<th>COMPLETE</th>
<th>INCOMPLETE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Me dijo que hacía calor</td>
<td>Me dijo qué vamos a comer</td>
<td>No sé qué vamos a comer</td>
</tr>
<tr>
<td>Yo sé si va a venir</td>
<td></td>
<td>Pregúntale qué vamos a comer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No me dijo qué vamos a comer</td>
</tr>
<tr>
<td>LACK OF KNOWLEDGE</td>
<td>-------------------------------</td>
<td>------------------------------</td>
</tr>
</tbody>
</table>

Table 4.30. Completeness of the proposition and lack of knowledge

When a complete proposition is included (or referred to by a pronoun) in the discourse, it necessarily means that the speaker has knowledge of the full proposition: the cell “complete-lack of knowledge” in Table 4.30 is empty because it is not possible for the speaker to refer to a complete proposition in the discourse and to lack knowledge about the content of that same proposition. However, lack of knowledge does not equal incomplete proposition: it is possible for a speaker to have knowledge about the complete content of a proposition and to include in the discourse an incomplete proposition. In other words, reference to an incomplete proposition in the discourse does not necessarily mean that the speaker does not know the complete proposition: the full proposition may not be being referred to, even when it is known by the speaker. In the sentences in the cell “incomplete-knowledge of the proposition” in Table 4.30, for example, the indirect interrogative sentences (qué vamos a comer [what we are going to eat] and si va a venir [whether he is coming]) refer to an incomplete proposition, but the speaker has knowledge of the full proposition and could, potentially, have chosen to include a complete proposition in the discourse. The sentences included in the cell “Incomplete -
lack of knowledge” in Table 4.30 illustrate examples where the incompleteness of the proposition referred to in the discourse coincides with the lack of knowledge of the speaker: the speaker does not know the content of the full proposition and, therefore, could not have included the complete proposition in the discourse.

Notice that only in the case of no sé can we be sure that there is lack of knowledge:

It is not impossible to create contexts in which sentences with “pregúntale + indirect interrogative” or “no me dijo + indirect interrogative” are uttered by a speaker who, actually, knows the complete proposition. This can be done by simply adding a continuation to these sentences such as Pregúntale qué vamos a comer. Yo ya lo he visto y no te lo voy a decir [Ask him what we are going to eat. I have already seen it and I am not telling you], or No me dijo qué vamos a comer, pero acabo de entrar en la cocina y lo he visto [He didn't tell me what we are going to eat, but I have just gone into the kitchen and I have seen it]. In most cases, with probably the only clear exception of the objects of no sé and sé, the information about whether the speaker knows or does not know the full content of the proposition needs to be retrieved from the discourse and, sometimes, will not be available at all for the interlocutor (or the linguist).

The hypothesis regarding the concept of lack of knowledge is that the null pronoun would be favored when the speaker lacks knowledge about the complete content of the proposition. This idea arose from observing that the contexts in which the higher

---

30 It seems possible to find the expression no sé when the speaker is actually defending his knowledge of the complement (or denying the lack of knowledge previously implied), in sentences like ¿Cómo que no sé dónde vive? or ¿Quién ha dicho que no sé dónde vive?, where the subordinate clause "que no sé dónde vive" is questioned (Rebeca Campos-Astorkiza, personal communication). Interestingly, in these cases the overt pronoun lo seems to be preferred over the null object: ¿Cómo que no lo sé? or ¿Quién ha dicho que no lo sé? vs ¿Cómo que no sé? or ¿Quién ha dicho que no sé?, and the explicit clausal object can be a declarative sentence (¿Cómo que no sé que Juan está viviendo en París?).
rate of null objects are found are contexts of a negated sentence in which the direct object pronoun refers to an incomplete proposition, especially when the host verb is *saber*. Besides the behavior of the *no (lo) sé* construction, the hypothesis of “lack of knowledge” can be supported by the different interpretation of the expression *no puedo decirte(lo)* [I can't tell (you)] or *no podría decirte(lo)* [I couldn't tell (you)]: across dialects, when *decir*, in these expressions, takes the clitic pronoun as a complement, the favored reading is that the speaker knows the content of the proposition, but can not say it for other reasons. On the other hand, when the complement of *decir* is a null pronoun, the interpretation that we get is that the speaker cannot tell because he does not know the content of the proposition. The following examples illustrate this:

(4.14)  A: ¿Y qué películas le gustaban más? ¿De qué tipo?  
        B: Bueno, pus habían varias, ¿no?: "Madre querida"... este... o...  
        C: "Santa."  
        B: "Chuco el roto" y... Bueno, había una que otra ¿no?, pero no le Ø puedo decir, ¿no? (Mex Pop)  
        A: and what movies did you use to like better? what kind?  
        B: well, there were several... "Dear mother", or…  
        C: "Santa"  
        B: "Chuco el roto" and, when, there were some, right? but I can't tell you Ø, right?, I

(4.15)  A: ¿Cuánto dinero han gastado?  
        B: No podría decirte Ø/no puedo decirte Ø (porque no lo sé)  
        B': No podría decirte(lo)/no puedo decirte(lo) (lo sé pero no puedo decirte(lo)  
How much have they spent?  
I couldn’t tell you Ø/I can’t tell you Ø (because I don’t know)  
I couldn’t tell you LO/I can’t tell you LO (I know, but I can’t tell you)
Example (4.14), from Habla Popular de Mexico, illustrates the use of "no puedo decir" with a null object indicating the lack of knowledge of the speaker: the speaker is not able to tell because he does not remember other movies that he liked. The constructed dialogue in (4.15) illustrates this contrast between the null pronoun and the overt clitic lo. In both cases B and B’, the direct object pronoun refers to an incomplete proposition (licensed by the interrogative sentence ¿Cuánto dinero han gastado?), but they differ in the speaker’s knowledge of the proposition: answer B, with a null pronoun, is interpreted as "I can't tell you how much money they have spent because I don't know", whereas version B’ with the overt clitic pronoun is appropriate when the speaker knows the answer but cannot tell the interlocutor, for other reasons. This contrast between responses B and B’ is perceived by speakers in both dialects. If it is clear in the discourse context that the speaker knows the content of the proposition and cannot tell for other reasons, like in (16), the null pronoun is less accepted in Peninsular Spanish than the overt lo, and, as expected, both anaphors are accepted by Mexican speakers:

(4.16) Le prometí a María que guardaría el secreto así que aunque me preguntes cuánto han gastado, no puedo decirtelo/no puedo decirte Ø.
I promised Maria that I would keep the secret, so even though you ask me how much they have spent, I can’t tell you LO/I can’t tell you Ø.

In fact, when the periphrasis is not poder decir but saber decir, like in (4.17), although both pronouns are possible, the null pronoun is preferred in both dialects:

(4.17)- la forma natural de todas esas rocas, que se forma así, a través de las filtraciones de agua que hay, y todas las piedras que hay... Bueno; hay muchas formas ahí. Y están...
- ¡Ay! Ha d'estar bien bonito. ¿Cuánto tiempo si hace di aquí allá?
… natural shape of all those rocks, that is formed like that, through the water filtrations, and all the stones that are there,… well, there are many shapes, and they are…

-it must be nice. How long does it take you to go there from here?
-I couldn’t tell you Ø, because…
-if you drove!

The following examples support the hypothesis of lack of knowledge: example (4.18) comes from Madrid Habla Culta, and the null pronoun is accepted:

(4.18)  - No, será entre distintas Facultades...
- Supongo que sí...
-...yo creo. Sí.
- Supongo que será así.
- A lo mejor intervenieses en el campeonato de España.
- ¡Je, je! No creo que tenga esa talla, no lo creas.
- Quizá, nadie Ø sabe.
- No, pero... no tengo resistencia. (Mad Cult)
- No, it will be among different schools…
- I guess, yes
- I think, yes,
- I guess it’s like that
- You may take part in the Spanish (National) championship
- Je, je!.., I don’t think I have that level,
- Maybe, nobody knows Ø
- O, but… I don't have that much stamina

In this example, nadie (nobody) includes the speaker: the speaker does not know either, and that would explain the acceptance of the null pronoun, according to the hypothesis of lack of knowledge. If this explanation is correct, we should expect that when nadie (lo) sabe refers to noone but the speaker, the null would not work as well, and the clitic would be preferred. I believe that the following example is evidence of this:
(4.19) Van a despedir a Juan. Pero no vayas a cotillear con los demás, ¿eh?, que nadie lo sabe.
They are going to fire Juan. But don’t gossip with everybody else, because nobody knows LO.

In (4.19) the speaker is not included among the group of people who don't know (the referent of nadie) and the clitic pronoun is preferred over the null object.

Finally, the idea of lack of knowledge would also explain why in the expression Quién sabe, used in both dialects and very frequent in the Mexican corpora, lo is never used in the data analyzed when this construction is used to express lack of knowledge on the part of the speaker:

(4.20) Me preguntaban que... que dónde trabajaba [mi padre]; pues yo les decía que debajo del portal, y decían: "Pues hay muchos portales. ¿Dónde es... el portal?"
-"Pos quién Ø sabe." Y no sabía.
They would ask me… where my father worked; and I would tell them that in the gateway, and they would say: “There are many gateways. Where is the gateway?” – Who knows Ø!” and I didn’t know.

(4.21) - Mi quehacer... ps todo lo que es... lo... lo que depende de una lonchería. Eso es todo lo que yo tengo que hacer: lo que... lo que... lo que le corresponde al lonchero.
- Sí. Es que le digo que, cuando usté falta, ps hace falta alguien. ¡Luego luego se ve!
- Ps ¡quién Ø sabe! Porque aquí habemos dos que sabemos trabajar,
- My work… everything that is... that depends on a “snack bar”. That’s all I have to do: what the snack bar guy attendant is responsible for.
- Yes, I tell you that, when you miss, they need somebody else. You can see that immediately!
- Who knows Ø! Because there are two people here who know how to work

31 For this reason, the construction Quién sabe is included within the exclusions listed in section 3.3.
Example (4.20) illustrates cases where *Quién sabe* responds to a question whose answer the speaker ignores, and in (4.21) *Quién sabe* is used to react to an assertion previously uttered, questioning whether the antecedent is in fact true.

A search of *Quién (lo) sabe* in the Corpus de Referencia del Español Actual (CREA) in both Mexican and Peninsular Spanish shows that the overt clitic pronoun is not impossible in this construction, but, as shown in Table 4.31, its occurrence is much less frequent than the null pronoun.

<table>
<thead>
<tr>
<th></th>
<th>QUIÉN LO SABE</th>
<th>QUIÉN SABE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MÉXICO</td>
<td>2% (1)</td>
<td>98% (63)</td>
</tr>
<tr>
<td>SPAIN</td>
<td>8% (15)</td>
<td>92% (174)</td>
</tr>
</tbody>
</table>

Table 4.31. The expression "*Quién (lo) sabe*" in CREA

In CREA, there is only one case of the expression *quién lo sabe* with the overt pronoun in Mexico, and the use of the null in this expression is almost categorical in this dialect. In the data from Spain, the clitic *lo* is more frequent than in Mexico but the null pronoun also occurs over 90% of the time in the expression *quién sabe*. Like in the case of *no sé*, the construction *quién sabe* is one in which the direct object is necessarily an incomplete proposition and the speaker does not know the complete content of the proposition and, as shown in Table 4.31, the null pronoun is overwhelmingly preferred over the clitic *lo*.

To summarize, the linguistic constraints commented so far, Dative pronoun, Type of antecedent and Polarity, are the factor groups that were selected as significant in the original analyses of both dialects. The results of the statistical analyses indicate that the
variation is conditioned, in part, by the same internal constraints in both dialects and that these linguistic factor groups have similar behavior in Mexican and Peninsular Spanish. Type of antecedent seems to be a powerful conditioning factor in both dialects. The distribution of the two variants according to the presence or absence of a dative pronoun and the mono or ditransitivity of the verb is also comparable in Mexican and Peninsular Spanish, and the favoring effect of negation on null objects seems to be dependent on the construction *no sé* in both varieties of Spanish. The next sections will be devoted to the differences found in the original multivariate analyses of Mexican and Peninsular Spanish.

### 4.5. Divergent conditioning

I will concentrate now on the differences between dialects found in the quantitative analysis. The first and more evident one is the average rate, already commented in section 4.1. Besides the average percentages of the null pronoun, the variation in Mexican Spanish is, according to the multivariate analysis, constrained by more internal (and external) factors than the variation in Peninsular Spanish. Selected in Mexican Spanish but not in Peninsular Spanish were two factor groups, Co-occurrence of a manner adverbial and type of sentence (host sentence). I turn to these now.
4.5.1. Manner adverbial

The factor group manner adverbial was selected in the multivariate analysis of Mexico, with the same range as Dative pronoun, i.e. in first place. The results of the Varbrul analysis for this factor group are repeated here:

<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA Absent</td>
<td>.53</td>
<td>86%</td>
<td>623</td>
<td>93%</td>
</tr>
<tr>
<td>MA Present</td>
<td>.13</td>
<td>37%</td>
<td>46</td>
<td>7%</td>
</tr>
</tbody>
</table>

*Range 40*

Table 4.32. Results of the Varbrul analysis for the group manner adverbial (ma) in Mexican Spanish for null object

Table 4.32 shows that when a manner adverbial is present in the sentence, only 37% of the propositional direct objects are coded as null pronouns in the Mexican data, which means that 63% of the objects are coded with the overt clitic *lo* when they co-occur with a manner adverbial. The multivariate analysis reveals that the presence of a manner adverbial strongly disfavors the null DO in Mexico (.13), but when no adverbial occurs, there is little effect on the variation (.53). The following examples illustrate the tendency to use an overt clitic when a manner adverbial expression co-occurs in the sentence:

(4.22) a. La última vez que fui, llorando, llorando, me dijo que ya se habían ido las madres de ahí, de su colonia; **llorando me lo dijo**, (Mex Cult)

The last time I went, crying she told me that the nuns had left the neighborhood, **crying she told me LO**
b. -porque así le pu... le puse: "Hermosa Luna". Porque hay muchas canciones relacionadas a la luna, ¿no? Tú lo sabes perfectamente- (Mex Pop)
Because I named it "Beautiful moon". Because there are a lot of songs related to the moon, you know LO perfectly well

In both examples, a manner adverbial (llorando in 4.22a and perfectamente in 4.22b) is present in the sentence hosting the propositional direct object, and these context favor the use of the overt clitic lo.

Although we don't have many tokens of manner adverbials, Table 4.33 shows that the presence of a manner adverbial is really affecting the form of pronouns with declarative antecedents:

<table>
<thead>
<tr>
<th></th>
<th>DECLARATIVE ANTECEDENT (301)</th>
<th>INTERROGATIVE ANTECEDENT (368)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NULL (%)</td>
<td>LO (%)</td>
</tr>
<tr>
<td>MA ABSENT (623)</td>
<td>74% (197)</td>
<td>26% (68)</td>
</tr>
<tr>
<td>MA PRESENT (46)</td>
<td>25% (9)</td>
<td>75% (27)</td>
</tr>
</tbody>
</table>

Table 4.33. Crosstabulation of manner adverbial (MA) with Type of antecedent in Mexican Spanish

We only have 46 cases of manner adverbials in the corpora from Mexico. From these 46 tokens, 36 (78%) occur when the antecedent is a declarative sentence and 10 (22%) when the antecedent is an interrogative sentence. What is interesting in Table 4.34 is that when the antecedent is a declarative and there is a manner adverbial, only 9 out of 36 objects are nulls (25%), which means that when there is a manner adverbial present, the lo becomes much more frequent than average in Mexico (75%, versus the 17% of overt lo in the complete data set in Mexico). When the antecedent is an interrogative sentence, even if there is a manner adverbial, the null is still preferred in the data that we
have (80%). However, since only ten object pronouns with interrogative antecedents co-occur in the Mexican corpora with a manner adverbial, the number is too low to draw any conclusions about the effect of the adverbial when the antecedent is an interrogative sentence, since the high percentage of null pronouns could be the effect of the interrogative antecedent overshadowing any adverbial effect.

The disfavoring effect of the manner adverbial on null DOs seems to be indicating that a modifying element (here, the manner adverbial) needs to have some overt linguistic material to modify (here, the DO that is part of the modified predicate). In fact, we find the same tendency with non-referential null DO in both Mexican and Peninsular Spanish, as shown in (4.23) and (4.24):

(4.23) - ¿Compraste café?
   - Sí, Ø compré
   - Sí, lo compré muy barato/de oferta
   - ?? Sí, Ø compré muy barato/de oferta
   - Did you buy (some) coffee?
   - Yes, I bought Ø (some)
   - Yes, I bought LO very cheap/on sale
   - Yes, I bought Ø (some) very cheap/on sale

(4.24) - ¿Alguna vez cocinas judías?
   - Sí, sí Ø cocino
   - Sí, las cocino al vapor
   - ?? Sí, Ø cocino al vapor
   - Do you ever cook beans?
   - Yes, I cook Ø
   - Yes, I cook LAS steamed
   - Yes, I cook Ø steamed

When a manner adverbial is included in sentences (4.23) and (4.24), the reading of the DO is interpreted as more specific and the overt clitic pronoun is preferred over the null object. In example (4.24), due to the present tense of the verb, the lack of overt
pronoun in *Sí, cocino al vapor* could receive a generic interpretation, but the overt clitic pronoun has the effect of restricting the reference of the DO.

Even though there are few tokens with a manner adverbial in the Mexican corpora (46), we find a wide variety of manner expressions.  

No special collocations are identified in the corpora but there is a difference between the manner adverbials that can co-occur with communication verbs such as *decir* or *explicar* and the manner expressions that coexist with *saber*: the former allow for a more varied range of adverbials (expressing the degree of certainty of the speaker [*con razón, por experiencia*] or other attitude of the speaker towards the utterance [*con confianza, en serio, con una naturalidad*], or referring to some characteristic of the context of utterance [*delante de mi mujer, en otro tono, llorando*] or the utterance itself [*en términos generales, detalladamente, textualmente*]), whereas *saber* is semantically restricted to co-occurring with adverbs related with *bien* [well]: *bien, muy bien, bien a bien, perfectamente, de memoria, exactamente*.

A last observation regarding the distribution of the manner adverbials in the data has to do with another of the factor groups included in the coding scheme but not included in the multivariate analysis: the presence or absence of the adverb *ya*. As explained in 3.4.11., the inclusion of this factor group in the coding scheme was based on my own observation that this adverb was very often found in the sentences under study.

---

32 The manner expressions found in the corpora are: *con confianza* (with confidence), *bien* (well), *llorando* (crying), *así* (like that), *como usted* (like you), *en términos generales* (in general terms), *en serio* (seriously), *con naturalidad* (naturally), *textualmente* (textually), *con razón* (with reason), *justo* (fair), *difícil* (difficult), *en serio* (seriously), *de una forma* (in such a way), *en función de las ventas* (according to the sales), *delante de mi mujer* (in front of my wife), *por experiencia* (from experience), *cómo* (how), *en otro tono* (in a different tone), *mejor* (better), *detalladamente* (in detail), *de otra manera* (in a different way), *exactamente* (exactly), *muy bien* (very well), *bien a bien* (very well), *perfectamente* (perfectly well), *de memoria* (by heart), *de una manera especial* (in a special manner).
and was spontaneously produced by speakers of both dialects in some of the examples constructed for the analysis, and that many of the examples mentioned in studies of null definite objects in Spanish contain the adverb *ya*. This factor group was later excluded from the internal variables included in the analysis in Tables 4.2. and 4.3. because it was not orthogonal with the factor group modal adverbial: in both Mexican and Peninsular Spanish, there were no sentences in which the propositional direct object co-occurred with both a manner adverbial and with the adverb *ya*.

A new analysis was conducted in which both factor groups, manner adverb and adverb *ya* were combined. Like the factor group manner adverbial, this new factor group was not selected as significant in the Peninsular analysis but was selected in Mexican Spanish. Table 4.34 shows the results of the multivariate analysis for this Factor Group.

<table>
<thead>
<tr>
<th>Adverb</th>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ya</td>
<td>.76</td>
<td>90%</td>
<td>63</td>
<td>9%</td>
</tr>
<tr>
<td>No manner/ya adverbial</td>
<td>.50</td>
<td>86%</td>
<td>560</td>
<td>84%</td>
</tr>
<tr>
<td>Manner adverbial</td>
<td>.13</td>
<td>37%</td>
<td>46</td>
<td>7%</td>
</tr>
</tbody>
</table>

Log likelihood = -202.475   p = .007

Table 4.34. Results of the factor group Adverb in Mexican Spanish.

33 In the Spanish data, besides this, only two sentences with the adverb "ya" had negative polarity.
34 The remaining factor groups selected as significant are the same factor groups contributing to the use of the null pronoun in the original analysis (Table 4.3). In this new analysis, the new factor group Adverb has the greatest magnitude of effect, followed by Dative pronoun, Type of antecedent, Sentence type and Polarity. Except for the fact that Adverb is now selected in first place, the ordering of the other factor groups by their magnitude of effect is the same, and the ordering of the factors within each factor group is also parallel to the analysis in Table 4.3. The weights are: Dative pronoun: ditransitive with dative pronoun: .55; monotransitive .53; ditransitive without dative pronoun: .16. Type of antecedent: interrogative: .67; declarative .29; Sentence type: interrogative: .72; declarative: .47; Polarity: negative .61; positive .38.
Out of the 63 occurrences of *ya* in Mexican Spanish, 8 occur in the collocation *ya le digo*; 11 in *ya me imagino* and 6 in *ya sé*. Although we have very few tokens of both co-occurring manner adverbials (N = 46) and co-occurring adverb *ya* (N = 63), the fact that both do not co-occur in the same sentence is of interest: my intuition is that the adverb *ya* and the manner adverbials have an opposite role in the discourse: the adverb *ya* tends to co-occur with telic predicates and, therefore, to focus the verbal action (Torres Cacoullos and Schwenter, forthcoming), and seems to defocus the rest of the predicate, in this case the (propositional) object. As García and Porter point out, it is commonly the case that when the focus of the sentence is turned to the verbal process itself, the object is more likely to be omitted (García and Porter 2002:12) and, in the case of propositional objects in Spanish, the presence of the adverb *ya* has precisely the effect of favoring a null object. On the other hand, in sentences with a manner adverb, the (propositional) direct object seems to be part of the focused element, since the whole predicate is being modified by the manner adverbial, and therefore the object tends to merit pronominal remention and the overt clitic *lo* is favored in the statistical analysis.35

4.5.2. Sentence type

Originally, the factor group Sentence type, referring to the host sentence, included four types of sentences: declarative, interrogative, exclamatory and imperative sentences, but in order to capture more general tendencies in the data, the factors were recoded as

---

35 This effect is comparable to the effect of the so-called completive or perfective particles in English, *up* and *out*, in phrasal verbs such as *as drink up, use up, seek out* or *work out*. These expressions are analyzed as serving to shift the focus of attention onto the result of the action, hence, onto the verbal object (Quirk et al 1985: 595), which explains why they are incompatible with object omission (García and Porter 2002:12)
either declarative and non declarative sentences. With this recoding, the distribution of the two variants according to sentence type and the results of the Varbrul analysis are shown in Table 4.35.

<table>
<thead>
<tr>
<th>SENTENCE TYPE</th>
<th>PROBABILITY</th>
<th>% Ø</th>
<th>TOTAL N</th>
<th>% OF DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non declarative</td>
<td>.72</td>
<td>91%</td>
<td>81</td>
<td>12%</td>
</tr>
<tr>
<td>Declarative</td>
<td>.47</td>
<td>82%</td>
<td>588</td>
<td>88%</td>
</tr>
</tbody>
</table>

Table 4.35. Results of the Varbrul analysis of the group Sentence type in Mexican Spanish

As these data show, when the host sentence is a declarative sentence, 82% of the propositional direct objects included in the Mexican corpora are coded as the null pronoun, and when the host sentence is a non declarative sentence, the use of the null pronoun is higher, 91%. The multivariate analysis reveals that, in fact, the factor declarative sentence slightly disfavors the use of the null pronoun although the effect is very small, but when the host sentence is a non declarative sentence, the null pronoun is clearly favored (.72). Example (4.26) illustrates this tendency:

(4.26) – No, because they were, we found out that they were giving these lands, we came, and here we made ourselves comfortable and we stayed here
- How did you find out Ø? or who told Ø you? or how did they tell Ø you?’
Table 4.36 shows the distribution of the two variants among the different non-declarative types of sentences:

<table>
<thead>
<tr>
<th>Type of Sentence</th>
<th>Null (%)</th>
<th>LO (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogatives (63)</td>
<td>90.5% (57)</td>
<td>9.5% (6)</td>
</tr>
<tr>
<td>Imperatives (10)</td>
<td>90% (9)</td>
<td>10% (1)</td>
</tr>
<tr>
<td>Exclamatory (8)</td>
<td>100% (8)</td>
<td>0% (0)</td>
</tr>
</tbody>
</table>

Table 4.36. Distribution of null pronoun and overt lo by non declarative sentence types in Mexico.

It can be seen in Table 4.36 that most of the non declarative sentences in the Mexican corpora are interrogatives (78%), and we only have 8 exclamatory sentences and 10 imperative sentences in the data. This table allows us to see, however, that the favoring effect of the factor Non declarative sentence on the null pronoun is shared in the three type of sentences collapsed in it, and it is not due to the sole effect of the more common sentence type within this group, interrogative sentences. The only case of an overt pronoun lo with an imperative sentence is the following:

(4.27) -Entonces le digo que ahí es donde yo me hago bolas; que no sé qué es lo más importante, porque, a veces...
   -Pero no lo digas... (Mex Cult)
   -So I tell him that there’s where I get confused, that I don’t know what the most important is, because, sometimes…
   -But don’t say LO!

In this example, the hosting verb is a ditransitive verb without a dative pronoun, a context that, as shown in section 4.2, strongly disfavors the use of the null pronoun.
It is worth remembering that, although the factor group "sentence type" was not selected in the original Varbrul analysis of the Peninsular Spanish data (Table 4.2.), it was selected as significant in this dialect when the no (lo) sé tokens were excluded from the data. The results of the factor group Sentence type in the analysis without no (lo) sé tokens in both dialects are repeated in Tables 4.37 and 4.38.

<table>
<thead>
<tr>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.72</td>
<td>91</td>
<td>81</td>
</tr>
<tr>
<td>Declarative</td>
<td>.45</td>
<td>76</td>
<td>432</td>
</tr>
</tbody>
</table>

*Range 27*

Table 4.37. Sentence type in the Multivariate analysis of Mexican Spanish data without “no (lo) sé”

<table>
<thead>
<tr>
<th>Probability</th>
<th>% Ø</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non declarative</td>
<td>.70</td>
<td>13</td>
<td>403</td>
</tr>
<tr>
<td>Declarative</td>
<td>.47</td>
<td>28</td>
<td>58</td>
</tr>
</tbody>
</table>

*Range 23*

Table 4.38. Sentence type in the Multivariate analysis of Peninsular Spanish data without “no (lo) sé”

In both cases, like in Table 4.35, non declarative host sentences clearly favor the null pronoun whereas declarative sentences slightly disfavor it. The fact that Sentence type was selected in the Peninsular analysis when the no lo sé tokens were excluded suggests that, in fact, this linguistic constraint should not be observed as a difference between the two dialects but as another similarity between the two Spanish varieties.
Non declarative sentences have been characterized, as opposed to declarative sentences, as being non-assertive contexts, or, in a functional linguistic perspective, sentences coding irrealis mode rather than realis. I will discuss the data offered here regarding the distribution of the two variants and sentence type in chapter 5.

4.6. Conclusion

In this chapter, I have presented the results of the statistical analysis and commented on the effect of the constraints selected as significant in the Varbrul analysis. This analysis reveals that the same three factor groups selected as significant in the Peninsular corpus are also affecting the variation in Mexican Spanish in very similar ways. The factor group that tests the effect of a dative pronoun (and the monotransitivity or ditransitivity of the verb) is selected as significant in both dialects: monotransitive verbs and ditransitive verbs with a dative pronoun slightly favor the null pronoun, whereas ditransitive verbs without dative pronouns highly disfavor it. The type of antecedent has been shown to play a significant role in the variation, declarative antecedents disfavoring the null pronoun and interrogative antecedents favoring it. Finally, the polarity of the sentence was selected as significant in both dialects but its effect has been shown to be mostly due to the no (lo) sé construction.

Despite these similarities, the Varbrul analysis reveals divergences in the dialects: the variation in Mexico is constrained by more internal factor groups, as well as external constraints. The presence of a manner adverbial has an effect on the variation in Mexican Spanish, in such a way that the presence of a manner adverbial clearly disfavors the use
of the null pronoun, but not in Peninsular Spanish and, although the factor group Sentence type was selected as significant in the original analysis only for the Mexican data, when the no (lo) sé tokens were excluded, the same effect (non-declarative sentences favoring null pronoun) was also found in the Peninsular data.

Besides the results of the statistical analysis, I introduced in this chapter some concepts that will be further developed in the remaining pages: I suggested that the concept of lack of knowledge, as described in section 4.4.1., seems to be playing a role in the variation studied here, and the effect of the factor group type of antecedent was related with the completeness of the propositional referent, a theoretical concept that will be discussed in the next chapter.

Chapter 5 will be devoted to the discussion of these results and the implications for the hypotheses and research questions posed in chapter 2.
CHAPTER 5

LINGUISTIC AND DISCOURSE CONDITIONING ON NULL/LO VARIATION

Although no previous studies have dealt specifically with the variation between the null object and propositional lo, the examination of the literature addressing the use of null objects in languages that, like Spanish, are not traditionally described as object drop languages, and the review of studies dealing with anaphora resolution, provided different hypotheses as to what linguistic factors could be playing a role in the distribution of the two variants in Mexican and Peninsular Spanish. To test these hypotheses, different factor groups were included in the statistical analysis and the results of this analysis were presented in Chapter 4.

In this chapter I will discuss the results of the quantitative analysis in relation with the hypotheses previously posed and the theoretical questions raised in chapters 1 and 2. In order to address those questions, the results of the statistical analysis will be complemented with qualitative observations. In section 5.1., I will review the hypothesis that the accessibility of the antecedent would play a role in the variation under study in this dissertation. Section 5.2. will be devoted to further develop the idea of the
completeness of the proposition and its role as a central concept in the distribution of the null pronoun and the propositional lo in Spanish. In section 5.3. I will relate the results of the quantitative analysis with the notion of transitivity. In section 5.3.1, it will be explained how the notion of "completeness of the proposition" interacts with the transitivity of the sentence.

5.1. Accessibility of the antecedent

One of the hypotheses that was drawn from previous literature on anaphora resolution was the idea that different anaphors are distributed in language mainly in relation to the accessibility of the antecedent, understood, broadly, as the cognitive (or discursive) status of the referent (see section 2.1.). Studies dealing with anaphora resolution share the idea that shorter anaphoric expressions are used when the antecedent is highly accessible, and longer, more complex anaphoric expressions refer typically to less accessible antecedents. Based on this observation, the logical hypothesis to test was whether the two variants under study, the overt pronoun lo and the null object, were distributed in the data relative to the accessibility of the antecedent. To operationalize the notion of accessibility, the factor groups Referential distance (Givon 1983), Number of times the proposition is referred to, and Turn were included in my coding scheme. Neither of these factor groups were discussed in chapter 4 because they were either not selected as significant (referential distance) or not included in the final Varbrul analysis due to crossovers in the data.\footnote{These crossovers are probably due to the fact that we coded for a lot of factor groups with a limited number of tokens available for the analysis in each dialect (Tagliamonte 2006:154). Specifically, regarding these factor groups, the group Number of times the proposition was referred to interacted with}
effect of these factor groups, and consequently of the notion of accessibility, in the variation between the null object and the propositional *lo*.

5.1.1. Referential distance

As explained in section 3.4., the notion of referential distance was proposed by Givon as a measure for topic continuity. In Givon's proposal, the degree of topic continuity or accessibility determines the choice of anaphoric expression: the more continuous the topic, the more accessible the antecedent and, therefore, the lesser the content required in the anaphor. With this starting point, the logical hypothesis regarding the variation between the propositional *lo* and the null pronoun was that the smaller anaphoric expression, i.e. the null pronoun, would be preferred when the referential distance is shorter and the overt anaphor, in this case the clitic *lo*, would be more likely used with greater referential distance.

The measure of referential distance is calculated counting the distance from the target anaphoric expression (here the propositional *lo* or the null pronoun) to the most recent prior mention of the same referent, counting the number of clauses and establishing an upper limit of 10 clauses, a distance quite larger than what is expected to be found in propositional anaphora according to some studies which suggest that, when dealing with higher order entities, the antecedent can only be found in the immediately previous sentence (Schiffrin 1985; Dahl and Hellman 1995). The distribution of the

---

Referential distance, in such a way that very few tokens were mentioned more than one clause before and more than one time in the previous discourse, and most of the tokens were situated in the cell of "referential distance of one sentence" and "referred to once in the discourse". The factor group turn interacted with type of antecedent: most of the interrogative antecedents were also in a different turn.
variants by referential distance is presented in Tables 5.1 and 5.2. Values higher than 4
(the last mention of the proposition took place four or more sentences before the sentence
containing the target pronoun) were collapsed because the numbers were very small.

<table>
<thead>
<tr>
<th>REFERENTIAL DISTANCE</th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sentence</td>
<td>85% (557)</td>
<td>32% (178)</td>
</tr>
<tr>
<td>2 sentences</td>
<td>9% (63)</td>
<td>22% (14)</td>
</tr>
<tr>
<td>3 sentences</td>
<td>4% (25)</td>
<td>16% (4)</td>
</tr>
<tr>
<td>4 or more sentences</td>
<td>2% (11)</td>
<td>9% (1)</td>
</tr>
</tbody>
</table>

Table 5.1. Distribution of null pronoun and overt clitic according to referential distance in Peninsular Spanish

<table>
<thead>
<tr>
<th>REFERENTIAL DISTANCE</th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 sentence</td>
<td>85% (572)</td>
<td>83% (478)</td>
</tr>
<tr>
<td>2 sentences</td>
<td>9% (58)</td>
<td>84.5% (49)</td>
</tr>
<tr>
<td>3 sentences</td>
<td>3% (22)</td>
<td>77% (17)</td>
</tr>
<tr>
<td>4 or more sentences</td>
<td>4% (29)</td>
<td>41% (12)</td>
</tr>
</tbody>
</table>

Table 5.2. Distribution of null pronoun and overt clitic according to referential distance in Mexican Spanish

Regarding the observation that high order entities are available for remention only
in the next sentence, these data show that most of the anaphoric uses of the *lo* and the null
pronoun in both dialects refer to entities introduced in the immediately previous
discourse, mainly in the previous sentence (85% in both dialects). Although reference to
entities introduced two or more clauses before in the discourse is not impossible, the
number of pronouns referring to entities mentioned for the last time three or more clauses
before is significantly lower. Example (5.1) shows one of the discourse fragments from the corpora in which a pronoun refers to a proposition last mentioned more than four clauses before:

(5.1) otr'anécdota, andábamos en, en, en Acapulco y en una de las albercas dond'estábamos, este, estaba yo solo, taba yo viendo a mis hijos nadar, tons veo yo al salvavidas y le digo al salvavidas, dije oye, aquel que anda ahí tirándose, no sabe nadar, dijo, no si ya me dijo que no sabe nadar, pero lo estoy viendo, bueno, o sea, se mete, ya, ya me metí al cuarto, bajé, como a la hora, gran confusión ahí en el, en el lobby, que habían sacado un ahogado de l'alberca, y que habían sacado un 'hogado de l'alberca, y lo que pasó fue que el salvavidas se confundió, andaban dos personas con el mismo color de traje, y él, en una de s-, brinco que dio el muchacho que no sabía nadar, lo perdió de vista, y como en esas albercas de los hoteles grandes en Acapulco tienen tres divisiones, tienen, tienen tres... secciones las albercas, tons yo le había recalcado, y cuando iba bajando, lo primero, que, que dijeron, ese señor fue'l que le Ø dijo, ¿verdá? ¿que usté le Ø dijo al salvavidas?, y yo sin, sin, sin querer hacerle un daño al pobre salvavidas (Monterrey)

Another anecdote, we were in Acapulco, and in one of the swimming pools where we were, I was on my own, watching my children swimming, and I saw the lifeguard, and I told him, listen, that man who is there diving, he can’t swim, and he said, no, he told me that that he can’t swim, but I am watching him, ok, I mean, he left, and I went to my room, I came down one hour later, and there was a great confusion in the lobby, that they had – a drowned person from the swimming pool, and what happened was that the lifeguard got confused, there were two people with the same bathing suit color, and in one jump of the boy who could not swim, [the lifeguard] lost sight of him, and since those big swimming pools in Acapulco have three divisions, have three sections, those swimming pools... I had told him... and when I was going down, the first thing they said: “this was the man who told Ø him, right?, you told Ø the lifeguard?” and I didn’t want to harm the poor lifeguard

Example (5.1) is one of the cases with greatest referential distance in the corpora: the antecedent of the null pronoun object of dijo, at the end of the paragraph, is the sentence aquel que anda ahí tirándose no sabe nadar, in the third line. Reference to a previously introduced proposition, as this example indicates, is not limited to entities introduced in the immediately previous sentence, but a very short referential distance is
significantly more frequent than reference to entities mentioned several clauses before. Regarding the distribution of the two variants according to referential distance, and keeping in mind that this factor groups was not selected as significant in the Varbrul analysis, there seems to be a progressive decrease of the null pronoun when the referential distance increases. This tendency is obvious in the Peninsular data but less clear in Mexican Spanish, where the distribution of the two variants between the factors "1 sentence" and "2 sentences" is very similar (the percentage of null pronouns being slightly higher for entities introduced two clauses before). This progression is more clearly represented in Figure 5.1:

![Figure 5.1](image)

**Figure 5.1. Percentages of null object by Referential distance (number of sentences).**

The lines in Figure 5.1 represent the percentages of null pronouns according to the number of clauses since the last mention of the proposition in the discourse (1, 2, 3 and 4 or more sentences, in the X axis) in Mexican and Peninsular Spanish. As Figure 5.1 shows, the overall rate of null pronouns is very different but the gradation pattern is
comparable across dialects: as the referential distance increases, the rates of null objects
decrease in both Mexican and Peninsular Spanish. Again, these results only reflect a
tendency in the data, since the factor group referential distance was not selected as
significant in the Varbrul analysis in any of the two dialects, but it would be interesting to
observe the effect of this factor group in a larger sample.

5.1.2. Number of times the proposition has been referred to

As we saw in section 2.1.2., in order to explain the fact that the demonstratives
this/that are considerably more frequent than the pronoun it referring to abstract entities
in English, it has been argued that a sentence that has just been introduced in the
discourse (and, therefore, only referred to once) only has the status “activated” in Gundel
et al.'s (1993) scale, and, for this reason, cannot be referred to with a pronominal form
(overt or null), since such a form requires that the antecedent have the status “in focus” in
the giveness hierarchy (Borthen, Fretheim and Gundel 2003; Gundel, Hegarty and
Borthen 2003; Hegarty 2003). According to these scholars, for a higher-order entity to be
“in focus” and, thus, available for reference with a pronoun, it has to be referred to at
least twice in the discourse.

If this theory were correct, and the idea that higher order entities do not acquire
the status "in focus" in Gundel et al.'s scale in the first mention and, therefore, the
pronoun is, in general, not accepted because the cognitive status of the abstract entity is
not accessible enough, it would be expected to work crosslinguistically: the differences
that, regarding the acquisition of a cognitive status, would exist between first order
entities and abstract entities should show up across languages, as argued in Gundel, Hegarty and Borthen (2003:281). To test if this explanation holds true in the Spanish data, we need to observe the number of demonstrative pronouns in Spanish and check if demonstrative pronouns are, like in English, more frequent than null or overt pronouns in the first anaphoric mention of a proposition. A sample of the data used for this dissertation was examined and the number of occurrences of demonstratives (esto, eso, aquello) and pronouns (lo and null) in two of the corpora is included in Table 5.3:

<table>
<thead>
<tr>
<th></th>
<th>PRONOUN (LO AND NULL)</th>
<th>DEMONSTRATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habla Culta Madrid</td>
<td>127</td>
<td>17</td>
</tr>
<tr>
<td>Habla Culta Mexico</td>
<td>78</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>205 (82%)</td>
<td>45 (18%)</td>
</tr>
</tbody>
</table>

Table 5.3. Occurrences of DO pronouns and demonstratives in Madrid and Mexico Habla Culta

These data show that in Spanish, unlike in English, pronouns (overt or null) are significantly more frequent than demonstratives to refer to propositions: most of the anaphoric references to propositions in DO position in the Habla Culta de Madrid and Habla Culta de Mexico corpora are done by means of pronouns (null or overt) (95% in Madrid and 74% in Mexico) and only a small percentage of the anaphoric direct objects referring to propositions (with the same verbs considered in this dissertation) are coded as demonstrative pronouns in this sample of the data (18%). In fact, out of the 205 pronouns referring to propositions in the two corpora, 176 (86%) refer to propositions that have

37 Only demonstrative pronouns in DO position are included in the table. In section 2.1.2 I argued that, although these studies do not differentiate between pronouns in subject and object position, the data that they provide consists mainly of subject demonstratives and pronouns.
only been mentioned in the discourse once and that, according to Borthen, Fretheim and Gundel (2003) and Gundel, Hegarty and Borthen (2003), would only have the status "activated" in the givenness hierarchy and would not be available for reference with a pronoun.

The explanation provided by these scholars for the distribution of English demonstratives and the pronoun "it" referring to high order entities presents the number of times an entity has been referred to in the discourse as a measurement of accessibility of the antecedent, suggesting that the higher the number of times an entity is referred to in the discourse, the more accessible it should be for the speaker; and the fewer the number of times, the lesser its accessibility. If the distribution of the null pronoun and the clitic lo were constrained by the accessibility of the antecedent in this sense, the prediction would be that the null pronoun would be linked to a higher number of mentions of the proposition, and the overt pronoun would be expected when the proposition is less accessible according to this measurement. The distribution of the two variants in Mexican and Peninsular Spanish by number of times the proposition has been previously mentioned in the discourse is presented in Tables 5.4 and 5.5.

<table>
<thead>
<tr>
<th>NUMBER OF MENTIONS</th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83% (424)</td>
<td>17% (86)</td>
</tr>
<tr>
<td>2</td>
<td>86% (94)</td>
<td>13% (15)</td>
</tr>
<tr>
<td>3 or more</td>
<td>76% (38)</td>
<td>24% (12)</td>
</tr>
</tbody>
</table>

*chi square 2.561 (not significant, p = .2779)*

Table 5.4. Distribution of null pronoun and clitic lo by number of times the proposition was referred to in Mexican Spanish

174
The data presented in Tables 5.4 and 5.5 show that, as with the data from Madrid Habla Culta and Mexico Habla Culta corpora in Table 5.3., propositions only mentioned once in the previous discourse can, in both dialects, be referred to by the null pronoun or the clitic *lo* in Spanish: if one wanted to affirm, following they idea of the giveness scale in Gundel et al. (1993), that for a pronoun to be felicitously used, the antecedent has to be in focus, the numbers in Tables 5.4 and 5.5 would indicate that the first mention of a proposition in Spanish is enough for a proposition to acquire that status, contrary to what was proposed for English.

The data in Tables 5.4 and 5.5 also indicate that, if the number of times an entity has been referred to in the discourse is a measure of accessibility, the use of the null pronoun and the clitic *lo* does not respond to the accessibility of the antecedent, at least not regarding this measurement: the null pronoun is not more often used with entities repeatedly referred to in the discourse; on the contrary, the rate of null pronouns is higher in Spain for the first anaphoric mention (33%) than it is when the referent has been previously mentioned several times (11%). In Mexico, the difference between first and second anaphoric mention is very small regarding the rate of null pronoun, but with

<table>
<thead>
<tr>
<th>NUMBER OF MENTIONS</th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33% (168)</td>
<td>67% (336)</td>
</tr>
<tr>
<td>2</td>
<td>22% (24)</td>
<td>78% (84)</td>
</tr>
<tr>
<td>3 or more</td>
<td>11% (5)</td>
<td>89% (39)</td>
</tr>
</tbody>
</table>

*chi square 13.047  p < .005*

Table 5.5. Distribution of null pronoun and clitic *lo* by number of times the proposition was referred to in Peninsular Spanish
entities mentioned three times or more, the rate of null pronoun is lower (76%). The hypothesis that the distribution of the variants under study could be constrained by the number of times the proposition has been referred to is, thus, not supported by the data and the predictions in terms of relative frequency of null vs \textit{lo} (the null object would be more likely to occur when the referent has been referred to several times) are not met in either of the two dialects.

5.1.3. Turn

The factor group Turn, which refers to whether the last mention of the proposition was made in the same turn of conversation or in a different turn of conversation, was included in the coding scheme as one of the linguistic factors that would determine, according to previous studies, the accessibility of the antecedent (Fox 1993; Clements 2006). When coding for this factor group, it was hypothesized that entities previously mentioned by the speaker would be more accessible and, accordingly, were expected to be more frequently coded as null pronouns if the variation under study here is to be explained in terms of accessibility of the antecedent. The frequencies of the two variants for this factor group are presented in Tables 5.6 and 5.7:

<table>
<thead>
<tr>
<th>TURN</th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>different turn</td>
<td>36% (141)</td>
<td>64% (247)</td>
</tr>
<tr>
<td>same turn</td>
<td>21% (56)</td>
<td>79% (212)</td>
</tr>
</tbody>
</table>

\textit{chi-square: 17.995 \quad p < .001}

\textbf{Table 5.6. Distribution of null pronoun and clitic \textit{lo} by Turn in Peninsular Spanish}
Table 5.7. Distribution of null pronoun and clitic *lo* by Turn in Mexican Spanish

<table>
<thead>
<tr>
<th>TURN</th>
<th>NULL</th>
<th>LO</th>
</tr>
</thead>
<tbody>
<tr>
<td>different turn</td>
<td>93% (334)</td>
<td>7% (26)</td>
</tr>
<tr>
<td>same turn</td>
<td>72% (222)</td>
<td>28% (87)</td>
</tr>
</tbody>
</table>

chi square: 51.904  p < .001

The data presented in these tables contradicts the aforementioned prediction: instead of finding a higher rate of null pronouns referring to entities mentioned by the same speaker in the same conversational turn, in both Mexican and Peninsular Spanish the null pronoun is more frequent (36% in Spain and 93% in Mexico) when the last mention of the proposition was made by another speaker, in a different turn, than when it was made in the same turn (21% in Spain and 72% in Mexico).

These results could be interpreted as saying that different turn favors the null pronoun in both dialects. This conclusion, however, should be taken very cautiously since, as mentioned before, this factor groups was not included in the Varbrul analysis due to interactions with the factor group kind of antecedent: very few tokens had an interrogative antecedent in the same turn. As we already saw in chapter 4, interrogative antecedents favor the null pronoun and, since most of the interrogative antecedents were also in a different turn, that could explain the high rate of null objects in different turn. It can not be concluded, with the data analyzed, what the effect of turn might be on the variation under study here, but the analysis does not seem to support the hypothesis that the null object would be favored when the antecedent is last mentioned in the same turn, and, therefore, more accessible in the discourse than propositions introduced in different
5.1.4. Propositional anaphora and accessibility

One of the questions posed at the beginning of this dissertation is whether the accessibility of the antecedent would condition the distribution of the two variants in this study, the null object and the propositional lo. In order to tackle this question, the factor groups referential distance, number of times the entity is referred to in the discourse and turn were included in the coding scheme, under the assumption that these discourse features affect the accessibility of the discourse entity in question; in turn, if the distribution of the two anaphors turned out to correlate with these measurements in the statistical analysis, we could conclude that the variation is constrained by the accessibility of the discourse referent. If this is the case, the prediction, based on the theories of anaphora resolution reviewed above, is clear: referents that are more accessible would be preferably coded with the smallest anaphor, in this case, the null pronoun, whereas the overt pronoun would be more likely used for referents that are less accessible.

The results of the statistical analysis for the factor group referential distance and the data provided in sections 5.1.2. and 5.1.3. for turn and number of times the proposition has been referred to show that the variation of the null and the propositional lo is not conditioned by these constraints: the factor group referential distance was not selected as significant in either of the dialects, although the numbers show a tendency of a decrease of null objects when referential distance increases; contrary to the prediction, the null pronoun is more frequent in our data when entities are only mentioned once in
the discourse than after repeated remention, and turn does not support the observation that entities introduced in the same conversational turn should be more accessible and, therefore, more likely coded as null pronouns.

Maintaining the assumption that these were adequate measurements of accessibility, we need to conclude that the variation under study is not constrained by the accessibility of the antecedent. This finding is consistent with Maes (1996), who finds that the theories of anaphora resolution (Gundel et al. 1993; Ariel 1988, 1990) do not account for the distribution of the abstract personal pronoun *het* ('it') and the demonstrative *dit/dat* ('this/that') in Dutch. These results are, however, somehow unexpected since the idea of accessibility has been considered central in the distribution of anaphoric expressions, and leave us with the question of what linguistic or discourse constraints explain this variation, and how would that be related, if it is related at all, with the notion of accessibility.

It is worth going back to Ariel's distinction between the main two factors that determine the accessibility (or salience, in Ariel's terms) of the referent: one of them is the nature of the relation between the antecedent and the referential expression, which includes the distance between the anaphor and its antecedent or the cohesion between the clauses containing these two expressions (Ariel 1994:28). The linguistic constraints just commented in sections 5.1.1., 5.1.2, and 5.1.3 would be included under this general factor in Ariel's proposal. The second main factor affecting the accessibility of the referent, according to Ariel, is the prominence of the antecedent and, under this factor, she includes some semantic characteristics of the referent of the noun phrase: mental
representation of the speaker and the hearer, or entities related to them, are relatively more prominent than the representations of other entities, and human noun phrases are, by their semantic content, more accessible than noun phrases referring to non-human entities, for example.\textsuperscript{38}

Although the data analyzed before suggests that the first of the two factors of accessibility in Ariel terms, i.e. discourse measures that define the relation between the anaphor and its antecedent, is not conditioning the variation between the null pronoun and the propositional \textit{lo}, it may be the case that the semantic content of the referent (which, following Ariel, could also contribute to its accessibility) would play a role in the variation between the null pronoun and the propositional \textit{lo}. The next section will be devoted to the discussion of the factor group Type of antecedent, the only factor group included in the analysis to capture the semantic content of the direct object referent. Once the effect of this factor group is commented, the discussion on the idea of accessibility and the implications that these data have for the theories of anaphora resolution will be taken up in the conclusions.

5.2. Type of antecedent and completeness of the referent

It is clear in the literature that the distribution of different anaphoric expressions, mainly the distribution of overt versus null pronouns in both object and subject positions, in Spanish as well as in other languages, is closely related to several semantic features of the NPs referred to by those pronouns, mainly animacy, definiteness and specificity. In Spanish, for example, overt subject pronouns can only refer to animate (mostly human) entities.\textsuperscript{38} Prominence is also related in this proposal with subject position and topichood.
entities and, in most dialects of Spanish, the null pronoun in direct object position is restricted to indefinite /non specific entities. Animacy is also a determining factor in the distribution of direct object pronouns in Spanish dialects that accept null objects referring to definite NPs: in these dialects, the null pronoun is restricted to (in Quiteño [Suñer and Yépez 1988] and Paraguayan Spanish [Choi 1998; Palacios 1998]) or highly preferred (Basque Spanish [Landa 1995] and Peruvian Spanish [Paredes 1996]) for inanimate entities and animate /human objects are preferably referred to with an overt clitic pronoun.39

These semantic features, as such, are of no use in the analysis of anaphora referring to high order entities, but it is a logical hypothesis that there should be some semantic distinctions among the group of "propositions" or high order entities that would possibly play a role as important as the semantic features aforementioned for NP anaphora.

Since very few studies have been devoted to high order entities from a linguistic point of view,40 and even less to the analysis of anaphoric processes related to propositions,41 we don't know what those semantic features might be. Previous literature dealing with high order entities has differentiated several semantic classes of abstract entities: events, propositions, facts and states (Asher 1993). Yet, this typology is not clear, nor is it clear whether, from a linguistic point of view, these distinctions are necessary, adequate or useful (Peterson 1982).

39 Frequently with the canonical dative pronoun le (Schwenter 2006:30)
40 More studies exist from a philosophical perspective: Vendler 1967; Parsons 1993.
In order to include a semantic difference in the referent of the object pronouns studied in this dissertation, I coded the extracted tokens for the type of antecedent, differentiating between declarative and interrogative sentence (and, originally, between yes/no questions and main and subordinate sentences). The results of the statistical analysis were reported in Chapter 4: In both Mexican and Peninsular Spanish, the factor group "type of antecedent" was selected as significant in the multivariate analysis, and, in both dialects, interrogative antecedents have a favoring effect on the null pronoun (.67 and .66 in Mexican and Peninsular Spanish, respectively) and a declarative antecedent disfavors this variant (.30 in Mexico and .41 in Spain). I suggested in chapter 4 that the explanation of these results has to do with the completeness of the proposition referred to by the DO pronoun: declarative antecedents typically (although not always) introduce complete, saturated propositions and interrogative sentences introduce open, or incomplete propositions. I will further develop this idea here, offering qualitative evidence of the importance of this semantic characteristic of the referent, i.e. completeness of the proposition, for the distribution of the null pronoun and the propositional lo in Peninsular and Mexican Spanish. I will argue that the distribution of these two pronominal forms is, in very important ways, constrained by the completeness of the proposition being referred to and that the feature of completeness has to be understood as a continuum or gradient feature. The distribution of the null pronoun and the overt clitic lo is related to the completeness of the proposition referred to by the DO in the following way: the more complete the proposition referred to by the DO, the more likely the propositional lo; the less complete the proposition, the more likely the null
object. More specifically, this continuum includes, at least, four different contexts, represented by capital letters A, B, C, D in Figure 5.2, which will be explained below:

<table>
<thead>
<tr>
<th>incomplete proposition</th>
<th>complete proposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;null&quot;</td>
<td>LO</td>
</tr>
<tr>
<td>A - Ø</td>
<td>B - Ø</td>
</tr>
<tr>
<td>C - Ø/lo</td>
<td>D -lo /(Ø)</td>
</tr>
</tbody>
</table>

**Figure 5.2. Completeness of the proposition**

In contexts A and B, only the null pronoun is accepted (Ø in Figure 5.2), I will argue, due to the incompleteness of the proposition. These contexts will be explained in section 5.2.1. Context C in Figure 5.2 includes sentences in which both the null pronoun and the overt propositional clitic are possible, and will be explained in section 5.2.2. Finally, context D is on the "complete proposition" end of the continuum and will be referred to in section 5.2.3.42

**5.2.1. Contexts of categorical null object**

Contexts represented as A and B in Figure 5.2 have in common that they both require a null pronoun: the clitic lo is not possible in either of these contexts, in Mexican or Peninsular Spanish (or, probably, in any dialect of Spanish). Both of these contexts fall

---

42 I consider that the proposition has to be understood as a gradient feature, unlike semantic features characterizing first order entities (such as animacy or definiteness), because the content of a proposition can be incomplete in different degrees. Evidence for this comes from the difference between examples like (5.2), where the propositional DO of dime has no content at all, and examples like (5.3), where a small part of the content of the propositional DO, namely the aboutness of the proposition (trips abroad, in this example), is present. How much of the information needed to calculate the truth value of the proposition is available seems to be a plausible measure of the completeness of the proposition. The idea of this continuum, however, needs further research: at this point, it is only defined by four different contexts, and one of them (C in Figure 5.2) includes different linguistic expressions that introduce incomplete propositions, and that could be argued to introduce propositions whose incompleteness is not always of the same type.
outside of the envelope of variation and were included in the list of exclusions provided in chapter 3. For this reason, the examples that will be used to illustrate these cases, when taken from the corpora, where not included in the statistical analysis. In both contexts A and B, the unavailability of the clitic *lo* has to do with the incomplete character of the proposition that would be referred to by the DO of the verb, but A and B have very different characteristics, explained below.

**A. NO CONTENT AT ALL**

Context A is situated at the "less content" end of the completeness continuum: in these examples, the null pronoun is required because the content of the direct object is completely unknown by the speaker. Example (5.2) illustrates this use:

(5.2)  - Marisa.
- *Dime Ø.*
- Y por fin el barco de Naxos, ¿a qué hora sale? (Corec)
- Marisa
- *Tell me Ø*
- And finally the bout to Naxos, at what time does it leave?

The clitic pronoun *lo* is inappropriate here because the null pronoun is not anaphorically referring to anything: there is no antecedent or licensor, in the discourse or the external context, which could guide us towards the identification of a referent for this pronoun. The answer *dímelo* could only be accepted in a very specific situation, where there is a complete proposition introduced in a previous discourse or in the discourse context, that is very salient to both interlocutors in such a way that, for example, the speaker who answers the phone is waiting for his interlocutor to call and give him a
certain specific information.

Interestingly, as the content of the pronoun's referent becomes known, the clitic pronoun becomes more acceptable.

(5.3) -Bien, ¿usted ha salido al extranjero alguna vez?
-Sí, un par de veces.
- A ver, ¡cuénteme Ø!
- La primera vez estuve en... Francia... no, la primera en Portugal fue. (Mad Cult)
-Well, have you ever been abroad?
- yes, a couple of times
- Let's see, tell me Ø!
- The first one I went to France… no, the first one was in Portugal

In (5.3) the null pronoun is still preferred over the clitic lo but the overt pronoun would be better than in (2), arguably because the previous discourse provides some information about the topic of the DO of contar, sus salidas al extranjero. Examples like (5.3), together with the contexts included in B and C, support the idea of considering "completeness of the proposition" as a gradient feature instead of a binary distinction: the referent of a DO when the speaker knows the aboutness of the proposition, like in (5.3), is arguably not as incomplete as the referent of the DO in examples like (5.2).

43 We will see in section 5.2.2.3. that the possibility of having a licensor that introduces the topic or part of the content of the DO is limited to contar and few similar verbs, like platicar and maybe explicar, but is not possible with other verbs like decir. Compare (5.3) with the following conversation:

A: Bien, ¿usted ha salido al extranjero alguna vez?
B: Sí, un par de veces.
A: #A ver, ¡digame(lo)!

This contrast will be commented in section 5.2.2.3.
B. INTERROGATIVE REFERENT NOT MENTIONED IN DISCOURSE

Another context excluded from the envelope of variation, and referred to in chapter 3, is the context in which the null object of the verb has to be interpreted as referring to an interrogative sentence (i.e. an incomplete proposition) that has not been previously introduced in the discourse. In this context, only the null pronoun is acceptable and the clitic *lo* is not possible. Example (3.17), repeated here as (5.4), illustrates this context:

(5.4) - Y le dimos el cuadro -
  - Allí cenamos todos.
  - y le encantó.
  - ¿Le encantó?
  - ¡Og!
  - ¿Sí?
  - Le encantó el cuadro.
  - Están entusiasmados con el cuadro.
  - Le habíamos regalado un edredón
  - Sí.
  - De la lista de bodas, ¿no?
  - Sí.
  - Le pregunté: "¿Qué quieres?", y tal, ¿no? que ya tenían lista de boda,
    ...
  - Pues - me dijo que tenía una - una lista de bodas y entonces, yo me fui al día siguiente a la lista de bodas y le regalamos una cosa de veinte mil pelas - y un cuadro. Y - y el de - y luego - "No; una sorpresa", decía ella, pero yo no Ø sabía, porque yo decía: "Si la regalo un cuadro a lo mejor pues no - " La apetece más otra cosa que le haga falta, quizá -
  - Sí.
  - Así le llevamos el cuadro de sorpresa.
  - Claro, es - eso no se lo regala nadie. (Corec)
  - And we gave her the painting (...) and she loved it
  - she loved it?
  - ooohh
  - did she?
  - she loved the painting.
  - They are delighted with the painting

186
- We had bought her a duvet
- yes
- from the wedding list, right?
- yes
- I asked her: "What do you want?", and so, right?, that they already had a wedding list... she told me that she had a, a wedding list and so I went the next day to the wedding list, and we bought her something for twelve thousand pesetas and a painting. And, then, "No, a surprise", she said, But I didn't know Ø, because I thought "If I give her a painting, maybe she prefers something else, that she needs, maybe"

The interpretation of yo no sabía in this discourse can be paraphrased as yo no sabía qué hacer/y no sabia si le gustaria [I didn't know what to do/I didn't know if she would like it]: the direct object of saber is necessarily an interrogative proposition, and the sentence cannot be interpreted in such a way that the null object refers to a proposition introduced in the previous discourse.

Examples (5.5), (5.6) and (5.7) from the corpora also illustrate the use of a null pronoun that is necessarily interpreted as referring to an interrogative proposition not introduced in the previous discourse:

(5.5) - De todas formas, bueno, est- nosotros nos iríamos a partir del.. veinti- siete, o así.
- Y... No sé, eso ya depende de cómo vayamos con la casa nueva y eso, ya t- ya te Ø diré.
- Ahá. Muy bien. ¿Y qué tal? Por lo demás, vosotros ¿cómo andáis? (Corec)
- Any way, well… we would leave starting the… twenty… seven, or so
- And, I don’t know, that depends on how we go with the new house and all that. I will tell you Ø
- Well. And, how is it going?, apart from that, how are you doing?

(5.6) - Eso es lo - lo bueno que tiene Estados Unidos, ¿no? Y lo malo: que apalean, pero que después denuncian, bueno -
- Peor sería que apalearan como en otro sitio y no quiero decir Ø y que no se entera nadie. (Corec)
That’s the good thing about US, right? And, the bad thing: that they beat you, but then they report it.

It would be worse that they beat you like in another place, and I don’t want to say Ø, and nobody realized.

Era la típica mujer española, que yo llamo, que la cosa maternal es más fuerte aún que la erótica, ¿no? no sé, un planteamiento mío, a lo mejor es equivocado pero que que quiero decir que - que era una hospitalidad y una cosa tremenda, generosa y tal y simpática -

Sí, sí. Es muy maja.

Lo pasamos muy bien con ella en Almagro. Luego nos cuidaba, no Ø sabes. Era una cosa de cuidados, de - de - de tal, de mostrarte (Corec)

She was the typical Spanish woman, that I call, whose maternal thing is even stronger than the erotic one, right? I don’t know, an approach of mine, it may be wrong but, I mean that… that it was such a hospitality, and generous, and nice…

Yes, she is very nice.

We had a very good time with her in Almagro. Then, she took care of us, you don’t know Ø. It was so hospitable.

In (5.5), the direct object of ya te diré is interpreted as ya te diré si vamos con vosotros de viaje o no [I will tell you if we go with you or not], or a similar interpretation licensed by the previous discourse but, crucially, as an interrogative sentence. Similarly, in (5.6) the sentence no quiero decir could be paraphrased as no quiero decir dónde [I don't want to say where] or no quiero decir qué sitio [I don't want to say what place], necessarily an interrogative complement of the verb decir. And, finally, no sabes in (5.7) is necessarily interpreted as having an incomplete proposition as its complement, which could be explicitly introduced in the discourse with an indirect interrogative clause (no sabes cómo nos cuidaba [you don't know how she took care of us]). I will argue that the use of a null object with the future of saber illustrated in the constructed examples (5.8) and (5.9), common in conversational speech and to a certain extent conventionalized, should also be included under context B in the continuum: the null pronoun is obligatory.
because the referent of the direct object is an incomplete proposition not mentioned in the discourse:

(5.8)  - Juan está enfermo
       - Pues él Ø sabrá
       - Juan is sick
       - (lit.) He'll know Ø

(5.9)  - No tengo ganas de ir
       - Bueno, tú Ø sabrás...
       - I don't feel like going
       - Well, (lit.) you'll know Ø

With the limited context provided in (5.8) and (5.9), there are not many clues for the interpretation of él sabrá and tú sabrás, but this interpretation is necessarily one in which the direct object of saber is an incomplete proposition, that would be expressed by means of an indirect interrogative sentence if it were explicit: él sabrá qué ha hecho para enfermarse [he will know what he has done to get sick], for example, in (5.8) and tú sabrás si vas o no/tú sabrás qué haces [you will know whether you go or not/you will know what you do] in (5.9).

5.2.2. Incomplete content of the proposition

The context identified as C in the continuum in Figure 5.2 includes cases where the referent of the DO pronoun is an incomplete proposition and there is variation between the overt clitic lo and the null pronoun in both Mexican and Peninsular Spanish. There are (at least) two sources of propositional incompleteness: a) there is some piece of information needed to complete a true proposition and b) the speaker lacks knowledge of
the truth of a proposition (Schiffrin 1994:65). The second type of incompleteness, i.e. lack of knowledge as to the truth of the proposition or, in Carlson's proposal, about the polarity of the proposition (Carlson 1983) is found in closed (yes/no) questions and in petitions or commands. The first type of incompleteness of the proposition is found in wh- questions, where the wh- word indicates the source of the incompleteness, that is to say, the part of the content that is unknown, and, I will argue, can also be found in the direct object of contar. I will refer to these different sources of incompleteness of the proposition in the following sections, showing that the null pronoun tends to be used when it refers to an incomplete proposition, in any of these ways, and the overt lo is favored with complete propositions.

5.2.2.1. Interrogative antecedents

The results of the quantitative analysis presented in chapter 4 showed that, in both dialects, the null pronoun is favored when the antecedent is an interrogative sentence (.67 in Mexico and .66 in Madrid), whereas it is disfavored when the antecedent is a declarative sentence (.30 in Mexico and .41 in Madrid). Remember, however, that in the quantitative analysis of Peninsular Spanish data, when the tokens of no (lo) sé were excluded, the factor group type of antecedent was not selected as significant, possibly because of the small number of tokens, but it still showed the same tendency (declarative antecedents: .48, interrogative antecedents: .61). I will offer here qualitative data that supports the importance of the type of antecedent and, more generally, of the semantic feature "completeness of the proposition".
As mentioned in section 4.3., the antecedent of the DO pronoun being an interrogative sentence is the clearest and probably most common case of the referent of the DO being an incomplete proposition: when the clause that licenses the DO anaphor is interrogative, part of the content of the referent of the DO pronoun is not available, namely the content corresponding to the wh- word or whether the proposition is true or false in a yes/no question. In this context, both variants are accepted and the null pronoun is favored in Mexican and Peninsular Spanish, as shown by the statistical analysis and illustrated with the constructed examples (5.10) and (5.11):

(5.10)  ¿Está fría la sopa?  
       Déjame que la pruebe y te lo digo/te Ó digo.  
       Is the soup cold?  
       Let me try it and I tell you LO/I tell you Ó

(5.11)  -Me interesa mucho saber si la sopa está caliente todavía o no, porque estoy probando este plato que he comprado y que dicen que mantiene la comida caliente por mucho tiempo.  
       -A ver, déjame que la pruebe y te Ó digo  
       -I am very interested in knowing whether the soup is still warm or not, because I am trying this new plate that I bought and that is supposed to keep food warm for a long time.  
       -Let me try it and I tell you Ó

In (5.10) and (5.11), according to speaker's intuitions, the null pronoun is accepted in Peninsular Spanish and preferred in Mexican Spanish. If we modify (5.11) in such a way that the DO pronoun can not refer to the interrogative sentence but only to the complete proposition introduced by a non-interrogative sentence, like in (5.12), the null pronoun becomes worse in Peninsular Spanish and the acceptability of the overt pronoun lo for Mexican speakers increases:
(5.12) -Me interesa mucho saber si la sopa está caliente todavía o no, porque estoy probando este plato que he comprado y que dicen que mantiene la comida caliente por mucho tiempo.
-A ver, déjame que la pruebe y, si está fría, te lo digo/te Ø digo
   -I am very interested in knowing whether the soup is still warm or not, because I am trying this new plate that I bought and that is supposed to keep food warm for a long time.
   - Let me try it and, if it is cold, I tell you LO/I tell you Ø

In (5.12), the sentence uttered by the second speaker was modified to include the conditional protasis si está fría, which causes that the DO of digo cannot refer to the interrogative proposition si la sopa está caliente todavía. In this context, where the DO of digo refers to a complete proposition (te digo que está fría), the overt clitic lo is preferred in Peninsular Spanish and, although the null object is accepted in Mexican Spanish, the overt clitic lo is better than in (5.10) and (5.11) for Mexican speakers and is, in fact, the preferred option for one speaker. In (5.13), a similar context in which only a declarative sentence is available as the antecedent of the DO of decir, the null pronoun is dispreferred in Peninsular Spanish and, interestingly, the sentence with the clitic lo is also preferred by the Mexican speakers consulted:

(5.13) Tienes que decirle que la sopa está buenísima porque ha estado cocinando toda la mañana
   Bueno, en cuanto la pruebe se lo digo/le Ø digo.
   You need to tell her that the soup is very good because she has been cooking the whole morning
   Ok, as soon as I try it, I will tell her LO/I will tell her Ø

---

44 This is not only the case with a conditional sentence. The same happens if the response does not have a conditional sentence, like "Mi madre a veces no se da cuenta de que sirve la sopa fría y yo se lo digo".
These qualitative observations support the idea that, even when the anaphoric expression is not *no (lo) sé*, the difference between interrogative and declarative antecedents plays an important role in the distribution of the clitic *lo* and the null pronoun in Peninsular and Mexican Spanish. Further evidence for the importance of the completeness of the proposition (due here to the declarative or interrogative form of the antecedent) is the fact that when two possible antecedents are available for the anaphoric DO, one of them being an interrogative clause and the other one a declarative clause, the null pronoun would be preferably interpreted as referring to the interrogative antecedent and the clitic *lo* would more likely be interpreted as referring to the declarative antecedent in both dialects. This is exemplified in (5.14) and (5.15):

(5.14)  

a. Teníamos la opción de comprar la casa más grande pero más lejos o la más pequeña cerca, y *yo no Ø sabía*, la verdad, porque...  
   a. We had the option of buying the bigger house that was further away, or the smaller one that was closer, and I *didn't know Ø*, because...

b. Teníamos la opción de comprar la casa más grande pero más lejos o la más pequeña cerca, y *yo no lo sabía*  
   We had the option of buying the bigger house that was further away, or the smaller one that was closer, and I *didn't know LO*

(5.15)  

Nos preguntaron qué casa preferíamos y *yo no Ø sabía/y yo no lo sabía*  
They asked us what house we preferred, and I *didn't know Ø/ I didn't know LO*

In these examples, two propositions, an interrogative and a declarative, are available as potential referents of the DO of *saber*, and in this situation, the null pronoun is more likely interpreted as having an incomplete referent whereas the clitic's preferred interpretation is that it refers to the complete, full proposition introduced by the declarative antecedent. In (5.14), the null pronoun is interpreted in both Peninsular and
Mexican Spanish as *no sabíamos qué hacer, cuál comprar* [we didn't know what to do, which house to buy], with an incomplete proposition licensed by the disjunction in the previous discourse, and the clitic *lo* would be more likely interpreted as *yo no sabía que teníamos esa opción* [I didn't know we had that option] or *yo no sabía que la casa más grande estaba más lejos y la más pequeña más cerca* [I didn't know that the bigger house was further and the smaller was closer], a complete proposition, in both Spanish varieties. Similarly, in (5.15) the null refers to the interrogative antecedent, *yo no sabía qué casa preferíamos* [I didn't know what house we preferred], whereas the preferred interpretation of *lo* is *yo no sabía que nos lo habían preguntado* [I didn't know they had asked us that], referring to a complete proposition introduced in the previous discourse, for Mexican and Peninsular speakers.

To summarize, then null object is accepted, and favored according to the statistical analysis, in both dialects when the antecedent is an interrogative sentence, and, in contexts where two clauses are available as possible antecedents of the DO pronoun, one of them being a declarative clause and another one an interrogative, the null object is preferably interpreted as referring to the incomplete proposition, and the clitic *lo*, as referring to the complete proposition.

These qualitative data, together with the quantitative data commented previously, provide evidence that the antecedent being an interrogative or a declarative sentence has an important effect on the choice of the two variants of propositional direct objects, in my proposal, because of the semantic completeness or incompleteness of the referent of the pronoun. Interrogative clauses, however, are not the only clauses that introduce
incomplete propositions. To further support the importance of the semantic feature of completeness of the proposition in the variation under consideration, I will turn now to other cases of incomplete propositions that, according to the qualitative data to be provided, also favor the use of the null pronoun over the clitic *lo*. Specifically, I will refer to pronominal DOs interpreted as petitions or commands, and to the DO of *contar*.

5.2.2.2. Petitions and commands

The possibility of using a null pronoun in direct object position in Peninsular Spanish and the high preference for the null object in Mexican Spanish, as well as the distribution observed between the null pronoun and the clitic *lo* when two propositions with different characteristics are available, (*lo* preferably refers to the complete proposition and the null form is more likely interpreted as referring to the incomplete proposition) are also found with incomplete propositions that would not be made explicit with an interrogative sentence but with a command or petition. Commands and petitions introduce incomplete propositions because, as yes/no questions, the speaker lacks knowledge about the truth of the proposition or, in other words, about which one of the propositions introduced will be true. Example (5.16) illustrates the use of a null object referring to a petition:

(5.16) Dice: "Bueno, pues no te preocupes porque... el problema lo sacamos entre los dos, y los temas yo los llevo, yo los llevo..., llevo chuletas de todos los temas y te los paso". Digo: "No, déjalo, si no me sale ya te Ø diría, ¿no?" (Mad Cult)

He said: "Well, don't worry because the problem, we can solve it together, and the topics I have them, I have them, I have cheat sheets of all the topics and I will pass them to you". I say: "No, no, don't worry, if I can't work it out, I would tell you Ø".
In this example from the Madrid Habla Culta corpus, the preferred interpretation of *ya te diría* in Peninsular Spanish is not *ya te diría que no me sale* [I would tell you that I can't work it out], which would be the interpretation if the null pronoun were to take the most accessible proposition as its referent, but *ya te diría que me pasaras las chuletas* (I would tell you to pass me the cheat sheets), a petition, i.e. an incomplete proposition, in this case licensed by *te los paso* (sic) in the previous discourse. On the other hand, if the clitic pronoun were to be used, *si no me sale ya te lo diría* (if I can't work it out, I would tell you), the preferred interpretation is *ya te diría que no me sale* (I would tell you that I can't work it out), in which the direct object pronoun takes a complete proposition as its referent.

There are no cases in the corpora analyzed where the DO refers clearly to a petition or command, and the few cases like (5.16) above found in the corpora were excluded from the statistical analysis because of the possible ambiguity regarding the interpretation of the null pronoun. It is possible, nevertheless, to offer constructed examples as evidence for the effect of this kind of incomplete propositions on the distribution of the two variants. Using the same context created before, example (5.17) provides two possible licensers of the DO of *diré*: one complete proposition, *está fría* (it's cold), and the other one, more distant in the discourse *te puedo calentar la sopa* (I can heat the soup).

(5.17)  A: Te puedo calentar la sopa, porque hace ya mucho rato que está en el plato
   B: Déjalo, si está fría ya te Ø diré
   B': Déjalo, si está fría ya te lo diré
   A: I can heat the soup for you, because it has been in the plate for a long time
   B: Don't worry, if it is cold, I will tell you Ø
B': Don't worry, if it is cold, I will tell you LO

Both the clitic lo and the null pronoun are possible in the answer (B and B'), but while the preferred interpretation of the clitic pronoun (B') is *ya te diré que está fría* (I will tell you that it's cold), the null pronoun is interpreted in Peninsular Spanish as an incomplete proposition, *ya te diré que me la calientes* (I will tell you to heat it for me), and it would be dispreferred to refer to the declarative clause, a complete referent.45 Examples (5.18) and (5.19) show this same contrast:

(5.18) A: Tenemos que hacer la mudanza este fin de semana  
B: Mi tío tiene una furgoneta. Si se la puedo pedir prestada, te lo digo /?te Ø digo  
A: We have to move this weekend  
B: My uncle has a van. If I can borrow it from him, I will tell you LO /I will tell you Ø

(5.19) A: Tenemos que hacer la mudanza este fin de semana  
B: Mi tío tiene una furgoneta, yo creo que nos la puede prestar  
A: Bueno, por ahora no lo molestes, pero si al final Luis no trae la suya, le Ø decimos  
A’: Ah, pues nos sería muy útil, ¿tú podrías decirle Ø/?por qué no le Ø decimos?  
A: We have to move this weekend  
B: My uncle has a van, I believe that he can lend it to us  
A: Well, don't bother him now, but if Luis doesn't bring his, we tell him Ø  
A´: it would be very usefull, could you tell Ø him/?why don't we tell him Ø?

In (5.18), the null pronoun is strange in Peninsular Spanish because it is not easy to retrieve an incomplete proposition from the discourse (or discourse context), while the clitic is accepted because there is a complete proposition available in the first part of the

45 This seems to be the case in dialogues, such as (18), as well as in monological discourses: in my opinion, the same contrast between the null and the overt pronoun found in (18), occurs in the following monological version:  
Hace mucho rato que la sopa está en el plato, así que te la puedo calendar. Pruébala y, si está fría, me dices/me lo dices  
The soup has been in the plate for a long time, so I can heat it. Try it and, if it's cold, you tell me/you tell me LO
conditional sentence, *te digo que se la puedo pedir prestada* (I tell you that I can borrow it). The Mexican speakers consulted preferred the overt *lo* in (5.18) but also accepted the null pronoun. In (5.19), the null object is possible in Peninsular Spanish in both A and A' because there is an available interpretation in which the null pronoun of *decir* refers to an incomplete proposition (*que nos la preste* [to lend it to us]), now licensed by *que nos la puede prestar* (that he can lend it to us). The null pronoun is clearly preferred by Mexican speakers in (5.19), and, interestingly, the use of an overt *lo* pronoun in (5.19A), *se lo decimos*, would be interpreted by one Mexican as "*van a chismorrearle al tío que Luis no trae la camioneta, no para pedirle la furgoneta prestada*" (they are going to tell the uncle that Luis does not bring the van, not to borrow the van).

To summarize, the pattern found when the direct object pronoun is interpreted as an interrogative proposition is also found when the incomplete proposition is a command or a petition: the null pronoun is accepted by Peninsular Spanish speakers and preferred in Mexican Spanish, and when an incomplete and a complete proposition are available as referents of the DO pronoun, the overt clitic *lo* preferably refers to the complete proposition whereas the null object is more likely interpreted as a command or a petition, an incomplete proposition.

A final case of direct object pronouns that encode an incomplete proposition and that are variable coded as a null pronoun and a clitic *lo* due to this incompleteness in the content of the DO referent is the direct object of *contar* and *platicar*, which will be commented in the next section.
5.2.2.3. The object of contar

As the contrast between (5.20) and (5.21) shows, the null pronoun seems to be more easily accepted in Spanish as the DO of the verb contar than as the object of other communication verbs, such as decir:

(5.20) A: Bien, ¿usted ha salido al extranjero alguna vez?
   B: Sí, un par de veces.
   A: Ver, ¡cuénteme Ø!
   -Well, have you ever been abroad?
   -yes, a couple of times
   -Let’s see, tell me Ø!

(5.21) A: Bien, ¿usted ha salido al extranjero alguna vez?
   B: Sí, un par de veces.
   A: #Ver, ¡dígame(lo)!
   -Well, have you ever been abroad?
   -yes, a couple of times
   -Let’s see, tell me (LO)!

The infelicity of (5.21) is due to the fact that it is odd, under normal circumstances, to ask the interlocutor to say what he just said. With contar, however, the discourse is felicitous. Interestingly, this is the case even when the antecedent of the DO of contar is a declarative sentence, like in the following examples from the corpora:

(5.22) - <<Pues mañana veréis porque la cosa va..., va a ser buena, ¿eh?, mañana porque me van a empezar a preguntar cómo se dice caballo y... pollo...>> ¡Bueno!, también se puede decir poyo. ¡Ya es que la gente ya empezó a... a hacer el indio con eso, y <<quiero vivir>> --empezábamos a decir, ¡je, je! ¡Qué noche pasamos! con la... la cosita esta.
   - No, si lo que quieren es que sea lo más espontáneo posible.
   - Claro; sí, por eso, pero...
   - Y luego ellos lo meten en el ordenador, nos Ø ha estado contando. (Mad Cult)

46 We could create a context in which this response would be appropriate just by adding a reason for B to repeat what he just said, for example, if speaker A needs to record B's utterance and, therefore, needs B to repeat what he just said because A could not record it before.
“You’ll see tomorrow, because it is going to be good, tomorrow, because they are going to ask me how to say horse and… chicken”, well! You can also say “poyo”. And people started to play the fool with that, and “I want to live”, and we started to say… jaja, what a night we spent! with this thing
- No, they want that to be as spontaneous as possible
- Sure, yes, but
- And then they put it in the computer, he has been telling us Ø.

(5.23) A: Pues vete, vete. Oye y ¿qué pasó el otro día en Cuatro Caminos?
B: Un secuestro.
C: Un viejo que había secuestrado a su médico. Y con razón.
A: ¿Cómo, cómo era eso?
B: A su médico no, a un forense.
C: ¡Ah! cuenta Ø. Es que yo no me enteré.
D: Le fueron a desahuciar. (Corec)
A: What happened the other day in Cuatro Caminos?
B: A kidnapping
C: An old man had kidnapped his doctor
A: how, how was that?
B: Not a doctor, a forensic doctor
C: Tell Ø! I didn’t realized that
D: They went to evict him.

I posit that the null pronoun is more accepted as the DO of contar in cases like (5.22) and (5.23) because the referent of this pronominal DO is also incomplete: even when its antecedent is a declarative sentence, the sentence that serves as the licensor does not include in the discourse context the complete referent of the DO pronoun. In this case, the incompleteness is not due to the lack of knowledge about the truth or falsity of the proposition, but to the fact that contar typically has direct objects composed of several propositions, and the declarative sentence in the discourse only provides part of the content that is or will be told, i.e. part of the content of the DO pronoun. Going back to examples (5.22) and (5.23), even though I identified the clause y luego ellos lo meten en el ordenador (and later they put it in the computer) as the antecedent of the null
pronoun in (5.22) for my quantitative analysis, it is very likely that the interpretation of
the null object of contar in this sentence does not refers exclusively to this proposition
but to something else, nos ha estado contando lo que hacen con estas grabaciones, una
cosa es meterlas en el ordenador, y hay otras (he has been telling us what they do with
these recordings; one of the things is to put them in the computer, and there are other
things). In (5.23), the sentence coded as the antecedent of the null object of cuenta is un
viejo secuestró a un forense (an old man kidnapped a forensic surgeon), but the
interpretation of cuenta is not cuenta que un viejo secuestró a un forense (tell that an old
man kidnapped a forensic surgeon) but cuenta cosas acerca de eso (tell things about
that).\textsuperscript{47}

The possibility of interpreting the null object of contar with this meaning explains
the difference between contar and decir that creates the contrast between (5.24) and 5.
(25):

(5.24) Me voy de viaje con mis padres esta semana. Ya te *lo/*Ø diré
I am traveling with my parents this week. I will tell you LO/Ø

(5.25) Me voy de viaje con mis padres esta semana. Ya te lo/Ø contaré
I am traveling with my parents this week. I will tell you LO/Ø

\textsuperscript{47} The interpretation of the null object, when used as the direct object of contar, platicar or explicar, can
be somehow compared to the partitive (indefinite) reading of the null pronoun in Spanish in sentences
like:

- Hay caramelos en esta cesta; si quieres, coge.
The comparison with this partitive null pronoun is, however, not exact: the referent of the null pronoun
in "Hay caramelos en esta cesta, si quieres coge" is interpreted as part of the set of elements that
conform the antecedent, whereas in the case of the propositional DO of contar, the antecedent is a part
of the referent of the null object. In other words, the null object referring to NPs is interpreted as a part
of the NP licensor and the null object referring to propositions with verbs like contar is interpreted as
the clause-licensor and more, or information of which the clause-licensor is part.
In (5.24), the second sentence, *ya te (lo) diré* is odd with any of the two DO variants referring to the first sentence, *me voy de viaje con mis padres* (I am going on a trip with my parents), because it is pragmatically strange to say that one will tell the interlocutor something that he has already told. The pragmatic oddness, therefore, is due to the fact that the referent of the DO of *decir* is exactly the proposition introduced in the previous clause. With *contar*, the situation is different: both variants are accepted in (5.25), and the null is in fact preferred, and the pragmatic oddness that we had in (5.24) does not exist here because the DO of *contar* is interpreted easily as a bundle of propositions and the interpretation that one can easily get from (5.25) is *ya te contaré cosas acerca del hecho de que me voy de viaje con mis padres* (I will tell you (things) about the fact that I am going on a trip with my parents): the DO of *contar* does not refer exactly to the previously introduced sentence or, in other words, the declarative sentence that licenses the DO does not provide all the content of that pronoun: the denotation of the DO pronoun is not complete, and, for this reason, the null object is favored.

### 5.2.3. Complete propositions

Finally, the end "complete proposition" in the continuum represented in Figure 5.1 is filled by context D, where the anaphoric DO refers to a complete proposition. Although, as we saw in section 4.3., a declarative antecedent does not necessarily imply a complete referent, the quantitative analysis revealed a disfavoring effect on the null pronoun when the antecedents is a declarative clause in both dialects. When the tokens of *no lo sé* (cases when a declarative antecedent would introduce incomplete referents) are
excluded from the data, declarative antecedents still disfavor the null in Mexican Spanish. In Peninsular Spanish, direct objects with declarative antecedents are coded as null pronouns 13% of the time (N = 48) and direct objects with interrogative antecedents, 24% of the time (N = 19). Furthermore, examples (5.14), (5.15) and (5.17) above showed that in the context of a discourse where one complete and one incomplete proposition are available as antecedents of the DO pronoun, the overt clitic *lo* is preferably interpreted in both dialects as referring to the complete proposition.

Unlike in the case of null objects, we don't find contexts where the use of the overt clitic *lo* is categorical, yet it is of interest that pronouns that refer to declarative antecedents and that have, presumably, complete referents are the pronouns more likely coded as *lo* in Mexican Spanish if they co-occur with a manner adverbial or a ditransitive verb without a dative pronoun.

The results of the statistical analysis, as well as the judgments elicited from some speakers, show that Mexican Spanish speakers highly prefer the overt clitic *lo* when the direct object co-occurs with a manner adverbial and when a ditransitive verb is used without a dative pronoun, as commented in section 4.2. What is of interest for the purposes of this discussion is that the crosstabulation of these factor groups and the group Type of antecedents shows that the disfavoring effect on the null pronoun is actually found when the antecedent is a declarative sentence and, most likely, the referent of the DO is a complete proposition. Tables 5.8 and 5.9 show the crosstabulation of the factor group Type of antecedent with Dative pronoun (Table 5.8) and Manner adverbial (Table 5.9) in Mexico.
Although the numbers are low, especially the tokens co-occurring with a manner adverbial (N = 17) and with a ditransitive verb without a dative pronoun (N = 30), Table 5.8. shows that, in Mexican Spanish, the overt clitic lo is favored with a ditransitive verb with no dative pronoun only when the antecedent is a declarative antecedent: in this context, only 31% of the tokens in the Mexican data have a null pronoun, and the remaining 69% of the objects with a declarative antecedent and a ditransitive verb without dative pronoun are coded as lo. Similarly, Table 5.9. shows that the low percentage of null pronouns in Mexican Spanish when the propositional direct object co-occurs with a manner adverbial is found when the antecedent is a declarative sentence (25%), and the rate of null objects when there is a manner adverbial but the antecedent is an interrogative sentence (80%) is similar to the percentage of null objects without a manner adverbial in the clause. The number of tokens, of course, is low and these
observations should be taken with caution

We don't find in Mexican Spanish contexts where the overt clitic *lo* is categorical, but the contexts that are close to categorical in quantitative terms, and where the speakers consulted prefer the clitic *lo* (declarative antecedents excluding the object of *no sé* in Peninsular Spanish and direct objects co-occurring with a manner adverbial or a ditransitive verb without a dative pronoun) are contexts in which the antecedent of the direct object pronoun is a declarative sentence and its referent, presumably, a complete proposition.

To summarize, although the percentages of null pronoun and propositional *lo* in Mexican and Peninsular Spanish are very different, the semantic feature "completeness of the proposition" is, according to this analysis, playing a very significant role in the distribution of the null and the overt clitic *lo* in both dialects. The categorical uses of the null pronoun in both dialects corresponds to contexts where the content of the DO referent is either unknown (context A in Figure 5.2) or incomplete and not mentioned in the discourse (context B). When the referent of the DO pronoun is an incomplete proposition, the null pronoun is highly preferred in Mexican Spanish and accepted in Peninsular Spanish and, when the referent of the DO pronoun is a complete proposition, the clitic *lo* is preferred in Peninsular Spanish and accepted by Mexican speakers.
5.3. Transitivity

Among the factor groups that were hypothesized to affect the variation between the propositional *lo* and the null pronoun, I included a series of linguistic characteristics of the host sentence and the host verb (type of sentence, polarity, verb person and verb tense). These factor groups were included in the analysis based on the observation that null pronouns in other dialects or languages (and often with other kind of antecedents) were more frequent in certain constructions. As I mentioned in chapter 2, these grammatical constraints coincide with some of the parameters that define the transitivity of the clause (Hopper and Thomson 1980; Lazard 2003), and one of the research questions that was posed at the beginning of this dissertation was whether the variation between the clitic *lo* and the null pronoun would somehow be related to the transitivity of the sentence.

If we go back to the effect of these linguistic constraints on the variation between the propositional *lo* and the null pronoun with propositional antecedents, a tendency can be observed such that the linguistic contexts in which the null pronoun is favored correspond to linguistic features associated with "low transitivity", whereas the linguistic contexts that favor the use of the overt clitic *lo* are linked to high transitivity of the sentence.

The results of the original statistical analysis for the FG polarity (Tables 4.2 and 4.3) showed that in both Mexican and Peninsular Spanish, negative sentences favored the null pronoun, while affirmative sentences disfavored it. Since polarity is considered one of the parameters of the transitivity of the sentence, these data would suggest that when a
sentence has lower transitivity (negative polarity), the null pronoun is favored and, on the other hand, when the sentence shows higher transitivity (affirmative sentence) the overt clitic pronoun is preferred.

As seen previously, the effect of the factor group polarity in the data analyzed was conditioned by the high frequency of the null pronoun in the no sé construction, which we associated with the necessary incompleteness of the propositional direct object of this expression. This, however, does not necessarily go against the idea that the null pronoun is favored in contexts of negation because of the low transitivity of the clause: first, although the factor group Polarity was no longer selected as significant when all the no (lo) sé tokens were excluded, the tendency of a null pronoun to be favored in negated sentences is still found in Mexico (positive: .47, negative: .54), and the non selection in the varbrul analysis could easily be due to the low number of tokens. It would be necessary to have more tokens to confirm whether this tendency is in fact meaningful in the analysis. Second, the effect of negation on the transitivity of the clause seems to be related to the characteristics of the direct object, and how the negation reduces the effect of the verb on the object: in my data, the effect of negation on the null object was (more) important when the direct object is incomplete, i.e. less individuated, which is consonant with Hopper and Thompson's observation about the effect of negation.

Hopper and Thompson (1980) do not explain in detail how polarity affects the transitivity of a clause: the only information provided, other than including polarity in the table that contains the parameters of transitivity, is that, in many languages, the object of

---

48 The weights of the factors under Polarity in Peninsular Spanish without the "no (lo) sé" tokens are: negative polarity: .49, positive polarity: .50.
a negated clause (especially if it is indefinite) appears in a form which shows that the action of the verb is deflected and less direct. Although not much is said about this, it seems that Hopper and Thompson’s explanation of the relationship between negation and low transitivity is that negation would reduce or cancel the effect that the action has on the object, especially if the object is indefinite, i.e. a less individuated object in a negated sentence receives the action of the verb less directly.

In a clearer and more sophisticated way than Hopper and Thompson, Næss’ (2007) approach also correlates the effect of negation in the transitivity of the clause with the characteristics of the object of a negated sentence: negation affects the transitivity of the clause because it modifies the semantic features that the arguments would have in the corresponding affirmative sentence. The consequence of this is that the semantic opposition between these arguments decreases and, consequently, the transitivity of the clause is also lower. For example, a typical transitive verb like "break" refers to the instigation of a certain event by the subject, with a certain effect on the object, in an affirmative sentence, but, when negated, it no longer refers to the instigation of anything and, presumably, does not refer to any perceivable effects on the object either. In this sense, the object is now less affected, and the subject is not an instigator, and the distinguishability of these arguments is decreased, affecting the transitivity of the clause. Which features are influenced by the negation depends on the semantics of the verb in question (frequently from [+affected] to [-affected] in the object and from [+instigator] to [-instigator] in the object), but Næss affirms that the effect of negation with respect to the feature specifications of the arguments of a clause is to reduce the semantic distinction
between arguments. As a result of this, negation reduces the transitivity of a clause (Næss 2007).

According to Hopper and Thompson, the effect of negation on the transitivity of the clause is clearer when the object is less individuated (indefinite in NPs). It can be argued that, in a similar way, the effect of negation on sentences with propositional objects is also more evident when the object has certain properties, in this case the property of incompleteness that, as I will argue in the next section, also characterizes the object as less individuated. The effect of negation on the transitivity of the clause is even more important when the object is an incomplete proposition, the context in which null pronoun rates are higher.

To summarize, the favoring effect of negation on the null pronoun is not clear from the data, but it is not ruled out either, and if it holds, it could be explained by the transitivity of the sentences: the null pronoun would be favored in clauses with lower transitivity regarding the parameter of polarity. The degree of transitivity of the clause is also affected by the Mode, irrealis versus realis, and low transitive sentences regarding this parameter are also contexts in which the null pronoun is favored.

In Mexican Spanish, the factor group Sentence type was selected as the fourth-most significant factor group in the statistical analysis, and it was also selected as significant in Peninsular Spanish when the tokens of no (lo) sé were excluded from the data (Table 4.25). In both cases, the null object is favored in non-declarative sentences and slightly disfavored in declarative sentences. Although not much is said in Hopper and Thompson about the parameter Mode, they include the distinction realis/irrealis in such a
way that realis mode is typically associated with higher transitivity and irrealis mode corresponds to low transitivity. Næss points out that the effect of the mode parameter on the transitivity of a clause is very similar to the effect of negation: "an event which is presented as not occurring, or as occurring only hypothetically or under non-real circumstances, cannot be conceived of as actually being instigated, or as having any effects on actually observable entities" (Næss 2007:17, 18). Mode reduces the overall transitivity of a clause because, like negation, it reduces the semantic opposition between arguments (subject and object).

Going back to the factor group "sentence type", declarative sentences would then be associated with high transitivity, and non-declaratives, with low transitivity, which turns out to reflect the same pattern found for polarity of the sentence: in Mexican Spanish, and in Peninsular Spanish if the no (lo) sé tokens are excluded, the overt clitic pronoun is favored when the clause is highly transitive, and the null is favored when transitivity of the clause is lower.

5.3.1. Transitivity and Completeness of the proposition

I will argue that the same pattern, i.e. low transitivity corresponding to the null pronoun and high transitivity associated with the overt clitic lo, is found for the factor group Type of antecedent. As explained before, the factor group Type of antecedent translates into completeness or incompleteness of the proposition being referred to by the DO pronoun, in the way discussed before. This factor group was included in the analysis as the only parameter targeted to test the effects of the semantics of the DO referent in the
If we go back to the components of transitivity in Table 2.1, only one of the parameters included in the list provided by Hopper and Thompson refers specifically to the characteristics of the object *per se*, namely the degree of individuation of the object: an object that is highly individuated correlates with high transitivity of the sentence, and an object that is non-individuated is associated with low transitivity. For the purposes of this dissertation, however, it is worth insisting that Hopper and Thompson, as well as other scholars who refer to this feature, elaborate the idea of individuation of the object considering exclusively objects expressed as nouns, or NPs. The authors explain that "the referents of *nouns* with the properties on the left below are more highly individuated than those with their counterparts on the right" (Hopper and Thompson 1980:253, emphasis added):

<table>
<thead>
<tr>
<th>INDIVIDUATED</th>
<th>NON INDIVIDUATED</th>
</tr>
</thead>
<tbody>
<tr>
<td>proper</td>
<td>common</td>
</tr>
<tr>
<td>human, animate</td>
<td>inanimate</td>
</tr>
<tr>
<td>concrete</td>
<td>abstract</td>
</tr>
<tr>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td>count</td>
<td>mass</td>
</tr>
<tr>
<td>referential, definite</td>
<td>non referential</td>
</tr>
</tbody>
</table>

Table 5.10. Individuation (Hopper and Thompson 1980)

The component of individuation refers to the distinctness of the patient from the Agent and to the distinctness from its own background. Arguably, a propositional direct object is even more distinct from the agent (especially from the animate agent that all the
clauses that I am considering have, since the agent of these verbs is always human) than
the less individuated NP. In this sense, Roegiest (1990:241) affirms that completive
clauses, which are inanimate and abstract, should be seen as an argument that intrinsically
manifests a minimum degree of individuation and, therefore, of participation in the
event.\textsuperscript{49} Besides the parameters of "animate-inanimate" and "concrete-abstract", the
remaining parameters offered by Hopper and Thompson under individuation are not
applicable to propositional objects. I posit that the semantic feature completeness of the
proposition indicates, for propositional referents, the degree of individuation of the object
in a way comparable to features such as referential or definite vs. non referential for NPs
in Hopper and Thompson's proposal, and that, therefore, propositional objects can also be
more or less individuated: complete propositions would be more individuated objects and
correspond, then, to high transitivity, and incomplete propositions are less individuated
objects, associated with low transitivity of the sentence. If we include the semantic notion
of completeness under the features of individuation, the results of the statistical analysis
are in keeping with the pattern found for polarity and type of sentence and transitivity:
the null pronoun would be associated with low transitivity (in this case, low individuation
of the object) and the overt clitic \textit{lo} is related to features signaling high transitivity, here,
individuation of the object.

Previously in this dissertation we noticed that some of the grammatical
characteristics of the sentence that were hypothesized to constrain the use of the null
pronoun coincide with parameters of transitivity, and from this arose the question of

\textsuperscript{49} As mentioned in section 2.5.1., Thompson and Hopper decide not to consider clausal complement as
objects of a complex sentence, but point out that, if these clausal complements were counted as objects,
the main clause would be very low in Transitivity (Thompson and Hopper 2001: 31).
whether the variation under study could be somehow related to the transitivity of the clause. It is suggested in this section that, based on the statistical analysis, the null pronoun would be favored in low transitivity clauses, although we need to keep in mind that this tendency is based only on three linguistic constraints: polarity, sentence type and individuation of the object, which would correspond to three of the various parameters affecting the transitivity of the sentence.

5.4. Conclusions

Several theoretical implications of the results presented in previous sections were discussed in this chapter. The first issue addressed in this chapter was whether the null/lo variation is somehow constrained by the accessibility of the antecedent. The distribution of the two variants by the factor groups included in the coding scheme to operationalize the notion of accessibility (referential distance, number of times the proposition was referred to, and turn) does not indicate that the null object and the propositional lo respond to the pattern found in previous literature dealing with first order anaphoric expressions: the factor group referential distance was not selected as significant in either dialect and the marginal results of the other two factor groups were either not significant or showed a distribution of the variants that does not support the prediction of the null pronoun favored by more accessible antecedents. The relationship between the notion of accessibility and one important constraint of the variation, namely the completeness of the proposition, will be addressed in the conclusions.
The second contribution of this chapter has been to show that the significance of
the factor group type of antecedent in the statistical analysis actually reflects an important
linguistic constraint in the variation under study: the semantic feature completeness of the
proposition. I argued that the distribution of the null object and the propositional lo is
conditioned, in important ways, by how complete or incomplete the content of its referent
is, and that this semantic feature accounts both for categorical and variable contexts.
Evidence for the significance of the completeness or incompleteness of the proposition
and for the gradient nature of this feature (unlike the semantic characteristics of first
order entities) comes from both naturally occurring and constructed examples in both
Mexican and Peninsular Spanish.

Finally, it was suggested that there may be an important relation between the
distribution of the null object and propositional lo and the transitivity of the clause. The
data presented in this study do not provide strong evidence for this relationship, but the
results of the quantitative analysis regarding polarity and sentence type point in the
direction that the overt clitic lo is linked with highly transitive clauses whereas the null
object is favored when the clause shows lower transitivity. The same pattern is observed
regarding individuation of the object if it is accepted that, as argued in the last section, the
completeness of the proposition should be included as one of the parameters of
individuation when the direct object is not a noun phrase but a clause.
Null objects in Spanish had proven in the prior literature to be a productive topic of study in Spanish and other languages, but these phonetically null pronouns were considered almost exclusively a feature of contact varieties of Spanish. This dissertation arose from the observation that in at least two monolingual varieties of Spanish, null objects were also possible when the referent of the direct object is a higher order entity.

This dissertation analyzed the variation between the null object and the clitic lo referring to clausally introduced entities in Mexican and Peninsular Spanish. This variable phenomenon was mentioned only in passing, if at all, in previous studies dealing with null objects in Spanish, despite being of great interest intra- and cross-dialectally. This dissertation provides the first empirical data on how widespread the use of null objects with propositional referents is in two monolingual varieties of Spanish and, particularly, in Mexican Spanish.
At the outset of this dissertation, I established as the main goals of this research to determine the linguistic constraints that condition the distribution of the two variants under study, the null pronoun and the propositional lo (characterized in Chapter 1), and to determine whether these linguistic constraints were in consonance with previous studies addressing the distribution of anaphoric expressions or the use of null objects in Spanish or other languages. Two main lines of research were relevant for my study, and this previous literature was reviewed in Chapter 2.

On the one hand, in studying the distribution of two anaphoric forms, it is essential to consider the literature on anaphora resolution, which posits explanations for the distribution of anaphors in relation to certain discourse characteristics which determine the accessibility (givenness, salience) of the antecedent. It was noted that most of this literature did not consider anaphora referring to abstract entities and, when it did, the proposal was not explanatory for the data observed here. On the other hand, literature dealing specifically with null objects in languages that are not considered "object drop languages" is reviewed. Literature in Spanish does not address the use of null objects with propositional antecedents, but provides some observations about linguistic constraints affecting the use of null objects with noun phrase antecedents. Studies of other languages point out that null objects are usually more frequent for propositional antecedents than for noun phrase antecedents and also provide several remarks about linguistic constructions in which the null is more often found.
Based on this literature review, several hypotheses were posed as to what linguistic factors would be playing a role in the variation between the null object and clitic *lo* with propositional antecedents and, to test these hypotheses, a variationist approach was chosen. This approach is appropriate because the phenomenon under consideration does not respond to categorical rules, i.e. the null and the clitic pronouns are not in categorical distribution, nor is there an apparent difference in meaning between the two forms in either dialect. The difficulty obtaining reliable intuitions from speakers also made it necessary to use naturally occurring speech as the data for analysis. Finally, the quantitative methodology used in a variationist analysis permits us to observe the effect of several constraints simultaneously conditioning the variation, and to compare the effect of the constraints across dialects, one of the goals of this dissertation. The decisions regarding the methodology, specifically the definition of the envelope of variation and the coding scheme used in this analysis were explained in Chapter 3.

The coded data were analyzed using the tool Goldvarb X, and the results of the statistical analyses were reported in Chapter 4. Keeping in mind that the number of tokens in each dialect is low, and consequently the conclusions should be taken cautiously, the independent statistical analyses reveal the linguistic conditioning of the variation in both dialects. A striking result of the quantitative analysis is the very high rate of null objects in Mexican Spanish, and, given that the phenomenon had not been previously noted in the literature, the mere occurrence of null objects in Peninsular Spanish is also surprising. Interesting findings come from the statistical analysis and the comparison of the factors affecting the variation in both dialects.
6.1. The variationist analysis and the comparison of Mexican and Peninsular Spanish

The variationist perspective was chosen for this dissertation because its methodology would allow us to identify what linguistic constraints condition the variation between the null pronoun and the propositional *lo* in Mexican and Peninsular Spanish, and to assess the relative strength and significance of these tendencies in the data when the factors are treated simultaneously, as well as to evaluate whether the linguistic conditioning is comparable in both varieties of Spanish.

Comparative variationist linguistics makes use of statistical analysis to compare the grammar systems of different linguistic varieties. The data to be compared is not the existence or rates of occurrence of a form in a language variety, but its distribution across different contexts, which is taken to represent the underlying grammatical structure (Poplack and Tagliamonte 1991:318; Tagliamonte 2001). The results of the statistical analysis regarding the effect of the internal factor groups in the variation offer a valuable tool to compare the grammars of different linguistic varieties. Specifically, observing the constraint ranking of factors (their significance, strength and ordering in the Varbrul analysis) offers insights to infer whether the two data sets share an underlying grammar and to what extent (Tagliamonte 2001: 731).

Table 6.1. shows the results of the independent analyses of Mexican and Peninsular Spanish.
Table 6.1. Internal factors contributing to the choice of the null DO in Mexican and Peninsular Spanish

Table 6.1. shows that the two dialects share linguistic conditioning with respect to some factor groups: three of the factor groups selected as significant at the .05 level when the complete data set (including no (lo) sé tokens) is analyzed coincide: dative pronoun,
type of antecedent and polarity. Their relative strength, indicated by the range, is comparable and, crucially, the constraint hierarchy and direction of effect within each of these factor groups is also parallel in both dialects. Similar constraint hierarchies evidence that the null pronoun does the same grammatical work in both varieties (Poplack and Tagliamonte 2001, 93). However, the results of the multivariate analyses also show some divergent linguistic conditioning: in Mexican Spanish, two other factor groups were selected as significant, which were not selected in Peninsular Spanish, and there is a great difference in the overall rate of null objects in both dialects, as reflected in the corrected means: .23 in Peninsular Spanish and .91 in the Mexican data. Although manner adverbial and sentence type were not selected as significant in Peninsular Spanish, the weights for the non-significant factor groups in Peninsular Spanish show the same direction of effect on the null pronoun, and it would be interesting to see the results for these two factor groups in a larger data set. A closer observation of the data revealed the important favoring effect on the null object of the construction no (lo) sé in both dialects. The analysis of the data set excluding the no (lo) sé tokens shows that the effect of negation in both dialects was epiphenomenal, due to the high frequency of null objects in this specific expression, and when the no (lo) sé tokens were excluded from the analysis, the factor group polarity was no longer selected as significant. Other changes in the multivariate analysis when the no (lo) sé tokens are excluded are that the factor group Dative pronoun is now selected in the third place and the constraint hierarchy within this factor group also changes: the presence of a dative pronoun favors the null object more than a monotransitive verb in both dialects.
The effect of the factor group dative pronoun in the analysis of the complete data set contradicts the observations made in previous literature regarding the presence of a dative pronoun for the null object to occur. These results show that the presence or absence of a dative pronoun with a ditransitive verb actually affects in important ways the null/lo variation, since ditransitive verbs without a dative pronoun highly disfavor the null object. The results for this factor group when a significant number of saber tokens are excluded (no (lo) sé tokens) indicate that the presence of a dative pronoun clearly favors the null object in both dialects.

It was also suggested in the discussion that the transitivity of the clause could be linked to the effect of sentence type and polarity (if the later is confirmed to affect the variation in a larger data set). The multivariate analyses presented in this dissertation do not provide strong support for this proposal but, I believe, suggest that the relation of null objects and the transitivity of the clause is one that would deserve further investigation.

Even though the factor group type of antecedent was not selected as significant in Madrid when the tokens of no (lo) sé were excluded from the analysis, probably due to the small number of tokens and, specifically, to the low rate of null objects in this data set, I argued that this factor group, and more precisely the semantic difference in the completeness of the referent that it reflects, is a crucial element in the distribution of the two variants. The conclusions regarding this theoretical contribution are presented in section 6.2 below.
The lack of significance for referential distance and the marginal results for the other two factor groups operationalizing accessibility of the antecedent deserve discussion. Section 6.3. will be devoted to this issue.

Lastly, one possible explanation for the differences in both dialects could be that the two varieties may represent different stages of a linguistic change in progress, since linguistic change is reflected synchronically in dialect differentiation (Sankoff, 1988a: 147). This issue will be addressed in section 6.4.

6.2. Completeness of the proposition

Since the distribution of null and overt pronouns referring to first order entities in object and subject position in Spanish is clearly constrained by certain semantic characteristics of the referent, such as animacy, person or definiteness, one of the questions posed in this dissertation was whether the semantics of the referent would also condition the use of the null pronoun and the propositional lo. One factor group was included in the statistical analysis to reflect the semantics of the referent, namely type of antecedent, and only two of the four possible factors included within this factor group were found in the corpora (declarative and interrogative sentences, but not exclamatory or imperative sentences). Although, probably due to the small data set and the large number of factor groups, the effect of the factor group type of antecedent in the Peninsular Spanish data turned out to be dependent on the no (lo) sé tokens, I argue that this is, in fact, a crucial constraint in the distribution of the null pronoun and propositional lo in both dialects.
Section 5.2. was devoted to arguing that the null pronoun is linked, in both dialects, to incomplete propositions and the overt clitic *lo* is favored when the referent is a complete proposition. Besides the results of the statistical analysis, evidence for the importance of the completeness of the proposition comes from qualitative data which shows that the null pronoun is accepted by Peninsular speakers when the referent of the object is an interrogative clause, a command or a petition or the object of *contar* whose referent is not complete in the previous discourse. Mexican speakers prefer the null object in these contexts and the clitic *lo* is less accepted than in contexts in which the anaphoric direct object refers to a complete proposition.

An interesting piece of evidence comes from sentences that were or would be excluded from the statistical analysis due to their potential ambiguity: sentences in which the anaphoric direct object has two potential referents in the discourse, one incomplete and one complete proposition. In these cases, the distribution of the two variants in both dialects is such that the overt propositional *lo* is preferred to refer to the complete proposition and the null object is preferably interpreted as the incomplete proposition.

The semantic feature completeness of the proposition, unlike the semantic features that describe noun phrases, is not a binary but a gradient feature. A first proposal of a continuum modeling this gradience was offered in Figure 5.1.

The analysis of the completeness of the proposition provides a good example of the need to combine quantitative methodology with qualitative observations in any study of language use: some of the tokens that were excluded from the statistical analysis due to categorical distribution (contexts A and B in Figure 5.1.) or to potential ambiguity, and
contexts not found in the corpora, provide in this case critical information for explaining the linguistic variation of the null object and the propositional lo.

6.3. Accessibility of the referent and the null/lo variation

In chapter 2, I reviewed the relevant theories of anaphora resolution which, with differences amongst them, revolve around the notion of accessibility (or similar notions such as salience or givenness) to explain the distribution of anaphoric forms in discourse. One of the theoretical questions addressed in this dissertation is to what extent this theories of anaphora resolution and the notion of accessibility can shed light on the variation between the null pronoun and the propositional lo.

As I briefly presented in section 2.1., the theories dealing with "the problem of anaphoric distribution in discourse" (Huang 2000), or how to account for the choice of a particular anaphoric form at a particular point in the discourse, share the idea that anaphors with more phonetic, morphological and semantic content are typically used to refer to less accessible entities, and vice versa (Givon 1980, Gundel et al. 1993; Ariel 1988, 1990). These proposals, however, differ in their approach to the notion of accessibility and in the way this accessibility should be measured, when they explicitly refer to that crucial issue. Following Huang, we can differentiate three main accounts: The topic continuity or distance-interference model (Givon 1983, 1985), the hierarchy model (Fox 1987), and the cognitive model (Gundel et al. 1993; Ariel 1988, 1990), briefly referred to in section 2.1. in this dissertation.
The difficulty in addressing the notion of accessibility in a variationist study such as this one is that it is necessary to operationalize the idea of accessibility with measures that can be used in the statistical analysis. In this study, three factor groups were included in the coding scheme to measure the accessibility of the antecedent: referential distance, turn and number of times the proposition was referred to. The underlying assumption was that these features affect the accessibility of the discourse entity and, in turn, if the distribution of the two anaphors turns out to be in relation with these measurements in the statistical analysis, we could conclude that the variation is constrained by the accessibility of the discourse referent.

Based on the distribution of the two variants with respect to these factor groups, it was concluded in section 5.1. that accessibility does not seem to be a conditioning factor in the variation of the null object and the propositional lo. Section 5.1.4. was closed alluding to the fact that, besides the relation between the anaphor and its antecedent (measured in this case by distance, structural units or number of times the entity was mentioned), it has been proposed that the accessibility of the referent can also be affected by its semantic content, in such a way that some entities, because of their semantic characteristics, are automatically more accessible than others (human than non-humans, entities related to the participants of the conversation than entities not related to them, etc.) (Ariel 1994: 28).

Based on the prominent role attributed to the factor Type of antecedent and, more generally, the completeness or incompleteness of the proposition being referred to, it could be argued that the variation under study here is in fact somehow constrained by the
accessibility of the antecedent, if we assume that this accessibility is determined (at least in part) by the semantic content of the referent; in this sense, complete propositions can easily be seen as more accessible than incomplete propositions, in the same way that human entities are usually more accessible to the speaker and hearer than non-human entities.  

If we assumed that completeness of the proposition is in fact a measure of accessibility in the way just explained, we would still have a disagreement with the main prediction of the anaphora resolution proposals: the analysis provided in this dissertation shows that, in both Mexican and Peninsular Spanish, more accessible entities (complete propositions) are, in fact, not linked to the null pronoun, as would be predicted by these theories, but to the overt propositional lo, and less accessible discourse referents (here, incomplete propositions) favor the null pronoun instead of the overt clitic lo. That is to say, if the semantic content of the proposition (in this case, its completeness) indicates accessibility, the generalization that more accessible entities are coded with smaller or simpler anaphors does not hold for propositional anaphora, since more accessible, i.e. complete, propositions have been shown to favor the overt pronoun "lo" while incomplete, less accessible entities are more likely coded as zeros.

It can be concluded that the notion of the accessibility of the referent, accepted as an accurate generalization able to explain the distribution of anaphors referring to first order entities, does not account for the data analyzed in this dissertation. The phonetically

---

50 It seems intuitively correct to assume that the correlation between the semantic content of the proposition and accessibility would be such that propositions with (more) complete content are more accessible and lesser accessibility correlates with (more) incomplete content, but this is one assumption that I will not try to justify at this point.
zero pronoun under study does not correlate with more accessibility when this notion is operationalized as discourse measurements such as referential distance or turn, nor is it preferably used to refer to the most accessible entities if accessibility is determined by the semantic completeness of the entity.

6.4. Variation and linguistic change: socio-demographic information

An interesting question to consider at this point is whether the variation under study represents a change in progress. Indeed, a crucial aspect of the study of variation is the fact that synchronic variation has been proven to be, in many cases, a reflection of a change in progress: when there is a linguistic change, a phase of variation in which two (or more) variants co-occur necessarily exists (Weinreich et al. 1969:188). An appealing question about the variation between the clitic lo and the null pronoun is whether it reflects a linguistic change in either or both dialects, i.e. whether this variation is a stage in the movement from one linguistic state to another (Chambers 2003:203). From a comparative point of view, dialect differences can reflect different stages in the linguistic change path and reveal insights into the details of the mechanism of linguistic change (Tagliamonte 2002:750). Although undeniable evidence that inter-variety differences are actually the result of language change happening at different rates only comes from diachronic data, indication of a change in progress can be obtained from the observation of socio-demographic constraints on the variation.
As mentioned in Chapter 4, some socio-demographic information was available in the corpora and was included in the statistical analysis: information about the sex \(^{51}\) and age\(^ {52}\) of the informants in both dialects and about education in Mexico. In the factor group age, speakers were distributed in three groups maintaining the division used in the Habla Culta project: under 34 years old, from 35 to 54 years old, and over 55. Regarding the factor group education, only two educational groups were differentiated, low and high education, because the sociodemographic information provided in some of the corpora was limited and further distinctions were not possible. I included in "high education" all the speakers from Mexico Habla Culta corpus and the speakers from the Monterrey corpus who had attended College (undergraduate and graduate studies, not necessarily completed), and the group "low education" comprises all the informants in the Mexico Habla Popular and speakers from Monterrey who were illiterate, speakers with primary education or who had not finished secondary education.

As Table 4.4 showed, the three social factors included in the Mexican coding scheme were selected as significant in the Varbrul analysis, whereas none of them were selected in the analysis of Peninsular Spanish when both internal and external variables were included in the analysis, and only age was selected as significant in Peninsular Spanish when only the social factors and corpus were included in the analysis.

---

\(^{51}\) This information was not available for 13 speakers in Mexican corpora and 1 in the Peninsular data.

\(^{52}\) Information about the age of the speakers was not available for 14 speakers in the Mexican corpora and 39 in Peninsular. For some speakers in COREC, the information about "age" is only approximate (marked as "circa" in the corpus).
What is of interest here is that the results of the multivariate analysis of Mexican Spanish for the social variables suggest a change in progress. First, regarding sex of the informants, women in Mexican Spanish favor the null pronoun (.61) and men disfavor it (.34). These results, together with the fact that the null pronoun is not at all stigmatized in Mexican Spanish, nor are speakers conscious of the variation analysed here, are of great interest, since it has repeatedly been shown in the sociolinguistic literature that women are most often the innovators in linguistic changes from below, i.e. changes that take place below the level of consciousness (Labov 1990). The distribution of the variants *lo* and null object according to the sex of the speakers leads to the inference that the variation under study is in fact a change in progress in Mexico, but the same effect of sex is not found in Spain.

The crosstabulation of the factor groups sex and age shows that in each age group, the women score higher than the men: women younger than 35 years old use the null pronoun 92% of the time, women in the second generation use the null pronoun 86% of the time, and the rate of null objects in women of the third generation is very close, 84%. Young women show the highest rates of null pronouns and 35-54-year-old men have the lowest percentage of null objects (65%). When the factor group sex is crosstabulated with education, we can see that women use the null pronoun more than men in both social classes: among high educated speakers, women use the null variant 81% of the time, and men 73%. In the group of low educated speakers, the percentage of null objects in women is up to 92% and 78% for men. The effect of sex in the variation, i.e. women favoring the null pronoun, is found across the data, independently of age and education of
the speakers.

Turning to the factor group education, the results of the Varbrul analysis show that low education favors the null pronoun (.58) while high education disfavors it (.40) in Mexican Spanish. The crosstabulation of sex and education shows a lower rate of null pronouns in high educated speakers of both sexes (women: 81% and men 73%) than in low educated speakers (women: 92% and man 78%). The favoring effect of the null pronoun by low educated speakers over high educated speakers can also be interpreted as indication of a linguistic change in progress, specifically a change from below. Although the notion of "change from below" refers to the level of consciousness of the speakers, it often correlates with a higher rate of the innovative variant in low educated speakers or lower classes.

Finally, the results of the Mexican Spanish analysis for the factor group age of the informant show that the youngest generation highly favors the null pronoun (.62), the eldest generation slightly disfavors it (.45) and the generation between 35 and 55, has a clear disfavoring effect (.37). These weights could be interpreted on first sight as suggesting a pattern of "age grading", but it is necessary to note that the data is not well distributed according to age: we only have 123 tokens in the group of old speakers and 85% of these tokens (104) are also from low educated speakers. The crosstabulation of the factor groups age and education shows that the "age grading" pattern suggested by these results is actually not found in low educated speakers in Mexico (first generation: 90%, second generation 86%, third generation 85%). Interestingly for the issue of language change, young speakers in both education groups show a high rate of null
pronouns (90% for low educated speakers and 86% for high educated speakers), which can be interpreted as indicating that the null object is very wide spread among young speakers, independently of their education, whereas the factor education is playing a clear role in the distribution of the two variants in older generations.

The sociolinguistic analysis just presented should be taken cautiously: the socio-demographic information that we possess is not very detailed; in some cases it is only based on the interviewer's "appreciations"; a greater number of speakers would be desirable and information about other social indexes or style is not available in the corpora; nevertheless, these results suggest the existence of a change in progress in Mexican Spanish in the direction of losing the propositional *lo* in favor of the null pronoun. Further studies with more complete socio-demographic data will be needed to confirm this impression. As for Peninsular Spanish, the social factors included in the analysis do not provide information that can lead us to infer that the same change in progress is taking place in this variety of Spanish: although the factor group age was selected in the analysis of the complete data set when only age, sex and corpus were included in the analysis, the effect of age turned out to be only an apparent effect and, when the *no (lo) sé* tokens were excluded, the marginal results show that there is no significant effect of age in the variation.

A very appealing aspect of contemplating the idea of a change in progress by which the propositional clitic pronoun *lo* is being replaced by a null pronoun are the parallels with the language change that took place in the DO system in Brazilian Portuguese. Null objects are now widespread in Brazilian Portuguese and Cyrino (1997),
analyzing diachronic data, shows that the first kind of DOs that were coded as zeros in Brazilian Portuguese were objects with propositional antecedents, i.e. the first DO clitic that was lost in DO position was the neuter *o* referring to abstract entities.

In order to test whether the direction of the change is, as it was in Brazilian Portuguese, toward the loss or limiting the use of *lo* in favor of the null pronoun, diachronic data would be needed. No previous study, as far as I know, has observed propositional *lo* and null objects from a historical perspective, and a diachronic study of this phenomenon is outside the boundaries of this dissertation, but Table 6.2 shows the results of a brief search in the data from Mexico in CORDE - Corpus Diacrónico del Español\(^{53}\) with some of the most common verbs that take the null pronoun or the clitic *lo* as its DO, and Table 6.3 shows the results of the search of the same expressions in the 17th century data of Spain in CORDE.

<table>
<thead>
<tr>
<th>se/le _ dije</th>
<th>LO</th>
<th>NULL</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>te _ dije</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>ya _ sabía</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>_ sabías</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 6.2. Null object and propositional lo in Mexican data from CORDE**

\(^{53}\) No dates were entered in the search but the results are from 17th, 19th and 20th centuries. These tables include all the cases found in the search.
null object and propositional lo in 17th century Spanish data

<table>
<thead>
<tr>
<th></th>
<th>LO</th>
<th>NULL</th>
</tr>
</thead>
<tbody>
<tr>
<td>se/le _ dije</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>te _ dije</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>ya _ sabía</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>_ sabías</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

This data is very limited but it indicates that the direction of the change is the progressive loss of the overt clitic pronoun *lo* in favor of the null object, the same direction of the linguistic change that took place in Brazilian Portuguese.

### 6.5. Future research

In this dissertation several important issues pertaining to the variation between the null object and propositional *lo* and the comparison of dialects with respect to this phenomenon have been raised. I have attempted to test the hypotheses and address all the questions through the statistical analysis as well as qualitative observations. Of course, not all the answers are provided and further research is needed concerning some of the issues raised in this dissertation.

First of all, the number of tokens analyzed in this study is low, and an analysis of this phenomenon in a larger data set would lead to more conclusive observations from the statistical analysis.

Very little is actually understood about anaphora with higher order entities and the distribution of anaphoric expressions referring to propositions, in Spanish or other languages. The contributions of this dissertation to the broad phenomenon of anaphora with abstract entities in Spanish is limited to the distribution of two anaphors in object
position, but further research should address the distribution of other anaphoric expressions, mainly demonstrative pronouns, and the distribution of these anaphors in both subject and object positions.

The main contribution of this dissertation to the Spanish dialectological literature is to establish the widespread existence of null objects in monolingual varieties of Spanish, and to suggest that this kind of anaphora should be considered independently of first order entity anaphora. I hope that this dissertation will be the first of many studies of this variation in other varieties of Spanish. In this respect, it would be of great interest to test whether there are differences between the distribution of null objects with propositional antecedents and null objects referring to NPs in the Spanish varieties that accept definite null objects, and in both monolingual and bilingual speakers. Similarly, I look forward to studies addressing the variation between null/propositional lo in other monolingual varieties of Spanish.

Lastly, it was suggested in this dissertation that the null/lo variation could be reflecting a change in progress, at least in Mexican Spanish. To verify this observation, a more detailed sociolinguistic investigation and a diachronic study of the null object and clitic lo with propositional antecedents would be of great interest.
REFERENCES


Bosque, Ignacio and Juan Carlos Moreno. 1990. Las construcciones con *lo* y la denotación del neutro. *Lingüística* 2.5-50.


237


Labov, William (to appear) Quantitative reasoning in linguistics. www.ling.upenn.edu/~wlabov/Papers/QRL.pdf


245


