Differential Object Marking in Spanish: A Quantitative Variationist Study

DISSERTATION

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

By

Ian Robert Tippets

Graduate Program in Spanish and Portuguese

The Ohio State University

2010

Committee:

Scott A. Schwenter, Advisor

Terrell A. Morgan

Rebeka Campos-Astorkiza
Abstract

This dissertation addresses the variable nature of the linguistic phenomenon known as Differential Object Marking (DOM) as it is manifested in Spanish. More commonly known in the literature as the personal a or the accusative a, this phenomenon has been attributed primarily to marking animate, predominantly human, direct objects. However, this marking is considerably more variable when observed in actual usage, permitting the marking of non-human referents as well as not always occurring with human referents.

This variation in marking is a prime candidate for variationist methodologies and the present dissertation quantifies the distribution of DOM in the spoken language of three dialects of Spanish in order to realize a multivariate analysis. The data is culled from three transcribed corpora, one from Buenos Aires, one from Madrid and one from Mexico City. To evaluate what factors are relevant in these dialects as well as to provide a general model of how the a-marking phenomenon works broadly, the multivariate statistical program Goldvarb is used to assess the data.

The principal findings of the dissertation include identifying a unifying concept, individuation, which is able to account for significant components of the data. The individuating factors found to be primarily responsible in accounting for observed variation in a-marking include animacy, definiteness, specificity and number. Additionally, the relationship between a-marking and topicality is explored more fully as
a result of implementing particular quantifiable metrics. It is shown that marking correlates with a relative notion of topicality wherein those objects which are frequent and have proximal coreferential expressions are more likely to be marked. Finally, dialectal distinctions are found when comparing the factors found to be significant in each dialect. The empirical data obtained through this quantitative approach to $a$-marking provide concrete evidence to show how Spanish Differential Object Marking is motivated by multiple factors.
Acknowledgments

This undertaking would not have been possible if it were not for the efforts of many people.

I would like to express my appreciation to my mentor and advisor Scott Schwenter for the incredible insight and knowledge he willingly shared throughout my graduate experience. His continual support and expectations for various projects and papers has enabled me to achieve my highest aspirations while keeping me appropriately grounded. It has been a remarkable and fascinating opportunity to be able to pursue my interests in Hispanic linguistics under the tutelage of someone so knowledgeable. Finally, I express my gratitude for the subtle awareness that he had for me as an individual in helping me develop my strengths and recognize my shortcomings.

I also would like to thank Terrell Morgan for his contributions to both the dissertation and to my academic preparation. I have observed his enthusiasm for and dedication to both linguistics and teaching in classrooms from Columbus to Guatemala and thank him for those opportunities.

I thank Rebeka Campos-Astorkiza for willingly participating on the committee. She provided invaluable suggestions and critical comments despite
being on different continents for the latter stages of the dissertation. The dissertation is stronger and more cohesive as a result of her efforts.

I have found incredible fulfillment in my graduate studies in no small part as a result of the dear friends and colleagues that have surrounded me. I want to thank Ryan Kowolski for his lasting friendship and kindly scowl. I have appreciated the opportunity to share so much of the graduate experience with Sarah Sinnott and the numerous roads we have been on together from Pennsylvania to Bolivia. Thank you Juliana de la Mora for your joy and also your patience with my perpetual preguntas. I have enjoyed the enthusiasm and energy of Meghan Armstrong in all our interactions. To these, my friends, I express my sincerest love and appreciation.

Finally, I want to thank Erin. I could not ask more from a wife, companion and friend than your unfailing support and love. You have allowed me to pursue my passions and accomplish my goals. I am eternally indebted to you for your generosity, thoughtfulness and dedication.
Vita

2003 ................................................. B.A. Spanish, Brigham Young University

2006 ................................................ M.A. Hispanic Linguistics, The Ohio State University

2004 to 2010 ................................. Graduate Teaching Associate, Department of Spanish and Portuguese, The Ohio State University

Fields of Study

Major Field: Spanish and Portuguese
Table of Contents

Abstract ......................................................................................................................... ii

Acknowledgments ....................................................................................................... iv

Vita ............................................................................................................................... vi

Fields of Study ............................................................................................................ vi

Table of Contents ....................................................................................................... vii

List of Tables ............................................................................................................... xi

List of Figures .............................................................................................................. xiv

CHAPTER 1: Introduction ......................................................................................... 1

1. Phenomenon ........................................................................................................... 1

1.1. Objectives and research questions ................................................................. 3

1.2. Significance of the project ............................................................................... 5

1.3. Organization of the dissertation .................................................................... 8

CHAPTER 2: PREVIOUS RESEARCH ON SDOM .................................................. 10

2. Overview of the Previous Research on SDOM ................................................. 10

2.1. Animacy in SDOM ......................................................................................... 13
4.1. Overall distribution of SDOM ................................................................. 100
4.2. Direct Object Animacy and SDOM .......................................................... 104
4.3. Definiteness and SDOM ............................................................................ 107
4.4. Relative Animacy in SDOM ...................................................................... 112
4.5. Clause Type and SDOM ........................................................................... 116
4.6. Ambiguity and SDOM ............................................................................... 118
4.7. Referential Distance and SDOM ............................................................... 120
4.8. Specificity and SDOM ............................................................................... 123
4.9. Number and SDOM .................................................................................. 125
4.10. Dialects and SDOM ................................................................................ 127
4.11. Conclusion ............................................................................................. 129

CHAPTER 5: DIALECTAL ANALYSES .............................................................. 130

5. Introduction .................................................................................................. 130
5.1. SDOM in Buenos Aires ............................................................................ 130
5.2. Multivariate analysis of HC Madrid .......................................................... 142
5.3. SDOM in HC Mexico City ......................................................................... 153
5.4. Conclusion ................................................................................................ 159

CHAPTER 6: DIFFERENTIAL OBJECT MARKING IN SPOKEN SPANISH ..... 161

6. Introduction .................................................................................................. 161
6.1. NP Features of SDOM................................................................. 162
6.2. Individuation.................................................................................. 171
6.3. Structures & Functions related to SDOM at the clause-level........... 175
6.4. Properties of the discourse domain: Topicality............................... 181
6.5. Dialects.......................................................................................... 185

CHAPTER 7: CONCLUSION........................................................................ 187
7. Summary............................................................................................ 187
7.1. Findings ......................................................................................... 187
7.2. Future Research............................................................................. 192

References............................................................................................. 195
List of Tables

Table 1 Information status (from Birner 2006b) ................................................................. 25
Table 2 Cross-classification of definiteness and specificity .................................................... 31
Table 3 Information status (from Birner 2006b) ................................................................. 86
Table 4 Overall frequency of SDOM ...................................................................................... 100
Table 5 Factors contributing to SDOM .................................................................................. 103
Table 6 Distribution of DO animacy before collapsing factors ............................................. 105
Table 7 Rate of marking for definite DOs .............................................................................. 108
Table 8 Rate of marking of definite DOs with human proper nouns excluded .............. Error!
   Bookmark not defined.
Table 9 Rate of marking according to definiteness and DO animacy ................................. 110
Table 10 Rate of a-marking according to definiteness and specificity ................................. 111
Table 11 Multivariate results for relative animacy .............................................................. 113
Table 13 Multivariate results for clause type .................................................................... 116
Table 14 Marking according to embedded clause and DO animacy ................................. 118
Table 15 Multivariate analysis of ambiguity ..................................................................... 119
Table 16 Multivariate analysis of total referential distance ............................................... 121
Table 17 Rate of marking according to human DO and total referential distance ....... 123
Table 18 Multivariate analysis of specificity ................................................................. 124
Table 19 Rate of marking according to specificity and DO animacy ............................... 124
Table 20  Multivariate analysis for number ................................................................. 125
Table 21  Rate of marking according to number and DO animacy ......................... 126
Table 22  Multivariate analysis for dialect ................................................................. 127
Table 23  Rate of marking according to dialect and DO animacy ............................. 128
Table 24  Factors contributing to SDOM in Buenos Aires ......................................... 131
Table 25  Comparison of rate of marking between DOs with greater animacy than subject and overall rate of marking for human DOs in the Buenos Aires data ..................... 133
Table 26  Rate of marking according to definiteness and DO animacy ..................... 134
Table 27  Rate of marking according to specificity and DO animacy ........................ 136
Table 28  Rate of marking according to definiteness and specificity ........................... 137
Table 29  Rate of marking for human DOs according to total referential distance ...... 139
Table 30  Multivariate analysis of total topic persistence .......................................... 140
Table 31  Rate of marking according to DO animacy and Total Topic Persistence ...... 141
Table 32  Rate of marking according to ambiguity and DO animacy ....................... 145
Table 33  Rate of marking according to definiteness and DO animacy ..................... 147
Table 34  Rate of marking according to definiteness and specificity ......................... 148
Table 35  Rate of marking according to specificity and DO animacy ........................ 148
Table 36  Rate of marking according to specificity and definiteness ........................... 148
Table 37  Multivariate analysis of DO animacy ............................................................. 155
Table 38  Rate of marking according to definiteness and DO animacy ..................... 156
Table 39  Rate of marking according to definiteness and definiteness ...................... 158
Table 40  Rate of marking according to number and DO animacy ............................ 158
Table 41  Individuation properties and groups selected as significant ...................... 164
Table 42  Properties of individuation selected as significant ..................................... 165
Table 44 Distribution of marking with DO animacy ......................................................... 167
Table 45 Rate of marking according to specificity and definiteness .............................. 170
Table 48 Rate of marking according to TTP and all human DOs in the data ............... 183
Table 49 Factor groups selected as significant in each dialect ................................. 186
List of Figures

Figure 1  Animacy scale ................................................................. 17
Figure 2  Extended animacy scale .................................................... 17
Figure 3  Yamamoto’s general animacy scale ..................................... 18
Figure 4  Referents susceptible to extension of animacy .................... 21
Figure 5  Definiteness scale (Aissen 2003:444) ................................ 26
Figure 7  Individuation scale ............................................................ 33
Figure 8  Implied scale of NP properties affecting SDOM .................. 34
Figure 9  Parameters of high and low transitivity ............................... 37
Figure 10 Properties of individuation ................................................. 163
CHAPTER 1: Introduction

1. Phenomenon

Spanish manifests a common yet interesting case of the linguistic phenomenon known as Differential Object Marking (DOM). DOM denotes a morpho-syntactic distinction between overtly marked and non-marked direct objects, where the variation between marking and non-marking is largely based in semantic, pragmatic and discourse features. Spanish provides a clear case of DOM in the variable marking of (usually) human direct objects. The marking observed in Spanish has been called by several names in the literature including personal a, accusative a, and prepositional accusative as well as the Spanish equivalents a personal, el acusativo preposicional. Having originated from the preposition a, the subsequent marking of accusatives in SDOM, the descriptive quality of the titles is apparent.

In the normative grammar of Spanish the prototypical case of DOM is to mark human direct objects with an ‘a’ while non-human direct objects are unmarked.

1a)  
\begin{align*}
\text{Yo veo a mi amigo} & \\
\text{I see A my friend} & \\
\text{‘I see my friend’} & 
\end{align*}

1b) *Yo veo mi amigo  
\begin{align*}
\text{I see my friend} & \\
\text{‘I see my friend’} & 
\end{align*}
2a) *Yo veo la película*
   I see the movie
   ‘I see the movie’

2b) *Yo veo a la película*
   I see A the movie
   ‘I see the movie’

Nevertheless, this analysis is prescriptive: it presumes categorical application of marking. In spoken Spanish, DOM is attested in numerous contexts that do not follow this over-simplified model of the grammar.

3a) *Es muy difícil de definir un porteño* (HC Buenos Aires)
   It's very difficult to define Ø someone from Buenos Aires

3b) *Me voy a quedar aquí una hora, esperando a las notas*
   (HC Madrid)
   I'm going to stay here an hour, waiting for A the grades.

In 3a we find a human direct object that is not marked, while 3b provides an example of marking with inanimates. The prescriptive explanation would preclude these uses of SDOM despite being actually attested in speech. While these examples are often overlooked, the view here is that these types of variable marking can be incorporated into part of the broader account of SDOM based upon multiple factors, not solely depending on the animacy of the direct object.

As a result of the imprecise summation of marking, numerous studies have sought to address inherent features (properties of the direct object referent) and contextual features (properties of the clause and discourse) that motivate DOM in Spanish i.e.,
Spanish Differential Object Marking (SDOM).\(^1\) In other words, there is no simple explanation that adequately accounts for the actual usage of SDOM in spoken Spanish. While animacy of the object remains the most prominent and frequently recognized factor in \(a\)-marking (Pensado 1995), it is neither necessary nor sufficient for SDOM when actual usage is considered. Other significant features include the semantic/pragmatic notions of specificity and definiteness of the accusative object (von Heusinger & Kaiser 2003), as well as the discourse-pragmatic concepts of discourse prominence (Dumitrescu 1997), "saliency" (Weissenrieder 1990, 1991) and topicality (Leonetti 2004). The distribution of \(a\) in Spanish shows that there is widespread variability between marking and non-marking, even in the purportedly stable realm of animacy. The fact that DOM is variable in Spanish presents an ideal situation for quantitative variationist study; there is observable variation that can be quantitatively and empirically measured to substantiate the factors that contribute to marking.

1.1. Objectives and research questions

The primary objective of this dissertation is to provide a quantitative account of the distribution of DOM in Spanish i.e., to account for the variation that exists in SDOM. Since no single factor accounts for the observed variation of \(a\)-marking, I assume the

\(^1\) As a convention throughout this dissertation I will maintain a distinction between the particular case of Spanish Differential Object Marking or “SDOM” and the general, cross-linguistic phenomenon of Differential Object Marking or “DOM.” In this way I am able to appeal to the general linguistic phenomenon of DOM and also distinguish the particular case of SDOM.
hypothesis that there is a multiplicity of factors that favor and disfavor\textsuperscript{2} \textit{a}-marking.

Entailed by the hypothesis of multiple factors is the way in which these factors may relate to one another in accounting for variable marking. Therefore an important objective of this dissertation is to show how these multiple factors affect marking with respect to one another. Among these factors, I also elaborate a quantitative measure for defining a relative notion of topicality. Finally, because the corpora encompassed in this dissertation represent a very general sampling of prominent Spanish dialects, a quantitative description of dialectal variation\textsuperscript{3} is an additional objective to be undertaken here.

Underlying these objectives are several crucial research questions that will be used to guide the investigation:

- Assuming the hypothesis that multiple factors are necessary to explain the observed variation of \textit{a}-marking, what are the relevant linguistic factors that govern SDOM?
- How do these multiple factors relate to one another?
- What is the envelope of variation for SDOM? In other words, under what linguistic conditions is the variation possible? Likewise, when (if ever) is there categorical marking or categorical non-marking?

\textsuperscript{2} It is not enough to identify where \textit{a}-marking occurs, but also where it does not occur. This presents a particular challenge in identifying and quantifying the phenomenon in its entirety. This challenge will be addressed in Chapter 3, Methodologies.

\textsuperscript{3} In my prior research on this point (Tippets and Schwenter 2007, Tippets forthcoming), there is evidence that suggests important dialectal distinctions in \textit{a}-marking between at least three widely-spoken dialects in Spain, Mexico and Argentina.
- What quantitative metrics can be used to address and test the claims regarding SDOM as a reflex of topicality?

- What are the patterns of linguistic factors shared across dialects i.e., what universal constraints can be identified for a-marking across dialects?

- What differences can be appreciated across dialects among the relevant linguistic factors? In other words, what are the dialectal “nuances” of SDOM?

These questions guide the research undertaken in this dissertation and are directed at addressing the principal claims and assumptions in the literature about SDOM.

1.2. Significance of the project

A detailed and extensive quantitative approach represents a novel approach to providing a linguistic account of SDOM. The summary of the literature in Chapter 2 outlines the many discussions and analyses of SDOM but it is important to note that none provides quantitative analysis of the actual distribution and variation of a-marking in the spoken language. The variationist approach adopted here presents a means to test quantitatively the assumptions and explanations made in the literature, namely those particular features/factors that have been claimed to motivate a-marking.

The innovative contribution of this dissertation is to address the multiplicity of factors that have been attributed to the variation of SDOM by means of multivariate analysis in order to measure quantitatively the effect and relevance of the purported
factors at play.⁴ Although there have been quantitative approaches in measuring SDOM (Weissenrieder 1990, 1991; Buyse 1998), particularly diachronically (Company Company 2002, 2003; Laca 2003; von Heusinger 2008; among others), these have had two primary drawbacks. First, in almost every case the data used for the quantitative analysis is based on written text. If we consider that the locus of variation and change lies in the spoken language the lack of quantitative spoken data represents a significant gap in the SDOM literature.

The second significant drawback is that in these quantitative studies, the fundamental assumption that these factors may affect markings simultaneously has not been fully corroborated. Multivariate analysis allows for variation in SDOM across multiple factors. That multiple factors account for marking is not new to the SDOM discussion, however testing them quantitatively in simultaneous fashion as potential factors with respect to one another is a significant and novel contribution of this dissertation. The intention of this methodology is to analyze the common factors cited in the literature as well as test other hypotheses and claims with innovative metrics. The central goal of this dissertation is then to identify and describe those features that are relevant to SDOM in the spoken language and how they coalesce and influence marking with respect to one another based upon empirical evidence.

Providing quantitative evidence of actual language usage presents a far more accurate and linguistically defensible explanation for SDOM. A central shortcoming for

⁴ Ironically, the SDOM and DOM literature is replete with claims of numerous factors contributing to marking however no study has been done to consider how these factors relate to one another.
the explanations and accounts for SDOM up to the present time has been the reliance on anecdotal evidence, constructed examples and “curio” analyses i.e., extracting seemingly interesting examples using technical manuals or stories as a database (Weissenrieder 1990, 1991). These “impressionistic” accounts can only ever present partial explanations: a qualitative analysis and explanation without the quantitative evidence. I believe this shortcoming results primarily from the challenges of syntactic variation and the legwork necessary to obtain sufficient data. The syntactic variable can generally be poorly distributed as well as infrequently realized in the data, i.e. the paucity of consistent examples of the phenomenon and the problem of accounting for the unmarked cases of SDOM represent significant obstacles in undertaking an extensive quantitative study (Lavandera 1978, Silva-Corvalán 2001). In response to these concerns I see arising from the SDOM literature, this dissertation seeks to contribute a significant empirical perspective on a-marking.

This endeavor also seeks to quantify what Laca (2006: 430) has termed “global properties” i.e., the properties of the context in which a-marking occurs. As will be discussed below, much of the recent literature has concluded that the variation in a-marking observed in spoken Spanish ultimately depends on broad (global) discourse-sensitive features of topicality or prominence that may override (local) features of the Noun Phrase and its referent. The primary contributions of these types of analyses to the discussion of SDOM reside in acknowledging the multiplicity of factors that motivate SDOM, and recognizing the relevance of the discursive context. Unfortunately, in so doing, these claims relegate variation to unique discursive or contextual situations with
only anecdotal evidence (at best). Thus another critical contribution of this investigation is to quantify the broad discourse properties as well as the referent-specific properties observed in the data.

Additionally, the broad scope of the current study will add to our knowledge of Spanish dialects. In this case, the research will give greater insight into the morphosyntactic characteristics across the Spanish dialect continuum by analyzing three principal dialects of Spanish: Buenos Aires, Madrid and Mexico City. Each of these dialects represent Spanish spoken by large populations in regionally distinct areas of the Spanish-speaking world i.e., Peninsular Spanish, Southern-Cone Spanish, etc. Additionally, these dialects were selected based on the few existing dialectal claims found in the literature (Dumitrescu 1997, von Heusinger 2003). The findings of this research provide detailed descriptions of the distribution of SDOM among these dialects of Spanish.

1.3. Organization of the dissertation

This dissertation is organized in the following manner. Chapter 2 presents a review of the literature including previous and recent studies on SDOM and DOM phenomena. Particular interest is paid to the discussions regarding how α-marking is manifest in Spanish. Central hypotheses and theories are outlined in order to identify the critical factors and claims to be tested in the empirical and multivariate analysis. Chapter 5 Clearly, I do not presume to claim that either Peninsular Spanish or Mexican Spanish is homogeneous, rather I feel that the chosen dialects provide a broad sampling of the dialectal variety in the Spanish-speaking world.
Chapter 3 operationalizes the theories and factors so that the data can be interpretable for multivariate analysis. This chapter also details the constraints I impose on the data (the envelope of variation) as well as a detailed discussion of the morpho-syntactic variable. Chapter 4 looks at the combined data of the three dialects to provide the trends and general patterns observable from multivariate and quantitative analysis. In Chapter 5, the dialects will be analyzed individually to provide a more precise description of the differences in the grammar of SDOM. Chapter 6 will focus on the discussion and the principal findings of the dissertation as they relate to current approaches to SDOM. The conclusion of this dissertation is found in Chapter 7. The findings are summarized with particular emphasis on the contributions of this dissertation to current and previous analyses of SDOM.
CHAPTER 2: PREVIOUS RESEARCH ON SDOM

2. Overview of the Previous Research on SDOM

This chapter is organized around the primary linguistic features tied to SDOM as discussed in the literature. While DOM phenomena may occur based on various linguistic features, there exists a central group of linguistic factors that tend to motivate (or are claimed to motivate) DOM cross-linguistically. Of these central features, SDOM is claimed to be sensitive primarily to animacy while definiteness, specificity, and related notions of individuation and discourse-pragmatic functions are postulated to varying degrees. As will be elaborated in this chapter, these different features are very much interrelated and as such the contributions to understanding the SDOM phenomenon are a mixed bag in that the general intuitions and insights identify important uses of SDOM but they also show tremendous overlap. For example, both the NP properties of definiteness and animacy are subsumed under the notion of individuation. As a result, the literature reveals a general recognition of what is involved in SDOM but the discrete and precise interplay among the potential and probable factors remains opaque.

The purpose of this dissertation is to quantitatively test the hypotheses and explanations for SDOM based upon the claims that have been made in the literature. In order to do so, it is imperative to sort through the claims that have been made and understand where the fuzzy or disputed areas lie (as well as to rethink/consider the prevailing assumptions regarding the motivating factors for SDOM).
This chapter begins with the primary features of the DO that the literature cites which are claimed to correlate with SDOM. Preeminent among these features is animacy, outlined in section 2.1. I discuss how animacy is conceived of cognitively and what effects it has in linguistic forms and features. Section 2.2 outlines definiteness as discussed in the SDOM literature. I review the varying notions of definiteness and argue that the prevailing notion of definiteness (as applied to a-marking) is actually a conflated term that entails both the NP form and an independent feature, specificity. I outline specificity as a separate referential property of NPs in section 2.3. These varied features are then summarized under the parameter of individuation which I explain in section 2.4. The chapter then shifts focus to those features relevant to SDOM that extend beyond the properties of the NP to the level of the clause. I begin by exploring transitivity in 2.5 and how it relates to SDOM, including the concept of relative animacy in DOM. In Section 2.6 I discuss the disambiguating function of SDOM as it relates to the broader disambiguating hypothesis of DOM. This discussion necessarily addresses the semantic roles of Agent and Patient within the transitive clause. Examples will be outlined discussing the likely conditions where ambiguity may arise in Spanish and how SDOM resolves the ambiguity. Section 2.7 discusses the claims of how verbal semantics affect marking and what, if any, categorical conditions exist for SDOM. In section 2.8 I address the most encompassing approach to SDOM dependent on the discourse and discourse-pragmatic concepts of topicality and referentiality. 2.9 briefly summarizes the content of the chapter.
Relevant Factors/Features of the Direct Object

A central assumption I maintain in attempting to account for SDOM is that there are potentially multiple factors that contribute to $a$-marking. These factors may reside at several levels ranging from inherent factors of the accusative NP, to the clausal level or the discourse level. Laca (2006) succinctly partitions features into two broad groups, *global factors* and *local factors*, where local factors address features of the accusative object (DO) itself and global factors address the greater syntactic and discourse pragmatic context of the DO i.e., the transitive verb construction and the discourse context.

That the types of factors that contribute to the variation in $a$-marking vary at different levels has been observed by other researchers. Weissenrieder (1991) analyzes how marking is manifest at the lexical (NP), sentence (clause) and discourse levels. Focusing on marking of inanimate DOs, she addresses many of the primary factors that will be discussed below.

These approaches capture a fundamental observation regarding the SDOM literature: $a$-marking cannot be reduced to any one feature of the DO or to any other syntactic or pragmatic feature of the discourse context to account for the observed variation. As a result, much of the SDOM literature presumes that both unique features of the marked or unmarked DO along with characteristics of the context (information status, topicality, transitivity) are necessary to account for the variable manifestations of SDOM.\(^6\)

---

\(^6\) Cases of SDOM that do not vary are not of immediate interest to this dissertation because categorical uses of $a$ fall outside the envelope of variation (cf. 3.4.)
2.1. Animacy in SDOM

The most prominent factor for explaining SDOM remains animacy. So transparent is the relationship between $a$-marking and an animate (human) direct object, that for most all initial encounters with SDOM, second-language learners are taught that it is the “personal $a$.“ As an initial pedagogical explanation, they learn that animacy is the strongest predictor in accounting for $a$-marking. Indeed no discussion of SDOM has tried to discount the effects of animacy. Nevertheless, $a$-marking does not quietly conform to the simple premise of marking [+human] accusatives. In spontaneous spoken Spanish a learner would be surprised to hear “exceptional usages” of $a$-marking with non-human direct objects, but also the lack of $a$-marking with direct objects. While this should not diminish the primacy of animacy in motivating SDOM, it is evident that variation occurs. Some of this variation results from the gradient nature of the linguistic conceptualization of animacy.

The linguistic approach to animacy appeals to two fronts. First, there is a cognitive awareness of the animacy status of entities (cf. Yamamoto 1999). Secondly, the linguistic behavior, whether it be morphological, syntactic, or pragmatic, as manifest cross-linguistically, directly results from the animacy status of the entity or entities involved. I therefore assume that animacy maintains a cognitive presence that is manifest in linguistic form. Unlike specificity/definiteness, animacy is an inherent feature of the lexical referent (de Swart 2007:174) and its relevance is found in the linguistic features of innumerable languages.
The uncontroversial argument that animacy is a central feature finds support in numerous cross-linguistic phenomena. While DOM is of preeminence for this dissertation, it is relevant to mention in passing that animacy effects can permeate a grammar. Croft (2003) outlines several linguistic patterns sensitive to animacy, especially pronominal and nominal distinctions found in languages (128-29). Dahl and Fraurud make the observation that “…there is a strong connection between the animacy of a referent and the choice between different ways of referring to it” (1996:56). Additionally, syntactic and morphological distinctions based on animacy are found in verb agreement (Maya-Quiché), the genitive (possessive) constructions in English and in case marking. It is within the area of case marking that Bossong (1985) originated the notion of Differential Object Marking in describing the common shared linguistic phenomenon of variably marking direct objects depending on their particular semantic or pragmatic features. One of the prominent features Bossong identified for motivating many cases of DOM more generally, and specifically for SDOM is animacy.

It is significant to note that the relevance of animacy in Spanish is not limited to the SDOM phenomenon. The relevance of animacy to the Spanish grammar is discernable in other areas of the language. Of particular interest is the subject pronoun constraints associated with animacy. Subject pronouns maintain a gender distinction in 3rd person that does not extend to the arbitrary gender assignment of non-human nouns. In Spanish, a 3rd person human referent may be represented by a subject pronoun él/ella or by a null subject. Non-human referents generally cannot be pronominalized and must either be realized as a full lexical NP or be realized as a null subject (cf. Jensen 1973).
It can be deduced from examples (1a-c) that animacy distinctions between human and animate, and between animate and inanimate are relevant and salient to the grammar of Spanish.  

Finally, the dialectal feature termed leísmo further demonstrates the sensitivity Spanish-speakers (and their underlying grammar) have to animacy. While there exist variations in the precise realization of leísmo, the most common realization of leísmo involves a third person accusative object with human, usually masculine reference (normally pronominalized as lo/los as shown in 2a) taking the dative pronoun le/les as seen in 2b, whereas non-human referents are pronominalized with the expected accusative lo/los as seen in 2c.

Further gradation in animacy may also be exemplified in the gender distinctions found among animals. Common animals in Spanish have both a male and female counterpart: perro/perra ‘dog/bitch,’ vaca/toro ‘cow/bull,’ caballo/yegua ‘horse/mare’ whereas less common animals have an arbitrary and invariant gender: la girafa ‘the giraffe (female or male),’ el gorila ‘the gorilla (female or male).’ It can be argued then that the level of animacy (common verses uncommon) among animate referents is observed in the grammar.
The aforementioned examples in Spanish show that animacy has an effect in the grammar. However beyond the evidence of animacy-based grammatical phenomena in Spanish, DOM in its cross-linguistic manifestations frequently correlates with the animacy of the DO. Animacy as a central property has been treated extensively in variable marking of objects across languages. (Comrie 1979, 1989; Aissen 2003; de Swart, 2007; Malchukov 2008). The general observation has been that animate, especially human objects are more marked by virtue of semantically and grammatically being more akin to subjects, particularly as a result of their animacy (in addition to being frequently definite). Objects that do not look like the prototypical DO, inanimate and indefinite are therefore marked relative to the properties and roles typically associated with DOs.

Assuming animacy is relevant to language, and particularly to Spanish, how should one categorize animacy? Yamamoto (1999) argues that humans perceive a cognitive distinction between animate and inanimate and also human and non-human entities. This basic distinction can be shown as a scale in Figure 1.
Human > Animate > Inanimate

Figure 1 Animacy scale

From the inherently anthropocentric view, human is considered “most” animate, non-human animates are less animate than humans but higher in animacy than inanimates. However, this basic notion belies the complexity of animacy and its relationship to DOM.⁸

Animacy-based phenomena cannot be explained in their totality using this tripartite distinction and more discrete models of animacy are needed. Dixon (1979:85) outlines a typology of animacy that crucially acknowledges person, referentiality and animacy as shown in Figure 2. (from Croft 2003:130).

First/second person pronoun < third person pronoun < proper names < human common noun < nonhuman animate common noun < inanimate common noun

Figure 2 Extended animacy scale

This extended scale stems from cross-linguistic evidence supporting the relevance of these categories in linguistic behavior. That is to say, languages manifest syntactic and morphological behavior that is sensitive to these discrete categories (cf. Croft 2003:5.2; Yamamoto 1999, Chapter 2).

⁸ One of the important cross-linguistic observations of DOM has been to show that marking may vary along the animacy scale with respect to what is sufficiently animate to warrant marking, a point that will be discussed when looking at the effects of animacy in the Spanish data of this dissertation (Section 4.2.)
Another discrete typology of animacy is found in Yamamoto (1999). The author presents a spatial representation he calls the General Animacy Scale shown in Figure 3 (1999:22).

Figure 3  Yamamoto's general animacy scale

As is readily apparent, the term “scale” can be a misnomer because Yamamoto constructs her animacy hierarchy following Lakoff’s radial prototype categorization (1987: Chapter 6). Yamamoto represents all potential entities within her General Animacy Scale and relies on the spatial representation to reveal the relative proximity to
the prototypical animate: the individual human being. It is unclear how to relate any of the central branches to one another i.e., is an anthropomorphized animal more or less animate than a supernatural being?, etc. Nevertheless, Yamamoto’s extensive representation succeeds in acknowledging the diverse ways in which animacy could be perceived.

Like Dixon and Croft, Yamamoto recognizes the relevance of grammatical person in animacy effects on language. Unlike Dixon and Croft, Yamamoto conceptualizes person in different terms; there is an interaction between the General Animacy Scale parameter and the Hierarchy of Persons i.e., 1st/2nd person > 3rd person. In addition to the Hierarchy of Persons, the other “core” interacting parameter is Yamamoto’s Individuation Scale,9 based on Dahl and Fraurud’s (1996) degree to which an entity is a “clearly delimited and identifiable individual.” Yamamoto’s individuation scale is based on two oppositions, the first between individualized humans and humans of certain categories and roles (akin to Croft’s human common noun) and the second between persons as individuals and persons as a mass. The author provides insight into the “grey area” where cognitive conceptualizations of animacy are not as clear cut as [+/- animate] or even human > animate > inanimate. An individual may be recognized by their given name, but also by their occupation, relationship and kinship. Yamamoto points out that a cognitive level of animacy of the referent is reflected in the NP forms used in a given

---

9 Note that the Individuation Scale Yamamoto discusses is never construed as a scale. It appears to be based on Hopper and Thompson (1980) but it is tenuously referenced. The general notion borrowed from Dahl and Fraurud may be completely distinct from Hopper and Thompson’s conceptualization of *individuation* (cf. Section 2.5)
context. Additionally, many NP forms are inherently inanimate but are comprised of animates (i.e. humans): faculty, family, team, company, etc. On a cognitive level these lexically forms can be perceived to behave as animate, sentient entities.

The efforts of Yamamoto, Dahl and Fraurud, Dixon, and Croft illuminate the complexity of animacy as it relates to language, providing valuable insight into the discussion of DOM and SDOM. Many of their observations contribute directly to the quantitative analysis of referents undertaken here. However, it is also important to note that certain components will not be taken into consideration because they are not relevant to the grammatical conditions of SDOM. Specifically person distinctions (1\textsuperscript{st}, 2\textsuperscript{nd}, etc.), while important for animacy-based linguistic phenomena, do not concern my analysis of SDOM because marking is invariable in relation to pronominal forms (cf. 3.4). Nevertheless, two important points arise from this discussion of animacy: animacy is a linguistically relevant feature of language and it has a necessarily cognitive dimension.

To further demonstrate the cognitive value of animacy one can analyze the metaphorical or metonymic extension of animacy (i.e. human animacy) to animate (non-human) and inanimate referents via personification, anthropomorphization, etc. In this sense, some variation in $\alpha$-marking may be accounted for and has been discussed in the SDOM literature. Discussing metaphorical extension, Pensado (1995: 31) states that prototypical cases of $\alpha$-marking are animates, and as such, marking endows the referent with human-like qualities. She further elaborates by proposing the inverse condition, where inanimate properties or characteristics may also be extended to animate entities.
indicated by non-marking i.e., dehumanizing the referent. However, she also asserts that metaphorical uses should not be thought of as systematic or regular. Torrego-Salcedo (1999:1800) also states that the metaphorical extension of animacy may account for some unanticipated *a*-marking. Dahl and Fraurud (1996) acknowledge metaphorical extension as one example where properties of animacy may be extended to different referents. Figure 4 summarizes other common referents that may be considered animate (from Dahl and Fraurud 1996:62).

- Metaphorical extensions: personification, anthropomorphization
- Metonymical extensions: *country* referring to its inhabitants
- Collective nouns: *family*
- Non-personal agents: *institutions, companies, associations*
- Mythological beings

Figure 4 Referents susceptible to extension of animacy

It is unsurprising to find that these referents readily permit *a*-marking in Spanish. The inherent risk arises in assuming that the animacy extension is only made when *a*-marking occurs; I do not assume that the extension of animacy is only observed via SDOM, merely that these referents can be conceptualized as animate, with or without marking.

On a broader level and to conclude the discussion of metaphorical extension, because extension of animacy is cognitive and not linguistic *per se* the opportunity to evaluate the animacy level with a metric such as SDOM presents a remarkable opportunity. To what

---

10 This provides an excellent topic for further research since little has been written on the subject.
degree there is variability in marking of these non-prototypical “animate” referents becomes another contribution of this dissertation.

To conclude this section, animacy most certainly correlates with SDOM. It has been shown that languages reflect sensitivity to the inherent animacy of a referent in varied linguistic phenomena. Within Spanish animacy effects manifest themselves in syntactic constraints (pronominalization), dialectal variation (leísmo) and most transparently in a-marking.

2.2. Definiteness in SDOM

Definiteness presents a particular challenge to this study not only for its primacy in the DOM and SDOM literature but especially as a result of the varied and disparate conceptualizations encompassed by the term. Further compounding these polemic circumstances is the frequently intertwined and potentially independent feature of specificity. I survey how definiteness has been applied in the SDOM literature. Additionally, I characterize the role of specificity in relation to definiteness. For the purposes of this dissertation I adopt an approach to definiteness that distinguishes the referring or (in)definite expression (i.e. the morphological form) from specificity, an independent feature of the referent determined by the discourse.

One of the fundamental conceptualizations of definiteness relies on the distinction a language makes between a definite article and an indefinite article, or more generally,
between definite and indefinite expressions. This most basic (and anglo-centric\textsuperscript{11})
dichotomy of definite and indefinite forms represents the basis for many discussions of
definiteness phenomena. In general, definite expressions include the following
grammatical forms: null NPs, demonstratives, possessives, personal pronouns, proper
names, universal quantifiers; indefinite expressions include the indefinite article, cardinal
numbers and existential quantifiers. These referring expressions have been the basis for
definiteness and I will return to their syntactic (structural) condition as part of my
quantitative methodology (section 3.5.5).

The conceptualization of (in)definiteness is inseparably linked to referentiality.
Much has been discussed regarding how and to what extent definite expressions are
referring expressions. The use of a definite expression generally signals that there is a
particular referent that is available in the discourse. Appropriately called the \textit{familiarity
hypothesis} (Christophersen 1939, Hawkins 1978), definiteness is identifiable to the
speaker and hearer. Familiarity may be achieved via several means: situation/context,
general knowledge (of how the world works), anaphora (the actual linguistic information
given) and/or bridging cross-reference/associative anaphora.

In her approach to definiteness, Heim (1988) formalizes the approach of
Christophersen and Hawkins: definites have familiarity presuppositions while indefinites
lack such presuppositions. Therefore definiteness is determined based on the familiarity

\textsuperscript{11} Many languages get along just fine without articles, let alone a distinction between
definite and indefinite articles. The “typical” characterization based on English becomes
a fundamental flaw to elaborating a unifying theory of definiteness. This is not to say
that these languages don’t have other expressions that distinguish (in)definiteness, but as
I will conclude, definiteness as commonly construed in DOM and SDOM literature may
be irrelevant anyway.
of a referent in the discourse. Lyons (1999) summarizes these familiarity approaches as the identifiability hypothesis, meaning that the definite expressions are related to the identifiability of the referent.

The assumption that speakers and hearers are readily able to identify referents belies the complicated circumstances regarding what is referentially available in the discourse. Beginning with the seminal work on information status and reference by Prince (1981), there have been many proposed typologies to account for the information status of referents in discourse (Gundel et al. 1993; Birner 2006a, 2006b). While this dissertation does not strive to redefine these possible paradigms of information status, it is imperative to identify a minimal level that can capture potentially meaningful distinctions in information status as related to SDOM.

For the scope of my quantitative approach, I find the typology of Birner (2006b) to be sufficient for the purposes and goals of this dissertation. Birner distinguishes four types of information status based on discourse new/old and speaker new/old as shown in Table 1. (cf. 3.5.8 for examples)
<table>
<thead>
<tr>
<th>Discourse-old:</th>
<th>Hearer-old: Evoked (inferentially linked and known to hearer)</th>
<th>Hearer-new: Bridging Inferrable (inferentially linked, but not known to hearer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse-new:</td>
<td>Unused (not inferentially linked, but known to hearer)</td>
<td>Brand-new (not inferentially linked, and not known to hearer)</td>
</tr>
</tbody>
</table>

Table 1 Information status (from Birner 2006b)

In this case Birner has focused on the referent as it relates to the hearer’s perspective and knowledge of the world and the status of the referent in the discourse. Discourse status describes whether the referent mentioned in the discourse has been previously established in the discourse (discourse-old) or has not been previously established (discourse-new). A referent is discourse old if it has been made explicit by prior mention or is inferrable from the prior discourse content (see pg. 25). Hearer status relates to whether the referent is new or old (already known) to the speaker. Hearer Old indicates that the information (and in the case of SDOM, the referent) is assumed to be known and identifiable by the hearer. In the case of hearer-knew information, it is impossible for the hearer to anticipate or know the referent until the speaker explicitly mentions the information/referent or the speaker’s contribution forces the hearer to infer the information/referent (i.e., a bridging inferrable).

Birner’s typology presents an appealing and straightforward approach that can be profitably applied to SDOM because it captures what I estimate as the four necessary distinctions of information status that I encounter in the data. This approach allows me to readily quantify DO referents as they potentially correlate to SDOM variation.
The purpose of this discussion of information status as it relates to definite expressions does not intend to resolve the debate regarding the ideal means for identifying what is presupposed or what is referentially available in discourse, rather to acknowledge that definiteness, in its principal conceptualizations relies upon referring expressions. Many of the means to determine (in)definiteness depend on referring expressions and predominate both the DOM (Aissen 2003) and particularly the SDOM literature.

Within the DOM literature, forms considered definite have generally been ordered on a scale from more to less definite. However, the relative discreteness of the different determiners has led to varied scales of definiteness depending on the aims (and languages) under consideration. A basic typological construal of definiteness is that of Croft (2003), who use a simple scale of the form (definite > specific > non-specific). Aissen (2003) argues that this “short-form” scale is insufficient to account for DOM systems based upon differing cross-linguistic sensitivities to these categories and (non)-marking. As a result, Aissen assumes the following definiteness scale:

Pronoun > personal name > definite > specific indefinite > non-specific (indefinite)

Figure 5 Definiteness scale (Aissen 2003:444)

Laca (2006) adopts the central tenet of Aissen: that SDOM can be characterized by the scalar factors of animacy and definiteness. While assuming the animacy scale of
Aissen, Laca reorganizes the definiteness scale to address other important categories. These include types of pronouns, universal quantifiers and existential quantifiers.

Personal Pronoun > Personal Name > Definite || Universal > Existential indefinite > no determiner

Figure 6 Definiteness scale (from Laca 2006:439)

In Figure 6, Laca elaborates Aissen’s scale, with particular attention to definiteness as it relates to Spanish: to include pronouns with non-lexical nominal heads (qué, quien, que, quien, el cual, etc.), bare quantifiers (alguien, nadie, algo, etc.) and pronominal uses of determiners (pg. 437-38). In her schema definites include the definite article, demonstratives, possessives and universal quantifiers (todos, cada, ambos). Indefinites consist of the indefinite article, numerals and existential quantifiers (algunos, pocos, muchos).

2.3. Specificity

Identifying and categorizing definiteness based upon referring expression has been the modus operandi for much of the research on DOM and SDOM (Kliffer 1984, 1995; Aissen 2003; Laca 2005; de Swart 2007; Shain 2009); unfortunately determining definiteness based on form becomes problematic particularly when it is evident that the common approaches to definiteness conflate the form of the referring expression and specificity. Many have assumed that specificity is hyponymous to definiteness in that specificity distinctions are only maintained with indefinite forms (Aissen 2003, Enç
The example *por excelencia* involves the ambiguity present in indefinite articles:

3a) *I'm looking for a painter (that was here last week)*

3b) *I'm looking for a painter (that is willing to do a portrait pro bono)*

Here the referring expression, the indefinite article, can have both a specific (3a) and non-specific interpretation (3b) and the additional discourse content provided in the parentheses is necessary to resolve the ambiguity.¹³

Enç (1991) proposes a Familiarity Condition and a Novelty Condition for defining definiteness in NPs. All indefinites in a sentence must be novel i.e., referents new to the discourse; and all definites must be familiar i.e., mapped onto a previous referent. Indefinites cannot have antecedents but definites must. This approach crucially depends on discourse to account for both specificity and definiteness.

Similar to the proposal by Enç is von Heusinger and Kaiser’s conceptualization of definiteness: “definiteness is a discourse-pragmatic property that indicates that the discourse referent associated with a definite expression can be identified with an already introduced discourse item…definiteness expresses familiarity in discourse” (2003:44).

¹² This approach therefore assumes that all definite expressions must therefore be specific, an assumption that will be questioned and challenged below.

¹³ Tipping my hand, specificity as I approach it in this research, is a discourse-pragmatic property that can only be determined in each unique context without necessarily depending on the referring expression.
This definition appears to be limited by the assumption that there must be anaphoric relationships, that definiteness is based solely on whether a referent is explicit (realized overtly) in the discourse.

Likewise, Spanish also encounters ambiguities between referring expressions and the specificity of the referent. The conditions associated with the indefinite article in English as exemplified in 3a) and 3b) apply equally well in Spanish. However, ambiguity regarding specificity also extends to definite articles in Spanish (and other languages):

4a) *El chocolate (que estamos comiendo) proviene de México*
   The chocolate (we are eating) comes from Mexico

4b) *El chocolate proviene (históricamente) de México*
   Chocolate historically comes from Mexico

Spanish frequently employs the definite article for referents that are not unique and identifiable, as in (4b) where there is a generic use of `chocolate.’ This is an especially important contrast to English where no determiner is used with generics. Thus the feature of specificity is not limited to indefinites but may also be expressed in definites.

One final example for the independence of referring expressions (traditionally “definiteness”) and specificity is impersonal you/tú. Shared equally by both English and Spanish, this construction relies on a personal pronoun (you/tú), assumed to be one of the most definite and therefore specific referring expressions, to express a generic and non-specific referent.
In these cases, it is possible to substitute the impersonal *you/tú* for the impersonal “one”. These variations in specific and non-specific readings indicate the insufficiency of morphological form in determining a discrete category of definiteness that also entails specificity. In other words, definiteness as one single category insufficiently accounts for what are two observably distinct features of a referent: the referring expression and specificity. I have therefore come to much the same conclusion as Abbott (2004:147) that definiteness fails to correspond neatly to a single formal and universal linguistic category. As a result of these observed conditions in Spanish, I will argue that relevant linguistic features of referents result from deconstructing the traditional concept of definiteness into two components: specificity of the referent and definiteness, meaning the form of the NP.

My insight is not necessarily novel; others have argued that definiteness and specificity are separate properties, albeit with distinct meanings for definiteness at times. Schwenter and Silva (2002), von Heusinger and Kaiser (2003) Pensado (1995: 32-3) and Laca (2006: 464-7) all argue that definites and indefinites (i.e. their referring expressions) may have specific or non-specific referents.
In their discussion of the distinct properties associated with SDOM, von Heusinger and Kaiser present support for their claim of the separate classification of definiteness and specificity with the following evidence in table 2:

<table>
<thead>
<tr>
<th>Specificity</th>
<th>[+definite]</th>
<th>[-definite]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+specific]</td>
<td><em>la mujer que sabe inglés</em></td>
<td><em>una mujer que sabe ingles</em></td>
</tr>
<tr>
<td>[-specific]</td>
<td><em>La mujer que sepa inglés</em></td>
<td><em>una mujer que sepa inglés</em></td>
</tr>
</tbody>
</table>

Table 2 Cross-classification of definiteness and specificity

The authors have availed themselves of mood to indicate specificity, where subjunctive indicates non-specific, and have used the articles to indicate definiteness. A few comments are in order regarding their argument.Ironically, von Heusinger and Kaiser later (2003) assume a discourse pragmatic concept of definiteness, distinct from what they are calling “definiteness” in this table. The table uses the generally held concept of definiteness referring to the lexical or morphological form of the determiner (D), however they are drawing attention to the independence of specificity from the article, i.e. the morphological form. This is the assumption that I make regarding SDOM. Mood (subjunctive vs. indicative) is one means to determine specificity, however it should not be assumed that it is foolproof, nor the only means to determine the specificity of the referent.\(^{14}\) Table 2 provides a useful paradigm to demonstrate that definiteness and specificity are indeed distinct and independent features of referents.

\(^{14}\) As established, I follow Lambrecht (1994) and Leonetti (2004) in assuming that specificity is a discourse-determined property.
From this review of the literature and subsequent discussion, in addition to animacy and definiteness, specificity makes up the third frequently cited feature that has been correlated to SDOM behavior. As argued, it is not the referring expression (definite or indefinite form) but contextual information that determines specificity. So while we may conceptualize referents as having a feature of +/- specificity, that feature is discourse-determined and not dictated based on the form i.e., the referring expression. Specificity derived from the discourse is an insight shared by several other researchers (Lambrecht 1994, Leonetti 2004) and has significant weight in my methodological approach to quantifying the data for this dissertation. The crucial perspective is that specificity is independent from referring expressions though there is frequent correlation between definite and indefinite forms/expressions and specific and non-specific referents, respectively.

2.4. Individuation

This discussion regarding properties of the DO, namely animacy, specificity and definiteness are characteristic of another important linguistic concept that is frequently associated with DOM: *individuation*. Related primarily to the transitivity relationship, individuation refers to the relative (scalar) distinctness or uniqueness of a patient (P), particularly in contrast to the agent (A). Hopper and Thompson (1980, 1995), define individuation as a scale of properties shown in Figure 7 (pg. 253):
**More Individuated** | **Less Individuated**
---|---
Proper/personal | Common
Human/animate | Inanimate
Definite | Indefinite
Referential | Non-referential
Singular | Plural
Count | Mass

Figure 7 Individuation scale

Individuation is a gradient notion, such that the properties in the left-hand column of Figure 5 are typical of highly individuated referents (NPs) while the properties in the right-hand column are typical of referents with low individuation. In this way referents can be categorized as more or less individuated with the presumption that linguistic behavior is sensitive to the individuation of the referent.

Because of the overlap in the properties of individuation and those attributed to SDOM it is unsurprising to find claims that highly individuated referents in Spanish correlate with *a*-marking, i.e. the individuation hypothesis. Kliffer (1995: 99, 102-05) makes claims about the relative strength of these features of individuation. Some of these claims can be constructed as the following scale of factors that form a hierarchy of strength favoring SDOM:
Animacy > definiteness > number > count/mass

Figure 8  Implied scale of NP properties affecting SDOM

Kliffer does not make claims nor provide examples about relative constraints for the other properties associated with individuation, hence the implied nature of the scale in Figure 8. This implied scale also illustrates Kliffer’s claim (shared by many) that animacy overrides all other properties of the NP in SDOM, definiteness may override number and count/mass properties but not animacy in SDOM. Number may override the count/mass distinction but not definiteness nor animacy, etc.

The underlying assumption is that NPs that are more highly individuated are more likely to be a-marked than NPs that are less individuated. Kliffer is not alone in his view of the role of individuation. Laca (1995:77-82) also argues that individuation is manifest in the distribution of SDOM, accounting not only for common contexts of animacy and specificity but also for marking or non-marking in generic, partitive constructions and quantity constructions. Cennamo (2003) has also commented that marking is likely to occur when the subject and object are highly individuated. Weissenrieder (1990) applies the properties of individuation to marked and unmarked non-human animates (i.e. animals) from a collection of Spanish texts. The correlation is not complete but only relative in that “while highly individuated animal nouns are not necessarily marked... where marking does occur, highly individuated nouns are more [frequently] marked than less individuated ones” (229).
The super-ordinate property of individuation circumscribes the critical characteristics of the DO that the literature indicates may affect SDOM. It comes as no surprise that researchers have availed themselves of individuation: it acknowledges the coalescence of multiple potential factors in explaining SDOM. The particular constellation of individuating properties of animacy, definiteness and specificity are tied to SDOM. Since individuation is a convenient term that allows for gradience, i.e. an object or a subject is measured in terms of being more or less individuated, it allows for the necessary flexibility in addressing the variation of SDOM. However, quantifying animacy, definiteness and specificity is crucial, in addition to the other properties, number, mass/count, etc. This would provide a more complete measure of, or at least a means to test, the validity of an argument for SDOM based on highly individuated NPs. The significance of the individuation hypothesis is the assumption that multiple features are involved in SDOM; there is a confluence of interrelated factors that collectively account for the variation in marking. Importantly the multivariate analysis presented in this dissertation is able to disentangle these factors and see the relative contributions of these factors (cf. Chapter 4 and Chapter 5).

The literature and research on SDOM has determined a particular group of features of the DO that are potentially relevant to the observed behavior of SDOM. Animacy is the most prevalent factor that explains SDOM, while definiteness and specificity have also been assumed to further motivate marking. The literature generally presumes that multiple factors are necessary to explain SDOM since no single feature can account for the observed variation. In even broader terms, the concept of *individuation*
provides a means to think about how animacy, definiteness and specificity along with other features of the DO may contribute to a general explanation regarding preferred DO features for *a*-marking. Kliffer (1984, 1995) hypothesizes about the potential relative strength among these several features, asserting that animacy is the one feature that can trump any other feature in affecting *a*-marking. This idea is echoed by de Swart (2007), albeit for slightly different reasons (see below).

In surveying the literature, the general intuition of the preeminence of animacy, and to a lesser extent other features of the DO, reflects the most observable cases of SDOM, i.e. marking human referents, which are usually also definite and specific. Accepting that there is more than one feature of the DO motivating SDOM is a critical impetus for ascertaining the factors that motivate variable *a*-marking. The next step is to consider the transitive structure and discourse context that are equally relevant to SDOM.

### 2.5. Transitivity

SDOM is a phenomenon associated with the transitive relationship: it marks an accusative objectand is not to be confused with dative and oblique NPs introduced by the preposition *a*. The transitive relationship presents insight into SDOM. In their seminal work, Hopper and Thompson (1980) establish important characteristics of transitivity as it operates cross-linguistically. Transitivity is a global property of an entire clause. It is a relationship that obtains *throughout a clause*. The defining properties of transitivity are discourse-determined. The property of transitivity is therefore a continuum, i.e. it is
scalar/gradient and determined by the high or low values of the parameters shown in Figure 9 (from Hopper and Thompson pg. 252):

<table>
<thead>
<tr>
<th>Parameters</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>2 or more participants A and O</td>
<td>1 participant</td>
</tr>
<tr>
<td>Kinesis</td>
<td>action</td>
<td>non-action</td>
</tr>
<tr>
<td>Aspect</td>
<td>telic</td>
<td>atelic</td>
</tr>
<tr>
<td>Punctuality</td>
<td>punctual</td>
<td>non-punctual</td>
</tr>
<tr>
<td>Volitionality</td>
<td>volitional</td>
<td>non-volitional</td>
</tr>
<tr>
<td>Affirmation</td>
<td>affirmative</td>
<td>negative</td>
</tr>
<tr>
<td>Mode</td>
<td>realis</td>
<td>irrealis</td>
</tr>
<tr>
<td>Agency</td>
<td>A high in potency</td>
<td>A low in potency</td>
</tr>
<tr>
<td>Affectedness of O</td>
<td>O totally affected</td>
<td>O not affected</td>
</tr>
<tr>
<td>Individuation of O</td>
<td>O highly individuated</td>
<td>O non-individuated</td>
</tr>
</tbody>
</table>

Figure 9 Parameters of high and low transitivity

These parameters provide an important insight into the gradient nature of transitivity.

Understanding transitivity in this detail provides a more informed approach to analyzing SDOM phenomena because we can consider the variable marking in light of the transitive *relationship* between subject and object.

Despite the seeming plausibility of an animacy-based account of SDOM, *a*-marking based on animacy alone does not occur categorically; even when accounting for metaphorical extension, in actual usage one finds both unmarked animate accusatives and
likewise, marked inanimate accusatives. This evidence strongly suggests that a more nuanced explanation of SDOM is needed that does not depend on animacy alone.

Animacy has been initially presented as a property inherent to referents, and within the scope of SDOM literature, the primary locus of animacy has been in the accusative NP. However the animacy of the accusative object should not be presumed to exist independently of the animacy of other referents and this is particularly the case when one considers that SDOM is inherently tied to the transitive relationship involving an agent and a patient. This is in fact one of the central insights of de Swart (2007) in outlining his cross-linguistic analysis of variation in DOM, the shared inherent feature of animacy by both subject and object making it an explanatory factor of variation in marking.

With the addition of acknowledging the animacy of the S, another means for thinking about SDOM emerges. As conceived, animacy is a scalar property and therefore it is possible to establish a metric comparing the animacy of the Agent (subject) and the Patient (direct object) known as relative animacy. This insight about the animacy levels between Agent and Patient has figured prominently in the discussion of DOM phenomena, albeit as part of broader theoretical approaches to DOM.

One of the central theoretical assumptions of SDOM and DOM in general is couched in the functional typology of markedness. Aissen (2003), Bossong (1991) and Comrie (1979, 1989) all assert that the behavior of DOM cross-linguistically demonstrates a preference for using a more marked construction i.e., DOM when an object manifests properties associated with agentive referents. It is in this context that
DOM has been referred to as “iconic” because there is an overt grammatical convention that indicates the “markedness” of the construction depending on the requisite features relevant to marking in a given language. Comrie posits the following explanation for markedness in DOM:

Although in principle either A [agent] or P [patient] can be either animate or definite, it has been noted that in actual discourse there is a strong tendency for the information flow from A to P to correlate with an information flow from more to less animate, and from more to less definite. In other words, 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. This has implications for a functional approach to case marking: the construction which is more marked in terms of the direction of information flow would also be more marked formally… (1989:128).

Relative animacy is then part of this theoretical perspective: it is the animacy of the object in relation to the animacy of the subject that potentially explains DOM. Theoretical assumptions aside, the potential relevance of relative animacy is all but absent from the SDOM literature with the exception of the related disambiguation hypothesis (see below).
2.6. Disambiguation

Disambiguation as a possible motivation for SDOM is commonly found in the literature. In her approach to transitivity, Næss (2007) presents a hypothesis based upon the transitive prototype that does not include the notion of markedness. Næss bases her hypothesis on empirical cross-linguistic data, proposing that all languages have a transitive prototype based on semantic properties, and more specifically, that the event type characterized by the transitive prototype is the marked two-participant event (pg. 25). The critical summary for the transitive prototype is the Maximally Distinguishable Arguments Hypothesis: “A prototypical transitive clause is one where the two participants are *maximally semantically distinct*\(^{15}\) in terms of their roles in the event described by the clause” (pg. 30).

A particular function of DOM immediately comes to mind relating to the Maximally Distinguishable Argument Hypothesis: the disambiguating function of SDOM where two arguments can be perceived as equal candidates for agent and patient given permissible OVS word ordering in Spanish.

\[(6a)\) (?))El arenal mueve el río
  \textit{The sand bank moves the river}

\[(6b)\) El arenal mueve al río.
  \textit{The sand bank moves to the river}.

\(^{15}\) Cennamo (2003) notes that marking occurs when you have two highly individuated referents as arguments. Essentially, two highly individuated referents are both potentially, likely and equally possible for S.
It is evident in examples (6a-c) that \textit{a}-marking disambiguates interpretation,\footnote{This is not a claim that SDOM must be used to disambiguate any potentially ambiguous context, only that the grammar makes this recourse available to disambiguate such contexts.} essentially case-marking by discriminating contexts where the semantic nature of potential referents is equally suitable as agent or patient. The quantitative nature of the present investigation presents an opportunity for determining the (non-)validity of a disambiguating function hypothesis primarily because it would show the extent of language change since SDOM originates precisely from a syntactic construction where ambiguity was impossible – with personal pronouns (cf. Company Company 2002, Laca 2006).

Returning to the ambiguity hypothesis, it is a combination of potentially viable referents that are equally likely to fulfill semantic roles of the verb that provides sound reasoning for the use of SDOM. But at what levels do ambiguities arise? An innovative approach to measuring the concept of ambiguity related to the verb is found in Shain (2009) and his analysis of DOM in Guaraní. His approach to ambiguity was based on grammatical function (GF), assuming ambiguity may operate at three levels: grammatical, thematic and contextual.

While the grammatical recourses for marking NPs are considerably more complex in Guaraní, Shain’s discussion relates well to \textit{a}-marking in Spanish. If the grammatical structure (syntax) does not force a particular reading (inflectional agreement, \textit{a}-marking,
etc.) for Subject and Object, there is potential for grammatical ambiguity. The thematic fit for Subject and Object is based on real world knowledge and expectations. A case of ambiguity would only arise in the event that the subject and object were equally likely to fulfill the roles of Subject and Object. The final ambiguity lies in contextual fit. It may also be the case that given a particular discourse context, the ambiguity may arise from or be resolved by what has been established in the prior discourse. The novel approach Shain takes provides greater traction in an attempt to explain the sources of ambiguity and what might potentially compel DOM, and specific to the focus of this dissertation, how ambiguity arises in Spanish.

2.7. Verbal semantics of SDOM

It is uncontroversial to argue that SDOM provides an overt syntactic means to disambiguate Subject and Object in the transitive relationship, however the fact that ambiguity can arise is in itself strong evidence for the relevance of the predicate and its arguments i.e., the semantics of the verb in relation to its complements within the transitive relationship. It is not merely the semantic referents but how these referents relate to the predicate i.e., the event type in Næss’ typology that determines relative transitivity. In other words, the semantics entailed by the predicate (the verb) and the relationship to its arguments must be considered when attempting to account for SDOM.

However, not everyone agrees that the verb contributes to the variability of DOM. It has been contended that the Disambiguation Hypothesis (the use of SDOM to disambiguate subject and object in particular cases), is irrelevant to the discussion of
SDOM variation. Kliffer (1995:99) asserts that SDOM is not constrained by “peculiar properties” of individual verbs at all. Likewise in a token count across limited corpora, Weissenrieder (1991:147-8) concludes that particular verbs do not categorically pattern with SDOM.

These claims are especially important because the empirical assumption of this dissertation is that a-marking varies. In order to identify what motivates the variation manifest in SDOM, it is critical to exclude conditions where marking is categorical or impossible. On one hand, the insights of Kliffer and Weissenrieder are important because they refute the claim that there is categorical marking by particular verbs in Spanish. On the other hand, disregarding the verb as a relevant factor in explaining SDOM is also problematic precisely because of the unavoidable intersection of both A and P via the transitive relationship entailed by the semantics of the verb. The task here is to quantify these claims regarding the (ir)relevance of the verb in accounting for SDOM.

In spite of these claims that there is no correlation between verbs and SDOM, the semantic nature of verbs is frequently cited in the literature on SDOM with varying degrees of relevance. These claims are most commonly built around the preferred or prototypical arguments i.e., there are expected types of subjects and objects associated with the verbs.

According to Pensado (1995: 33- 4) the typical semantic/propositional meaning associated with the verb can account for the variant nature of SDOM. There is a “tendency” for those verbs that normally affect an animate object (matar ‘to kill’,
obedecer ‘to obey’, etc.) to co-occur with ‘a’ irrespective of the animacy status of the DO. Those verbs that normally affect inanimates avoid a-marking regardless of animacy.\footnote{However, there is ample evidence that marking does occur with verbs that prototypically have inanimate DOs. De la Mora (forthcoming) looks at the variable use of \textit{a}-marking with verbs of ingestion such as \textit{comer} ‘to eat,’ \textit{beber} ‘to drink.’} The generalization regarding unexpected or exceptional \textit{a}-marking would then result from the inherent verbal semantics associated with prototypical verbal complements.

Torrego-Salcedo (1999: 1784-93) outlines numerous restrictions on \textit{a}-marking related to the semantic properties of the predicates themselves. According to Torrego-Salcedo, the few transitive verbs that allow marking of an inanimate DO in addition to an animate (human) DO will never manifest SDOM when the subject is also inanimate; verbs such as \textit{esconder} ‘hide’, \textit{merecer} ‘to merit’, \textit{buscar} ‘to look for’ will only potentially mark objects if the subject is animate (pg. 1785).\footnote{Torrego-Salcedo uses \textit{esconder} to exemplify her claim: \begin{itemize} \item a. Este abogado escondió a muchos prisioneros \item b. Esta montaña escondió (*a) muchos prisioneros. \end{itemize} The difficulty I have with this claim is determining what verbs truly permit inanimate subjects, versus simple metaphorical extension of agentivity associated with the verb to a particular inanimate subject. It is my suspicion that this claim cannot be borne out.}

Citing Bello (1847), who argued that those verbs that normally govern inanimate DOs will frequently avoid marking human objects, Torrego-Salcedo contends it is actually the agentivity of the subject carrying out the action that constrains SDOM (1999:1800). Subjects more agentive in realizing the action favor overt marking which is very much along the lines of animacy: human or animate subjects are understood to carry out actions. Conversely less agentive subjects will disfavor overt marking. For the
majority of cases, the subject of a verb that takes DOM is clearly agentive or causative. The principal shortcoming of such an analysis is the means to actually quantify the level of agentivity of the subject.

Unlike Torrego-Salcedo, von Heusinger (2008) approaches the relevance of verbs to SDOM by assessing the preferred type of object. Whereas Torrego-Salcedo focuses on the agentivity of the subject, von Heusinger determines three classes of verbs based upon their predisposition to select animate, both animate & inanimate, or inanimate DOs. Using these broad categories, von Heusinger shows the gradual encroachment of SDOM into the domain of inanimate marking with these three classes of verbs diachronically (cf. Company Company 2002, 2003; von Heusinger 2008).19

As stated previously, the approach in this dissertation assumes that SDOM results from a multiplicity of factors that may operate in several domains. The perspectives of Torrego-Salcedo and von Heusinger regarding the relationship between the verb and SDOM are important because they outline why the agent and patient are relevant to the outcome of SDOM in the clausal domain; the A and the P crucially contribute to how a particular verb is interpreted. Extending the approach of von Heusinger (2008), it is equally relevant to ask what the prototypical or preferred subject is for verbs as related to SDOM.

Significant disagreement regarding the relevance of the verb to SDOM most certainly results from imprecise characterizations of verbal properties. What the

19 Von Heusinger also follows the same approach for the Romanian DOM pe, showing similar general trends but with distinct developments regarding specificity with these verbs (von Heusinger and Onea Gáspár 2008)
literature has provided are analyses based primarily on intuitions regarding the predicate and the relationship between arguments. Unfortunately, it remains consistently problematic to conceptualize verbs within a unified taxonomy. Being able to discretely categorize verbs along semantic lines exceeds the scope of this dissertation, although the quantified nature of the data may provide useful insight to potential verb typologies. The diverse conceptual capacity and extension that verbs may undergo in language makes this task seemingly impossible. Nevertheless, there have been attempts to generalize verb-types along grammatical and aspectual lines that lend themselves to more clearly defined categories. If one assumes that this is the case, and much of the literature does embrace this account of SDOM relating to verbs, it is still very much an intuition perpetuated in the literature and generally unsubstantiated quantitatively.

2.8. SDOM in the domain of discourse: topicality

The discourse domain figures as the common denominator among the vast majority of recent work in SDOM. Most researchers have come to the conclusion that discourse context and discourse-pragmatic information are necessary to fully account for the variable use of $a$-marking. It is insufficient and limiting to only view SDOM in terms of properties of the NP or even the clause level (cf. Dumitrescu 1997: 317).

Laca (1995) claims that individuation may only be partially accounted for based on the semantic and grammatical features of the NP: the totality of individuation is a discourse pragmatic property (81-2). SDOM in discourse marks an entity as part of the
“expressive intention of the interlocutor” (82). Enç (1991: 7) invokes the information status of the discourse referents to account for definite and indefinite readings. Dumitrescu (1997) ambitiously attempts to account for unexpected a-marking (primarily of inanimate or indefinite referents) through constrained discourse principles. She proposes that distinct dialects (Madrid and Buenos Aires) employ a-marking in distinct but systematic ways using the discourse principle she calls “Discourse Prominence.” Weissenrieder (1991) explores the discourse-grounded features of SDOM including referentiality and topicality. Even Leonetti, an antagonist of specificity as an NP-level property, argues that the inference of specificity is only made available in the discourse, permitting a specific reading (2003: 110).

Laca (2006) presents a concise organization of the tension between properties that correspond to the domains of the referent, the clause and the discourse context. She partitions all those properties that contribute to a-marking into two generalized factors: global factors and local factors. Local factors are those properties pertaining to the object NP while global factors have to do with contextual features i.e., clausal and discursive features (pg. 429-30). She claims that animacy and definiteness are primary among local factors. Local factors may be dominated by global factors including: lexical nature of the verb, presence of a secondary predication referring to the object, preverbal position of the object, and clitic doubling. Other factors may also contribute to the high

20 It appears she is trying to give a name to an intuition that marked, individuated objects in context will be interpreted as more relevant, or more “present” (her word) than other properties.
topicality of a referent and/or the co-reference with prior mention to account for the unanticipated use of SDOM with inanimate objects.

One intuition that reflects the global or discursive feature of SDOM is the frequent appeal to topicality. The consensus among many researchers is that in some way the phenomenon marks a referent as “topical.” The notion of topicality that is generally assumed in the SDOM literature seems to interpret topic or topicality as the “aboutness” that a referent maintains in the discourse context. I will review the relevant literature on topicality, especially that which pertains to SDOM.

By way of introduction to SDOM and topicality, Lyons states “…topic-subject identity as the unmarked situation is extremely widespread and probably universal” (1999: 230). While he probably did not have DOM phenomena in mind, the non-prototypical case of an object being more topical, or at least being introduced as topical lends itself to DOM phenomena quite aptly. Marking may indicate the topicality\textsuperscript{21} of the DO, or that the object is more topical than the subject. I am deliberate in my claim that marking a DO is only a relative measure of topicality. Those referents that are frequent in the discourse are relatively topical by virtue of their frequency in relation to other referents; marking does not mark a topic \textit{per se}.

The seminal work of Givón (1983) outlines how topicality operates in discourse. In Givón’s framework, the central unit of discourse is a thematic paragraph (which may be combined into larger discourse units) and is the basic unit for analyzing and discussing continuity in discourse. Givón measures topicality via topic/participant continuity in

\textsuperscript{21} As presumed in this and other research, marking may occur for other reasons, disambiguation, syntax, etc. Topicality may be one of the functions of SDOM.
discourse. Topic availability is determined by the binary property of definiteness, \(^{22}\) definites being identifiable for the hearer (as assumed by the speaker) or indefinites, which are presumably topics introduced by the speaker. The topics that are generally available all the time he calls generics.

To quantify his notion of topicality, Givón establish two metrics: Referential Distance (RD) and Topic Persistence (TP). RD looks back in the discourse and is a “measurement [that] assesses the gap between the previous occurrence in the discourse of a referent/topic and its current occurrence in a clause…” with an upper bound of 20 clauses. TP measures the potential topicality of a referent as it persists in the discourse. TP counts the number of times the referent continues in the following discourse in any kind of mention.

This is a very straightforward means to quantify Givón’s concept of topicality and is a practical tool for general measure of how referents are distributed in discourse. What follows from these measurements is a critical insight for gauging topicality; it is not a binary property (± topical) but a gradient property. In other words the frequency of the referent and the proximity of the next prior or following mention of the referent are ways to track and quantify what more topical referents relative to less topical referents in the discourse.

Others have approached topicality in different terms. The central tenet of Laca (1995) is that topicality is the unifying property that accounts for SDOM (both synchronically and diachronically). She defines topicality as the “psychological subject –

\(^{22}\) Here Givón uses definiteness to mean specificity as interrelated to the definite or indefinite NP.
what is talked about” (pg. 86). She reiterates that individuation is a discourse-pragmatic concept; SDOM marks something as being individuated. An individuated object indicates it may serve as a probable topic. Her examples are repeated here as (8a-b):

(8a) Juan mató a un tigre  
*John killed A a tiger*  

(8b) Juan mató un tigre  
*John killed a tiger*

Her intuition is that the marked case (8a) indicates there will be more information regarding the object NP i.e., it will be a probable discourse topic. Her conclusion assumes that at some level speakers make use of SDOM to mark the referent as highly topical and independent within its predication (pg. 89).

Melis (1995) defines topic at two levels: sentence and thematic paragraph, based on Givón (1983). At the sentence level a topic is definite, individuated, agentive and the focus of interest and attention in the interior of the predicative sequence i.e., within a thematic paragraph. At the sentence level the topic is considered the given information. The central participant of a thematic paragraph frequently coincides with the “given” element in several sentences.

Torrego-Salcedo (1999) draws a connection between information status, syntactic position and SDOM. Constituents at the beginning of a sentence are accompanied by morpho-syntactic elements associated with givenness (definite article, possessives and other features of individuation). This is also the case with right and left dislocated
constituents. Because these syntactic positions have a topicalizing function and a-marked objects can be expressed by the same means with the same informative content, SDOM corresponds to topicality. Unfortunately, this line of reasoning is circular: marking makes a referent topical and topical referents are marked.

Additionally, Torrego-Salcedo argues that a-marking may also occur with inanimates as a means to “destacar” (emphasize, make stand out) the accusative referent.\(^{23}\) This description, while unquantifiable, appeals to the idea that marking serves discursive functions (foregrounding, topicalization, etc.). While this claim appears intuitively plausible, if not probable, again Torrego-Salcedo and others lack a quantifiable means to demonstrate such a function, instead presenting solely constructed and “token” examples.

Weissenrieder (1991) draws a similar conclusion to Laca (2006) and her local and global factor analysis. Weissenrieder states the following:

…the association of marking with a particular verb, syntactic structure, or case marking is only part of the picture, however. Correlates that normally favor or disfavor marking at lower levels of analysis seem to take a back seat to the nouns’ overall role in the larger context. Animate and inanimate nouns, which one would otherwise expect to be marked, remain unmarked when non-referential. At the other extreme, highly unpredictable, “stylistic” occurrences of a follow from

\(^{23}\) Torrego-Salcedo appears to qualify this statement, attributing this marking of inanimates as a result of discursive and sentential elements (while saying nothing about what those elements may be).
general patterns of discourse-related phenomena such as topicality. Thus it is impossible to state that certain lexical items...appear categorically with the preposition, or that others...occur randomly with the preposition “according to taste” (1991:155).

It can be seen from the literature that the overwhelming consensus is that if there is any hope of a unifying account for SDOM, it must be understood from the discourse notion of topicality. Ironically Lambrecht, who approaches information structure and topicality at the clausal/propositional level, states that one cannot always identify topic based on syntax alone; the discourse context is necessary (1994:120).24

A further refinement of Givón’s RD and TP metrics is presented by Shain (2009). Shain employs seven metrics for quantifying topicality in Guaraní (pg. 77). In addition to using RD and TP, Shain astutely observes that the two metrics are incompatible for providing one complete measure of topicality since they measure distinct aspects in opposing directions (number of previous clauses between mention of the referent vs. number of occurrences of the referent in the following clauses). Noting that Givón did not preclude extending the RD and TP measurements in the opposite direction, Shain extends RD forward and TP backwards in the discourse to produce measures of Total Referential Distance (T-RD) and Total Topic Persistence (T-TP). Reiterating the fact that these measurements show that topicality is not a binary property (± topical) but a

---

24 Unfortunately, this commentary is equally problematic because there is no heuristic to determine what is then topical.
gradient one, this approach presents an intriguing hypothesis for SDOM: we would expect to find a correlation between increased α-marking and greater topicality evinced by lower T-RD and higher T-TP.

I believe that such a quantitative notion of topicality presents the ideal solution for explaining SDOM phenomenon to its fullest because topicality operates at a “global” i.e. discourse-pragmatic level. A discourse-pragmatic measure of topicality provides sufficient flexibility to encompass the multiplicity of potential factors motivating α-marking: lexical and semantic features of the referent and the features associated with the clausal relationship inherent in transitivity.

2.9. Summary

This chapter has discussed the primary literature regarding SDOM. The principal focus of this discussion has been to illuminate the critical components that have been attributed to SDOM. The purpose of this chapter has been to review the previous studies and theories regarding SDOM and DOM phenomena. The organization of this chapter follows the various factors that have been attributed to SDOM and how these factors extend from characteristics of the NP (animacy, definiteness, specificity), to the clausal (relative animacy, disambiguation, etc.) and to the discourse domain (topicality). These discussions, analyses and assumptions form the foundation for the quantitative methodology used in approaching the data, which is outlined in Chapter 3.
3. Overview of Theoretical Framework and Methodology

The theoretical framework for this dissertation assumes that language is inherently variable and dynamic. This perspective on language is relevant to the circumstances of SDOM phenomena because the survey of the literature and research shows that no single explanation can account for all manifestations of \textit{a}-marking (or lack thereof). A static perspective on language, a hypothesized and idealized structure for explaining language is limited by its inability to accommodate linguistic phenomena that are potentially motivated by multiple and varying factors. In fact, the assumption that there exist discrete rules that dictate precisely how language functions precludes the possibility that multiple factors may motivate linguistic phenomena in variable ways. Furthermore, these static analyses of language are bereft of the critical context of language. While this static approach to language can serve particular purposes in linguistic analysis, the non-categoricity of \textit{a}-marking in Spanish represents a phenomenon that is best approached through theoretical and methodological approaches that assume multiple causation precisely because no single factor sufficiently circumscribes the observed behavior of SDOM. While animacy initially appears to provide the most straightforward account, the exceptional marking that is not infrequent, and the additional arguments regarding the function and features of DOM lead one to reject a simplistic account based on animacy alone.
The variation indisputably manifest in SDOM and the several potential motivating factors attributed to the phenomenon present ideal conditions for the application of variationist methodologies. This chapter discusses the theoretical background and details the methodological approach used in this research to illuminate SDOM. Section 3.1 reviews the concept of the linguistic variable and how this notion is applied to syntactic variation i.e., the conceptualization of the syntactic variable. Section 3.2 presents the corpora selected for the dissertation and discusses the advantages and motivations for using those corpora. In Section 3.3 my methodological approach is outlined. The cases of SDOM that fall outside the variable contexts are outlined in Section 3.4, the envelope of variation. In Section 3.5 I review how I operationalize the factor groups and factors for suitably coding the data required for Goldvarb analyses. Section 3.6 reviews how the quantified data alone is insufficient, requiring an analysis informed by linguistic theory. Section 3.7 is a summary and conclusion to the chapter.

3.1. The Linguistic & Syntactic Variable

SDOM variation is binary in that the linguistic variable is either marked or unmarked; given the appropriate context of a transitive verb with an overt DO, this DO may or may not be a-marked. However, the assumption that there are two ways of saying the same thing, while taken for granted in many fields of linguistics, has not been unquestioned and has required careful and thoughtful circumscription. This is particularly the case with the concept of the syntactic variable. Because SDOM is a (morpho) syntactic variable and I analyze SDOM distribution as it relates to semantic,
pragmatic and syntactic factors, a summary of what is meant by linguistic variable and the more specific concept of the syntactic variable are in order.

The formulation of the linguistic variable has arisen from the initial observation that language varies. Language offers multiple means for communicating and in turn, speakers communicate through these varied means with a sensitivity to external and internal factors. While observed language variation had been well documented in the case of lexicographers and dialectologists, it was Labov (1966) who first formulated the notion of the linguistic variable. Labov systematically approached the variation in language by presuming that speakers have linguistic options readily available and that these options are not randomly selected; rather, the varied linguistic realizations both correlate to internal and external factors and have social significance. These circumstances were deftly described as ordered heterogeneity (Weinreich et al. 1968). This ordered heterogeneity can only be arrived at through the quantitative measurement of variance. Without quantifying the linguistic variable i.e., what linguistic feature is varying and how it is varying, the variation cannot be understood. Without the careful and meticulous observation and quantification of the variable, the language variation is relegated to intuition and impressionistic accounts.

The notion of the linguistic variable was well received as it pertained to phonological variation. Labov showed quite convincingly and irrefutably that the phonetic realization of underlying phonemes varied consistently with social factors. The social identity of speakers, i.e. the social factors of class, ethnicity, occupation, education, etc. correlated with language use. The variation of a given phonological phenomenon
could be explained by social factors. Assumed was an analogous relationship between a phonological variable and a syntactic variable (Weiner and Labov 1977, 1983). However this assumption drew significant but ultimately constructive criticism. Lavandera (1978) proposed a “referential vs. non-referential” argument against the syntactic variable. Phonological items are non-referential but can have ‘style’ or social meaning whereas syntactic items are inherently referential: there is semantic content unique to the syntactic construction or form. Romaine (1984) echoed these sentiments. In addition she objected, or at least raised the concern about the syntactic variable on semiotic grounds. The pairing of sign and sound is arbitrary (ergo phonology is arbitrary) while syntax is iconic; it is interpreted anew each time. The structure of an equation is iconic; the symbols are arbitrary but the relation between them is not (pg. 6-7). This raised the question if in fact it was legitimate to claim that two distinct syntactic structures were indeed “the same.”

What is particularly insightful in Romaine’s argument is her appeal to pragmatic and semantic perspectives. Both Lavandera and Romaine had questioned the ability and feasibility of correlating social factors with syntax. In outlining the substantial distinctions between syntactic and phonological phenomena, Romaine characterized the very means to identify syntactic variables via internal factors e.g., syntactic structures, semantics, morphology that are independent from the external factors (social factors). Ultimately Romaine contributed significantly to the concept of the linguistic variable because she recognized that different types of variables have different conditioning factors (see pg. 15). She identified four types of variables including phonological, morphophonemic, morpho-syntactic and ‘pure’ syntactic variables. Relevant to the
present scope of this dissertation is her distinction between morphi-syntactic variables that are affected by both syntactic (language internal) and social (language external) factors and ‘pure’ syntactic variables that are based upon syntactic factors alone. Wolfram (1993) and Winford (1984) both support this distinction between syntactic variation resulting from social factors and variation resulting from purely syntactic factors. Wolfram clarifies that the linguistic variable must be a “…linguistically-defined unit…” (pg. 196), meaning that the variation must be delimited by consistent, recognizable and logical linguistic features.

A more informed perspective on both the syntactic variable and the linguistic variable has resulted from this careful and warranted scrutiny. The discussion and critiques summarized in this section provide a more constrained notion of the linguistic variable, particularly the syntactic variable. For my research, I assume the notion of a ‘purely’ syntactic variable and the case of DOM in Spanish represents a context where speakers have two options for “saying the same thing.” I assume that SDOM is not a case of free variation, nor is marking and non-marking found in complementary distribution: speakers have a linguistic choice regarding SDOM. I will assume that SDOM is conditioned by particular linguistic factors i.e., internal factors. Bayley (2002, 2004; Young and Bayley 1996) argues that a crucial principle guiding variationist endeavors is the “principle of multiple causes.” The linguistic variable “[is] subject to not one, but many contextual conditioning factors.” Operating under this hypothesis, it is the purpose of this endeavor to show the multiplicity of factors that condition and explain SDOM.
In order to approach the multiple factors that account for SDOM I will rely on variationist analytical tools. One of the primary and innovative objectives of this undertaking is to provide quantitative evidence that accounts for SDOM in Spanish. I employ Goldvarb (Sankoff et al 2005, Tagliamonte 2006) as the primary means to organize and interpret my data. It is important to note that as with any quantitative variationist approach, the results are probabilistic (Sankoff 1988, Guy 1991). The probabilistic perspective on relevant features and factors reflecting language variation falls in line with the crucial assumption within variationist approaches to language that the distribution of linguistic phenomena is non-categorical, or at the very least, that variation may be introduced into the system. In effect, if SDOM could be explained categorically, there would be nothing to quantify. Fortunately it is not the case that SDOM is categorical as can be attested in the Spanish-speaking world.

I have outlined the merit and support of the dynamicity of language. I argue that language is not a closed and static system; while certain features may appear static synchronically, in the long run (literally) the diachronic perspective of language reveals that syntax, and even functional words, show variation and change. This model of language, supported by empirical evidence, is crucial theoretical validation for my approach to SDOM. This undertaking assumes that large-scale quantitative analysis captures meaningful variation in spoken language, and avoids the limited explanations provided by anecdotal or contrived data.

One final comment is necessary regarding my assumptions of the SDOM phenomenon. While this research is indeed variationist and falls within the purview of
sociolinguistics, it will not reflect the socialness of SDOM per se because no external factors are incorporated. However, as will be noted in the following sub-section about the relative uniformity of the Habla culta corpora, there is essentially an indirect social measurement resulting from the interacting factors of education and social class tied to the scope of the Habla culta project: these data give us a perspective on SDOM in educated, upper-class speech. Nevertheless, the social information is insufficient and in most cases there is no counterpart, i.e. a lower class or less-educated corpus with which to make valid comparisons, the exception being the Habla culta de la ciudad de México (1971) and Habla popular de la ciudad de México (1976). In the case of Mexico City speakers, a second project was undertaken, with the data being provided by lower class speakers. Unfortunately, this was not replicated by other Habla culta projects, thereby preventing an additional social component from being included in this current investigation.

3.2. Corpora

One of the important contributions of this investigation is to characterize the use of SDOM in spoken Spanish. As outlined in Chapter 2, much ink has been spilt regarding prescribed use of SDOM and generalizations regarding when and why it is

---

25 While there have been no claims that SDOM is sensitive to external (social) factors, further research regarding social factors represents an important future expansion of this project with potentially unprecedented findings regarding social factors and SDOM.
used, however the minuscule quantification of SDOM has been limited to written texts without any attempt to quantify the phenomenon in speech. Furthermore, this dissertation evaluates the potential variation cross-dialectally.

The investigation includes several dialects from the Spanish-speaking world. The *Habla Culta* project (see Lope Blanch 1971, 1976) encompassed numerous recorded interviews made in principal cities from many Spanish-speaking countries. These spoken data were transcribed by the authors and made available in the public domain. I selected these corpora for several important reasons. First, they presented a readily available source of transcribed spoken data from which to extract tokens. An individual undertaking by me to travel and record the requisite interviews was unfeasible and impractical when considering temporal and financial constraints. Additionally, and imperative for quantitative comparison, these corpora present homogenous and uniform data by virtue of the selecting criteria: that the speakers were raised locally and college educated. The primary topics in the interviews were relatively similar including education, occupation and family. The *Habla Culta* corpora consist of interviews that followed the same protocols and were carried out by trained fieldworkers and linguists, further ensuring linguistic fidelity and comparability.

The corpora selected for this study were: *Habla culta de Buenos Aires* (1987), *Habla culta de la Ciudad de México* (1971), and *Habla culta de Madrid* (1981). These

---

26 Also known as the *Norma culta* project, the *Habla Culta* project included Barcelona, Bogota, Buenos Aires, Caracas, Havana, La Laguna/Santa Cruz de Tenerife, La Paz, Lima, Madrid, Mexico City, Montevideo, Panama City, Quito, San Jose, San Juan de Puerto Rico, Santiago de Chile, Santo Domingo and Seville.
three corpora provided the sources from which to compare “grammars” of SDOM since these dialects are recognized for varying phonological and syntactic areal features (Canfield 1981, Lipski 1995, et al.). These dialects were deemed sufficiently distinct and with large representative numbers of speakers. In addition claims have been made regarding dialectal distinctions of SDOM among these varieties: Dumitrescu (1997) claims that important discourse pragmatic features distinguish a-marking in Buenos Aires and Madrid while von Heusinger & Kaiser (2003), von Heusinger (2005), assume that marking cannot occur with inanimates in Peninsular varieties and that such exceptional marking i.e., marking of inanimates is a feature of Latin-American dialects.

Regarding the concept of dialect, certainly the corpora from which tokens were culled are not representative of the true dialectal diversity manifest in the Spanish-speaking world, nor their respective countries. Moreover, the corpora can be characterized as educated, metropolitan varieties and should in no way be considered representative of the city. Unfortunately, it is the case that often the corpora are assumed to represent the dialect of a particular country. At no point do I assume that the speech of a particular city represents the speech variety (or varieties) of an entire country when it is the case that dialectal features are rarely, if ever, confined to arbitrary geo-political boundaries. This project will make use of the term dialect in its weak form, referring simply to the variety spoken in each of the three selected cities, and representative of one relatively homogeneous social class. However, when used in this weak form, dialect does indeed characterize the comparison I am undertaking. Specifically, by comparing
the same type of speaker across broad varieties of Spanish, this investigation is able to contribute greater detail to the descriptions of these dialectal grammars.

3.3. Methodology

Investigating variable syntactic phenomena presents particular challenges. Foremost among them is the relative frequency or infrequency of a variable. The debate regarding the “sameness” of two (or more) syntactic constructions has been discussed in 3.1. Finally, there is often an assumption that syntactic structure is limited in how much or if it truly varies. While this has been shown to not be the case, it still remains that finding sufficient cases of a syntactic variable, or eliciting sufficient examples is a challenge. It is my estimation that these have been significant factors in deterring an extensive quantitative analysis of SDOM prior to this dissertation, particularly in spontaneous spoken contexts.

To overcome these particular challenges associated with syntactic variables, I initially implemented a methodology to provide a reliable heuristic for quantifying SDOM. In Tippets & Schwenter 2007 and Tippets (forthcoming) the envelope of variation was determined inductively, by extracting overtly $a$-marked direct objects culled from the corpora. Each corpus was searched for all cases of DOM with careful attention placed on avoiding all cases of non-accusative uses i.e., prepositional uses of $a$. The verbs that occurred with overt marking then provided a basis to search for the un-marked or $a$-less tokens of these same verbs. This method allowed for identifying unmarked tokens in a feasible and relevant manner that had the advantage of being readily replicable.
Despite this step in quantifying SDOM, there are ways to improve this methodology. I limited the variation by tying it to the verb forms resulting in, my initial methodology indirectly incorporating the verb as an independent variable. Additionally, it somewhat skews the true distribution of marking for all accusative objects by only considering those that happened to have both marked and unmarked cases manifest in the corpora. The advantage of this initial methodology was the ability to compare marked and unmarked forms with the same verb. However, this approach limits the ability to look at the behavior of SDOM for all accusative objects. This methodology precluded any tokens that didn’t have marked forms, the short-coming being that despite there being no example with a-marking, this should not entail that marking can never occur. Including the cases where non-marking could occur but did not is critical for completely quantifying the SDOM phenomenon. These data are equally important in assessing the features that are responsible for, in this case, the non-occurrence of a i.e., the non-marking of DOs. Furthermore, by including the non-marked cases I am able to give a much more accurate description of the actual rate of a-marking as a subset of the all DOs in the corpora. Given this significant expansion in my approach to quantifying a-marking, it was critical to circumscribe the variable context and to carefully consider in which contexts variation was possible.

3.4. Defining the Envelope of Variation

27 This methodology was partly implemented for convenience but also out of necessity since Goldvarb requires that there be variation with respect to the dependent variable. 28 This is not to say that every unmarked accusative DO could potentially be marked. See section 3.4 regarding those unmarked cases that were excluded from the token count.
Using the refined methodology outlined in 3.3, I circumscribed the contexts of potential tokens i.e., I determined my envelope of variation. Because of the probabilistic nature of measuring the linguistic variable, it is critical to delimit what can and cannot vary. Otherwise, the quantitative data are compromised, weakening any and all conclusions derived from the data. In the present study, it was necessary to identify and delimit the variable contexts of a-marking. Likewise those contexts that were either invariably marked or invariably unmarked needed to be excluded from the data. Specific parameters were set in order to identify the particular cases where variation could occur and to rule out those cases where variation did not occur or where variation did not apply. The corpora were culled for all overt accusative objects regardless of marking. While this undertaking was substantial, it was necessary to give an accurate description of the true distribution of accusative objects as they related to the various semantic and pragmatic factors potentially correlating with SDOM. From among all these direct objects, invariant cases of marking or non-marking were excluded in addition to cases where a potential ambiguity existed regarding the a. I outline below the cases and contexts where the potential marking fell outside the envelope of variation.

All cases where a-marking is invariable were necessarily excluded because no variation is manifest. Personal pronouns were excluded since obligatory a-marking occurs in clitic doubling where a redundant emphatic pronoun is always marked as seen in the examples below:

29 While it is the case that there exist null objects in Spanish (cf. Schwenter 2006), it is never the case that these are variably marked. Therefore, the first parameter for identifying tokens was that they be overt accusative objects.
1a)  *Me vieron a mí*
    They saw me

1b)  *Me vieron mí*
    They saw me

1c)  A mí me vieron
    They saw me

1d)  *Mí me vieron*
    They saw me

Since there is never a case where these pronouns appear without marking, all cases of personal pronouns as DOs were excluded. In addition to pronouns, all idiomatic expressions were not considered as potential cases of SDOM variation since they are lexicalized with respect to marking or not. These included phrases like “el ojo del amo engorda el caballo,” “tomar clases,” etc. All cases of marking with *nadie, quien* and *alguien* were excluded since they are categorically marked when they function as DOs and in the last case with *alguien*, marking is irrelevant regardless since DOM would be phonetically indiscernible.

In addition to excluding pronouns and idiomatic expressions, there have been claims that certain cases of *a*-marking are semantically lexicalized with the verb (Pensado 1995:33-34; Torrego-Salcedo 1999:1787-8). In other words, it is argued that there indeed exist examples where a DO is invariably marked based upon its predicate alone. Those cases were excluded from the envelope of variation (*ayudar, involucrar, etc.*).
A non-categorical claim regarding SDOM is that there are particular verbs whose semantics favor *a*-marking (Pensado 1995, Torrego 1999, García 1995) but that both marked and unmarked objects occur with these verbs. In general the semantics of the verb are such that the subject and/or object are expected to be animate (e.g. *matar*, *citar*, *conocer*, *odiar*, etc.) leading to a predisposition for *a*-marking. I follow the weaker claim of Pensado and others and assume that while there may be a tendency for marking with particular verbs, variation is still found. I therefore included these verbs while excluding those verbs with invariant marking or non-marking from the data.

Closely associated with claims that there is categorical *a*-marking with particular verbs, it is relevant to comment on the fact that there are semantic meanings tied to SDOM. The most common cases where the verbal semantics depend on the presence or absence of *a*-marking include *tener* and *querer*. However, this also extends to other verbs like *terminar a* (to kill) *dibujar a* (to draw a portrait of someone). These examples of SDOM were not included because the marking is not truly variable; there is a specific semantic function of SDOM in these cases.

A number of verbs were excluded based upon the potential ambiguity arising from dative referents that are closely tied to the semantics of the verb. *A*-marking with these verbs is infelicitous precisely because the dative interpretation of the referent is the

---

30 As noted by Terrell Morgan in discussing this dissertation, the case of semantic distinctions between *tener* and *tener a* are possibly transparent precisely because there is a measurable distinction in the transitivity of the verb. *Tengo dos hermanos* is far more existential in interpretation while *tengo a dos hermanos en casa* functions far more transitively.
preferred reading with the presence of *a*, therefore precluding SDOM from occurring. These included *escribir, contestar, enseñar, comentar, proponer, pedir, alquilar* etc.

Furthermore, there are cases where a verb may have restrictions tied to their semantics and related dative/accusative objects. Such is the case with the infamous *llamar*, where the meaning ‘to call someone or something by a particular name or title’ presents a polemic question of case. While this construction is indeed variably marked with an *a*, it is debated as to whether the object is a dative or accusative, a condition further complicated by the inescapable ambiguity of *a* in so much as it could be a dative *a* or an accusative *a*. These cases were therefore excluded. However, *llamar* ‘to call’ was counted because of the clearly transitive properties of this verb in conjunction with observed variable marking with *a*.

The verb *haber* was excluded for potential marking following Torrego-Salcedo (1999:1785) who asserts that there is a restriction for the possibility of *a*-marking with predicates that have a subject with the semantic value of agent. This found additional empirical support in that no marking was ever found with *haber* in the three corpora.

Cases where impersonal *se* and passive constructions with *se* were ambiguous (3\textsuperscript{rd} person singular) fell outside the envelope because it was impossible to definitively determine if *a*-marking was possible. However, cases where the transitive relationship is clear i.e., the referent is accusative were counted:

2a) \textit{Se pesca esas corvinas}  
\textit{They fish for those drum fish}  
(HC Buenos Aires)
2b) A cada persona se censa
They take a census of A each person
(HC Buenos Aires)

In these cases, there is a clear DO that can be quantified, therefore they fell within the envelope of variation.

The tokens that fell outside the envelope of variation included all contexts where the possibility of a-marking was indeterminable or potentially ambiguous since it is only in the cases of marking or the lack of verbal agreement that indicates the referent is indeed accusative and not nominative.

Other cases that arose in the coding were excluded based upon the following:

Ambiguity of animacy:

3a) Voy a aceptar uno
I am going to accept one
(HC Buenos Aires)

Ambiguity in animacy and specificity:

3b) Claro, porque el sistema de Él es preparar la clase en la clase...
Of course, because his system is prepare the class in the class.

(HC Buenos Aires)

Referent is semantically underspecified or non referential:

3c) …la comedia es palliata, por su origen, porque imita a lo griego
…the comedy is in its origin, because it imitates greekness.

(HC Buenos Aires)
Confusion of DO referent:

3d) no conocemos su nom... su...sus obras.
   (HC Buenos Aires)
   We don’t know ??...his/her/their works.

Self-correction:

3f) Nosotros dedicamos el domingo íntegro desde las ocho o nueve de la mañana que se despiertan las chicas... a las chicas,
   (HC Buenos Aires)
   We dedicated all Sunday from 8:00 or 9:00 in the morning, from when the girls woke up…when they were woken up.

Incomplete, omitted or redacted information:

3g) ¿Vos lo viste a [.........] que...
   Did you see A [.........] that…
   (HC Buenos Aires)

Finally, tokens where a-marking precedes a word-initial ‘a’ were also excluded, because they would be phonetically indiscernible in the phonetic environment and thus “heard” by the transcriber alone. With no means to verify their presence (or absence) of SDOM these contexts were excluded.

3.5. Independent variables.

Having excluded all contexts where SDOM is invariant or indeterminable, the tokens extracted from the remaining variable contexts were coded for factors that potentially motivate SDOM. These factors have been derived from the literature on SDOM (see Chapter 2). Because of the quantitative nature of this endeavor, the numerous
hypotheses, including the proposed features and factors relevant to \(a\)-marking, needed to be operationalized for Goldvarb. In regression analyses these individual factors are the independent variables that may potentially affect the dependent variable; in this research the dependent variable is SDOM, whether a particular DO is marked or unmarked with \(a\). Each token is necessarily coded for the same independent variables in order to be interpretable by the variable rule program Goldvarb X (Sankoff, et al. 2005). Every token, both marked and unmarked, was coded for the independent variables that are described in sections 3.5.1-3.5.10. These independent variables are organized into factor groups in the Goldvarb program, and the various options within each factor group are called factors.

All information for the tokens was recorded using the Excel software program for uniformity, organization and ease of use in managing the data. While not measurable in regression analysis, the immediate context of the token, at minimum the entire clause where the marking or non-marking occurred, was also recorded to assist in identifying tokens in the spread sheet. Additionally, the location in the corpus of each token and the verb whose accusative argument was the marked or unmarked referent were also recorded. The subject (when possible) and the object were also recorded for reference.

The modus operandi of quantitative analysis of language variation is to cast as wide a net possible regarding the potentially motivating factors of a phenomenon. In other words, it is methodologically preferable to be maximally inclusive in the possible motivating factors in order to capture the relevant factors. From unpleasant first-hand experience, I know that it is always easier to go back and eliminate or combine Factor
Groups than to revisit every token in the data to code for additional features. With this in mind, I detail all of the initial Factor Groups and their Factors used for coding so that my methodology and considerations are completely transparent and explicit in the following section. In Chapter 4 and Chapter 5, I discuss my analyses and detail the circumstances and decisions that motivated recoding the data through collapsing and eliminating Factor Groups and Factors.

3.5.1 Animacy of the Direct Object

Appropriately accounting for the variation inherent in SDOM necessitates a quantifiable concept of animacy. As discussed in 2.1., animacy can be conceived of in several ways, however the linguistic evidence for relevant distinctions made in language leads to a scale based on the following distinctions: human, animate, inanimate. However, I chose to divide the animate category into both animal and plant in part because of the possibility that there might be distinctions despite their infrequent occurrence. In addition to these clear-cut cases of animacy, additional factors were needed to account for the extension of animacy into other NPs that are not literally or directly human or animate referents (as seen in Figure 4.) I included a factor for metaphoric referents i.e., collective and non-personal agents, representing referentially animate but lexically inanimate entities (la facultad, el populacho, etc.) and cases of personification or anthropomorphism. The initial Fs within the FACTOR GROUP Animacy of DO are the following ordered from most to least animate:
a. human:

4a) *Me gustaría verla a Susana ahora.* (HC Buenos Aires)
I would like to see A Susana now.

b. impersonal:

4b) *lo expulsan a uno* (HC Buenos Aires)
They expel one/you

c. metaphoric:

4c) *Debo admitir que el tango ya no representa a la ciudad.*
(HC Buenos Aires)
I must admit that the tango does not represent A the city anymore.

d. animal:

4d) *...y este muchacho, pues le veo que le gusta el campo, que es capaz de conocer a un animal a distancia.* (HC Madrid)
…and this kid, well, I see that he likes the countryside and he is good at recognizing A an animal from a distance.

e. plant:

4e) *En la vegetación, las plantas anuales han desaparecido, han dejado únicamente la semilla que perpetúa su especie.*
(HC Mexico City)
In the vegetation, the perennials have disappeared; they have only left their seed that perpetuates their species.

f. Inanimate:

4f) *Es decir, determinados elementos van a atrapar, digamos, a un gas; lo van a enclaustrar ¿verdad?* (HC Mexico City)
That is to say, specific elements are going to trap, you might say, A a gas; they are going to capture it, right?
While it is not the case that animacy guarantees marking (i.e. it is not necessary, nor sufficient), Kliffer (1995:99) argues that animacy is the single feature of the DO that can neutralize or override any other. Leonetti (1999:866) makes a similar, albeit more focused claim stating that animacy predominates over specificity. It stands to reason, and is corroborated by any basic observation of Spanish, that animacy is significant and relevant to the discussion of SDOM. However by operationalizing animacy for Goldvarb this dissertation will be able to quantify the perceived levels of animacy as they relate to SDOM (see 2.1.). More importantly, the advantage of using regression analysis in approaching this phenomenon is to be able to view animacy as a factor motivating SDOM while also taking into account other possible factors that may explain the observed patterns of marking.

3.5.2 Animacy of the Subject

In addition to the animacy of the DO, I also coded for the animacy of the grammatical subject (S). The S was coded for the same Fs as the Animacy of the DO. Importantly, there exists an additional F for Subject animacy that is not possible for DO: the subject animacy may be indeterminate because the S is clausal or an otherwise unidentifiable agent that functions as S. For example:

*indeterminable animacy:*

5) *Educar a un niño es importante.* (HC Mexico City)
   *Educating A a child* is important.
In this case the entire phrase *educar a un niño* functions as the subject and no agent can be identified. These cases were coded simply as an additional F indicating that the animacy of the S referent was indeterminable. Similar to indeterminable animacy, impersonal S constructions were coded separately.

**impersonal:**

6a) Che, *uno* a... a Vicente López y Planes- -- lo conoce (HC Buenos Aires)  
Man, *one* knows… A Vicente López – Planes

6b) …(∅) *han detenido* como a setenta y tantas personas (HC Buenos Aires)  
…(*they*) have detained about A 70 or so people.

To summarize, the initial Fs within the FACTOR GROUP Animacy of S were the following in rank from most animate to least animate: *human, impersonal, metaphoric, indeterminate, animal, plant, inanimate.*

### 3.5.3. Relative Animacy

Another means to capture the importance of animacy is the factor group *relative animacy.* The motivation for this factor group arises from the insight made by Comrie (1989) regarding language universals and DOM. Based upon his analysis of many distinct languages, he hypothesizes that languages that make use of DOM, do so as a means to indicate an unexpected and atypical relationship between A and P. In other words, DOM is likely to occur when the prototypical semantic roles of A and P are
switched. Comrie relies on the general assumptions of the semantic characteristics of A and P to account for the markedness approach to DOM. I would add that this perspective may be sufficient for DOM, but the animacy of the subject may also play a role as well. Thus the animacy of the direct object relative to the subject is also critical because of the implications it has for prototypical conceptualizations of the transitive relationship between A and P and the underlying motivations for DOM. I have quantified this concept of relative animacy in previous research (Tippets and Schwenter 2007, Tippets forthcoming), and have found that in some dialects of Spanish, it is not the animacy of the accusative NP rather the relative animacy between the subject and object that is the more significant factor favoring marking. While in terms of overall occurrence irrespective of the animacy of S, the overwhelming number of cases are marked human DOs, the overall rate of marking was in fact higher and statistically more significant for those contexts where the DO was equal to or higher in animacy (usually human) than the S. To operationalize relative animacy, the animacy of subject and object were compared to one another in terms of where they fall on the animacy scale: human, impersonal, metaphoric, animal, plant, inanimate, indeterminate.

Quantifying this factor was straightforward since there are only four possible relations: the subject is higher in animacy than the object, the object is higher in animacy than the subject, the subject and object are equal in animacy than the object or the relative animacy cannot be determined as a result of the subject lacking identifiable animacy (no subject, clausal S, etc.). Examples of each relation are shown below:
a. Subject animacy higher:

7a)  *Yo cuidaba a los gatos.* (HC Mexico City)
I cared for A the cats.

b. Object animacy higher:

7b)  *La riqueza distancia al hombre.* (HC Mexico City)
Wealth distances A man.

c. Equal in animacy:

7c)  *Vimos a las voluntarias.* (HC Mexico City)
We saw A the volunteers.

d. Indeterminate relative animacy (example repeated from above):

7d)  *Educar a un niño es importante.* (HC Mexico City)
Educating A a child is important.

Because the factor group Relative Animacy is a metric of where the agent and patient of a transitive clause fall on the animacy scale in relation to one another, corresponding to SDOM, it serves as a litmus test for Comrie’s claim (see section 2.5.) that the propensity for DOM should be rooted in the clausal relationship between Agent and Patient when the animacy level of the Patient approaches, equals or surpasses the Agent. It is important to recognize that the Factor Group Relative Animacy subsumes the Factor groups Animacy of DO and Subject Animacy  because it is measuring the animacy of both Factor groups. There is a partial redundancy with this Factor Group in that animacy of both Subject and DO are already individual Factor groups. Nevertheless it is the potential effect of S and O animacy *in relation to one another* as it pertains to SDOM that
needs to be quantified. This observation is a novel contribution of this dissertation and will be elaborated in section 4.4.

3.5.4. Definiteness & Specificity

While animacy is indeed the most frequent and common factor associated with the SDOM, I argue that it is neither necessary nor sufficient for SDOM. Other factors have been proposed to help account for the apparent variation in a-marking, and preeminent among them is definiteness.

The primary notion of definiteness in the SDOM literature essentially conflates the morphological form (the NP) of the referent and the identifiability of the referent i.e., specificity. Definiteness must be accounted for in order to both address previous research and to clarify my approach to definiteness as it relates to SDOM. This will make transparent the subsequent operationalization of definiteness and specificity in coding and analyzing the data.

As discussed in section 2.2., definiteness suffers from several distinct meanings and specificity may or may not be entailed or even conflated with the notion of definiteness (see Aissen 2003, Laca 2006). My quantitative approach to SDOM requires absolute values for specificity and definiteness and I distinguished two factor groups: specificity and determiner form (see 3.5.5); what has traditionally been associated with ‘definiteness’ is quantified based upon what type of determiner is used.
3.5.5. Quantifying specificity

The factor group specificity consisted of two factors: specific and non-specific. In this case, specific objects consist of uniquely identifiable referents from the speaker’s perspective based on the discourse context. Because of the complexity in determining specificity I used several heuristics to ascertain whether a referent was specific or non-specific.

Following along the lines of Shain (2009:17), my measurement of specificity depends on the de dicto vs. de re distinction. If referents are selected from and stand apart from a particular set, they are specific. Non-specific referents are those that are not unique among the members of their sets.

a. Specific

8a) \[y\ nos fuimos a León a ver a una hermana suya que tiene casada.\] (HC Madrid)
...and we went to León to see A a sister of his that is married.

8b) \[En algunas ocasiones, también, llegué a entrenar a los muchachos de primera fuerza, de acuerdo a los conocimientos que había yo adquirido.\] (HC Mexico City)
On some occasions, I also was able to train A the first level youth according to the knowledge I had acquired.

b. Non-specific

8c) \[...si yo encuentro a algún amigo gamberro, pues me dedicaré a hacer de extraterrestre por ahí.\] (HC Madrid)
...if I find A some hooligan friend, well I’ll dedicate myself to act like an alien around there.

8d) \[Si somos religiosas, cómo vamos a vivir con esta abundancia, ¿no? Vámonos a buscar a otras monjitas.\] (HC Mexico City)
If we are religious, how are we going to live with this abundance, huh? We’re going to find other nuns.

The *de-dicto vs. de-res* basis for determining specificity was crucial because my perspective regarding specificity is that it is ultimately a property of discourse. While determiners (possessives, articles, etc.) and the noun form (*e.g.* personal pronoun) can be strongly indicative of specific or non-specific referentiality, it is evident that linguistic form alone is not sufficient to determine specificity, ergo my discourse-based disposition (*cf.* 2.3.).

Ultimately determining the specificity of a given DO referent was based upon the context. In addition to evaluating the determiner and the noun form, a number of discourse-context heuristics were used. Contextual information was used including anaphoric reference and relative clauses of the accusative object to determine specificity. Crucially, just as the type of determiner or noun form are not categorical in determining specificity of the referent, the existence of anaphoric/cataphoric reference or relative clauses were not categorical in determining specificity *i.e.*, it is possible to find these phenomena with non-specific referents:

---

31 One example for consideration is the impersonal 2nd person singular, interchangeable with the impersonal “one.” Conveniently used in both English and Spanish, this personal pronoun is typically specific, based upon its deictic nature. Nevertheless, it is readily enlisted for impersonal constructions. It is examples like this and the notorious fallibility of definite and indefinite articles in Spanish that compel me to reject approaches to specificity based on form alone.
9) Sí-- de vez en cuando las investigaciones en vez de hacerlas con animales, cuando se puede, la... las hacemos con enfermos eh... porque eh... lo que más se cuida es siempre es no eh... no molestar al enfermo y no--- en fin--- hacer algo que pueda perjudicarlo; solamente se prueban cosas que no lo molesten en absoluto o muy poco--- y--- que puedan dar estÉ...información que no se pueda obtener con animales. (HC Buenos Aires)
Yes---from time to time the research we do it on animals, when able, we do it with sick people um, because um...the thing we care about is to always not, um, not bother A the sick person and no---in the end---do something that might injure him; we only test things that absolutely don’t injure him or very little--- and --- that can dive us information that can’t be obtained with animals.

In the case above, while there is anaphoric reference, there is no unique, identifiable sick person to whom the doctor is referring, therefore the referent is counted as non-specific.

Further expanding on these types of conditions, hypotheticals are clear cases of non-specific referents in my analysis of the data:

10) “...si tengo que defender a... en un juicio commercial a dos personas que para mí son exactamente las mismas... eh... si son dos comerciantes que discuten por cuestiones de intereses que, en el fondo a mi me es totalmente indiferente.” (HC Buenos Aires)
If I have to defend A in a public court A two people that for me, are exactly the same, …um…if they are two businessmen that are arguing over personal interests, deep down, it is totally indifferent to me.

In these cases, referents are established in the discourse and behave as if they were specific. However, because they do not have a unique, existential referent, they are coded as non-specific.
Finally, the previous examples are particular cases of the generalization that in any and all cases where the DO referent is non-existent, meaning the referent is not uniquely identifiable in the present world, the referent was counted as non-specific. Thus other contexts that established non-specificity included *ningúno/a/s* and *cualquier*.

3.5.6. Definiteness

The category definiteness is generalized along the lines of Abbott (2006). However, definiteness only speaks to the morphological form that is used. In this sensed my conceptualization of definiteness is slightly distinct from *definiteness* as it is presumed in much of the research on SDOM (Aissen 2003, Laca 2003, Melis and Flores 2009, et al.) in that I assume specificity is independent definiteness. This FACTOR GROUP includes three distinctions in definiteness: *definites, indefinites* and bare NPs. Definites include:

a. *Definite articles*

11a) *...he terminado por odiar, así en términos generales, a la burguesía...* (HC Buenos Aires)
...I’ve ended up hating, in general terms, A the bourgeoisie...

b. *Demonstrative adjectives*

12b) *Mira, mi hijito: se ve que adoras a esta niña ¿verdad?* (HC Mexico City)
Look, mi child; it’s obvious that you adore A this girl, right?
c. Possessive adjectives

12c) El caso fue que ella mató a sus hijos para que no le estorbaran. (HC Mexico City)
The case was that she killed her children so that they didn’t get in her way.

12d) Sabíamos conocer a Libia. (HC Madrid)
We knew how to know Libya

Indefinite expressions include:

a. Existential quantifiers and indefinite articles

12e) La conocieron como pudieron conocer a una prima. (HC Mexico City)
They knew her as well as they could know a cousin.

b. Bare NP:

12f) ... las obras que representaban ∅ personajes griegos y que sucedían en el ámbito de Grecia... (HC Buenos Aires)
...the works the represented ∅ Greek characters and that took place en the Greek surroundings...

This is a departure from the usual conceptualization of definiteness found in the SDOM literature because this operationalization strips out specificity as an entailed property of definiteness, reflecting only the morphological form of the NP. This approach to definiteness is based upon the assumptions by Enç (1991) that there can be
specific and non-specific indefinites and the assumption by Lyons (1999) that there can be specific and non-specific definites.

a. Specific definite:

13a) *Teníamos que defender a nuestra familia* (HC Madrid)
We had to defend our family.

c. Non-Specific definite:

13b) *...tratábamos a las mamás y a los niños* (HC Mexico City)
...we treated the moms and the children

c. Specific indefinite:

13c) *Yo defendí a un muchacho que había sido acusado...* (HC Mexico City)
I defended a guy that had been accused...

d. Non-Specific indefinite:

13d) *...ve usted a un obrero vestido igual que un señor* (HC Madrid)
...you see a worker dressed the same as a gentleman.

In addition to my distinction of specificity and the determiner form traditionally associated with definiteness, my distinction has the particular advantage of being able to account for bare NP examples which are possible in Spanish:

a. Specific Bare NP:

14a) *...este páramo ha sido el que ha formado a España* (HC Madrid)
…this plateau has been the one that has formed A Spain

b. Non-specific Bare NP:

14b) …deberían de empezar a poder aceptar a ingenieros, a las maestrías o doctorados (HC Mexico City) …they ought to start to be able to accept A engineers in the master’s or doctoral [programs].

3.5.7. Clause Type

Another factor that has been cited in the literature as coinciding with SDOM is embedded clauses (Torrego-Salcedo 1999, Roegiest 2003, Laca 2003). These authors have pointed out that in the embedded clause construction, the accusative object of the matrix verb functions as the subject of the infinitival embedded clause. The result is the conflation of both agent and patient roles on this DO. Therefore I include a category to evaluate marking in this particular syntactic construction.

a. Main clause

15a) (El doctor) atiende al enfermo. (HC Buenos Aires) The doctor attends A to the ill patient

b. Embedded clause

15 b) (El doctor) hace pasar al enfermo. (HC Buenos Aires) The doctor makes A the ill patient come in.
3.5.8. Discourse status

Discourse status is the factor group I used to measure the information status of the referent in the discourse. Relating to SDOM this FACTOR GROUP has unfortunately been called ‘definiteness’ by Enç (1991) and von Heusinger & Keiser (2003); however, to disambiguate, I will use the term discourse status. While there are several potential and elaborate ways to discuss the information status of a referent in a given discourse (cf. Prince 1981, 1992; Gundel et al. 1993), I opt to follow the model outlined by Birner (2006b). Birner’s approach provides four clearly distinguishable levels of information status that can be coded for and were observed in the data I was coding.

<table>
<thead>
<tr>
<th>Discourse-old:</th>
<th>Hearer-old:</th>
<th>Hearer-new:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evoked (inferentially linked and known to hearer)</td>
<td>Bridging Inferrable (inferentially linked, but not known to hearer)</td>
<td></td>
</tr>
<tr>
<td>Discourse-new:</td>
<td>Unused (not inferentially linked, but known to hearer)</td>
<td>Brand-new (not inferentially linked, and not known to hearer)</td>
</tr>
</tbody>
</table>

Table 3 Information status (from Birner 2006b)

The following are examples taken from the data that illustrate the information status of referents based upon Birner’s model:

\[ a. \text{ Evoked} \]

16a) Y estaba el Sagrado Corazón, y abajo... una estatua de Beethoven; y yo muriéndome con la niña que venía, y viendo a Beethoven. Cuando veo a Beethoven digo: ¡Ay! Aborrezco a Beethoven.
b. Bridging Inferrable

16b) Eso, siempre y cuando los padres deseen proyectar a los hijos hacia un futuro mejor que el que uno ha tenido. (HC Mexico City)
That, always and when the parents desire prepare their children for a better future than the one that one has had.

c. Brand-new

16c) Mi mamacita se fue al teatro, y dejó como seguridad en la casa a un primo con su esposa. (HC Mexico City)
My girlfriend went to the theater, and for safety, left a cousin and his wife at home.

d. Unused

16d) ¿la vieron a Niní Marshall? (HC Buenos Aires)
Did you all see A Niní Marshall?

In 16a the referent Beethoven is introduced to the discourse as the speaker describes the room where she was giving birth. Beethoven is now hearer-old and discourse-old because of the explicit mention of Beethoven. Beethoven is now an evoked referent in the discourse and the subsequent mentions of Beethoven, all a-marked incidentally, are coded as evoked.

16 b is an example where the introduction of referent in the discourse makes certain knowledge or assumptions salient that hearers can infer by making a connection,
i.e. a bridging inferable. In this case, because the speaker chose the term *padres* ‘parents,’ the referent *niños* ‘child’ can be immediately inferred; there is universally understood connection between ‘parent’ and ‘child’ because the conditions necessary to be a parent entail that one has one or more children.

*Brand new* represents those referents that are both *discourse new* and *hearer-new*. At no previous point in the discourse had the speaker made mention of her cousin. Therefore it is new to the discourse. There is nothing in the prior discourse to indicate that the referent ‘cousin’ could be inferred. Finally, the means by which the ‘cousin’ is introduced using the indefinite articles strongly suggests that the referent is *unused*. If the existence and identity of this cousin in relation to watching over her house had been known to the speaker’s interlocutors, we would expect a possessive or even the cousin’s name.

*Unused* describes the condition where a referent is introduced to the discourse in a way that indicates that this referent is previously known to the hearer or the speaker assumes the hearer is aware of the referent, hence *hearer-old*. However there has been no explicit mention of the referent in the prior discourse i.e., it is not *evoked*, nor is there anything in the prior discourse that would lead the interlocutor to infer that such a referent is relevant to the discussion i.e., it is not a *bridging inferable*. This is the condition of being *discourse new* and *hearer-old*. In the case of 16 d, Niní Marshall was a TV celebrity of the time and this question serves as a topic-change in the interview instigated by one of the participants. Nothing in the prior discourse would provide reasonable cause to assume this is a case of a bridging inferable. However, the way the
speaker introduces the referent indicates that Niní Marshall is assumed to be known to all participants. This is corroborated in two ways. First, the way in which the speaker introduces the referent in her question indicates that the speaker assumes that Niní Marshall is known to all interlocutors. Second, the following responses by the speaker’s interlocutors indicate that they were familiar with Niní Marshall.

3.5.9. Number

Kliffer (1984, 1995) discusses the relevance of highly individuated object referents to SDOM. Specifically, he notes that there is a propensity for highly individuated objects to be marked where individuated objects tend to have characteristic features. One of these features is number with the assumption that singular referents are more individuated than plural referents. The intuition is that singular objects are likely specific (non-generic) and potentially more patient-like than plural objects precisely because there is one individual referent functioning as the P. As a result, I coded for number: singular vs. plural to test this assumption. However not all nouns are count-nouns: mass nouns were coded as non-count.

a. Singular

17a) Oye, yo a las seis también me quiero ir. Que me gustaría ir a buscar a mi madre a casa. (HC Madrid)
Hay, at six I also have to go. I’d like to go look for my mom at home.
b. Plural

17b) *Pero no me puedo despertar muy tarde--- porque tengo que ir a casa temprano, cambiarme, buscar las cosas para la Alianza.*
(HC Buenos Aires)
But I can’t wake up late, because I have to go home early, change, look for the **things** for the Alliance.

c. Non-count

17c) *Era muy, muy estricta para eso, y además curiosa: ¡le encantaba ver a la gente!* (HC Madrid)
She was very, very strict regarding that, and really curious as well: she loved to see the **people**!

3.5.10. Topicality

Quantitative analysis of topicality in Spanish has been demonstrated by Bentivoglio (1983) following Givon’s (1983) metrics of Referential Distance (RD) and Topic Persistence (TP) (see 2.8). Briefly, RD is a measure of the number of clauses between a referent and the most recent overt realization of its antecedent, i.e., looking back to the most recent manifestation of the referent with an upper bound of 20 clauses. TP is a count of the number of times a referent persists in the discourse in the following 10 clauses i.e., looking forward in the discourse. Schwenter (2006) suggests Bentivoglio’s findings are indirect evidence that SDOM is indeed a marker of topicality because Bentivoglio shows that those NPs that are more topical are also animate. Therefore, any quantitative analysis that hopes to measure topicality must operationalize topicality with uniformity in all contexts.
I expanded upon Givon’s RD and TP metrics (1983) by following Shain (2009: 77) and the measures of topicality he proposes. Shain acknowledges the challenge of identifying a single complete measure of topicality or “total topicality” since RD and TP are mutually exclusive in how and what they measure. Shain’s solution consists of extending RD forward and TP backward to arrive at metrics called Forward RD (F-RD) and Backward TP (B-TP). The sum of RD and F-RD provides Total Topicality in terms of RD (T-RD). Likewise the sum of TP and B-TP results in a measure of Total Topicality in terms of TP (T-TP).

One of the key advantages of these additional metrics is that they are able to accurately reflect the topicality of referents that are either chain-initial topics or chain-final topics instead of non-topic referents. Whereas the referent of a chain-initial topic will appear non-topical from an RD perspective i.e., there are no prior mentions and likewise a chain-final topic for the standard TP measurement, these “total” measures still capture that the referent is topical when the entirety of the discourse is taken into account. Shain’s approach is tremendously useful for operationalizing topicality in my quantitative approach to SDOM. Because the only evidence readily presented to support a topicality function of SDOM are syntactic constructions that purportedly reflect higher topicality, and therefore represent a circular argument, Shain’s methodology deftly relies on the discourse context to determine the topicality of a referent.

However, an important disclaimer must be made. This measure of topicality is purely relative in as much as the metric indicates the frequency of a particular referent in the discourse compared to others. It is not a means for determining a topic per se, merely
that a higher frequency would indicate the “aboutness” or relevance of a referent to the
discourse as a whole. In this fashion, I was able to provide quantitative data for a relative
notion of topicality built upon Givón’s approach with the metrics of T-TP and T-RD.

The counts for RD and TP were obtained by counting back each full clause and
determining if the referent was mentioned in the clause or not. The count was a
maximum of 1 per clause thus clitic doubling or verb inflection with overt subject
pronouns were still only considered once. In addition to inflection and nominal referents,
adjectives including possessive adjectives and descriptive adjectives were also considered
as referential but only in the case that no verbal or nominal reference could be made.

The uniformity of the referent was also critical to this count. Synonyms for the
same referent in a given discourse were considered manifestations of the same referent.
While absolute semantic synonymy is not an assumption that I make, it is the case that
the same referent may have different names. In one excerpt of the discourse, an
informant refers to her husband by name, and also by using the terms marido ‘husband’
and esposo ‘spouse/husband.’ Clearly the referent is the same regardless of terminology
and these were consider instantiations of the referent and counted accordingly for the
measures of TP and RD. This type of complication extends to deictic expressions as
exemplified below:

19)  

Int. La midi no me gusta. Bah- - - vos, ¿esto es midi, no? Disculpá [..........]
In the examples above, the interviewer asks about the skirt the informant is wearing while discussing the pros and cons of mini-skirts and longer (maxi) skirts. The informant refers to her own skirt with a different word, pollera, but it is clear that the referent of “esto” and “pollera” are the same and thus each counts in the measure of TP and RD.

This last example also brings up important issues for what counts in measuring TP and RD. As mentioned earlier, I excluded tokens from the speech of the interviewers because they are not reliably representative. However, the function of TP and RD is to measure how the referent appears in the discourse, which necessarily includes its occurrence in the speech of the interviewer. As a result, the measures of TP and RD include any occurrence of the referent irrespective of speaker. Additionally, when a referent has a sufficient and obvious synonym (sharing that same referent) mentioned in the discourse, it is also included in the TP and RD counts.

3.5.11. DO Placement

Related to the notion of topicality has been the placement of DOs with respect to the verb. Left dislocation is the prototypical structure for topicality (Downing and
Downing 2007). While I have already mentioned the circularity of determining topicality via a particular structure (left-dislocation) and using that same structure as evidence for topicality, I believe that it is relevant and potentially informative to consider correlation between SDOM and displacing the DO from its prototypical post-verbal position. Von Heusinger (2008:16) alludes to as much in a footnote, hypothesizing that SDOM may result from any displacement of the DO, left-dislocation or otherwise. To test this claim, I accounted for the placement of the DO in relation to its governing verb with the possible locations being preposed (i.e., left-dislocated), postverbal (prototypical) and adjacent to the verb or post verbal with something intervening where some element intervenes between the transitive verb and the DO. Below I have underlined the verb to emphasize its relation to the position of its DO complement.

a. left-dislocated

Entonces, nosotros a los niños no los vamos a educar, a guiar por ninguna religión, dejarles completamente... (HC Mexico City)
So, we aren’t going to educate them, to guide them towards some religion, let them completely...

b. postverbal adjacent

Bueno, personalmente conozco a unos. (HC Madrid)
Well, personally I know A a few.

c. postverbal with something intervening

Mi mamacita se fue al teatro, y dejó como seguridad en la casa a un primo con su esposa. (HC Mexico City)
My girlfriend went to theater, and left A a cousin and his wife to protect the house.
3.5.12. Ambiguity

To address the ambiguity hypothesis, verbs were categorized according to the possible ambiguity that may arise when A and P are equally likely candidates for A. This FACTOR GROUP targets the ambiguity hypothesis of SDOM, where unexpected or exceptional marking results from the need to disambiguate two equally probable referents per the semantics of the verb. This FACTOR GROUP was based upon Shain (2009) and his Thematic Fit category, “a non-linguistic measure based on intuitions about the way the world works” (p. 75) with binary classification referring to whether the arguments of a transitive clause are equally likely to fill the semantic role of A or P (+ ambiguous) or if there is predisposition for a particular A or P reading (- ambiguous). The “fit” was established by determining if either referent could likely fill the roll of A or if there was a clear interpretation of A and P per the verbal semantics and knowledge of the world.

Ambiguity possible:

*Horacio sigue fielmente a Aristóteles* (HC Buenos Aires)
Horatio faithfully follows Aristotle

*El pop ha tomado a la publicidad* (HC Buenos Aires)
Pop has taken over publicity

Grammar precludes ambiguity:

*Ah, la fui a ver a Laura K* (HC Buenos Aires)
I went to see Laura K

*...a ésta siguieron otras y otras* (HC Buenos Aires)
…others and others followed A this one

It should be noted then, that ambiguity in this analysis can only be based upon the semantic ambiguity that arises when the grammar does not readily disambiguate the subject and the object (i.e., person agreement, number agreement, relative clauses, case, etc.) and the subject and object referents being equally likely to fulfill either agent or patient roles given one’s understanding of the world. Therefore, the only cases where ambiguity could potentially exist was in cases where either referent is equally possible as A or P and the grammar does not disambiguate the two.32

3.6. Statistical Analysis

Having established the envelope of variation in Section 3.4, the data that did fall within the envelope of variation were analyzed and evaluated using various statistical tools and techniques. One of the primary and innovative objectives of this undertaking was to provide quantitative evidence that accounts for SDOM in Spanish. I relied on Goldvarb (Sankoff et al. 2005, Tagliamonte 2006) as the primary means to organize and interpret my data. It is important to note that as with any quantitative variationist approach, the results are probabilistic (Sankoff 1988, Guy 1991). This falls in line with the crucial assumption in variationist approaches to language that the distribution is non-categorical and not evenly distributed. The insights gained from a probabilistic approach to language are derived from a true-to-life account of language use as opposed to

32 As will be seen in Chapter 4 and Chapter 5, these conditions for ambiguity were extremely infrequent in the data.
anecdotal or contrived data. The advantage of this approach is in capturing the dynamic conditions of variation related to SDOM as manifest in spontaneous spoken language.

However, in undertaking a quantitative analysis of SDOM it is imperative to acknowledge that the simple implementation of sophisticated statistical tools for quantifying linguistic variation in and of itself does not provide an account of SDOM or any other linguistic phenomenon. It is incumbent upon the researcher to interpret the data, to provide a linguistically-informed analytical perspective in exploring and explaining the potential models of the SDOM phenomenon. Goldvarb allows for discrete and fine-grained analyses of the data and the ability to recode and reorganize the data in order to understand how the data are distributed and to correctly characterize the phenomenon. It is the researcher’s responsibility to understand the data and how the data relate to the factor groups that have been used for regression analysis. Crucially, these types of variationist analyses can provide empirical evidence to support or recast generalizations that have been made regarding the factors that motivate variation in SDOM. Chapters 4 and 5 will discuss in detail the results of the Goldvarb analyses and present the distribution of the data. I detail the selections and decisions I made in analyzing the data and discuss my assumptions and choices based upon linguistic precedence and observation.

3.7. Conclusion

In this chapter I have outlined the theoretical and methodological approach of this dissertation. The variationist approach to linguistic data was discussed as well as the
crucial assumptions of the linguistic variable. The variable nature of SDOM requires both a theoretical perspective and methodological approaches that can account for this observed variation. A discussion was then presented regarding the validity of a syntactic variable. The focus of this research is to quantify the variation of the syntactic variable of SDOM as it is manifest in three dialectal corpora of spoken Spanish from the Habla Culta Project. The dialects of Buenos Aires, Madrid and Mexico City provided the diversity of varieties of Spanish to evaluate SDOM. The envelope of variation, i.e. the context where variation is possible, was circumscribed and the invariant, categorical cases where marking or non-marking occurred were ruled out. The procedure for extracting tokens was outlined and the coding scheme was made explicit. This included detailed descriptions of each factor group and the subsumed factors in each.
Chapter 4: Quantitative Analysis

4. Overview of Quantitative Analysis of Spanish Differential Object Marking

This chapter presents the findings of SDOM resulting from the methodology outlined in Chapter 3. This includes the results of the multivariate analysis of all data obtained from the three dialectal corpora. This macro-level multivariate analysis described in the following section is based upon those factors selected as significant in explaining how a-marking appears in the data. The sections of this chapter are organized to present a detailed analysis of each significant factor group from the multivariate analysis. Section 4.1 presents the overall distribution of the collective data from the three corpora. This is followed by analysis and discussion of animacy in Section 4.2. Following animacy is the discussion of definiteness in Section 4.3. The following section (4.4.) focuses on the particular problems encountered with relative animacy. Section 4.5 analyzes the effect of clause structure on marking in the data. In Section 4.6 the disambiguating function of SDOM is discussed in light of the findings from the data. Referential distance is the topic of 4.7. Section 4.8 presents the effect of specificity on SDOM. In Section 4.9 I analyze how number correlates with marking. This is followed by the findings regarding dialect as an independent factor. A summary of all of Chapter 4 is presented in Section 4.10.
4.1. Overall distribution of SDOM

Following the methodology of Chapter 3, a total of 2090 tokens were culled from the three dialect corpora.\textsuperscript{33} I first consider these data as one large data set to capture broad generalizations regarding the distribution of SDOM and the factors that motivate the observed variation.

<table>
<thead>
<tr>
<th>\textit{a}-marked</th>
<th>27% (569)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmarked</td>
<td>73% (1521)</td>
</tr>
<tr>
<td>Total (N)</td>
<td>2090</td>
</tr>
</tbody>
</table>

Table 4  Overall frequency of SDOM

Depending on one’s perspective regarding the frequency of DOs in discourse and the frequency of SDOM, these figures may seem high or low. The critical perspective on this frequency is to remember that these frequencies are essentially “adjusted.” The 2090 tokens culled from the data do not represent every DO in the data. DOs (both marked and unmarked) were excluded because they fell outside the envelope of variation as presented in Section 3.4. The primary group of marked DOs excluded was personal pronouns. When one considers the use of DO pronouns in spoken Spanish, the actual frequency of DOs is obviously much higher than the 2090 tokens obtained from the data. In addition to pronouns, the other frequent tokens in the data not considered for SDOM variation were DOs that appeared in the speech of the interviewer.

\textsuperscript{33} In actuality, more than 2300 tokens were initially obtained from the data. However careful review of the data allowed me to remove those tokens that did not fall within the envelope of variation.
The tokens that fell within the envelope of variation provided the data for quantitative analysis. Each token was coded per the factor groups outlined in Section 3.5 in order to be interpretable for the Goldvarb statistical program. The multivariate analyses required extensive examination of the data. Numerous reconfigurations of the data were necessary to identify the most accurate means for interpreting the data. This was primarily accomplished in two ways; some factors (F) within factor groups (FG) were necessarily collapsed because of the (poor) distribution of the data and other groups were excluded from successive multivariate analyses because it was evident they were not statistically significant. This recoding of FGs and subsequent reanalysis of the data is discussed where applicable in the following sections.

The results of the multivariate analysis of the entire data collected are found in Table 5. The format of Table 5 and subsequent multivariate analysis tables in this and the following chapters allow a straightforward and uniform presentation of the data, providing the critical information to calculate distributions and interpret the Goldvarb results. The Goldvarb results select the FGs that are statistically significant in accounting for the data and assign a probability for each F within the FG in terms of a factor weight. This factor weight is based upon the assumption (i.e. the null hypothesis) that there is no factor (independent variable) that accounts for an observed linguistic phenomenon (the dependent variable) and the likelihood that the phenomenon will occur is 50% i.e., it is at chance. The factor weight assigned to this is .5; when a factor weight is greater than .5 the factor favors the phenomenon (i.e. overt a-marking) and if it is less than .5, the factor disfavors the phenomenon. These factor weights are understood as probabilities that a
particular factor (the independent variable) is motivating the behavior of the linguistic phenomenon (the dependent variable). For this dissertation the Goldvarb program evaluates SDOM (the dependent variable) per the specified factor groups (the independent variables) and assigns a probabilistic measure of whether a factor favors a-marking or disfavors a-marking. In Table 5 each factor group is listed in the left-hand column with the individual factors that make up each factor group. The probability that a particular factor favors or disfavors SDOM is then provided in the 2nd column in terms of a factor weight. The 3rd column provides the percentage of marked tokens as it relates to the 4th column of Total tokens (both marked and unmarked) of that particular factor. Finally, the total number of tokens for a given factor in the 4th column is shown as a percentage of the entire data in the last column. The factor groups are arranged vertically based upon the range. The range is determined by calculating the difference between the highest and lowest factor weights (probabilities) within each factor group. Ordering the factor groups by range captures the effect of a factor group in terms of the largest disparity between factors favoring and disfavoring a-marking. The factor groups that failed to obtain statistical significance in the multivariate analyses are not included in these tables because they do not explain nor improve an account of SDOM.

In reference to the Goldvarb program and multivariate analysis, factor weight is synonymous with probability i.e., the probability shown in the probability column of the multivariate analysis table is the same as the factor weight.
<table>
<thead>
<tr>
<th>Probability</th>
<th>% *-marked</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Object Animacy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>.91</td>
<td>72%</td>
<td>681</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>.68</td>
<td>41%</td>
<td>60</td>
</tr>
<tr>
<td>Inanimate</td>
<td>.23</td>
<td>4%</td>
<td>1349</td>
</tr>
<tr>
<td><strong>Definiteness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite</td>
<td>.62</td>
<td>33%</td>
<td>1434</td>
</tr>
<tr>
<td>Indefinite</td>
<td>.36</td>
<td>20%</td>
<td>418</td>
</tr>
<tr>
<td>Bare DO</td>
<td>.12</td>
<td>7%</td>
<td>238</td>
</tr>
<tr>
<td><strong>Relative Animacy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object greater</td>
<td>.80</td>
<td>76%</td>
<td>116</td>
</tr>
<tr>
<td>Equal animacy</td>
<td>.65</td>
<td>64%</td>
<td>554</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>.65</td>
<td>26%</td>
<td>324</td>
</tr>
<tr>
<td>Subject greater</td>
<td>.35</td>
<td>4%</td>
<td>1054</td>
</tr>
<tr>
<td><strong>Clause Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small clause</td>
<td>.87</td>
<td>86%</td>
<td>28</td>
</tr>
<tr>
<td>Normal</td>
<td>.49</td>
<td>26%</td>
<td>2062</td>
</tr>
<tr>
<td><strong>Ambiguity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>.78</td>
<td>80%</td>
<td>78</td>
</tr>
<tr>
<td>No</td>
<td>.49</td>
<td>25%</td>
<td>2012</td>
</tr>
<tr>
<td><strong>Total Referential Distance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-20 Clauses</td>
<td>.61</td>
<td>50%</td>
<td>161</td>
</tr>
<tr>
<td>21-39 Clauses</td>
<td>.60</td>
<td>36%</td>
<td>641</td>
</tr>
<tr>
<td>40 Clauses</td>
<td>.41</td>
<td>16%</td>
<td>1127</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.58</td>
<td>32%</td>
<td>1178</td>
</tr>
<tr>
<td>Non-specific</td>
<td>.40</td>
<td>22%</td>
<td>912</td>
</tr>
<tr>
<td><strong>Number</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singular</td>
<td>.55</td>
<td>28%</td>
<td>1000</td>
</tr>
<tr>
<td>Plural</td>
<td>.42</td>
<td>21%</td>
<td>698</td>
</tr>
<tr>
<td><strong>Dialect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexico City</td>
<td>.58</td>
<td>34%</td>
<td>622</td>
</tr>
<tr>
<td>Madrid</td>
<td>.52</td>
<td>26%</td>
<td>467</td>
</tr>
<tr>
<td>Buenos Aires</td>
<td>.44</td>
<td>23%</td>
<td>1001</td>
</tr>
</tbody>
</table>

Total N = 2090, Input: 0.109 (27% *-marked), Log Likelihood = -545.414, p =.018

Factor Groups not selected: DO placement, Discourse status, Subject Animacy, RD, FRD, TP, BTP, TTP

Table 5 Factors contributing to SDOM
4.2. Direct Object Animacy and SDOM

Unsurprisingly, Direct Object Animacy (DO Animacy) is the FG that accounts for the greatest amount of variation in a-marking in Spanish. The basic generalization regarding what motivates SDOM is supported by the multivariate analysis; human DOs strongly favor marking with a probability of .91 while inanimate DOs strongly disfavor marking with a probability of .23. The effect of DO animacy is inescapable. Furthermore, the extent of its relevance in the other FGs will be seen shortly, providing better insight into the robustness of DO animacy in accounting for SDOM.

Animacy of the DO has few relevant categories. Despite my efforts to establish a detailed animacy scale (cf. 3.5.1.), the data distribution and the results of numerous Goldvarb analyses show that DO animacy is basically sensitive to a binary Human/Non-human distinction. While I believed the animacy scale elaborated in Chapter 3 would provide some fine-grained description to SDOM, the data provided too few tokens to provide the detail I had presumed to show. Quite simply, the corpora did not provide sufficient numbers of marked and unmarked DOs beyond the binary classification of human and inanimate DOs.

Table 6 presents the exceptionally poor distribution of DO Animacy in the data despite combining all three corpora.\footnote{This will be especially prominent in the dialectal distributions as these dialects are subsets of data.}
From the start, 94% (1968/2090) of the tokens are already definitively human or inanimate. As the data show, the remaining 6% do not provide much data to be distributed among the other measures of animacy. These discrete animacy distinctions are clearly low in frequency with the unfavorable result that two groups, Impersonal DO and Plant, show no variable marking.

Because of these extreme disproportions, I was forced to recode by collapsing groups. Several configurations were attempted but what provided the best results involved collapsing the FG Impersonal with Human and collapsing Animate and Plant with inanimate. Collapsing Impersonal with Human was a straightforward decision since these constructions, while non-referential, entail humanness of the DO. I had presumed that non-sentient organisms (plants) would be treated as inanimates, however I was surprised to see virtually no variation with animates. There is one token of marking with animates in the Buenos Aires corpus, one example in the Madrid data and no cases of
animates being marked in the Mexico City Corpus. With just 4% (2/49) of non-human animates being marked, it appears that animates are treated like inanimates and I assume as much for the purposes of the analysis.

With these adjustments to the factors Human and Inanimate, the log likelihood improved each time and so only metaphorical DOs remained. Metaphorical DOs present an interesting situation since they include personified and metonymic DOs. Since they are semantically non-human but assigned human-like properties in a given context, I felt they should remain an independent factor within the factor group of animacy. The multivariate analyses bore this out; collapsing metaphorical DOs with Human and then Inanimate only worsened the analysis in each case.

From these attempts to more discretely define animacy, the data show that there are two primary animacy distinctions in Spanish as they relate to a-marking: human and non-human. The statistical and regression analyses provide evidence that SDOM is overwhelmingly sensitive to the animacy of the DO. The third relevant category of animacy, metaphorical DO, felicitously straddles the other two categories: inanimate DOs have human-like properties extended to them within the context, and as such it is

36 HC Buenos Aires: “Envolviendo al animal... este... es una riqueza.” ‘Wrapping up A the animal...um...is a pleasure.’

HC Madrid: “Es capaz de conocer a un animal a distancia.” - ‘He’s capable of recognizing A an animal from distance.’

37 The effect (and importance) of frequency on the data can be demonstrated by observing the fact that the overall marking of inanimates in the data is higher (4%) than that of the exceedingly limited number of animates (0.2%). Following a fallible line of logic, marking would appear to prefer inanimates to non-human inanimates based on these numbers alone.
unsurprising to observe significant variation of $a$-marking in these cases. In other words, more variance in the marking observed with metaphorical uses appears to support an argument that the very nature of metaphorical uses (i.e. extensions of animacy by some means to inanimates) would lead to these DOs being considered very much a transanimate category. The robust quantitative evidence from the multivariate analysis supports DO animacy as the principal factor in motivating SDOM. While more discrete categories of Animacy are not sufficiently represented in the data, this paucity of data provides useful insight into the relevant and representative categories of animacy: the primary animacy distinction is human vs inanimate. What will soon become evident with closer inspection of the data is the persistence and pervasiveness of DO animacy in the other factor groups.

4.3. Definiteness and SDOM

The multivariate analysis shows that definiteness is selected as significant in accounting for SDOM in the data. While it may initially seem unsurprising to find the prominence of definiteness as a feature related to SDOM, the results of definiteness become far more interesting under closer inspection.
What immediately stands out is the disproportionate number of definite direct objects. Per my methodology and much along the lines of Abbott’s distinctions between definites and indefinites (2004), the class of definites includes definite articles, demonstrative adjectives and pronouns, possessive adjectives, proper names and universal quantifiers. The frequency of occurrence of definites in the data is more than triple that of indefinites. It is important to note that certain types of indefinites were excluded for phonetic reasons, namely algún/alguna ‘some’, alguien ‘someone’ and algo ‘something’ (cf. Section 3.4.). Potentially including these cases would still not account for the large difference between definite and indefinite tokens found in the data however. There are 1016 more definite tokens (N = 1434) than indefinite tokens (N = 418). These data reveal the overwhelming pervasiveness of definite DOs in Spanish relative to indefinite DOs and especially relative to Bare DOs.

The third component which appears to be frequently overlooked when discussing definiteness is the case of Bare DO referents. While there are very few of these cases in the data at only 11% (238/2090) they do occur and demand separate consideration.

<table>
<thead>
<tr>
<th>Definiteness</th>
<th>Marked</th>
<th>Unmarked</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite</td>
<td>(33%) 470</td>
<td>(67%) 964</td>
<td>1434</td>
</tr>
<tr>
<td>Indefinite</td>
<td>(20%) 83</td>
<td>(80%) 335</td>
<td>418</td>
</tr>
<tr>
<td>Bare DO</td>
<td>(7%) 16</td>
<td>(93%) 222</td>
<td>238</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 84.246 \quad p < .0001 \quad df = 2 \]

Table 7  Rate of marking for definite DOs
because they cannot be classified as definite or indefinite expressions. By coding for these additional tokens, the data reveal how strongly Bare DOs disfavor marking with a probability of .120. It is this disfavoring probability that results in the large range (50) found in the Goldvarb results.

Table 4.4 clearly indicates that a-marking coincides with this pervasiveness of definite NPs. Part of this might be accounted for by considering the class of proper nouns that fall within the category of definiteness, since many proper nouns are human referents. It may be that the increased a-marking in definiteness is influenced by the presence of human referents that are proper pronouns. Of the 1434 definite tokens, 12% (160) are proper nouns. Of the 160 proper pronouns in the data, 101 are human referents. Although proper nouns are most certainly definite expressions, if we were to remove the human proper nouns from the data, we would still see that definite expressions are still marked more frequently.

<table>
<thead>
<tr>
<th>Definiteness</th>
<th>Marked</th>
<th>Unmarked</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite</td>
<td>(28%)</td>
<td>(72%)</td>
<td>1333</td>
</tr>
<tr>
<td>Indefinite</td>
<td>(20%)</td>
<td>(80%)</td>
<td>418</td>
</tr>
<tr>
<td>Bare</td>
<td>(7%)</td>
<td>(93%)</td>
<td>238</td>
</tr>
</tbody>
</table>

$\chi^2 = 57.08$  
$p < .0001$  
df = 2

Table 8  Rate of marking of definite DOs with human proper nouns excluded
This leads to a broader question regarding how animacy, namely human referents, may be interacting with the factor group Definiteness. Including all definite DOs, they can be compared with DO Animacy where the effect of animacy can be immediately observed.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Definite</th>
<th>Indefinite</th>
<th>Bare DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>83% (402)</td>
<td>58% (74)</td>
<td>21% (15)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>44% (23)</td>
<td>29 % (2)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>5% (45)</td>
<td>2% (7)</td>
<td>1% (1)</td>
</tr>
</tbody>
</table>

Table 9 Rate of marking according to definiteness and DO animacy

The data confirm the relevance of definiteness to SDOM, but the effect of DO Animacy and most strikingly, the humanness of the referent, is unmistakable. The a-marking is a result of animacy as shown in Table 9 above: the majority of marked objects in the data are definite and human: 72% (402/569). Likewise, indefinite forms and those referents lacking a determiner are rarely marked. However, the table also shows that definiteness increases marking in all cases, regardless of animacy. In addition to definite human DOs being marked in majority of cases, the metaphoric definite were marked also marked more frequently with 23/25 marked metaphoric DOs being definite. Finally, the few inanimates that were marked happened to also be definite (45/53). This is important
evidence supporting (S)DOM arguments that consider definiteness to be a critical
dimension of marking (Comrie 1979, 1983; Company-Company 2002, 2003; Kliffer

In addition to DO Animacy, specificity needs to be analyzed as it relates to
definiteness. A central assumption of this dissertation is to treat definiteness and
specificity as independent factors. To show this, I present the cross-tabulations in Table
4.7.

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Definite</th>
<th>Indefinite</th>
<th>Bare DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>34% (336)</td>
<td>21% (30)</td>
<td>12% (6)</td>
</tr>
<tr>
<td>Non-specific</td>
<td>30% (134)</td>
<td>19% (53)</td>
<td>5% (10)</td>
</tr>
<tr>
<td>Total</td>
<td>33% (470)</td>
<td>20% (83)</td>
<td>7% (16)</td>
</tr>
</tbody>
</table>

Table 10 Rate of a-marking according to definiteness and specificity

The table critically shows that the number of tokens (336) and the rate of marking (34%)
are highest in the case of both definite and specific DOs. What can also be observed is
the contrast in marking with indefinite and non-specific DOs, where marking is only at
19%. Only Bare DOs are marked less frequently, a trend that has already been pointed
out. When we consider the relevance of definiteness to marking, it is also clear that
marking is notably higher for definite DOs than indefinite DOs in both specific and non-
specific cases. Interestingly, while specificity does favor marking, the observed effect in
the rate of marking is less than the effect of definiteness on marking. This corroborates
the probabilities of the multivariate analysis, where definiteness falls much higher in the
constraint ranking hierarchy of Table 5, indicating that the definiteness of the NP has a strong effect in SDOM variation. The effect of definiteness and specificity on marking in the data provides an important evidence to support the multiple causation hypothesis of this dissertation. Furthermore, these two factors will be shown as critical evidence corroborating the individuation analysis and hypothesis that has been assumed in some SDOM literature.

4.4. Relative Animacy in SDOM

A critical component of this quantitative study is to evaluate the effect of animacy between the subject and the DO. This factor group is in part inspired by Comrie’s work on functional case-marking (cf. 1979, 1983, 1989). Comrie asserts that agents (normally the grammatical subject) tend to be animate and definite while patients (normally the grammatical direct object) tend to be inanimate and indefinite (1989:128). If either the subject or object does not coincide with these “natural kinds of transitive constructions” there will be a preference for marking. I extended this analysis to consider how the varied animacy of subject and object relative to each other may affect marking through the metric of relative animacy.

Building on Comrie’s approach, I hypothesized that in addition, DOM can be expected when these roles are inverted; when objects appear more like subjects (animate/human) or subjects appear less like subjects (inanimate). I have investigated the effects of this FG in previous research (Tippets and Schwenter 2007, Tippets forthcoming) and wanted to compare previous findings with those of this dissertation.
using a more inclusive but also restricted means of identifying tokens.\textsuperscript{38} The Goldvarb results are found below in Table 11.

<table>
<thead>
<tr>
<th>Relative Animacy</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object greater</td>
<td>.80</td>
<td>76%</td>
<td>116</td>
<td>6%</td>
</tr>
<tr>
<td>Equal animacy</td>
<td>.65</td>
<td>64%</td>
<td>554</td>
<td>27%</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>.65</td>
<td>26%</td>
<td>324</td>
<td>15%</td>
</tr>
<tr>
<td>Subject greater</td>
<td>.35</td>
<td>4%</td>
<td>1054</td>
<td>52%</td>
</tr>
</tbody>
</table>

Table 11  Multivariate results for relative animacy

When the object is greater in animacy than the subject, there is a very high probability that marking will occur. In 76\% of these cases where the object was greater than the subject, we find SDOM. This result partially supports Comrie’s assumptions regarding the expected function of marking; the object is more subject-like by virtue of its animacy (humanness). Additionally per my hypothesis, these cases are more likely to be marked because the subject is less subject-like by virtue of lower relative animacy.

The quantitative insight also corroborates Comrie’s initial assumptions regarding the expected characteristics of agents and patients. There is a “strong tendency” for typical agents to be animate and patients to be inanimate. We find the number of “agent-

\textsuperscript{38} In Tippets and Schwenter (2007) and Tippets –(forthcoming), I used an inductive methodology that considered all cases of SDOM and compared them with unmarked cases based up the verbs where marking occurred. This dissertation has been more restricted regarding which tokens of SDOM are considered and more inclusive in the fact that any DO of a transitive verb was initially considered as a possible site of variable marking.
like” patients is very few at 6% (116) while more than half the data follows the prototypical roles with 52% (1096) of all tokens following the pattern of a subject greater in animacy than the DO.

The relevance of the relationship of animacy between subject and object is also evinced in the other factors. When a subject and object are of equal animacy $a$-marking is favored. Again the expected condition is encountered: human subject and inanimate object is not found and marking is favored. Furthermore, in the same way that the prototypical roles are inverted or altered in the case of objects that are equal or greater in animacy to the subject, we also find that those contexts that have no identifiable referent as the subject favor $a$-marking as well.

However there are several important concerns that this factor group and its multivariate results raise. Given that the metric of Relative Animacy is in part derived from DO animacy it is incumbent to establish whether these two factor groups are independent of each other. It is evident that such independence is not attained.

Looking at Table 5 there is a cross-over effect. This is observable when the factor weights of Equal Animacy and Indeterminate Animacy are compared with their rate of marking: the probabilities do not reflect the percentage of marking observed in the data. While Equal Animacy and Indeterminate Animacy both have the same probability (.65) the observed rates of marking are not remotely similar at 64% for the former and 26% for the latter. This mismatch or “cross-over” indicates that the multivariate analysis is measuring an effect in one factor group that it is not independent from another factor group or groups. This interaction between factor groups can be shown using cross-
Cross-tabulations of DO Animacy and Relative Reveal the interactions shown in Table 12.

<table>
<thead>
<tr>
<th>Relative Animacy</th>
<th>Human</th>
<th>Metaphoric</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object greater</td>
<td>78% (82)</td>
<td>67% (6)</td>
<td>[ ]</td>
</tr>
<tr>
<td>Equal animacy</td>
<td>72% (340)</td>
<td>73% (8)</td>
<td>10% (7)</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>68% (65)</td>
<td>55% (6)</td>
<td>6% (13)</td>
</tr>
<tr>
<td>Subject Greater</td>
<td>[ ]</td>
<td>17% (5)</td>
<td>3% (33)</td>
</tr>
</tbody>
</table>

Table 12 Rate of marking for cross-tabulate DO animacy and relative animacy

From the table it is evident that marking is coinciding with the DO being human. In fact 72% (491/681) of all the marked DOs in the data are human. When compared with this overall rate of α-marking for human DOs at 72%, the increase in marking can be observed at 78% when the DO is higher in animacy than its subject.

This may provide evidence that relative animacy does have some relevance in accounting for α-marking; however the effect of DO animacy persists. Notice that in the case of equal animacy between subject and object, the rate of marking is exactly in line with the overall rate of marking for human DOs at 72%. Equal animacy also represents the most frequent marked context of relative animacy where the subject is also human (N = 340). In this respect, there is no way to determine if this condition of equal animacy between subject and object is distinct from merely marking human DOs.

With the majority of tokens coinciding with human DOs and the limited tokens of metaphoric and inanimate tokens, the source of the interactions in the multivariate
analysis is clear. The regression analysis is detecting a shared effect of one factor, human DO, in two different factor groups. Lacking ideal distributions of the data across all cells of DO animacy and relative animacy, any definitive claims about relative animacy at a general level will remain speculative (however see the Buenos Aires results in 5.1.2.). From these observations, relative animacy as a general factor accounting for SDOM is removed from the multivariate analysis of the collective data set.

4.5. Clause Type and SDOM

Despite the minimal number of tokens with embedded clauses, the multivariate analysis was able to identify Clause Type as having a significant effect in SDOM.

<table>
<thead>
<tr>
<th>Clause Type</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embedded Clause</td>
<td>.87</td>
<td>86%</td>
<td>28</td>
<td>1%</td>
</tr>
<tr>
<td>Main</td>
<td>.49</td>
<td>26%</td>
<td>2062</td>
<td>99%</td>
</tr>
</tbody>
</table>

Table 13 Multivariate results for clause type

Although there are exceedingly few examples of the small clause, this factor group manifests the value of quantitative data. First, the challenges of studying syntactic data (cf. 3.1) reveal how truly infrequent syntactic phenomena can be. Regardless of its scarcity, the embedded construction does indeed show a very high probability favoring SDOM (.81). Furthermore the quantitative data of this study reveals that SDOM is essentially at chance (.49) in prototypical (i.e. non small-clause) constructions. By virtue of the extremes in the distribution of this factor (99% of the tokens are prototypical), the
rate of marking for the prototypical construction is exactly in line with the overall frequency of marking at 26%.

This factor group proves to be significant and the explanation for this can be found in the entailed syntax and semantics. Roegiest (2003) makes the observation that particular syntactic constructions may favor, if not motivate particular cases of SDOM. Specifically, Roegiest suggest that embedded clauses favor marking: “Mr Witt veía llorar a algunas mujeres con el rostro apenas visible” ‘Mr Witt saw some women weep with their faces hardly visible’ (pg.311). Because the object of the matrix verb is cast as the subject of the following embedded clause, there is an inherent juxtaposition of the role of both agent and patient. As the DO object of the matrix verb, it is clearly a patient. However, the same DO functions as the agent of the following embedded infinitival clause. Given this semantic duality, the DO is far more agentive than the prototypical DO because the structure entails that the DO have the role of agent. Taking into account the premise that marking occurs with more agentive or subject-like objects, the selection of this structure favoring marking is unsurprising.

The data has corroborated the assertion that the embedded clause favors marking, however this assumption is not supported when we observe the effect of animacy in these constructions.
<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>% a-marked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>96% (24/25)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>0% (0/1)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>0% (0/3)</td>
</tr>
</tbody>
</table>

Table 14  Marking according to embedded clause and DO animacy

Of the 28 embedded-clause tokens found in the data, 25 of them have Human DOs. In the three cases where we find non-human DOs, none are marked. The animacy of the DO correlates precisely with all cases of marking. This is sufficient evidence to acknowledge that although the FG is selected as significant by the multivariate analysis, any definitive claims regarding the structure and marking cannot be made. The embedded clause construction cannot be considered independent of the effects of the animacy of the DO based on the data obtained in this study. This is not to say that there is not something interesting or significant; the human tokens behave as expected being overwhelmingly marked in the data. The critical test is to see the variation in marking with a greater number of tokens of non-human DOs. The data for this dissertation do not provide those tokens in sufficient numbers to conclude if this structure does indeed affect marking across the animacy scale, and remains a topic for further study.

4.6. Ambiguity and SDOM

Also selected as significant in the multivariate analysis is the effect of possible ambiguity between the subject and the object.
In coding this factor group, it became immediately apparent that in the overwhelming number of transitive constructions no ambiguity arises between subject and object. The majority of these cases are resolved by morphological means (number and person agreement) or by general assumptions and knowledge about how the world works.

*Morphologically unambiguous:*

1a) *Vimos a los pescadores.* (HC Buenos Aires)
   ‘We saw A the fishermen.

*General knowledge:*

1b) *El abogado tiene que defender a... a su defendido.*
   (HC Buenos Aires)
   ‘The lawyer has to defend A…A his defendant.’

The end result was a handful of tokens where ambiguity did arise. Very much in line with what was observed regarding the factor group *clause type*, the large number of cases where no ambiguity arises show a probability that is essentially at chance (.49) while the ambiguous cases favored marking (.78).

The factor group *ambiguity* also shows interaction with DO animacy. 77% (60/78) of the ambiguous tokens have human DOs. The marking of human DOs where
ambiguity arises between subject and DO is not categorical; nevertheless human DOs are marked at a very high rate (56/60).

\[\text{Ambiguous:}\]

1e) \textit{La mujer pesca al hombre.} (HC Mexico City)

‘The woman catches A the man.’

In the case of inanimate DOs marking is much less frequent (4/12). This proportionally coincides with the marking observed with ambiguous contexts for metaphorical DOs (2/6). From the data it appears that when ambiguity arises, the overwhelming number of occurrences of SDOM correlates with the DO being human. Crucially we can readily observe the difference in rate of marking between ambiguous human DOs at 93% (56/60) and the overall rate of marking for human DOs at 72% (491/681). This provides quantitative support for the disambiguating hypothesis that is often cited in accounting for SDOM (Torrego-Salcedo 1999, Company Company 2002, Laca 2006): \( a \)-marking would be expected in the cases where an ambiguity arises between potential subjects and objects.

4.7. Referential Distance and SDOM

The next factor group selected as significant was total referential distance. This was the sum of the referential distance score, in clauses, to the nearest prior mention of the variably marked DO and the score of the nearest mention forward in the discourse with an upper bound of 20 clauses in either direction.
Table 16 Multivariate analysis of total referential distance

<table>
<thead>
<tr>
<th>TRD</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-20 clauses</td>
<td>.61</td>
<td>50%</td>
<td>161</td>
<td>15%</td>
</tr>
<tr>
<td>21-39 clauses</td>
<td>.60</td>
<td>36%</td>
<td>641</td>
<td>31%</td>
</tr>
<tr>
<td>40 clauses</td>
<td>.41</td>
<td>16%</td>
<td>1127</td>
<td>54%</td>
</tr>
</tbody>
</table>

Table 16 shows the final groupings of the multivariate analysis. Initially there were five measures, 1-10 clauses, 11-20 clauses, 21-25 clauses, 26-39 clauses and 40(+) clauses. The numbers were not arbitrarily determined. For the upper bounded limit of 20 prior and following clauses, a referent with a TRD value of 40 meant that in none of the prior or following discourse is there mention of the referent. Additionally, the TRD value of 21 was also critical because this number captures a frequent occurrence in the data: a referent being either chain-initial (new to the discourse and therefore lacking any prior mention) or chain-final (no further mention of the referent). In these cases, the maximum upper bound of 20 clauses is reached without the reoccurrence of the referent in one, but not both directions in the discourse (cf. 4.7)

These groupings are an attempt to capture the assumptions of Givón (1983) regarding topicality and the proximity of a particular referent. Those referents that have closer referential distance via anaphoric (and in my methodology, also cataphoric) referring expressions are more topical than referents with more distant or no previous or following reference. This quantifiable conceptualization of topicality was used to test the
assumption that a more topical DO will favor marking. The results here corroborate these assumptions.

Total Referential Distance was found to be significant in the multivariate analysis. These multivariate results were improved by collapsing the initial five distinctions into the three observed in Table 16. These groups result from collapsing the factors with few attested tokens: TRD value of 11-20 (N = 25) and TRD value of 26-39 (N = 53). As hypothesized, marking is favored in the case of prior or following mention and disfavored by no prior or following mention. What is surprising is the nearness in probabilities between marking that occurs with a TRD value of 1-20 (.61) and those with a TRD value of 21-39 (.60). Collapsing these factors did not improve the multivariate analysis, and there are important distinctions preserved by maintaining the three.

Table 16 also provides important insight into the discourse features of marking. A TRD value of 20 or less necessarily indicates that the referent has both prior and following reference and that the distance is relatively minimal. In these cases marking occurs 50% of the time. Meanwhile, the TRD value of 21-39 captures cases where the referent has reference in the discourse that also favors marking (36%), whereas those cases with no prior or following mention (TRD value of 40) disfavor marking (16%).

From these data we can form a generalization regarding referential distance as a relative

---

39 An additional insight taken from Table 16 is the fact that the vast majority of DO referents in the discourse are indeed being “introduced” i.e., they are new to the discourse. While this may not be unexpected, the follow up to the comportment of DO referents in discourse is that the majority of DOs are not topical in the sense that they do not persist in the discourse whatsoever. In my data 54% (1127/2090) of the tokens were one time occurrences.
means to quantify topicality and SDOM: DOs with reference in the discourse are more likely to be marked than those that are one-time occurrences.

This insight must again be compared with the interaction of DO animacy.

Fortunately, it is evident that these correlations between TRD and marking hold up under scrutiny.

<table>
<thead>
<tr>
<th>TRD</th>
<th>% a-marking of human DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-20 clauses</td>
<td>85% (153/179)</td>
</tr>
<tr>
<td>21-39 clauses</td>
<td>76% (193/255)</td>
</tr>
<tr>
<td>40 clauses</td>
<td>59% (145/247)</td>
</tr>
</tbody>
</table>

Table 17 Rate of marking according to human DO and total referential distance

When the results of Table 17 are compared with the overall rate of marking of human DOs at 72% (491/681) we can see the independence of TRD and DO animacy. Moreover it is evident that this factor group captures an important aspect of how SDOM interfaces with discourse referentiality that has not been previously discussed in the SDOM literature.

4.8. Specificity and SDOM

In line with what has been claimed regarding the marking of DOs, it is evident from the multivariate analysis that specificity of the object does favor marking. While it is selected as statistically significant, I had anticipated a stronger effect in the data following claims in the SDOM literature (Pensado 1995, Torrego-Salcedo 1999, von...

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>.58</td>
<td>32%</td>
<td>1178</td>
<td>56%</td>
</tr>
<tr>
<td>Non-specific</td>
<td>.40</td>
<td>22%</td>
<td>912</td>
<td>44%</td>
</tr>
</tbody>
</table>

Table 18 Multivariate analysis of specificity

While it is clear that there is an appreciable pattern of marking with specificity: specific DOs favor marking (.58) and non-specific DOs disfavor marking (.40), the actual probabilities are not particularly robust.

The effect of specificity can be better observed by evaluating how specificity intersects with object animacy.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Specific</th>
<th>Non-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>83% (312)</td>
<td>58% (179)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>54% (22)</td>
<td>16% (3)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>5% (38)</td>
<td>3% (15)</td>
</tr>
</tbody>
</table>

$\chi^2 = 7.317$  $p < .02577$  $df = 2$

Table 19 Rate of marking according to specificity and DO animacy

As in the other FGs, Table 19 illustrates that the effect of animacy in marking is unmistakable: human DOs strongly favor marking compared to inanimate DOs. Equally
unmistakable is the effect of specificity in the rate of marking human DOs. Human DOs are more likely to be marked than their human, non-specific counterparts. This marking of specific human DOs (83%) is also higher than the overall marking of human DOs (72%). Much like definiteness, specificity has an observed effect in increasing the rate of a-marking of non-human DOs, however no appreciable difference between specific and non-specific inanimates is found in the data.

4.9. Number and SDOM

The multivariate analysis of the data also selected Number as a significant FG. The results for Number are shown in Table 20.

<table>
<thead>
<tr>
<th>Number</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>.55</td>
<td>28%</td>
<td>1000</td>
<td>48%</td>
</tr>
<tr>
<td>Plural</td>
<td>.42</td>
<td>21%</td>
<td>698</td>
<td>33%</td>
</tr>
</tbody>
</table>

Table 20  Multivariate analysis for number

These results are interesting because they reflect an assumption expressed by the concept of *individuation* as it pertains to SDOM: objects singular in number (as opposed to plural) are considered more individuated and individuated objects are more likely to be a-marked. Though the actual probabilities are not particularly robust, the individuation assumption of Kliffer (1995) and others regarding number and SDOM finds support in the data.

125
Originally, I had coded for non-count nouns. However, it became apparent in the multivariate analyses that there were interactions with these tokens. The factor weights exhibited cross-over effects in the analyses, interacting with DO Animacy. This results from the fact that 84% (122/145) of non-count DOs were human. These cases were of primarily proper names or the high-frequency, human non-count DO *gente* ‘people.’

Comparing Number and DO provides further evidence for multiple factors contributing to *a*-marking. While I had been initially skeptical of any potential findings in coding for number, when number is cross-tabulated with DO Animacy the data reveal the following pattern in Table 21.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>78% (239)</td>
<td>63% (130)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>31% (9)</td>
<td>20% (2)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>5% (30)</td>
<td>3% (14)</td>
</tr>
</tbody>
</table>

Table 21  Rate of marking according to number and DO animacy

While the limited number of tokens for metaphoric and inanimate DOs is expected, the effect of number on SDOM with human DOs is apparent; singular human DOs are more frequently marked than plural human DOs. Particularly prominent is the decrease in *a*-marking with plural human DOs. At 63% marking this is nearly 10% lower than the

---

40 These interactions became apparent when I compared the probabilities with rate of *a*-marking. These so-called “crossover effects” present a contradiction when the probabilities for the factors do not reflect the actual rate of marking. This indicates that one or more factor groups overlap, i.e. factors from different factor groups are not independent from one another.
overall rate of marking for human DOs at 72%. The data provide quantitative evidence that support one of the mainstream claims regarding SDOM, namely that individuated objects are more likely to be marked.

4.10. Dialects and SDOM

The results of table 5 represent the multivariate analysis of all the data collected from the three corpora. However, I also coded the data for dialect to see if the dialects individually would behave differently with respect to SDOM. The multivariate analysis did indeed show a measurable distinction among the dialects of Buenos Aires, Madrid and Mexico City.

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico City</td>
<td>.58</td>
<td>34%</td>
<td>622</td>
<td>30%</td>
</tr>
<tr>
<td>Madrid</td>
<td>.52</td>
<td>26%</td>
<td>467</td>
<td>22%</td>
</tr>
<tr>
<td>Buenos Aires</td>
<td>.44</td>
<td>23%</td>
<td>1001</td>
<td>48%</td>
</tr>
</tbody>
</table>

Table 22: Multivariate analysis for dialect

These results are interesting for several reasons. First, they provide initial evidence for considering that the observed variation in SDOM reflects dialectal variation as well. Therefore, in considering what potential factors are motivating SDOM, it is crucial to consider that the particular factors involved may vary from dialect to dialect. Table 22 is also surprising when one considers the few (but common) assumptions regarding dialectal variation of SDOM. While primarily anecdotal, Buenos Aires is cited as a place
where exceptional *a*-marking occurs (Dumitrescu 1997, von Heusinger and Kaiser 2003) while Madrid, a peninsular dialect, is considered more normative with *a*-marking occurring with specific and human DOs and non-marking with non-specific humans and inanimate DOs (von Heusinger and Kaiser 2003). This is found not to be the case in the data. Especially surprising is that fact that in the Buenos Aires data *a*-marking is actually disfavored overall (.44), contradicting the anecdotal claims regarding variation in *a*-marking among porteños (cf. Dumitrescu 1997, von Heusinger 2003).

Upon analyzing the data a surprising result is found with respect to dialectal distinctions: Buenos Aires marks human DOs less frequently than the other two dialects as shown in Table 23.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Buenos Aires</th>
<th>Madrid</th>
<th>Mexico City</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>65% (195)</td>
<td>77% (108)</td>
<td>78% (188)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>39% (11)</td>
<td>38% (5)</td>
<td>47% (9)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>4% (28)</td>
<td>3% (9)</td>
<td>4% (16)</td>
</tr>
</tbody>
</table>

Table 23  Rate of marking according to dialect and DO animacy

In this regard, Madrid and Mexico City are far more similar in their rate of marking human DOs. Additionally, we find that all three dialects essentially mark inanimates at the same rate. Finally, due to the low number of tokens, I will not comment on the rates for metaphoric DOs across dialects. This initial observation of divergence in SDOM behavior across dialects, supported by the multivariate analysis and other potential areas of divergence will be the focus of the following chapter.
4.11. Conclusion

This chapter has presented the combined data obtained from the three corpora. The large number of tokens led to a very complex description of the data. The resulting multivariate analysis confirms the central hypothesis of this dissertation: there exists a multiplicity of factors that contribute to the observed variation of $a$-marking in Spanish. The multivariate analysis shows the relative effect of multiple factors in marking. Importantly through careful observation of the results from the multivariate analysis, interactions among particular factors, namely total topic persistence, relative animacy and DO animacy were identified. This rigorous analysis of the data has provided insight regarding the general trends in $a$-marking. The underlying and most robust factor that accounts for marking is the animacy (humanness) of the object. However, the data also show this factor does not account for all cases of marking, confirming that other factors contribute to the variability of $a$-marking and that these factors operate across NP (definiteness, specificity), clausal (clause type, disambiguation) and discursive domains (total referential distance).
CHAPTER 5: DIALECTAL ANALYSES

5. Introduction

This chapter examines each dialect individually. It follows the same format as the previous chapter with the same organization and presentation of tables. Whereas in Chapter 4, the macro analysis provided over-arching trends for SDOM, this chapter will delve into the micro-analyses and explore more fully the particular distributions of the data within each dialect. Section 5.1 consists of the data analysis of the HC Buenos Aires corpus. Section 5.2 will analyze the HC Madrid corpus. Section 5.3 will focus on the HC Mexico City corpus results. Section 5.4 concludes the chapter.

5.1. SDOM in Buenos Aires

The Buenos Aires corpus provides an important source for addressing claims and assumptions regarding SDOM. It is the largest of the three corpora and provided essentially half of the tokens for this dissertation. The multivariate results are presented in Table 24 below.
<table>
<thead>
<tr>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DO ANIMACY</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>.89</td>
<td>65%</td>
<td>301</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>.69</td>
<td>39%</td>
<td>28</td>
</tr>
<tr>
<td>Inanimate</td>
<td>.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Relative Animacy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object greater</td>
<td>.84</td>
<td>76%</td>
<td>50</td>
</tr>
<tr>
<td>Equal animacy</td>
<td>.68</td>
<td>55%</td>
<td>266</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>.67</td>
<td>20%</td>
<td>177</td>
</tr>
<tr>
<td>Subject greater</td>
<td>.31</td>
<td>3%</td>
<td>508</td>
</tr>
<tr>
<td><strong>Definiteness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite</td>
<td>.62</td>
<td>28%</td>
<td>694</td>
</tr>
<tr>
<td>Indefinite</td>
<td>.29</td>
<td>14%</td>
<td>177</td>
</tr>
<tr>
<td>No determiner</td>
<td>.21</td>
<td>10%</td>
<td>130</td>
</tr>
<tr>
<td><strong>Clause Type</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small clause</td>
<td>.89</td>
<td>80%</td>
<td>10</td>
</tr>
<tr>
<td>Normal</td>
<td>.50 (.495)</td>
<td>23%</td>
<td>991</td>
</tr>
<tr>
<td><strong>Ambiguity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>.75</td>
<td>78%</td>
<td>37</td>
</tr>
<tr>
<td>No</td>
<td>.49</td>
<td>21%</td>
<td>964</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.60</td>
<td>30%</td>
<td>573</td>
</tr>
<tr>
<td>Non-specific</td>
<td>.36</td>
<td>15%</td>
<td>428</td>
</tr>
<tr>
<td><strong>Total Referential</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-20 Clauses</td>
<td>.61</td>
<td>42%</td>
<td>134</td>
</tr>
<tr>
<td>21-39 Clauses</td>
<td>.60</td>
<td>32%</td>
<td>307</td>
</tr>
<tr>
<td>40 Clauses</td>
<td>.42</td>
<td>14%</td>
<td>560</td>
</tr>
</tbody>
</table>

Total N =1001, Input .091, (23% a-marked). Log Likelihood = -265.889, p = .034

Factor Groups not selected: DO placement, Discourse Status, Subject Animacy, Number, RD, FRD, TP, BTP.

Table 24  Factors contributing to SDOM in Buenos Aires
Considering that the Buenos Aires data (N = 1001) contributes nearly half the tokens of the data, it is unsurprising to find that the factors selected for SDOM are similar to the factors selected as significant for all dialects (cf. Table 5). However, compelling distinctions can be observed which lead to a more detailed understanding of the distinct conditions motivating SDOM in the HC Buenos Aires data.

5.1.1. DO Animacy

The effect of DO animacy is attested in the case of the Buenos Aires data. Animacy of the direct object has the greatest relative strength among all factor groups. Although this result is expected, as noted in 4.1.9., the rate of marking for human DOs (65%) is less than the overall marking of human DOs across all dialects (73%) and less than the rate of marking human DOs in both the Mexico City (78%) and Madrid (77%) dialects. This is reflected in a smaller range (60) as well as the actual probability of marking human DOs (.88) being slightly less. Meanwhile the constraint ranking (the ranking of the factors within a given factor group based upon their probabilities) is exactly as expected; Human DOs most favor marking followed by Metaphoric DOs with a probability of .69, whereas Inanimate DOs strongly disfavor marking with a probability of .29. Despite the slightly reduced effect of DO Animacy in the Buenos Aires corpus, the relevance of DO animacy should not be underestimated. However, this will have critical statistical relevance when we test for independence with several other factor groups, including the factor group of next highest strength, Relative Animacy.
5.1.2. Relative Animacy

Second only to DO animacy, relative animacy has a strong effect in determining SDOM with a range of 54. This is the first departure from what was observed in the overall data of Table 5.: Relative Animacy has a stronger effect in the multivariate analysis in the Buenos Aires data than in the overall data set. Furthermore, the independence of this relative animacy from DO animacy is far more transparent in the Buenos Aires data.

<table>
<thead>
<tr>
<th>% a-marking</th>
<th>DOs animacy greater</th>
<th>Human DOs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>76% (34)</td>
<td>65% (195)</td>
</tr>
</tbody>
</table>

Table 25  Comparison of rate of marking between DOs with greater animacy than subject and overall rate of marking for human DOs in the Buenos Aires data

Isolating the Buenos Aires data reveals the relevance and effect of Relative Animacy. As noted, the overall rate of marking for human DOs is unexpectedly lower at only 65% (195/301). However the rate of marking for human DOs that are also higher in animacy than their subjects is more than 10% higher at 76% (38/50). The Buenos Aires corpus shows an independent effect of relative animacy in favoring a-marking. While this effect is not as pronounced as previously found (Tippets forthcoming), the data still corroborate the validity of relative animacy in accounting for variation in SDOM in HC Buenos Aires.
5.1.3. Definiteness

Definiteness is also selected as a significant factor group, revealing the same constraint hierarchy shown overall in the data: definite expressions favor marking (.62) while indefinite expressions (.29) and Bare DOs (.21) disfavor marking. Not only do definite DOs appear most frequently in the data at 69% (694/1001) but definite DOs also have the highest rate of marking at 28% (196/498). Following the pattern observed in the combined data of the corpora, indefinite and Bare DOs are infrequently marked relative to definite DOs, the former marked at 14% (25/177) and the latter at 10% (13/130).

The effect of Animacy can be observed within the factor group of Definiteness; the majority of marking that occurs coincides with greater animacy i.e., the humanness of the DO as seen below.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Definite</th>
<th>Indefinite</th>
<th>Bare</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>79% (162)</td>
<td>40% (21)</td>
<td>27% (12)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>39% (9)</td>
<td>50% (2)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>5% (25)</td>
<td>2% (2)</td>
<td>1% (1)</td>
</tr>
</tbody>
</table>

Table 26 Rate of marking according to definiteness and DO animacy

Table 26 shows that the overwhelming number of marked DOs in the data happen to also be human 83% (162/195). Among these marked human DOs, it is evident that definiteness has an independent effect: the rate of marking of definite human DOs (79%) is much greater than the overall rate of marking for human DOs in the Buenos Aries data (65%). Likewise, human indefinite DOs are less frequently marked (40%) than the
overall rate, as are the sparse bare human DOs (27%). The poor distribution of data in the non-human cells precludes any further generalizations regarding definiteness effects on marking. However the data from Table 26 clearly show the effect of definiteness within the class of human DOs.

5.1.4. Clause Type.

Despite the size of the Buenos Aires corpus, only 10 embedded-clause tokens were obtained. Although minimal, the embedded clause proves to be measurably significant in the multivariate analysis with a rate a-marking of 80% (8/10) and a very high probability of .89. Again, however, this result is deceiving when these embedded-clauses are cross-tabulated with DO animacy. Of the 10 DO tokens culled from the data, nine are human with the one non-human DO remaining unmarked. Without sufficient tokens to observe if there is variation based upon the clause type that that does not also coincide with DO Animacy, this factor group cannot substantiate the independence of clause type and DO animacy within the present data (c.f. 4.1.4.).

5.1.5. Ambiguity

The Buenos Aires data show that contexts with ambiguity between subject and DO favor a-marking. Of the 37 cases where ambiguity is found in the corpus, 78% (29) are marked. The multivariate analysis corroborates this distribution with a factor weight of .76 for ambiguous contexts. In cases where there is no ambiguity, the factor weight indicates that marking is at chance (.49). Cross-tabulations with DO Animacy reveal the
Ambiguity does interact when the DO is human. 73% (27/37) of the ambiguous tokens are human. The rate of marking for these ambiguous human DOs is higher at 96% (26/27) than the overall marking of human DOs in the Buenos Aires data at 65% (195/301). This would indicate that there is independence between the DO Animacy and ambiguity, in spite of the inherent preference of SDOM to mark human DOs.

5.1.6. Specificity

The HC Buenos Aires corpus confirms the relevance of specificity in accounting for SDOM. The relative strength of specificity is less than the previous factors but its relevance becomes apparent when cross-tabulated with DO Animacy.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Specific</th>
<th>Non-specific</th>
<th>Overall rate of α-marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>80% (140)</td>
<td>43% (55)</td>
<td>65% (195)</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>52% (9)</td>
<td>18% (2)</td>
<td>39% (11)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>5% (20)</td>
<td>3% (8)</td>
<td>4% (28)</td>
</tr>
</tbody>
</table>

Table 27 Rate of marking according to specificity and DO animacy

As expected, human DOs favor marking. However the rate of marking for specific human DOs is noticeably higher (80%), while non-specific human DOs are marked at a lower rate (43%). This pattern is observed across all three measures of DO animacy, with the caveat being the greatly reduced number of tokens, limiting broader generalizations. What emerges from the quantitative data is the insight that specific human DOs are far more likely to be marked. This has been presumed by many to be the case (von
Heusinger and Kaiser 2003, Aissen 2003) and the HC Buenos Aires data provide the evidence to support this argument.

The effect of specificity provides further and unmistakable support for the assumption that multiple factors must be taken into account in understanding the variable nature of a-marking. This is manifest in the theoretical implications of the individuation hypothesis as it relates to SDOM. Specificity is purported to be relevant for marking in general as cited above. It is also one of the core features of individuated DOs. The data reveal the relationship of highly individuated objects and marking by showing that specificity is indeed a contributing factor among several that explain SDOM.

Having considered definiteness and specificity independently, let us now look at how they intersect.

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Definite</th>
<th>Indefinite</th>
<th>Bare DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>31% (149)</td>
<td>20% (14)</td>
<td>22% (6)</td>
</tr>
<tr>
<td>Non-specific</td>
<td>22% (47)</td>
<td>10% (11)</td>
<td>7% (7)</td>
</tr>
<tr>
<td>Total</td>
<td>28% (196)</td>
<td>14% (25)</td>
<td>10% (13)</td>
</tr>
</tbody>
</table>

Table 28 Rate of marking according to definiteness and specificity

---

41 Hopper and Thompson (1980) and Kliffer (1984, 1995), who bases his individuation parameters on Hopper and Thompson, use the term referential to mean specific, although with a tacit assumption that actual referring expressions i.e., definiteness correlates with specificity. In her discussion of individuation, Laca (1999) makes explicit the relationship of specificity to referentiality, identifying specific DOs as individuated and non-specific DOs as non-individuated.
In the Buenos Aires data it can be observed that the largest number of marked DOs and the highest rate of marking (31%) are those DOs that are specific and definite. From Table 29 it is clear that both specificity and definiteness also favor marking independently. Thus specific indefinites and non-specific defines also mark at a higher rate despite smaller numbers. When the DO is both indefinite and non-specific, we see a very low rate of marking (10%) accompanying very few tokens of this type (N = 11). The fact that specific definites are more frequently marked in the Buenos Aires data supports the claims regarding the coalescing of individuating factors in accounting for SDOM.

5.1.7. Total Referential Distance

The final factor group selected for significance in the HC Buenos Aires corpus is Total Referential Distance (range = 19). The Buenos Aires data follow the pattern observed in the overall data (cf. Table 5). Marking is more likely with DOs that have a referent in the prior or following discourse as indicated with a probability of .60. This probability is slightly improved (.61) when the DO has a referent in both the prior and following discourse. The condition of having no referring expression in the prior or following discourse disfavors marking of the DO (.42).

42 As discussed in 3.1, the methodology for measuring TRD was to sum the RD and FRD with an upper limit of 20 clauses. Mathematically, any value for TRD that falls between 2 and 20 entails that the DO has a preceding and following referring expression in the discourse.
The cross-tabulation of DO Animacy and TRD reveals that again, animacy intersects with the various factor groups. Table 30 does show an independent effect where the rate of marking is higher than the overall rate of marking.

<table>
<thead>
<tr>
<th>TRD</th>
<th>% <em>a</em>-marked human DOs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-20 clauses</td>
<td>84% (51)</td>
</tr>
<tr>
<td>21-39 clauses</td>
<td>70% (81)</td>
</tr>
<tr>
<td>40 clauses</td>
<td>51% (63)</td>
</tr>
<tr>
<td>Overall Rate of <em>a</em>-marking</td>
<td>65% (195)</td>
</tr>
</tbody>
</table>

Table 29  Rate of marking for human DOs according to total referential distance

As observed in the overall data in Table 30, marking occurs with those DOs that are referred to in the prior or following discourse, especially when it is so referred to in both cases. These metrics are used to determine the relative topicality of a marked or unmarked DO in the discourse. The data show that preference for marking DOs whose reference in the discourse is proximal while non-referenced DOs disfavor marking.

5.1.8. Interactions with Total Topic Persistence

In the numerous multivariate analyses, Total Topic Persistence also was selected as significant. However, comparing the constraint hierarchy of TTP with the actual distribution in the Buenos Aires data revealed crossovers. Further investigation led to the immediate discovery that the factor group TTP was interacting with DO Animacy, causing the factor weights of TTP to fluctuate greatly in the regression analysis.
Introducing other factor groups into the regression analysis further exacerbated the contradicting TTP factor weights. This also led to the discovery that TTP was interacting with TRD. The multivariate analysis for the HC Buenos Aires data revealed a crossover effect in the TRD factor group, flipping the relative strength of factor weights between the factors 1-20 clauses and 21-39 clauses. Since this contradicted the rate of marking as well as the base-line factor weights it was quickly discovered that the crossover occurred when TTP was added to the regression analysis. In light of the fact that TTP interacts with several factors, but particularly because it interacts with DO Animacy, whose primacy in SDOM is unquestioned, TTP was eliminated from the multivariate analysis.

Despite the havoc wrought in the multivariate analysis, the data present compelling insight into the potential topicality of marked objects. The base-line factor weights (probabilities) for TTP in the regression analysis of HC Buenos Aires are shown below, including TTP distribution in the corpus.

<table>
<thead>
<tr>
<th>TTP</th>
<th>Probabilities</th>
<th>% a-marked</th>
<th>N</th>
<th>% of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mentions</td>
<td>.39</td>
<td>15%</td>
<td>572</td>
<td>57</td>
</tr>
<tr>
<td>1-5 mentions</td>
<td>.64</td>
<td>30%</td>
<td>365</td>
<td>37%</td>
</tr>
<tr>
<td>6-20 mentions</td>
<td>.82</td>
<td>56%</td>
<td>64</td>
<td>6%</td>
</tr>
</tbody>
</table>

Table 30 Multivariate analysis of total topic persistence
In accordance with the assumption that SDOM is used with more frequently-occurring, hence “topical,” DOs, the distribution of marked objects corroborates the increase in rate of marking with an increase in frequency of occurrence of a referent in the discourse.\footnote{My claims regarding topicality only extend what I call a \textit{relative} conceptualization of topicality in so much as the more frequently a referent appears in the discourse, we can say that it is more topical relative to other referents which are less frequent in the discourse.}

To further test the relevance of TTP in accounting for \textit{a}-marking in Spanish, cross-tabulations are performed with DO Animacy.

<table>
<thead>
<tr>
<th>TTP</th>
<th>Human</th>
<th>Metaphoric</th>
<th>Inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Mentions</td>
<td>54% (71)</td>
<td>33% (4)</td>
<td>3% (13)</td>
</tr>
<tr>
<td>1-5 Mentions</td>
<td>70% (92)</td>
<td>42% (5)</td>
<td>6% (13)</td>
</tr>
<tr>
<td>6-20 Mentions</td>
<td>84% (32)</td>
<td>50% (2)</td>
<td>9% (2)</td>
</tr>
<tr>
<td>Total</td>
<td>65% (195/301)</td>
<td>39% (11/28)</td>
<td>4% (28/672)</td>
</tr>
</tbody>
</table>

Table 31 Rate of marking according to DO animacy and Total Topic Persistence

Despite the preference for marking human DOs, the relative concept of topicality in accounting for SDOM finds support in the quantitative data in the Buenos Aires corpus. Those human DOs that have a high frequency of occurrence in the data are marked at much higher rate (84\%) than the overall rate of marking of human DOs (65\%). Additionally, those human DOs that do not persist in the discourse, are marked at a lower rate (54\%). These results present provocative evidence suggesting that the frequency of occurrence of the referent in the discourse correlates with the probability of \textit{a}-marking.
5.1.9. Summary of SDOM in Buenos Aires

What has been observed in the Buenos Aires data is that the factor groups selected as significant and the ordering of these factor groups according to their relative strength of effect are very similar to what the overall data show (cf. Table 5, Table 24). This is primarily a result of 1) the size of the Buenos Aires corpus, providing half of all tokens and 2) the likelihood that the assumed underlying factors of SDOM are indeed explanatory for much of the variation observed. Despite these conditions, there are several crucial distinctions manifest in the Buenos Aires corpus. Foremost among them is the lower rate of marking of human DOs. Also of interest is the prominence and independence of Relative Animacy in affecting a-marking. The Buenos Aires data also provide consistent evidence of the independent effects of several factors including Definiteness and TRD. Finally, although not quantifiable in the multivariate analysis due to interactions with just about everything else in the data, Total Topic Persistence, when evaluated independently, appears to capture potentially the intuition regarding the relative topicality of a referent and its propensity for a-marking.

5.2. Multivariate analysis of HC Madrid

The multivariate analysis of the HC Madrid corpus provides additional evidence to support the hypothesis that multiple factors are necessary to account for SDOM. Although smaller than the other two corpora, multiple factors are also selected as
significant for $a$-marking in the Madrid data. However, more collapsing and analysis was required, yielding the following multivariate analysis

<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
<th>% $a$-marked</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO Animacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>.97</td>
<td>77%</td>
<td>140</td>
<td>30%</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>.82</td>
<td>39%</td>
<td>13</td>
<td>3%</td>
</tr>
<tr>
<td>Inanimate</td>
<td>.17</td>
<td>3%</td>
<td>299</td>
<td>67%</td>
</tr>
<tr>
<td><strong>Range: 80</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>.93</td>
<td>64%</td>
<td>14</td>
<td>3%</td>
</tr>
<tr>
<td>No</td>
<td>.48</td>
<td>25%</td>
<td>453</td>
<td>97%</td>
</tr>
<tr>
<td><strong>Range: 45</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definiteness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite</td>
<td>.57</td>
<td>31%</td>
<td>313</td>
<td>67%</td>
</tr>
<tr>
<td>Indefinite</td>
<td>.36</td>
<td>16%</td>
<td>154</td>
<td>33%</td>
</tr>
<tr>
<td><strong>Range: 21</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.59</td>
<td>31%</td>
<td>266</td>
<td>57%</td>
</tr>
<tr>
<td>Non-specific</td>
<td>.39</td>
<td>19%</td>
<td>201</td>
<td>43%</td>
</tr>
<tr>
<td><strong>Range: 20</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total $N = 467$, Input: .104, (26% $a$-marked). Log Likelihood = -114.841 $p = .043$

Factor Groups not selected: DO placement, Discourse Status, Subject Animacy, RD, FRD, TRD, TP, BTP

Table 32 Factors contributing to SDOM in Madrid
The multivariate analysis selected these four groups to be significant in favoring SDOM. The overall rate of marking in the Madrid data is 26% (122/345). This is in line with what has been observed in the overall rate of marking for all dialects together at 27% (569/290). The patterning of factor groups for the Madrid data in terms of relative strength is similar to what has been seen in the Buenos Aires multivariate analysis and the overall analysis. Additionally, the results show there are measurable and observable distinctions in the functioning of SDOM.

5.2.1. DO Animacy

It continues to be evident that DO animacy critically accounts for much of the variation in the data. The strength of the factor group itself is noteworthy with a robust range of 80 resulting from the considerable difference between Human (.97) and Inanimate (.17) factors. Metaphorical uses also favor marking with a weight of .81. The hierarchy of constraint rankings is as expected from the previous multivariate analyses. Unlike the Buenos Aires data, the rate of marking of human DOs is more in line with the overall rate of marking observed in Table 5 as well as the HC Mexico City corpus to be seen below. The effects of DO Animacy will again be shown to pervade the other selected factor groups.
5.2.2. Ambiguity

The disambiguating function of the SDOM hypothesis finds support in the multivariate analysis of the Madrid data. As previously mentioned, actual ambiguity between subject and object continues to be shown as an infrequent occurrence with 14 total cases of ambiguity in HC Madrid representing only 3% of the total data. Of these 14 cases of ambiguity between subject and object, 64% (9) are marked. The small numbers are sufficient for the multivariate analysis to select this factor as significant, ambiguous contexts favoring marking with a probability of .89 and non-ambiguous contexts show an extremely minimal effect of disfavoring marking (.48).

Given the propensity of DO Animacy to be positively correlated to marking, Ambiguity and DO Animacy are cross-tabulated. The results of this cross-tabulation are shown below as the rate of marking.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Ambiguous</th>
<th>Non-ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>88% (7)</td>
<td>77% (101)</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>[ ]</td>
<td>38% (5)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>33% (2)</td>
<td>2% (7)</td>
</tr>
</tbody>
</table>

Table 33  Rate of marking according to ambiguity and DO animacy

Looking at the weak distribution of the data, the validity of the claim lies on a small handful of ambiguous tokens. As discussed in 4.6, the potential for ambiguity between subject and DO to arise is greatly mitigated by primarily morphological means, although syntactic constructions and general knowledge of the world disambiguate most other
contexts. Despite the rate of marking for animate DOs being higher at 88% (compared to the overall marking in the Mexico City data of 77%), the limited number of ambiguous contexts found in the Madrid data compel me to be somewhat reserved in arguing the relevance of ambiguity as an important contributing factor in the analysis.

5.2.3. Definiteness

The effect of definiteness is the next factor selected as significant in the multivariate analysis of Madrid. In order to successfully use the Goldvarb program to evaluate definiteness, it was necessary to collapse the Bare DO tokens because they were all unmarked (N=16). With this lack of observed variation in the data, Bare DO tokens were collapsed with indefinites since they obviously disfavor marking (an assumption supported by the observed rate of marking in the other dialects). The data from the Madrid corpus show that definite DOs favor marking (.57) while indefinite forms do not (.36). Considering that no Bare DOs are marked in the Madrid corpus, we might also add that these strongly disfavor marking, despite not being testable via the multivariate analysis. An additional insight garnered from the table is that the frequency of definite DOs to the indefinite DOs is exactly 2:1 (331/154), following the general tendency for definite DOs to be far more frequent in the data.

The results of cross-tabulating definiteness with DO animacy can be seen below in Table 34.
In the case of definite human DOs, the rate of marking exceeds the overall marking of human DOs (77%). Furthermore, the effect of indefinite DOs with human reference is even more pronounced with a decreased rate of marking at 58%. The data corroborate the multivariate analysis showing that definiteness is a feature that captures meaningful variation in $a$-marking.

5.2.4. Specificity

Following definiteness, specificity is selected as significant in the multivariate analysis of the Madrid data. More than half the DOs are specific, favoring marking with a probability of .59. The non-specific cases disfavor marking with a probability of .39. The frequency of marking supports these probabilities with 31% of specific DOs being marked while only 19% of non-specific DOs being $a$-marked.

To test for the independence of specificity in the Madrid data, Specificity and DO Animacy were cross-tabulated as shown below.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Definite</th>
<th>Indefinite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>84% (87)</td>
<td>58% (21)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>42% (5)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>3% (6)</td>
<td>3% (3)</td>
</tr>
</tbody>
</table>

Table 34 Rate of marking according to definiteness and DO animacy
The data continue to show the perpetual preference for marking DOs that are human. Like the conditions of definiteness in the Madrid data, there is an important difference between the overall rate of marking human DOs and the rate of marking specific human DOs. The Madrid data corroborate those claims that SDOM is sensitive to specificity.

To further understand the relevance of specificity and definiteness, cross-tabulations are undertaken. The results of comparing them can be seen in Table 36.

<table>
<thead>
<tr>
<th>Definiteness</th>
<th>Specific</th>
<th>Non-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite</td>
<td>33% (74)</td>
<td>27% (24)</td>
</tr>
<tr>
<td>Indefinite</td>
<td>21% (9)</td>
<td>14% (15)</td>
</tr>
</tbody>
</table>

Table 36 Rate of marking according to specificity and definiteness

The majority of marked DOs are both definite and specific (N =74). While the overall rate of marking is not particularly high at 33%, it is consistently higher than any other combination of definiteness and specificity. Further supporting this claim is the lowest rate of marking with indefinite, non-specifics at (15%). From these distributions it can be observed that definiteness and specificity have an effect on marking: specific and/or
definite DOs are more frequently marked than their non-specific and/or indefinite counterparts.

5.2.5. Interactions and potentially significant factor groups in Madrid.

The relatively limited number of tokens obtained in the Madrid data led to particular cases of poorly distributed data, namely the fact that of the marked tokens extracted from the data, only 11% (14/122) are non-human DOs. These disparities in the data prohibited successfully incorporating particular factor groups into a multivariate analysis. These factor groups were *Relative Animacy*, *Total Topic Persistence* and *Number*. The underlying reason for these omissions is their interaction with DO Animacy.

Indisputably maintaining the strongest effect on the SDOM variation observed in the data, DO Animacy has consistently been shown to have relevance in all of the other factor groups. With the aforementioned factor groups, the fundamental conflict with the animacy of the DO compelled me to remove them from the data. The fluctuation of factor weights in building regression analyses resulting from the interaction with DO Animacy precluded an accurate assessment of the true effect of these factor groups on SDOM variation within the multivariate analysis. I have presented these factor groups in Table 37 to show the tantalizing potential relevance of these factors in accounting for SDOM. Note however that the probabilities listed are the base-line probabilities for each factor group: they are not the result of a multivariate analysis because they are not
relative to one another nor are they relative to the factor groups selected as significant in the best multivariate analysis of the Madrid data.\textsuperscript{44}

<table>
<thead>
<tr>
<th>Relative Animacy</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object greater</td>
<td>.95 (.948)</td>
<td>74%</td>
<td>23</td>
<td>5%</td>
</tr>
<tr>
<td>Equal animacy</td>
<td>.95 (.946)</td>
<td>73%</td>
<td>108</td>
<td>23%</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>.77</td>
<td>27%</td>
<td>67</td>
<td>14%</td>
</tr>
<tr>
<td>Subject greater</td>
<td>.33</td>
<td>3%</td>
<td>269</td>
<td>58%</td>
</tr>
</tbody>
</table>

Range: 62

<table>
<thead>
<tr>
<th>Total Topic Persistence</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5 mentions</td>
<td>.67</td>
<td>34%</td>
<td>183</td>
<td>39%</td>
</tr>
<tr>
<td>6-20 mentions</td>
<td>.47</td>
<td>56%</td>
<td>25</td>
<td>5%</td>
</tr>
<tr>
<td>0 mentions</td>
<td>.38</td>
<td>17%</td>
<td>259</td>
<td>56%</td>
</tr>
</tbody>
</table>

Range 29

<table>
<thead>
<tr>
<th>Number</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>.61</td>
<td>28%</td>
<td>215</td>
<td>46%</td>
</tr>
<tr>
<td>Non-count</td>
<td>.53</td>
<td>31%</td>
<td>85</td>
<td>18%</td>
</tr>
<tr>
<td>Plural</td>
<td>.34</td>
<td>20%</td>
<td>167</td>
<td>36%</td>
</tr>
</tbody>
</table>

Range: 27

Table 37 Groups selected as significant but eliminated due to interactions with DO animacy

\textsuperscript{44} The base-line probabilities are the factor weights of the factors of a single factor group, considered alone. This is essentially a singlevariate analysis, where it is assumed that a single factor group will account for the observed variation. While this does not produce a very good model of the data, i.e. the “fit” of the model to the data is poor; the baseline weights are useful guide for how a particular factor group affects the dependent variable.
Relative Animacy could not be disentangled from DO Animacy in the Madrid data. Looking at the base-line probabilities and the rate of \(a\)-marking for the factor group, Relative Animacy does appear to be capturing an important correlation: when the object is equal or greater in animacy than the subject, marking is greatly favored whereas objects paired with a subject greater in animacy will strongly disfavor \(a\)-marking. This pattern falls in line with Comrie’s assumptions as outlined earlier and what I had observed in previous studies (Tippets forthcoming). However the distribution of the data (better said, the lack of distribution in the data) leads to severe interactions with DO Animacy making it impossible to show independence from DO Animacy. Specifically, at 78%, the rate of marking for objects equal or greater in animacy than the subject is virtually the same as the overall rate of marking for human DOs (77%). The significant effect that the regression analysis is detecting is evidently the humanness of the DOs. 84% (91/108) of the cases where the object is equal or greater than the subject are also cases where the DO is human. Without additional tokens, particularly in the underrepresented categories of the animacy scale (animate non-human, metaphoric, etc.) I cannot make any definitive claims regarding relative animacy as it affects marking in Madrid.

Interactions with Total Topic Persistence

As observed in the Buenos Aires data, the Madrid data show the same interaction between DO Animacy and TTP. The base-line weights and the rate of marking reflect
what is expected in terms of marking indicating relative topicality. DOs that do not persist in the discourse disfavor \(a\)-marking (.39) whereas the greater frequency with which a referent is mentioned, the more likely they DO will be marked with a probability of .60 for 1-5 mentions and a probability of .82 for 6-20 mentions. This evidence supports the hypothesis that topical referents are more likely to be \(a\)-marked DOs, that is to say, \(a\)-marking is likely to occur with those DOs whose referent occurs with high frequency in the limited confines of the immediately preceding and following discourse.

Nevertheless, these initial quantitative data belie the pervasive effect of DO animacy. 100% (14/14) of the marked tokens with 6-20 mentions are human while all of the unmarked tokens are inanimate. The tokens for 1-5 mentions, while not so vastly skewed, cannot be necessarily considered an improvement; 86% (54/63) of these tokens are also human. Given these conditions in the Madrid data, it is not possible to show independence between TTP and DO Animacy, with the result that TTP is precluded from the multivariate analysis.

**Interactions with Number**

In the Madrid data it is evident that interactions occur between Number and DO Animacy. This again is a result of the poorly distributed data, only 11% (14/122) of the marked DOs are non-human i.e., metaphoric (5) or inanimate (9). Further complicating this is the factor *non-count*. 81% (22/27) of the marked non-count tokens were human DOs. These non-count nouns are essentially of two varieties, personal names (e.g. *Roberto*) or the non-count animate *gente* ‘people.’ Despite excluding the non-count
factor from additional multivariate analyses, such skewing of the data make it impossible to overcome the interactions between number and DO Animacy.

5.2.6. Summary of Madrid

The multivariate analysis of data from the HC Madrid corpus provides a subtle shift from what is observed in the Buenos Aires data. In accounting for what motivates marking in the Madrid data, similar factor groups become apparent, namely DO Animacy, Definiteness, Ambiguity and Specificity. The precise strength and relative ordering is not identical showing dialect-specific aspects with regard to SDOM behavior. DO Animacy has a much stronger effect in a-marking than observed in the Buenos Aires data and occurs with higher frequency.

The Madrid data also introduce particular challenges to multivariate analysis. The relatively small number of tokens culled from the corpus leads to poor distributions in the data. This has the particular result of magnifying the effect of DO Animacy in several potentially significant factor groups including Relative Animacy, Total Topic Persistence and Number.

5.3. SDOM in HC Mexico City

The multivariate analysis for the Mexico City data suggests a straightforward explanation of the motivating factors. Like Madrid, the Mexico City data show a higher rate of overall a-marking than Buenos Aires at 33%. But as seen in the table below, there are very few factors required to provide a concise description of the SDOM data.
<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DO Animacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>.95</td>
<td>78%</td>
<td>240</td>
<td>39%</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>.66</td>
<td>47%</td>
<td>19</td>
<td>3%</td>
</tr>
<tr>
<td>Inanimate</td>
<td>.12</td>
<td>4%</td>
<td>363</td>
<td>58%</td>
</tr>
<tr>
<td><em>Range: 83</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Definiteness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite</td>
<td>.67</td>
<td>41%</td>
<td>426</td>
<td>69%</td>
</tr>
<tr>
<td>Indefinite</td>
<td>.28</td>
<td>25%</td>
<td>138</td>
<td>22%</td>
</tr>
<tr>
<td>No determiner</td>
<td>.06</td>
<td>5%</td>
<td>58</td>
<td>9%</td>
</tr>
<tr>
<td><em>Range: 61</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singular</td>
<td>.61</td>
<td>38%</td>
<td>297</td>
<td>48%</td>
</tr>
<tr>
<td>Plural</td>
<td>.35</td>
<td>28%</td>
<td>220</td>
<td>35%</td>
</tr>
<tr>
<td><em>Range: 26</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total N = 622, Input: .183, (34% a-marked). Log Likelihood = -173.064  p =

Factor Groups not selected: Ambiguity, Specificity, Subject Animacy, Relative Animacy, Clause Structure, DO placement, RD, FRD, TRD, TP, BTP, TTP.

Table 38 Factors contributing to SDOM in Mexico City

5.3.1 DO Animacy

DO Animacy is the predominant factor group favoring *a-marking* in the Mexico City data. Observing the factor weights for human (.95) and inanimate (.12), we arrive at

154
the robust range of 83, making DO Animacy the overwhelming factor in favoring a-marking, precluding several other factors. This is in line with what I had previously observed in the Mexico City data (Tippets forthcoming). Because DO Animacy accounts for so much of the data, we also observe the effect that the majority of factor groups were eliminated. The overall distributions below show how animacy affects a-marking in the data.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>.95</td>
<td>78%</td>
<td>240</td>
<td>39%</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>.66</td>
<td>47%</td>
<td>19</td>
<td>3%</td>
</tr>
<tr>
<td>Inanimate</td>
<td>.12</td>
<td>4%</td>
<td>363</td>
<td>58%</td>
</tr>
</tbody>
</table>

Table 39  Multivariate analysis of DO animacy

The data show the expected results: the majority of human DOs are marked (188/240) while inanimate DOs are rarely marked (16/363). The robustness of animacy in explaining marking will also be made clear in the other significant factor groups selected by the multivariate analysis.46

45 Particularly reaffirmed is the finding that the metric of relative animacy provides no measurable difference in explaining SDOM. The rate of marking for objects higher than animacy in the HC Mexico City corpus is exactly the same as the overall marking of human DOs at 77%.
46 Interestingly, my data do not show the frequency of inanimate marking that would be expected based upon the diachronic evidence of Company-Company (2002). Despite using the same corpus, this discrepancy may find an explanation in that certain tokens may have been excluded per my methodologies in the HC Mexico City corpus. Also, she used several other texts to gather marked inanimate DOs and I presume this explains for the difference (cf. 2002:153). I take up the conditions of marked inanimates 6.2.1.
5.3.2. Definiteness

As in the other dialects, we find that definiteness is selected as significant in favoring a-marking. Although relative to one another, the effect of definiteness (Range of 61) does not approach the effect of DO Animacy (83), yet definiteness clearly provides explanatory power to the observed variation of SDOM. Definite DOs favor marking with a probability of .67 and a rate of marking of 41% whereas indefinite DOs disfavor marking with a probability of .28 and a rate of marking of 25%. Particularly stark is the effect of Bare DOs, which are almost never marked.

Due to the pervasive nature of animacy observed in the data, the effects of definiteness on marking must be cross-tabulated to determine if these effects on SDOM are truly independent.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>definite</th>
<th>indefinite</th>
<th>No determiner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>88% (153)</td>
<td>67% (32)</td>
<td>15% (3)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>53% (9)</td>
<td>0% (0)</td>
<td>[ ]</td>
</tr>
<tr>
<td>Inanimate</td>
<td>6% (14)</td>
<td>2% (2)</td>
<td>0% (0)</td>
</tr>
</tbody>
</table>

Table 40 Rate of marking according to definiteness and DO animacy

From cross-tabulating Definiteness and DO Animacy, it is clear that definiteness is reflecting the animacy of the DO. Nevertheless, the effect of definiteness on human DOs is independent when we consider the overall rate of marking for human DOs at 78%.
The rate of marking for definite human DOs is 10% higher, the rate of marking in human indefinites is 11% lower, and an even lower rate of marking at only 15% for Bare DOs. Further examining the effects of definiteness, it is evident that marking varies based upon the definite expression as well. Definite human NPs are marked with extreme frequency (88%) while human indefinites are less frequently marked (67%) by comparison. Therefore the multivariate analysis captures a meaningful effect of definiteness in accounting for a-marking in the Madrid data.

5.3.3. Number

Throughout the various regression analyses performed on the Mexico City data, the factor group number was consistently selected as significant. Marking is favored by singular DOs with a probability of .61 while marking is disfavored by plural DOs having a probability of .35.

The original factor group for number also contained a third factor, non-count noun to account for nouns that have referents that are neither singular nor plural.

However, in the multivariate analyses it became apparent that there were interactions. The factor weights for non-count nouns would swing from significant to non-significant, contradicting the initial base-line weight of .54. This occurred immediately as a result of interactions with DO Animacy. Looking at the cross-tabulation of DO Animacy and Number it became apparent that the interactions arose from the poor distribution of data for non-count nouns: 85% of marked non-count noun tokens are human (i.e. names of people) whereas there is only one marked inanimate token (2% of non-count nouns).
Given this poor distribution, the non-count noun factor group was removed for the final multivariate analysis shown above.\textsuperscript{47}

With factor group number adjusted, it is prudent to see how animacy interfaces with singular and plural DOs. Table 41 illustrates the effect of animacy with the factor group number.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>Single</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>85% (96)</td>
<td>65% (58)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>40% (4)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>6% (11)</td>
<td>3% (4)</td>
</tr>
</tbody>
</table>

Table 41 Rate of marking according to number and DO animacy

As expected, virtually all marking in the Mexico City data occurs with human DOs. When the overall rate of marking human DOs in the Mexico City data is considered (78%), the effect of number is apparent: singular DOs favor marking while plural DOs disfavor marking. When human DOs are singular they are more likely to be marked than when they are plural.

5.3.4. Summary of the Mexico City data

The results of the multivariate analysis provide a thorough and detailed description of the specific factors motivating SDOM in the Mexico City data. Despite

\textsuperscript{47} The difference in log likelihoods between the two analyses was minimal. The best log likelihood with the observed crossover in the multivariate analysis was -172.689 and the multivariate analysis removing non-count nouns is -173.064.
the difficulties regarding number, the primary factor groups of DO Animacy, Definiteness and Number provide a very defined model of how SDOM is manifested. The strength of effect of these factor groups, particularly of DO Animacy and Definiteness, indicate how efficiently they account for the SDOM variation observed in the HC Mexico corpus. The robustness of DO Animacy in accounting for marking is to the detriment of all other factors, with the aforementioned exceptions. In the case of definiteness, however, it is observed that there is an independent effect: for the class of definite human DOs, the rate of marking is significantly higher and significantly lower for indefinite and bare human DOs. Finally, number has the same observable effect; while human DOs overwhelmingly favor marking, number maintains an independent effect in marking.

5.4. Conclusion

The chapter outlined the multivariate analyses of the individual corpora. It was observed that there are particular patterns shared across the three dialects with respect to SDOM, namely the similar effects of DO Animacy and Definiteness. In addition, the data show that there are observable differences among the factors selected as relevant in accounting for marking in these three dialects. This can be seen primarily in the differing levels of complexity of factors selected as significant across the dialects: Buenos Aires and Madrid had more factor groups selected than Mexico City indicating that there are more nuanced factors contributing to the marking.
Despite the pervasiveness of DO animacy in motivating $a$-marking, the data show the independence of numerous factors in accounting for increased or decreased rates of marking, providing strong evidence to support the hypothesis undertaken in this dissertation: multiple factors motivate a-marking. The following chapter addresses the findings of Chapter 4 and 5 with respect to the previous and current studies of SDOM and the contributions of these multivariate analyses to the literature.
CHAPTER 6: DIFFERENTIAL OBJECT MARKING IN SPOKEN SPANISH

6. Introduction

The variable nature of Spanish Differential Object Marking has been the focus of this dissertation. Because no single factor is able to account for the variation of a-marking in Spanish, I have assumed that there must be multiple factors that motivate this variation. This central hypothesis is based upon the previous literature regarding the relevant factor(s) of SDOM (cf. Chapter 2). In order to test this hypothesis, my approach has been to quantify the myriad potential features by observing how SDOM is used in spoken Spanish. To this end the findings of this dissertation contribute to the general analyses regarding SDOM by providing quantitative data to corroborate their assumptions and assessments of a-marking (e.g. animacy, definiteness, individuation, etc.). Furthermore, the findings also recast particular assumptions regarding marking (topicality, clause-level properties). The organization of the present chapter centers on the various factors that have been shown to account for SDOM. Furthermore, the individual sections confirm the secondary hypothesis that relevant features of SDOM operate at different levels. In Section 6.1 I discuss the relevant NP properties favoring marking, substantiating the individuation hypothesis. In Section 6.2 I focus on the relevance and utility of the individuation parameter. Section 6.3 focuses on the clause-
level properties that are relevant to marking. I then discuss how SDOM is manifest at the discourse-level through a quantified conceptualization of topicality in Section 6.4.

6.1. NP Features of SDOM

What is evident in the quantitative data obtained in this project is the relevance of particular linguistic factors to DO marking. The multivariate analyses tested for the significance of several features particular to the NP and found that there are clear correlations between NP-level properties of the DO namely animacy, specificity, definiteness, number, and a-marking. These NP features coincide precisely with the fundamental features of the parameter known as individuation (cf. Table 5.1). Therefore in selecting these features the multivariate analyses has provided evidence that the parameter of individuation favors SDOM.

As discussed in 2.4., individuation is a collection of features that characterize the object of a transitive clause. Originally conceived by Hopper and Thompson (1980), individuation is a gradient property of the accusative NP, where a DO is considered more or less individuated depending on the clustering of these binary features. In Figure 10 I repeat the binary properties of individuation as presented by Kliffer (1984)\textsuperscript{48}:

\footnotesize

\textsuperscript{48} Kliffer actually reorganizes the original categories of Hopper and Thompson slightly as outlined in his endnote (cf. Kliffer 1984: 214) to address his Spanish data.

\addcontentsline{toc}{section}{References}
<table>
<thead>
<tr>
<th>Individuated</th>
<th>Non-individuated</th>
</tr>
</thead>
<tbody>
<tr>
<td>proper</td>
<td>common</td>
</tr>
<tr>
<td>human</td>
<td>inanimate</td>
</tr>
<tr>
<td>definite</td>
<td>non-definite</td>
</tr>
<tr>
<td>referential</td>
<td>non-referential</td>
</tr>
<tr>
<td>(specific)⁴⁹</td>
<td>(non-specific)</td>
</tr>
<tr>
<td>singular</td>
<td>plural</td>
</tr>
<tr>
<td>count</td>
<td>mass</td>
</tr>
</tbody>
</table>

Figure 10 Properties of individuation

A brief survey of the properties of individuation in Figure 10 immediately brings to mind particular features that have been attributed to SDOM in previous studies. It should therefore be unsurprising to hypothesize that the variation in marking should correlate with more or less individuated DOs.

If we compare the properties of individuation with the factor groups selected as significant from the multivariate analysis, we see this is the case: these properties of highly individuated DOs coincide with NP properties favoring marking.

⁴⁹ I follow Kliffer’s precedent for reorganizing the typology; his use of the term ‘referential’ is precisely my distinction between specific and non-specific.
<table>
<thead>
<tr>
<th>Individuated DOs</th>
<th>Factor groups selected in the multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proper</td>
<td>Yes (part of FG definiteness)</td>
</tr>
<tr>
<td>Human</td>
<td>Yes</td>
</tr>
<tr>
<td>definite</td>
<td>Yes</td>
</tr>
<tr>
<td>referential (specific)</td>
<td>Yes</td>
</tr>
<tr>
<td>singular (number)</td>
<td>Yes</td>
</tr>
<tr>
<td>count</td>
<td>(Not tested independently of number)</td>
</tr>
</tbody>
</table>

Table 42  Individuation properties and groups selected as significant

What the multivariate analysis has permitted is a description of *a*-marking that demonstrates that multiple factors contribute to the marking. The central premise of this dissertation is the hypothesis that multiple factors can best account for marking and the data show that there is a particular constellation of NP factors that characterize marked DOs, individuation. By virtue of the NP properties coded in the data, individuation has been quantified and the multivariate analysis essentially has allowed me to unpack the parameters of individuation. In addition to validating the effect of these factors on SDOM, the multivariate analysis allows us to see both how the factors are distributed in the data as well as how they relate to one another.
<table>
<thead>
<tr>
<th>Factor Group</th>
<th>Probability</th>
<th>% a-marked</th>
<th>Total N</th>
<th>% of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DO Animacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>.91</td>
<td>72%</td>
<td>681</td>
<td>33%</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>.68</td>
<td>41%</td>
<td>60</td>
<td>3%</td>
</tr>
<tr>
<td>Inanimate</td>
<td>.23</td>
<td>4%</td>
<td>1349</td>
<td>64%</td>
</tr>
<tr>
<td><strong>Range:</strong> 68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Definiteness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definite</td>
<td>.62</td>
<td>33%</td>
<td>1434</td>
<td>69%</td>
</tr>
<tr>
<td>Indefinite</td>
<td>.36</td>
<td>20%</td>
<td>418</td>
<td>20%</td>
</tr>
<tr>
<td>Bare NP</td>
<td>.12</td>
<td>7%</td>
<td>238</td>
<td>11%</td>
</tr>
<tr>
<td><strong>Range:</strong> 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.58</td>
<td>32%</td>
<td>1178</td>
<td>56%</td>
</tr>
<tr>
<td>Non-specific</td>
<td>.40</td>
<td>22%</td>
<td>912</td>
<td>44%</td>
</tr>
<tr>
<td><strong>Range:</strong> 18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singular</td>
<td>.55</td>
<td>28%</td>
<td>1000</td>
<td>48%</td>
</tr>
<tr>
<td>Plural</td>
<td>.42</td>
<td>21%</td>
<td>698</td>
<td>33%</td>
</tr>
<tr>
<td><strong>Range:</strong> 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 43  Properties of individuation selected as significant

Remarkably, the relative strength of the factor groups reflects hypotheses new (Melis and Flores 2009) and old (Kliffer 1984, 1995) regarding the relationship among these factors.
While no one would dispute the preeminence of animacy in affecting the likelihood of marking, Kliffer, as well as Melis and Flores (cf. pg. 279), postulate a broader spectrum of factors potentially construed as a hierarchy *relative* to each other. It is precisely these observations that are validated by the quantitative approach to SDOM I have used in this dissertation.

The individuation hypothesis finds considerable evidence from the data. In Table 42 we have seen that a particular clustering of factors favors marking. What makes the individuation hypothesis so tenable, in addition to the overwhelming quantitative evidence is the fact that it assumes variability. Individuation is a relative property: DOs may be more or less individuated. An analysis that does not impose categorical constraints on SDOM is far more favorable precisely because it can account for the variation observed in the data. The greater the clustering of these individuating features on a DO the more *probable* that marking will occur. Foremost among these features is animacy, the overriding factor (see Kliffer’s observations below) in what might be considered the initial probability in marking.\(^{50}\) The data show the overwhelming effect of animacy and its pervasive nature in other factors associated with marking.

### 6.1.1. Animacy

As Kliffer asserts “Of all the individuation properties detected, only ‘human’ can override every competing criterion…” (1984: 200). There is no factor that can better

\(^{50}\) As played out in the data, being human or not human is the single greatest factor in explaining marking. However see 6.2.1. referring to the probability of marking when the subject is inanimate.
account for marking. This is the case as shown in the multivariate analysis: 88% (491/569) of the marked tokens are human DOs. However, the quantitative data provided an opportunity to evaluate the effect of animacy in the data and test for the independence of other factors. I will now discuss what the data tells us about the nature of animacy.

Based on the data, animacy is essentially a binary distinction between human and non-human. Despite culling a substantial number of tokens from three corpora, the breadth of animacy distinctions I had anticipated was disappointingly not observable.

<table>
<thead>
<tr>
<th>DO Animacy</th>
<th>% a-marked</th>
<th>% Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>72% (486)</td>
<td>32% (676)</td>
</tr>
<tr>
<td>Impersonal</td>
<td>100% (5)</td>
<td>0.2% (5)</td>
</tr>
<tr>
<td>Metaphoric</td>
<td>41% (25)</td>
<td>3% (60)</td>
</tr>
<tr>
<td>Animate</td>
<td>4% (2)</td>
<td>2% (49)</td>
</tr>
<tr>
<td>Plant</td>
<td>0% (0)</td>
<td>0.4 (8)</td>
</tr>
<tr>
<td>Inanimate</td>
<td>4% (51)</td>
<td>62% (1292)</td>
</tr>
</tbody>
</table>

Table 44 Distribution of marking with DO animacy

Clearly, Human and Inanimate are the two categories best represented in the data, with insufficient representation of other measures of animacy for evaluation.\(^{51}\) It would appear that despite the time and effort to obtain the data of this study, even greater amounts of data would be necessary to eventually get sufficient tokens to feel confident

\(^{51}\) To circumvent the infrequency of particular cases of non-human marking previous studies have intentionally sought non-human animates in literary data (Weissenrieder 1990, 1991).
in determining the conditions for variation with respect to animals (domesticated vs. wild), insects, and plants as manifest in spoken data. Therefore, these discrete categories have been relegated to the most reasonable animacy level. Considering the lack of marking observed with non-human animate (2/49) DOs, I am skeptical that non-human animates should be treated any different than inanimates.

The only other group that provides sufficient tokens is the class of metaphoric uses. This category is particularly compelling because it straddles the gap between animate (human) and inanimate. The metaphoric group consists of inherently inanimate referents that have animacy extended to it through various means including metonymy, metaphor, anthropomorphization, etc. (cf. Figure 4).\textsuperscript{52} It would appear that the distribution of the metaphoric class of DOs reflects the fact that the class is indeed straddling the two categories.\textsuperscript{53} Thus we find examples of both:

\begin{quote}
1a) \textit{El pop ha tomado a la publicidad,...como elemento, así...de ejemplo} (HC Buenos Aires)
‘Pop has taken publicity… like an element, like…as an example ’

1b): \textit{Tiene, además, un intento- - - de despertar sentimientos de ternura} (HC Buenos Aires)
‘It has an attempt, infact, to awaken the feelings of tenderness.’
\end{quote}

\textsuperscript{52} As I began coding it quickly became apparent that such distinct categories for extending animacy to inanimates would prove too discrete.
\textsuperscript{53} The distribution of marked vs. unmarked cases and a probability that slightly favors marking indicates that metaphorical extension can keep its feet in both the inanimate and animate group as it were. Since the direction of the extension is towards animacy (which favor marking), it is unsurprising to see that overall, marking is favored in metaphorical contexts.
The fundamental feature for explaining for variation in $a$-marking has been presumed to be animacy (humanness). The quantitative data shows that animacy is indeed at the core of most all marking in the data. The fact remains that this is neither a necessary, nor sufficient condition for marking as attested by the non-categorical marking of human DOs in the data and the marking of non-human DOs (cf. 6.2.1.).

Animacy is pervasive in the data. Given that 88% of the marked tokens are human, this may initially seem like a deterrent to identifying other relevant factors. However, it can be seen that animacy is only one element, albeit the predominant one, in the constellation of factors that account for marking.

6.1.2. Definiteness, Specificity and Number

Following the animacy of the DO, the data show convincingly that definiteness is an important factor in accounting for marking. While the marking is reflecting the animacy of the DO, the difference in rate of marking for definite human DOs is significantly higher than the rate of marking for indefinite DOs. Why should it be that definiteness favors marking? Definiteness (like specificity) is derived in the discourse whereas animacy is an ontological state; a property of the entity referred to by an NP.

The motivation for this preference for marking can be understood initially by considering what falls within the category of definites. Proper names are among the many NPs that are definite and animate. However, definite forms are tied to specificity, an additional factor of individuation. If we compare the marking of specific definite DOs in the data, a trend for overt marking becomes quite apparent as shown in Table 45.
Table 45 Rate of marking according to specificity and definiteness

<table>
<thead>
<tr>
<th>Definiteness</th>
<th>Specific % a-marking</th>
<th>Non-specific % a-marking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite</td>
<td>34% (336)</td>
<td>30% (70)</td>
</tr>
<tr>
<td>Indefinite</td>
<td>21% (30)</td>
<td>19% (53)</td>
</tr>
<tr>
<td>Bare DO</td>
<td>12% (6)</td>
<td>5% (10)</td>
</tr>
</tbody>
</table>

While animacy and definiteness have the strongest effects of the individuation properties, specificity does show that it favors marking more frequently than non-specific cases.

Specificity as determined in this dissertation is selected as significant, at least partial confirmation of its relevance, despite most frequently being assumed as hyponymous to definiteness. Furthermore, the data have corroborated the assumption (and often prescriptive explanation) that specific human DOs are marked. Specificity was shown to have an independent effect on marking, which simultaneously strengthens the claims that multiple factors are at work in accounting for SDOM variation and that highly individuated DOs are ideal candidates for marking.

The final component of the features that make-up individuation is number. It is shown to be significant. Remarkably, Kliffer also presumed that number would be significant, “Number, then, can be significant when it is not overshadowed by definiteness or humanness” (pg. 207). Section 4.8. has shown how number affects marking. This factor forms part of the features characteristic of individuated DOs, and it has been shown to correlate with marking. Based upon the multivariate results, number 170
is far less robust than animacy and definiteness. As will be seen, taken together these features of definiteness, specificity and number do have tangible effects on marking.

6.2. Individuation

The concept of individuation captures an important generalization about the features of marked DOs: there are series of factors that become particularly relevant in accounting for the variation in SDOM. These factors are NP features associated with more individuated DOs. Based upon the relative strength of these factors in the multivariate analysis, we can hypothesize a particular ordering of relevance among these factors. Depending upon which and how many features a DO has, we can make assumptions about what types of DOs are more or less likely to be marked.

Animacy has continually asserted its preeminence in favoring marking, however the likelihood of marking increases with the addition of other features, namely when the DO is definite, specific and singular. Assuming the individuation parameters are capturing something meaningful, we can postulate that the DO that is most likely to be marked will be: Human, definite, specific and singular. If this is the case, it would then be expected that this particular constellation of highly individuated features would be the most frequently marked of all marked DOs. This is precisely what we find in the data. When these four factors are considered, the DOs with this particular clustering of features are the most frequently marked (N = 205).\(^{54}\) More than 1/3\(^{rd}\) \((205/569)\) of all marked

\(^{54}\) Note that if we only considered animacy i.e., humanness, then animacy would be the most frequent single factor that accounts for marking in the data. However, a single factor cannot account for the observed variation and it is clearly evident that other
tokens in the data are “maximally” individuated: human, definite, specific, individuated. The rate of marking for this pattern of features is 89% (205/230). For some perspective, the next most frequent constellation of “highly” individuated DOs: human, definite, specific, plural has only 78 occurrences with a lower rate of marking at 83%. This pattern continues as shown in Table 46.

<table>
<thead>
<tr>
<th>Individuation level</th>
<th>% a-marked</th>
</tr>
</thead>
<tbody>
<tr>
<td>human, definite, specific, single</td>
<td>89% (205/230)</td>
</tr>
<tr>
<td>human definite, specific plural</td>
<td>83% (78/94)</td>
</tr>
<tr>
<td>human definite, non-specific, single</td>
<td>77% (78/101)</td>
</tr>
<tr>
<td>human definite, non-specific, plural</td>
<td>72% (41/57)</td>
</tr>
</tbody>
</table>

Table 46  Individuation level for marked human DOs

Note the general pattern of Table 46, the percentage of marking decreases as the DO becomes less individuated. This also coincides with a decrease in the overall frequency of these marked DOs in the data. Taken together, there is strong evidence that the complex concept of individuation captures meaningful patterns of marking. This will be further substantiated in the case of inanimate DOs.

features also have measurable effect on marking. Therefore, looking at only one factor misses the dynamic coalescing effect of the properties of individuation on SDOM.
6.2.1. Inanimate marking and individuation

What is to be said for inanimates that are marked? There has been much made of the extension of *a*-marking into the realm of marking non-human animates and inanimates (cf. Weissenrieder 1990, 1991; Pensado 1995; Company Company 2002; Laca 2006).\(^{55}\) The data here show that marking is still consistently low with a relatively uniform frequency across dialects (3% Buenos Aires, 4% Mexico City, 3% Madrid).

Addressing the previous literature and theories on SDOM, there are at least two perspectives based on what is found in the data.

The assumption that “exceptional” *a*-marking is in any way incipiently extending into the language in any significant or measurable way at least as evinced in the synchronic data of this dissertation is brought into question. Diachronic analyses of SDOM do show a persistent encroachment and expansion of *a*-marking into the ever-increasing contexts that diverge from the canonical marked DO: definite specific human (Company Company 2002; Laca 2003, von Heusinger 2008; Melis and Flores 2009).

The more optimistic perspective is the fact that we have quantified, documented evidence that SDOM does indeed occur with DOs that what would initially seem to be inexplicably marked based on animacy (and presumably not attributable to disambiguating function or metaphorical extension to the DO). Nevertheless, they are known to occur despite what prescriptive (and even some descriptive) texts may claim. This current endeavor also revealed marked inanimate DOs in the data, returning to the question as to why this marking may occur.

\(^{55}\) Curiously, von Heusinger and Kaiser presume that marking inanimates in the standard (read Peninsular) variety is ungrammatical (2003:42).
Remarkably, the answer is precisely the same as for marking with animate DOS: highly individuated inanimate objects also favor marking. There are 53 marked inanimate DOS found in the corpus. Among these inanimates the group that is most frequently marked in the data consists of definite, specific and singular DOS (N = 22). That is to say, in every other way these marked inanimate DOS are identical to their marked animate counterparts – they are highly individuated. The second largest group of marked inanimates consists of definite, specific, plural DOS with 10. If we group together these two sets of highly individuated DOS (definite and specific), they account for 60% (32/53) of all marked inanimates.

<table>
<thead>
<tr>
<th>Individuation level</th>
<th>% of total marked inanimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>inanimate, definite, specific, single</td>
<td>42% (22/53)</td>
</tr>
<tr>
<td>inanimate definite, specific, plural</td>
<td>19% (10/53)</td>
</tr>
<tr>
<td>inanimate definite, non-specific, single</td>
<td>19% (10/53)</td>
</tr>
<tr>
<td>inanimate definite, non-specific, plural</td>
<td>6% (3/53)</td>
</tr>
<tr>
<td>Total</td>
<td>85% (45/53)</td>
</tr>
</tbody>
</table>

Table 47 Individuation levels for inanimate marked DOS

Table 47 shows what the vast majority of marked inanimate DOS the share properties of individuated objects. There are several insights that can be gleaned from these results. First, despite not being animate, definiteness appears to be a critical component of individuated DOS. As the DOS descend in level of individuation in the table, the decrease
in their relative frequency and rate of marking is obvious. From these findings it can be argued that the minimum critical mass for marking inanimates is approximately being definite and single.

The data shows that marking for animates and inanimates is higher with the accumulation (or clustering) of additional individuating factor. Likewise, the fewer individuating factors a DO has, the fewer the cases and lower the rate of marking. This is compelling evidence that the parameter of individuation is capturing a meaningful effect in marking.

However individuation has an equally relevant correlation with other factors that have been associated with marking. The particular cluster of features that make a DO highly individuated (human, definite, specific and singular) also correlate with referents that are more likely to be topical. Topicality has been primarily associated with animacy, one of the individuating features. The observation for most has been that humans are commonly topical (Bentivoglio 1983, Leonetti 2004). Laca (1995) and Schwenter, 2006) however, extended this relationship between animacy and topicality to potentially accounting for SDOM. We will return to this shortly in 6.3.

6.3. Structures & Functions related to SDOM at the clause-level

In the context of SDOM, individuation presents a hypothesis for accounting for marking; the features that make a DO more individuated coincide with the features that favor marking. Although the properties of individuation are characteristic of the NP, the origins of individuation stem from the clausal relationship of transitivity, according to
Hopper and Thompson, “Individuation…refers both to the distinctness of the patient from the A[gent] and to its distinctness from its own background” (1980:253). Despite the features of animacy, definiteness, specificity and number being attributed to the DO, the cumulative effect of these factors is to individuate the DO in terms of the broader context of the transitive clause.

If we are to assume that individuation does appropriately capture those DOs that favor marking, these DOs are also individuated with respect to the greater clausal context. It follows then that SDOM is not just a feature of DOs, but is in fact, also a clausal property (de Swart 2007). This finds support in the structures and functions of SDOM that operate at the clausal level. These clause-level structures and functions are primarily manifest in the disambiguating function, embedded clause structure and relative animacy relationship. SDOM is said to depend on the conditions, properties and participants contained in the clause itself. I will discuss how these clause-level conditions for SDOM are manifest in the data and what they tell us about the claims in the literature.

6.3.1. The Disambiguation Hypothesis

As has been discussed in the literature (cf. 2.6), SDOM may occur in contexts where an ambiguity arises between the subject and the object. My methodology specifically accounts for these conditions and the quantitative data provide some important insight to inform the discussion on ambiguity.
To properly discuss disambiguation, let us contextualize the concept within the data. Ambiguous contexts are exceedingly rare, occurring in 4% (78/2090) of the data. Because the subject and object can be determined and ambiguity circumvented by numerous means (morphological, syntactic and semantic), it is unsurprising to find so few truly ambiguous contexts. This means that ambiguous contexts represent a particular circumstance where no other linguistic recourse is available. What the data show is that in 77% (60/78) of conditions where ambiguity arises, human DOs are involved. And of these 60 cases, 93% (56/60) were marked. From this result the assumption is that the most likely context for disambiguation is where both subject and DO are human.

Comrie asserts that the function of overt marking is to capture the *difference* between subjects and direct objects (1979:19). In the case of ambiguity, that difference is lost. The resulting conditions would then lead to preferential marking. In ambiguous conditions we have atypical objects\(^{56}\) that favor marking and whose direct effect is to confound the roles of agent and patient. This only occurs at the clause level. Comrie also assumes that those features of the prototypical subject (animate, definite, etc.) are marked when they are attributed to objects. In essence, Comrie has detailed two different motivations for SDOM, albeit one that results from the other. Objects that look like subjects favor marking and the fact that they look like subjects leads to ambiguity, motivating the differentiating or disambiguating function of SDOM in the case of Spanish.

\(^{56}\) Per the markedness assumptions of Comrie (1981) and others (Silverstein 1976)) regarding Agents and Patients, objects are prototypically inanimate and indefinite.
We can construe this as two different motivations for marking supported by the quantitative data. First, the general observation is that, in the data, human DOs are clearly the most frequently marked at 72% (491/569). Second, in the relatively few cases of ambiguity, primarily arising from an ambiguity between human NPs, the function of SDOM is to disambiguate. This is corroborated by the fact that the marking of DOs in such cases is disproportionally higher as observed in the rate of marking at 93% (56/60). Therefore disambiguation is indeed a function of marking.

Let us now consider the cases where ambiguity arises among non-humans. Despite the tendency in many studies to present examples of disambiguating functions of SDOM with inanimates, the data only provides 12 such cases (of 2090 tokens total). Among them only 33% (4/12) were actually marked. With such sparse results, I will limit my speculation to the following observations. These types of ambiguous constructions are frequently cited, and primarily consist of constructed examples. However note that in these cases we do not have the same circumstances as ambiguity arising as before. The object is supposed to be unmarked because of its (prototypically) inanimate condition. What would be expected per the assumptions of Comrie (1989) and Aissen (2003) is Differential Subject Marking since the subject would be marked with respect to those features of the object: inanimate, indefinite, etc.\textsuperscript{57} However this theoretically presumption is not substantiated since Spanish does not mark subjects differentially. What is observed for ambiguous contexts with human or inanimate NPs is the function of SDOM to disambiguate. It is a syntactic solution to a syntactic and

\textsuperscript{57} Typologically, Differential Subject Marking occurs with very low frequency in natural languages (\textit{cf.} de Hoop and de Swart 2008).
semantic problem that arises when the particular properties of the NP, the clause and/or the discourse context fail to make the subject and object discrete. This connection between characteristics of one level or domain (the NP) and another level or domain (the clause level) provides important evidence for understanding that marking depends on multiple factors from distinct domains.

6.3.2. Relative Animacy

In precisely the same fashion, Relative Animacy demonstrates how SDOM operates as a clausal property. The NP features of Subject and Object are relevant in the clausal domain. In previous research (Tippets forthcoming) I noticed a prominent and measurable effect for Relative animacy and a-marking. Having refined the methodology for this dissertation, I was somewhat disappointed to see that relative animacy failed to have quite the robust effect in the data as I had observed previously. In the case of the Mexico City data, my previous observations were confirmed: relative animacy does not have an independent or measurable effect on SDOM for that corpus. However, the Madrid data proved problematic. Despite being selected as significant, the regression analysis showed interactions and the independence of relative animacy from animacy could not be teased apart. Nevertheless, the clause-level property of relative animacy is shown to have an independent effect when all the data were evaluated (cf. Table 5) and in the Buenos Aires data. The increased rate of marking observed in cases where the animacy of the subject and the object are atypical corroborates Comrie’s claims about the preference for marking by virtue of the reversal of features expected for subject and
object (1981: 128). Furthermore, this provides evidence for my argument that SDOM operates in the clausal domain, affected by clausal-properties.

6.3.3. Embedded clauses

Embedded clauses are particularly telling in showing how marking is a result of properties in the clause domain. As noted in previous studies (Torrego-Salcedo 1999, Roegeist 2003, Laca 2003, von Heusinger and Kaiser 2008), this structure favors marking precisely because the object of the matrix verb functions as the subject of the following infinitive. Given that there are already selective properties involved with this construction, namely that the object of the matrix verb will tend to be agentive as the subject of the embedded clause, it is logical to conclude that marking will be favored. The selectional properties that are presupposed by the structure (a preference for agentive objects) would lend themselves to the aforementioned claims of Torrego-Salcedo, Roegeist and others favoring marking. Marking is indeed favored based upon the data gathered. However, upon scrutinizing the data, in the 24 cases where marking occurred, the object was also human. Furthermore, the three unmarked cases of the embedded construction were all inanimate DOs as shown below.

2a)  *Yo he visto pasar por acá todos los colectivos* (HC Buenos Aires)  
‘I’ve seen all the buses pass by here’

2b)  *(yo)*prefería estar viendo correr los pájaros. (HC Madrid)  
‘I preferred to be watching the birds run.’

2c)  *Tritón hizo... este...brotar el agua.* (HC Mexico City)  
‘Triton made, um, the water shoot up.’
As a result of these findings, it is impossible to distinguish the effects of animacy and the structure of an embedded clause itself. Given that this structure is very infrequent any claims for this configuration are speculative. Because the data have not shown variation in marking independent of the animacy of the DO, I was compelled to exclude this factor group in my final analysis of the data.

What I have explored in this section are the clausal level properties that have been shown to affect SDOM. Disambiguation, relative animacy and embedded clauses were incorporated into my quantitative analysis of SDOM with the purpose of testing for their relevance in accounting for marking. Crucially, they each function (and can only be measured) at the clause-level. From these findings, I conclude that SDOM is sensitive to clause-level properties, particular in the case of disambiguation and, although less apparent, relative animacy.

6.4. Properties of the discourse domain: Topicality

One of the primary claims that has gained support in explaining the variable nature of SDOM is its function at the level of discourse. Since it is clear that marking is non-categorical outside of particular semantic and verbal contexts, appealing to the discourse provides the flexibility, to allow for the observed variable marking to be ultimately a result of discourse pragmatic conventions (Weissenrieder 1991, Dumitresu 1997, Torrego-Salcedo 1999, Laca 2006). The conclusion drawn is that there are supra-nominal and supra-sentential factors in the domain of discourse that also contribute to a-
marking, such as Laca’s global properties (2006). Weissenrieder states “…no description based solely on NP characteristics, lexical verbs, or case assignment, will account for all the uses of [SDOM]. Although these semantic elements are discourse sensitive, there are other discourse-motivated and sometimes discourse-exclusive elements which condition the preposition’s appearance” (1991:151).

An important component of my analysis of SDOM from the beginning was to find a means to test for discursive features. Namely I wanted to be able to test the assumption shared by many that some marking is a result of topicality considerations, i.e. when DOs are topical, they will favor marking. Based on the metrics of Givón (1983) and Shain (2009) I was able to quantify a relative measure of topicality using referential distance and topic persistence.

As seen in Section 4.7, one measure of topicality, total referential distance was shown to correlate with marking. In addition to being selected as significant in the multivariate analysis, the distribution of marked DOs in relation to total referential distance showed the likelihood of marking increased with the proximity of co-referencing expressions. The data also show that those DOs that are (more) topical as a measure of TTP i.e., frequency in the data are more likely to be marked. A-marking is likely to occur with those DOs whose referents occur with high frequency in the limited confines of the immediately preceding and following discourse. However, TTP failed to reach significance in the multivariate analyses for all dialects and the combined data. While this causes some initial consternation, a detailed analysis presents critical insight into how SDOM operates at the domain of discourse.
Despite interactions the data show a promising correlation between TTP, SDOM and relative topicality.

<table>
<thead>
<tr>
<th>TTP</th>
<th>% $a$-marked human DO</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-20 mentions</td>
<td>84% (93)</td>
</tr>
<tr>
<td>1-5 mentions</td>
<td>77% (239)</td>
</tr>
<tr>
<td>0 mentions</td>
<td>61% (158)</td>
</tr>
<tr>
<td>Overall rate</td>
<td>72% (491)</td>
</tr>
</tbody>
</table>

Table 48 Rate of marking according to TTP and all human DOs in the data

As can be observed, when a marked human object has no TTP i.e., it is a one-time occurrence and the absolutely least topical DO, its rate of marking (61%) is well below the overall rate of marking human DOs (72%). On the other hand, the human DOs with the highest frequency (in terms of number of mentions in the data) in the data are marked at a greater rate (84%) than the overall rate.

We can also see the relationship between topicality and individuation. 68% (63/93) of high-frequency marked human DOs are highly individuated (human, definite and specific). Note however that without a quantified conception of topicality, the assumption that subject-like objects are indeed topic-worthy, or more topical, is impossible to prove. Consider that in the data there are 86 highly individuated marked human DOs that are one-time occurrences. They cannot be topical per total referential distance and total topic persistence. Therefore, an assumption that a marked, highly
individuated DO must be relatively topical in the sense that it is high frequency and proximal is untenable based upon the data.

This tandem of individuation and topicality are significant quantitative contributions towards understanding variable a-marking. The quantitative approach of this dissertation has brought a crucial insight to a-marking as it relates to the concepts of topicality and individuation that appear in some of the SDOM literature: being highly individuated greatly increases the likelihood that a DO will be marked in general, but it does not mean that this DO is necessarily topical. With the independent measure of TTP, it is shown that those DOs that are most topical are more likely to be individuated which is not the same claim. The strong argument based on the data then is that being an individuated DO is a necessary but not a sufficient condition for topicality: the DO must be observed to be topical in the discourse per the TTP. This insight of the relationship between individuation and topicality both supports the premise of this dissertation: there are independent factors that contribute to marking, and shows these factors operate across domains.

Lastly, one of the significant challenges addressed here has been to find a way to quantify and measure topicality in relation to a-marking. TRD and TTP confirm that SDOM can function as a property of the discourse, corroborating the unsubstantiated assumptions of Weissenrieder (1991) and Laca (2006). Crucially, a measurable conception of topicality independent of SDOM shows how variable marking reflects

\[58\text{ Topical as I have defined this concept in using the metrics of TTP and TRD.}\]
relatively topical DOs vs. relatively non-topical DOs based upon their frequency in the data (TTP) and the relatively proximal nature of their prior reference in the data (TRD).

6.5. Dialects

The multivariate analyses of each dialect show evidence of similarities and differences in what factors contribute to SDOM. The advantages deriving from these individual analyses are several. First, these analyses provide the ability to identify general trends i.e., the similarities in the distribution of SDOM. Broad generalizations based on these trends are possible as a result of the uniformity of the methodology and the uniformity of how the data is reported in the multivariate analyses. Second, the individual analyses provide a means to identify precisely how these dialects diverge relative to one another. The individual multivariate analyses show that each dialect has its own unique constellation of factors. These distinctions are summarized in Table 49 below.
<table>
<thead>
<tr>
<th>FACTOR GROUP</th>
<th>Buenos Aires</th>
<th>Madrid</th>
<th>Mexico City</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO ANIMACY</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Definiteness</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ambiguity</td>
<td>X</td>
<td>X</td>
<td>[ ]</td>
</tr>
<tr>
<td>Specificity</td>
<td>X</td>
<td>X</td>
<td>[ ]</td>
</tr>
<tr>
<td>Relative Animacy</td>
<td>X</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Number</td>
<td>[ ]</td>
<td>[ ]</td>
<td>X</td>
</tr>
<tr>
<td>Total Referential Distance</td>
<td>X</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>Total Topic Persistence</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

( X = selected as significant, [ ] = not significant)

Table 49  Factor groups selected as significant in each dialect

By virtue of having more factor groups selected, the Buenos Aires and Madrid data have more complex and nuanced conditions for α-marking than the data from Mexico City. Excluding the shared significant factors of DO Animacy and Definiteness, the analyses of Buenos Aires and Madrid also select Ambiguity and Specificity as significant. Based on these findings a tentative claim is that the Buenos Aires and Madrid dialects are similar based on their shared significant FGs. In addition we might characterize SDOM in Buenos Aires as relatively more complex with respect to SDOM in the Madrid and Mexico City dialects based on the number of factor groups that contribute to marking. It is hoped that this insight into one particular area of morpho-syntactic variation may be profitably applied to the analysis and description of these dialects more generally.
CHAPTER 7: CONCLUSION

7. Summary

This dissertation has presented a quantitative analysis on Differential Object Marking in Spanish. The variation in $a$-marking has been the focus of many productive studies and continues to be a point of departure for significant research and analysis. This dissertation has ascribed the variable nature of SDOM to multiple underlying features, factors and functions. Importantly, this endeavor has been able to substantiate the motivating factors that are responsible for the majority of variable marking observed in the data with quantitative evidence.

The undertaking of this dissertation was guided by the central hypothesis that multiple factors account for the variable nature of SDOM. In order to test for the relevance and effect of these potential factor groups, a quantitative approach was selected. The use of the Goldvarb multivariate statistical program proved invaluable because it facilitated the testing of the hypothesis. The ability to analyze the relevant factors with respect to each other not only provides critical insight regarding how these factors relate to one another, but speaks directly to the claims that these factors have varying levels of influence on determining SDOM (Kliffer 1984, 1995; Weissenrieder 1991; Laca 2006).

7.1. Findings
Among the features purported to motivate $a$-marking in previous studies, I observed that these factors operate across different domains, namely at the level of NP properties, the clausal domain and the discursive domain. Therefore, one of the additional goals of this dissertation has been to show evidence that SDOM does indeed operate in these different domains as evinced by the observed effect on $a$-marking by these factors.

The principle findings of this dissertation are outlined below:

- Multiple factors are responsible for accounting for the variable nature of SDOM
- Individuation is identified as a critical component in explaining marking, providing the most efficient account of the data
- The relevant factors found in the data also operate across different domains; SDOM variation is beholden to properties at the NP-level, the clause-level and the discourse-level
- A quantifiable conception of topicality is shown to be relevant in characterizing a relative topical function of SDOM
- SDOM varies along dialectal lines

In order to test and ultimately prove the multiplicity of factors hypothesis, quantitative evidence was obtained from three corpora of transcribed spoken Spanish. The Goldvarb multivariate statistical program provided the necessary statistical modeling capabilities to analyze multiple potentially relevant factors and the variation manifest in $a$-marking. Multiple factors were shown to have independent effects in the quantitative analyses of SDOM. The principle factors that coincide with $a$-marking include animacy,
definiteness, specificity and number. These factors happen to coincide with several significant features related to transitivity postulated by Hopper and Thompson (1980). These features have figured prominently in the discussion of transitivity, DOM and SDOM. It is unsurprising that these features that reflect a maximal distinction in transitive relationships between subject and object (pg.253) should also be relevant to a particularly “marked” grammatical and linguistic phenomenon like DOM.

Individuation is a valid and meaningful category for identifying the conditions in which the variable nature of a-marking can be explained. Individuation provides a more precise characterization of the factors motivating marking than animacy (+/-humanness) alone. Given what is observed in the data, the relevance of animacy cannot be underestimated. However, it has been shown that animacy of the DO combined with other features of the NP like specificity, definiteness and number, show a greater likelihood of marking and provide a more complete account of the variable marking in the data of this dissertation. Individuation conveniently captures these relevant categories under one generalization.

It is worth noting that this generalization is particular convenient and apt for the phenomenon. Due to the variable nature of SDOM, categorical claims regarding marking will necessarily fail. The utility of individuation is in its gradient and relative nature. Accounting for variable marking necessitates that gradience. Fortuitously, the gradience in marking was quantifiable using the particular cluster of individuating properties. By coding for those individuating properties, it is possible to see how they coalesce and favor marking, or if there is no clustering of the features, the lack of marking. The
passing yet, specific assumptions of Kliffer (1984, 1995) and Laca (2006) regarding the multiple factors contributing to the variation in SDOM have been verified. Critically, the idea that a hierarchy among the potential factors could be teased out was made possible using multivariate statistical software.

In exploring the extent of individuation in marking, it was discovered that not only does individuation account for marking of prototypical DOs (human, definite, specific, etc.) but it is equally adept at accounting for the alleged “exceptional” cases of a-marking. As discussed in 6.4.1, inanimate DOs that had the other features associated with highly individuated objects, especially definiteness and specificity, were marked far more frequently and constituted more than half of all marked inanimates. The literature has presumed that much of the inanimate marking that occurs is a result of extending animacy metaphorically, or as a result of the verb itself. What can be seen in this data here is these peripheral features of definiteness, specificity and number do in fact contribute to marking, an observation that also reiterates the premise that multiple factors affect marking. In addition to being selected as significant and contributing to marking per the multivariate analyses, the case of inanimate marking provides additional confirmation that these features do favor marking independent of animacy.

The findings of the quantitative analyses indicate that SDOM does indeed operate at different levels. Essentially a-marking is sensitive to properties of the NP, the clause

59 Describing the other features as peripheral does not down play their significance, but I do think it is an apt description when the data from this endeavor is considered. Animacy is absolutely pervasive in the marked DOs and is clearly the single factor that best explains the data. However, one of the primary findings is that marking is more probable in the case of higher individuation; the clustering of several features in addition to animacy increases the likelihood of marking.
and the discourse. The data show how these factors operate in relation to one another: they interface with one another. This is particularly evident regarding the parameter of individuation and how it relates NP properties to the clause-level principles of disambiguation and the clause-level property of relative animacy. Furthermore, individuation is shown to relate to the discourse-level property of topicality.

Perhaps the most challenging concept to quantify, topicality also represents one of the principal contributions of this dissertation. Frequently the catch-all explanation for marking, the generalized and polemic concept of topicality has nevertheless been a common solution for explaining the variation in SDOM: speakers are at liberty to mark object in the discourse as a means of making it prominent, or to make the DO standout because it is somehow topical. One of the challenges addressed in this dissertation was to quantify a discourse-based notion of topicality in relation to a-marking. The findings provided quantitative evidence showing that SDOM can function as a means to mark a DO that is topical relative to its frequency and distribution in the data. This is a deliberate distinction in terms of quantifying topicality, based on the metrics of Total Referential Distance and Total Topic Persistence. This quantifiable and repeatable metric of topicality presents a promising avenue for further analysis of a-marking in discourse. Furthermore, by addressing topicality in a quantifiable manner, it has provided further evidence supporting the multiplicity-of-factors premise of this dissertation.

To emphasize the point regarding the slippery nature of topicality, this explanation for topicality is circular: topical DOs favor marking and marked DOs are topical. The argument for topicality is that those DOs whose referent is topical as a measure of frequency and distribution in the data are more likely to be marked. This does not mean that marking a DO makes it topical, since many marked objects are not topical (cf. 5.3.3).
Like other morpho-syntactic features that vary along dialectal varieties of Spanish, measurable distinctions in how SDOM functions were observed in the Buenos Aires, Madrid and Mexico City corpora. Dialectal variation in marking was observable both in the multivariate analyses and in the overall distributions of the data. By incorporating dialects as part of this dissertation, I am able to add to the description of these Spanish dialects.

7.2. Future Research

Affectedness is one of the primary characteristics of transitivity (cf. Hopper and Thompson 1980: 252) and a property of the NP, This notion has recently gained ground in describing and accounting for cross-linguistic DOM phenomena (Næss 2004, 2007; Malchukov 2008; von Heusinger 2009). It is closely tied to the parameter of individuation because those DOs that are highly affected are also claimed to be highly individuated. What we find then is a somewhat circular pattern of explanation. Because most all of this discussion regarding affectedness is centered on examples and contrived data, finding a way to quantify and measure affectedness presents an avenue for additional variationist-type studies.

Finally, an ample area for future quantitative research is in testing the possible external factors that may correlate with marking. Only one external factor was actually measured in this study: dialect. I was able to show that the underlying grammar of SDOM does show subtle but tangible variation across dialects in terms of what motivates marking. Outside of this dissertation, a buried footnote, or a brief generalization made in
passing, are the extent of the quantitative studies that detail dialectal variation in the realization of SDOM. Spanish has notably contributed its share of research on morpho-syntactic variation, particularly along the dialectal dimension. Quantitative studies of these dialects may provide some insight into the general grammar. As an aside, the Habla Culta (Norma Culta) project has more than 10 dialects that are very uniform in informants, topics, and protocols. These corpora are transcribed, digitalized and accessible. I have attempted to make my methodology as transparent as possible so that further dialectal comparisons may be repeatable and comparable among these shared corpora.

There may be particularly interesting findings regarding $a$-marking in dialects and other components within their grammars. The structuralist claim of *tout se tient*, that change in part of a language system leads to other changes in that system is another avenue for further inquiry. Considering the number of factors and domains in which SDOM functions it would not be a large-stretch to presume that the dialectal peculiarities of $a$-marking correlate with other components of the grammar. This might include any number of morpho-syntactic, semantic and pragmatic features as manifest including case marking, clitics, anaphora and null objects (*cf.* Schwenter 2005).

One last consideration regarding external factors is the social function of $a$-marking. To my knowledge there have been no claims made regarding how SDOM may vary along social dimensions. In terms of how SDOM may correlate to age, gender,

---

62 García García (2007) has already proposed a pragmatic account of SDOM, based upon implicature and informativeness.
education, etc., I find it difficult to formulate a hypothesis. How salient is SDOM for native speakers? What innovative developments and changes can be detected and what or who is driving the change? In this regard, the social variation of SDOM is a prime candidate for future research.

Tied to the social evaluation of SDOM variation are the studies designed to observe change-in-progress. Per the claims that $a$-marking is incipient and slowly encroaching from the numerous diachronic studies (Pensado 1995; Company Company 2002; Laca, 2003, 2006; von Heusinger and Kaiser 2005; von Heusinger 2008; Melis and Flores 2009) there are ever growing resources from which to originate new research endeavors and to test diachronic data with observed synchronic data.
References


200


