# THE DISTRIBUTION OF DIFFERENTIAL OBJECT MARKING IN PARAGUAYAN GUARANÍ

THESIS

Presented in Partial Fulfillment of the Requirements for

the Degree Masters of Arts

in the Graduate School of The Ohio State University

By

Cory Adam Shain, B.A.

* * * * * * * *

The Ohio State University
2009

Approved By:

Master's Examination Committee:

Dr. Judith Tonhauser, Advisor

Dr. Peter W. Culicover, Advisor

_____

Dr. Scott A. Schwenter

_____

Advisors
Linguistics Graduate Program

# ABSTRACT

Differential Object Marking (DOM), the marking of some but not all direct objects in a given language, is active in over 300 languages around the world. I examine the distribution of DOM in Paraguayan Guaraní by evaluating transitive clauses in a corpus of naturally-occurring Guaraní data with respect to several factor groups. I find animacy and topicality to have the strongest relationship with object-marking frequency and show that human topical objects are frequently marked, while non-human and/or non-topical objects are infrequently marked. I then examine a Guaraní corpus published in 1640, prior to substantial Spanish contact, and do not find any clear instances of DOM. This supports the hypothesis that DOM evolved in Guaraní as a result of Spanish influence, though I also show that DOM is not distributionally equivalent in Spanish and Guaraní today. The relevance of topicality to DOM in Guaraní is unaccounted for by the two main theories of DOM, what I call the "markedness approach" and the "transitivity approach," which only appeal to animacy and definiteness. Thus the current study motivates incorporating topicality into these theories.

To my wife

# ACKNOWLEDGEMENTS

I owe an enormous debt of gratitude for the completion of this thesis to a number of wonderful people, without whom this would have been a much more difficult or downright impossible task. Perhaps the largest portion of thanks goes to the unnamed Guaraní consultants who supplied Dr. Tonhauser with the data that I have had the privilege of analyzing. I've also been very lucky to have such dedicated, responsive, and knowledgeable advisors in Professors Judith Tonhauser and Peter Culicover. I'm thankful for their generosity and patience in sacrificing an hour a week for advisory meetings, responding in detail to what must have seemed like an endless stream of inquisitive emails, and meticulously reading and commenting on multiple evolutions of the material presented here. Without their invaluable instruction and feedback this project simply could not have gotten off the ground, and its contents have been refined immensely as a result of their investment. Thank you both very much. All remaining oversights, confusions, and outright mistakes are entirely my responsibility.

Another very important individual in this process was Scott Schwenter, who essentially served as a third advisor and was kind enough to sit on my thesis committee. His dedication to the quantitative study of Differential Object Marking was a fundamental inspiration for the methodology employed here, and without his expertise regarding both Differential Object Marking and variationist research the quality of this work would have suffered tremendously. I'm grateful for his sacrifice in meeting with me on a number of occasions, his commitment to read and comment on the the material I sent him, and his helpful direction in performing corpus-based research, which was novel to me prior to this investigation.

A number of other groups and individuals made important contributions to this research. My thanks go out to all those who were involved in the former Computing Research Laboratory (CRL) at New Mexico State University, which compiled large bodies of text from a number of underdocumented languages, including Guaraní. I'm grateful for those who currently labor to maintain the CRL's website, so that these resources remain available to myself and others, and for Jim Cowie of NMSU, who was the human face of the CRL for me and was kind enough to answer several of my questions and point me to useful re-

# VITA

December, 2008 ....................................... B.A. Linguistics, The Ohio State University

April, 2009 ........................................... Presented *Differential Object Marking in Paraguayan Guaraní* at the Ohio State University College on Hispanic and Lusophone Linguistics

May, 2009 ............................................ Presented *Differential Object Marking in Paraguayan Guaraní* at the Denman Undergraduate Research Forum

May, 2009 ............................................ Received Ohio State's *Best Undergraduate Paper in Linguistics* award for *Differential Object Marking in Paraguayan Guaraní*

## FIELDS OF STUDY

**Major Field:** Linguistics

# TABLE OF CONTENTS

# LIST OF TABLES

# ABBREVIATIONS

1 .....................................1<sup>st</sup> person
2 ...................................2<sup>nd</sup> person
3 ...................................3<sup>rd</sup> person
A .........................set A verb agreement
ABS ............................ absolutive case
ACC ........................... accusative case
ALT ......................alternative question
ANTIPASS ........................ antipassive
B .........................set B verb agreement
CAUSE ..............................causative
CL ...................................classifier
CLI .........................................clitic
COMP .............................compound
COMPL .....................completive aspect
CONJ ........................... conjunction
CN ............................. common noun
DAT ............................... dative case
DEF ...................................definite
DEM .......................... demonstrative
DES ...................... desiderative mood
DET ...............................determiner
DIM ...............................diminutive
DIST ..................................... distal
DOM ............. differential object marking
DYN ................................. dynamic
EMPH ............................. emphatic
ERG ............................. ergative case
Excl ................................. exclusive
F .....................................feminine
FAC ....................................factive
FOC .....................................focus
GEN .............................genitive case
GUI ...............ablative postposition *-gui*
IMP ..........................imperative mood
IMPERF ...................imperfective aspect
Incl ...................................inclusive
IND ..........................indicative mood
INDEF .............................indefinite
INSTR ...........................instrumental

INTR ..............................intransitive
INTS ............................... intensifier
ITG .................................intangible
JE ............. reflexive and middle prefix *-je*
KUE ........terminative nominal aspect *-kue*
KURI ....... anaphoric past morpheme *-kuri*
M .................................. masculine
NEG ................................. negation
NMLZR .........................nominalizer
NOM ....................... nominative case
NPST ...............................non-past
OBJ ...................................object
OBL ...................................oblique
OM .............................object marker
PART ............................partitive case
PE .........................the morpheme *-pe*
PERF .......................perfective aspect
Pl ...................................... plural
POL ....................................polite
POSS ...............................possessive
PRES ...........................present tense
PRO ................................. pronoun
PROG .................... progressive aspect
PST .................................past tense
PURP ..............................purposive
QU ...................................question
RA ........... prospective nominal aspect *-rā*
RAE .... non-expected evidential marker *ra'e*
RC ............................. relative clause
REFL .................................reflexive
REHE ..............the morpheme *-re/ -rehe*
REP .................reportative evidentiality
REQ ..............................requestative
Sg ................................... singular
SM ...........................subject marker
(Sp) ...................... Spanish loan word
SUBJ ................................ subject
TEL .....................................telic

# CHAPTER 1

# INTRODUCTION

Differential Object Marking (DOM) refers to variation within a single language as to the morphosyntactic realization of direct objects (various definitions appear in the literature, to be discussed in §2.1). It is quite robust typologically, appearing in a large number of languages around the word from diverse language families (Bossong, 1985a). The question of what conditions the distribution of object marking variants, both cross-linguistically and language-internally, has drawn quite a bit of attention in the literature (see Bossong; Aissen, 2003b; Hopper & Thompson, 1980; Enç, 1991; Kliffer, 1982; Leonetti, 2004; Lazard, 1982; de Swart, 2007; de Hoop & Malkuchov, 2008; Næss, 2004, among many others). A variety of language-specific analyses of the distribution of DOM have been provided (see §2.2), as well as theories of DOM which make predictions about typology (see §2.3). Taken together, these accounts collectively draw on descriptive tools from a broad swath of linguistic subfields, including morphology, syntax, semantics, pragmatics, sociolinguistics, historical linguistics, discourse analysis, even language processing. The study of DOM has potential ramifications for our understanding of transitivity, markedness, prominence, and topicality, among other things. Thus the attention it has received is perhaps unsurprising.

The purpose of this thesis is to turn this attention to a language that has been studied fairly little with respect to DOM: Paraguayan Guaraní (hereafter Guaraní), a Tupí language spoken by about four million people, and a national language of Paraguay. With the exception of Bossong (1985b) and an earlier thesis of mine (C. Shain, 2008) the coverage of DOM in Guaraní has been fairly cursory in the descriptive literature on the language. Furthermore, the existence of DOM in Guaraní is not often noted in the DOM literature, and (to my knowledge) only the aforementioned analyses have been provided to date. In this study, I analyze the distribution of DOM in Guaraní by examining a corpus of naturally-occuring Guaraní data, collected by Judith Tonhauser. Hereafter I distinguish between this corpus and a larger "extended"

1

corpus, which I refer to but do not include in my statistical analysis (for details about the corpus, the extended corpus, and other data sources used here, see the Appendix). I annotate this corpus with respect to a number of potentially relevant factor groups that have been proposed as conditions on DOM in other languages. I conclude that the object marker in Guaraní is essentially a marker of topical human objects, and that objects that are either non-human or low in topicality disprefer marking.

The object marker in Guaraní is the morpheme *-pe*, which I call a "postposition" here. This is a terminological convenience and not a claim regarding the question of whether *-pe* is a case marker or an adposition (for a brief discussion of the category of *-pe* in Guaraní, see §3.1.2.4).[1] Object marking in Guaraní is differential in that some direct objects (DOs) appear with *-pe* as an object marker, indicating their object grammatical function (GF), while other DOs have no such marker. By way of example, compare (1a) and (1b).

(1)    (Adapted from Gregores & Suárez, 1961, p. 223)[2]

    a.   A-hecha   **ne-róga**.

        A1.Sg-see **2Sg.POSS-house**

        'I see **your house**.'

    b.   A-hecha   **ne-rú-pe**.

        A1.Sg-see **2Sg.POSS-father-PE**

        'I see **your father**.'

Both sentences in (1) contain the same transitive verb *hecha* 'see.' Yet the direct object in (1b) is marked by *-pe*, while the direct object in (1a) takes no marking. This pair exemplifies the differentiality of object marking in Guaraní, and variations such as this one are attested throughout the Guaraní corpus. The question of why clauses like (1a) exhibit no object marking while clauses like (1b) do is the primary focus of the current work.

The discussion is structured as follows. In Chapter 2 I survey the existing literature on DOM, reviewing the proposed definitions and distributional accounts of the phenomenon. I take these distributional proposals as motivation for several hypotheses of factor groups potentially relevant to DOM in Guaraní.

---

[1]There seems to be a pattern in the DOM literature of avoiding commitment as to the syntactic category of object markers cross-linguistically. With respect to the Spanish object marker *a*, for example, Leonetti (2004) uses both labels: "case marker" and "preposition," Aissen (2003b) uses "case marker," and Laca (1995) uses "preposition."

[2]Throughout this thesis, examples cited from previous works were labeled as "adapted" for one of two reasons. First, some notations were adjusted to maintain uniformity and avoid confusion. For example, some sources use the abbreviation "CL" to mean "classifier" and others to mean "clitic." Here, clitics were glossed as "CLI" to resolve the resulting ambiguity. Second, additional details were added to some glosses in order to clearly bring out the relevant differential object marking contrast. For example, (19) in the original presentation by Comrie (1989) contains only the sentence in Russian and the translation into English, with no second line in the gloss. In order to clearly illustrate what (19) has to do with DOM, the second line was added.

In Chapter 3 I lay out my methodology, specifying the criteria according to which (i) clauses in the corpus were considered relevant/irrelevant to the current study, and (ii) clauses were classified with respect to the factor groups under consideration. In Chapter 4 I discuss the results of the corpus study, and present an analysis of the distribution of object marking in Guaraní. In Chapter 5 I provide a comparative study of the modern Guaraní corpus and de Montoya (1640b), a Guaraní catechism published in 1640, prior to significant Spanish contact. I provide evidence that DOM did not exist in the Guaraní of de Montoya, lending credence to the claim in Bossong (1985b, n.d.) that DOM in Guaraní arose primarily as a result of contact with Spanish. In Chapter 6 I conclude, discussing the theoretical implications of my findings and offering some hypotheses to be tested by future research.

# CHAPTER 2

# DESCRIPTION AND THEORY IN THE LITERATURE ON DOM

As I mentioned in the previous chapter, quite a lot has already been said about DOM and what conditions it. This literature is a rich source of insight as to what kinds of hypotheses to test with respect to the DOM system of Guaraní. For this reason, in the current chapter I outline in brief some of the relevant descriptive generalizations and theories about the distribution of DOM. In §2.1 I survey the range of definitions that have been given to "DOM" in the literature, which are diverse, and point out that the $\pm pe$ alternation in Guaraní qualifies as DOM according to all of them. In §2.2, I discuss several different underlying conditions on the distribution of DOM across languages, which will serve as the basis for my own hypotheses about the relevant factor groups to test in this analysis. In §2.3 I review two competing syntheses of these generalizations into predictive theories about the distribution of DOM cross-linguistically. In §2.4 I summarize and conclude the literature review.

## 2.1   Definitions: What is Differential Object Marking?

The term "Differential Object Marking" (DOM) first appeared in Bossong (1985a), the author of which has done extensive research on the topic (see Bossong, 1983–1984, 1985b, 1991, 1998, among others). The specifics of its definition vary from publication to publication, but at least one common thread can be traced through the literature as to the range of phenomena that we may classify as "differential object marking": variation in the morphosyntactic realization of constructions containing an object. We may rephrase "object" more precisely according to the taxonomy of Dixon (1979), who labels the single argument of an intransive verb "S" and the two arguments of a transitive verb "A" and "O." The A-argument is the one formally encoded like S in a nominative/accusative system while the O-argument is the one en-

coded differently, and the O-argument is the one formally encoded like S in an ergative/absolutive system while the A-argument is the one encoded differently. Thus A corresponds to the subject of a transitive verb and O corresponds to the object. By way of illustration, Dixon provides the following diagram:

(2)  NOMINATIVE  $\begin{cases} A & \text{ERGATIVE} \\ S \\ O \end{cases}$ ABSOLUTIVE

ACCUSATIVE     O                              (Dixon, 1979, p. 61)
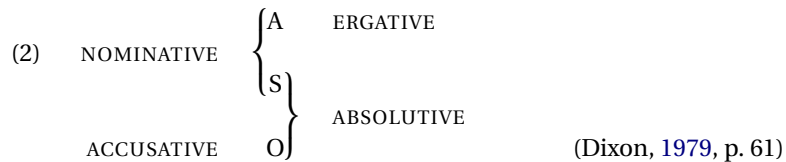
Overwhelmingly, the literature on DOM concerns itself with the argument O. Thus, when I refer above to the realization of "objects" as relevant to the study of DOM, I mean the encoding of the O-argument, excluding patientive (object-like) S-arguments and the third argument of 3-place predicates (the indirect object), since no study that I know of includes them as potentially differentially marked objects. Hereafter, I use the terms "O," "O-argument," "object," and "direct object (DO)" interchangeably.

Although the grammatical function relevant to DOM is generally agreed upon, there are a variety of definitions of it in the literature. Many studies restrict the phenomena under consideration to an alternations between overt and null case on objects alone. For example, in the introduction to her well-known account of DOM cross-linguistically, Aissen (2003b, p. 435) provides one of the more restrictive definitions of DOM in the literature, limiting it to cases in which "languages with overt case-marking of direct objects … mark some objects, but not others, depending on semantic and pragmatic features of the object." This definition excludes any morphosyntactic variation that does not involve alternation between overt and null case on the object (such as verb agreement), as well as any conditioning factor besides semantic and pragmatic features of the object (such as clausal features). Take the following pair from Catalan (Indo-European),[1] for example, in which the accusative case marker $a$[2] marks O in (3a) but not in (3b).

(3)  (Adapted from Aissen, 2003b, pp. 451-2, citing Comrie, 1979, p. 15)

    a. No m'-havien    vist **a   mi**.

       NEG CLI-they.have seen **ACC 1Sg**

       'They had not **seen me**.'

    b. No havien    vist **l'-alcalde**.

       NEG they.have seen **the-mayor**

       'They had not seen **the mayor**.'

---

[1] All language classification information from Gordon (2005): http://www.ethnologue.com/.

[2] As was the case for -*pe* in Chapter 1, my labeling of *a* as a case marker is not intended to be a commitment to the syntactic analysis of *a* as a case marker rather than a postposition when it marks objects. I treat object case- and adposition-marking as essentially identical for the purposes of this discussion.

Since this alternation involves object case-marking vs. null marking, conditioned by whether the object is a personal pronoun (according to Aissen, 2003b), Catalan fits Aissen's definition of a differential object marking language.

However, not all definitions of DOM include overt vs. null case alternation alone. Bossong (1983–1984, p. 8), for example, characterizes DOM as "the subcategorization of direct objects, or, more precisely, of transitive patients ($Z_2$),[3] depending on the semantic properties of the object noun phrase." I am not sure exactly what morphosyntactic restrictions this definition imposes on which kinds of alternations may be considered DOM. It nevertheless does not overtly appeal to an overt/null case alternation as Aissen (2003b) does, and might therefore include other morphosyntactic alternations beyond the presence or absence of an object case marker. Note that his restriction on the underlying conditions is similar to that of Aissen, namely, semantic/pragmatic properties of O. Schwenter and Silva (2002, p. 578) state their definition of DOM this way: "the phenomenon of distinct realization of different types of direct objects." This definition (as I understand it) encompasses a broader range of formal phenomena than that of Aissen, such as, eg., the presence vs. absence of an object agreement affix on the verb (at least such alternations are not explicitly ruled out). The inclusiveness of this definition is consistent with the position of Morimoto (2002), who includes such overt vs. null verbal head agreement with the object in her characterization of DOM. She does this on the basis of evidence that the same factors underly overt/null agreement in the Bantu languages she surveys and overt/null case in languages traditionally analyzed as exhibiting DOM. The following pair from Chicheŵa (Niger-Congo) exemplifies this:

(4)  (Adapted from Morimoto, 2002, p. 295, citing Bresnan and Mchombo, 1987, pp. 744-5)

a. Njûchi zi-ná-lúm-a        alenje.
   bees   SM-PST-bite-IND hunters
   'The bees bit the hunters.'

b. Njûchi zi-ná-**wá**-lúm-a        alenje.
   bees   SM-PST-**OM**-bite-IND hunters
   'The bees bit them, the hunters.'

The sentence in (4b) contains a verb with the object agreement prefix *wá*, which does not appear in the otherwise identical counterpart in (4a). Regardless of what the distributional constraints on the appearance of this agreement prefix are, Chicheŵa does not count as a DOM language according to Aissen's

---

[3]His notational equivalent of "O."

definition, since it doesn't not exhibit overt/null object case alternation. However it does fit the definition of DOM in Schwenter and Silva (2002), so long as "realization of ... direct objects" is taken to include object agreement . The inclusiveness of Schwenter and Silva's definition is also well adapted to their own data, which show that the contrast between overt and null objects in spoken Brazilian Portuguese is based on similar conditions to the contrast between overt and null object case in Spanish, a language with DOM.[4] Notice also that Schwenter and Silva's definition makes no appeal to distributional constraints (in contrast to Aissen, who defines DOM such that it must be conditioned by properties of the object). Other definitions in the literature, such as those of Næss (2004) and García García (2005), similarly do not constrain the distributional properties of DOM.

Though neither De Swart (2007) nor Malchukov (2008) give an explicit definition of DOM, both of them include alternations between two different overt cases in their discussions of the phenomenon, which Aissen (2003b) and García García (2005), by only appealing to overt and null case-marking, seem to rule out as relevant morphosyntactic contrasts to DOM. Finnish (Uralic) is an example of a language with such a system (Chesterman, 1991). In the Finnish examples below, the object in (5a) is marked by the accusative case, while the object in (5b) is marked by the partitive case, both of which are overt.

(5)  (Adapted from Chesterman, 1991, p. 92)

    a.  Henry rakensi talo-**n**.

       Henry built    house-**ACC**

       'Henry built a/the house.'

    b.  Henry rakensi talo-**a**.

       Henry built    house-**PART**

       'Henry was building a/the house.'

Thus Finnish would be considered a differentially object marking language by de Swart (2007) and Malchukov (2008), but not by Aissen (2003b).

In one of the less restrictive definitions in the literature, Lima (2006) includes case-marking, verbal head agreement, object incorporation, verbal (in)transitivity marking, antipassivization, word order variation, clitic doubling, object overtness variation, and passivization as morphosyntactic alternations that fall under the umbrella of DOM.[5] Thus Lima (2006) labels some languages as differentially object marking

---

[4]As far as I know, regardless of definition, this is a universally agreed-upon characterization of Spanish (see Schwenter & Silva, 2002; Bossong, 1991; Weissenrieder, 1990; Lima, 2006; Aissen, 2003b; Torrego Salcedo, 1999; Leonetti, 2004; von Heusinger & Kaiser, 2005, and many others).

[5]Although she calls these Differential Object Behaviour (DOB) in Lima (2003).

that would not qualify as such under the definition in Aissen (2003b). The following are examples of a few of them. All are adapted from Lima (2006), with the sources she cites presented for each.

(6) *Object Incorporation* (Kusaiean; *Austronesian*; Hopper and Thompson, 1980, p. 259)

    a. nga ɔl-læ        **nuknuk** ɛ

       I    wash-COMPL **clothes  the**

       'I wash **the clothes**.'

    b. nga owo  **nuknuk** læ.

       I    wash **clothes**  COMPL

       'I wash **clothes**.'

(7) *Verbal (In)Transitivity Marking* (Selayarese; *Austronesian*; Finer, 1997, pp. 679-80)

    a. ku-alle-i   doeʔ-iñjo

       1Sg-take-3 money-the

       'I took the money.'

    b. **(a)ng**-alle-kang doeʔ

       **INTR**-take-1Pl  money

       'We took (some) money.'

(8) *Clitic Doubling* (Macedonian; *Indo-European*; Haspelmath, 2001, p. 57)

    a. **Ja**  citam    kniga-ta.

       **it.F** 1Sg.read book.F-the

       'I am reading the book.'

    b. **(*Ja)** citam    kniga.

       **(*it.F)** 1Sg.read book

       'I am reading a book.'

(9) *Word Order Variation* (Yiddish; Indo-European; Diesing, 1997, pp. 389-90)

    a. Max hot **dos bukh** mistome/nekhtn/keyn mol nit geleyent.

       Max has **the book** probably/yesterday/never    read

       'Max has probably/never read **the book** (/yesterday).'

    b. *Max hot **a bukh** mistome/nekhtn/keyn mol nit geleyent.

       Max  has **a book** probably/yesterday/never    read

       Intended: 'Max has probably/never read **a book** (/yesterday).'

Counterbalancing her definitional inclusivity in terms of morphosyntax, Lima (2006) is very restrictive in terms of the permitted distributional properties of DOM. She limits the qualifying alternations to those conditioned by animacy, definiteness, or specificity. Thus if any of the alternations exemplified above were conditioned by, say, topicality or GF-assignment ambiguity (see §2.2), they would presumably fail to qualify as DOM.

The space of variation in the definition of DOM has led to some misunderstanding between scholars. For example, as I pointed out above, Malchukov (2008) seems to include alternation between two types of overt case in his characterization of DOM, while Aissen (2003b) does not, defining DOM instead as alternation between overt and null case. But Malchukov does not take this definitional difference into account in his critique of Aissen. He points out that her theory of DOM, which is in part based on an appeal to the need to differentiate A and O efficiently in a transitive clause, fails to capture alternation between overt cases. The reason for this is that, unlike null case, both overt cases differentiate the object from the subject equally well, i.e. neither alternant is preferred according to a differentiation hypothesis, and no predictions are made. However, as is clear from her definition, Aissen does not set out to account for alternation between overt cases in the first place. One may of course argue that such an alternation *should* be included in the definition of DOM, but Malchukov does not fault Aissen for improperly defining DOM, but rather for failing to predict certain alternations. Thus the critique as it stands is of a straw man.

All this definitional variance is perhaps because "DOM" is essentially a synthetic category — as opposed to a directly observable natural language phenomenon — which is of necessity defined cross-linguistically according to our intuitions about what kinds of properties alternations need to have in order to qualify. To the extent that those intuitions differ from scholar to scholar, the definition of DOM is essentially a matter of taste. Nevertheless, these definitional differences complicate the effort to develop a typology of DOM.

Despite the subjectivity inherent in defining DOM and comparing definitions of it, some approaches nevertheless seem to lead to counterintuitive results. For example, while some of the definitions provided above constrain the qualifying distributional conditions on DOM, I am not sure whether it is desirable to do so. That is, I am not convinced that it is intuitively satisfying to commit oneself to the position that a language which exhibits a qualifying morphosyntactic alternation but non-qualifying distributional properties does not, in fact, exhibit DOM. For example, Aissen (2003b) only allows null/overt case alternations conditioned by properties of the object to count as DOM. Taking Spanish as a starting point, many instances of object *a*-marking variation (those based on definiteness/specificity, animacy, topicality, etc.)

would be DOM. However, García García (2005) points out that telicity also plays a role in *a* marking, to the effect that certain telic propositions may only realize an *a*-marked (not an unmarked) DO.[6] Since telicity is not an object property, Aissen's definition commits us to the position that telicity-related *a*-marking alternations are not DOM, while, e.g., specificity-related *a*-marking alternations are. While this is certainly a *possible* position to take, it does not accord with my own intuition, at least, that DOM should include *all* *a*-marking variations in Spanish, and not just the ones that have nothing to do with telicity, or other non-object properties like relative animacy/definiteness or ambiguity (to be discussed in §2.2.3 and §2.2.4, respectively). Furthermore, Aissen gives no indication that she excludes the latter alternations from consideration in her own study of Spanish DOM. Similarly, as discussed above, Bossong (1983–1984) limits the conditions on DOM to "semantic properties of the object" (p. 8). Thus he would also need to exclude not only telicity-based *a*-marking, but also topicality-based *a*-marking (see Leonetti, 2004), since this is a discourse, not semantic, property of objects. Again, to me this is not intuitively desirable. These counterintuitive consequences are due to the building of distributional constraints into the definition of the phenomenon itself. Perhaps a good way of avoiding them would be to limit the definition of DOM to the domain of morphosyntax, and leave the question of distribution to subsequent analysis.

As shown above, a frequent intuition that appears in the literature is that looking strictly at alternation between overt and null case-marking on objects will probably result in missed generalizations, and that at least some of the aforementioned morphosyntactic alternations are different manifestations of something which is, at some level, a single phenomenon.[7] To the extent that we find it desirable to unify our analysis of these alternations, we should formulate a definition of DOM that includes them. For reasons of simplicity and generality, it is preferable to execute this inclusion in terms of principled criteria according to which alternations may be included/excluded, rather than a disjunctive list of alternations such as that provided by Lima (2006).

In light of all this, I do not intend to muddy the waters by proposing a definition of my own. As the reader can verify, the object marking system in Guaraní exemplified in (1) qualifies as DOM according to even the most restrictive definitions listed above: it consists of an alternation between overt and null object case-marking, which, as I will show, is based almost entirely on object properties. In the ensuing brief survey of DOM cross-linguistically and its conditioning factors, I assume overt/null case-alternation to be a sort of "prototypical" DOM, since it is not excluded under any definition, and discuss its mani-

---

[6]I discuss García García (2005) and telicity more thoroughly in §2.2.6.

[7]For instance, while Aissen (2003b) is very restrictive about what kinds of morphosyntactic alternations she labels "DOM," she includes a broader range of phenomena under the label "Differential Coding" in Aissen (2003a).

festations without comment. Whenever other potentially DOM-related alternations (i.e. verb agreement, object incorporation, etc.) are under discussion, I note them in light of the fact that considering such alternations to be DOM does not meet with universal acceptance. To the extent that the reader disagrees with the classification of a certain alternation as DOM, the discussion of it can be ignored. Nothing in this analysis depends crucially on an inclusive definition.[8]

## 2.2   DOM Cross-Linguistically: Some Descriptive Generalizations

Bossong (1985a) claims he has found DOM to be active in over 300 languages around the world, occupying diverse language families. An increasingly large literature has appeared in response to the question of the distributional conditions on DOM within and across these languages, and a variety of conditions have been identified. The two overwhelmingly prevalent underlying factors appear to be the animacy and definiteness of the O-argument (see Comrie, 1989). As an example of animacy-related DOM, take the following examples from Kannada (Dravidian).

(10)   (de Swart, 2007, pp. 178-9, citing Lidz, 2006, p. 11)

   a.   Naanu **sekretari(-yannu)\*** huDuk-utt-idd-eene.

        I.NOM **secretary(-ACC)\***    look.for-NPST-be-1Sg

        'I am looking for **a secretary**.'

   b.   Naanu **pustaka(-vannu)** huDuk-utt-idd-eene.

        I.NOM **book(-ACC)**        look.for-NPST-be-1Sg

        'I am looking for **a book**.'

As exemplified in (10), animate objects in Kannada must take an accusative marker, while inanimate objects may be unmarked (de Swart, 2007). An example of a language with definiteness-based DOM is Mandarin Chinese (Sino-Tibetan),[9] which has the object marker *ba*.[10]

---

[8]However, I note in Chapter 6 that an interesting area of future research might be the distribution of null vs. overt objects and incorporated vs. non-incorporated objects in Guaraní. While variation in case-marking, variation in object overtness, and variation in object incorporation have been discussed for various languages in the DOM literature, they have not, to my knowledge, been explored *within a single language* that exhibits them all. A thorough examination of whether the conditions underlying all of these phenomena are related could therefore prove quite revealing. If in fact the distributional constraints on these three types of object coding are similar/related to each other, it might motivate a definition of DOM that includes them all.

[9]As de Swart (2007, p. 183) shows, DOM in Mandarin is also a function of animacy. The examples provided are simply to illustrate how definiteness can interact with DOM, and thus only highlight the definiteness aspect of DOM in Mandarin.

[10]There is difference of opinion in the literature on Mandarin as to the syntactic status of *ba* (van Bergen, 2006). I, like de Swart (2007), refer the reader to Yang and van Bergen (2007) in justifying the position that it is an object case marker.

(11)   (Adapted from De Swart, 2007, p. 183, citing van Bergen, 2006)

   a.  Ta **(ba)** na-ge    pingguo chi le.

       He **(OM)** that-CL apple     eat PRT

       'He ate that apple.'

   b.  Ta **\*(ba)**  yi-ge pingguo chi le.

       He **\*(OM)** a-CL apple     eat PRT

       'He ate an apple.'

The object in (11a) has a definite determiner *na* 'that' and may be realized with or without the object

marker *ba*. However, the object in (11b) has an indefinite determiner and obligatorily takes *ba*.[11] Thus,

Kannada exhibits an animacy-based DOM alternation while Mandarin exhibits a definiteness-based DOM

alternation.

Yet it has been fairly well-established that animacy and definiteness alone are insufficient to account

for the full range of DOM cross-linguistically, and a several additional factors have been identified for par-

ticular languages (see de Swart, 2007; Lazard, 1982; Leonetti, 2004; García García, 2005; Tippets & Schwen-

ter, 2007; Shokouhi & Kipka, 2003; Mardale, 2008, among many others). In this section, I provide a sum-

mary of the literature on each of several major factors underlying DOM cross-linguistically, along with

data from a sample of languages to which they pertain.

### 2.2.1   Animacy

Many studies related to DOM appeal to a non-linguistic conceptual hierarchy on which entities in the

universe are ranked according to their aliveness/sentience/capacity for independent motion, i.e. how

"animate" they are (Silverstein, 1976; Comrie, 1979, 1989; Aissen, 2003b; de Swart, 2007; von Heusinger &

Kaiser, 2003; de Hoop & Malkuchov, 2008, among many others). It is usually talked about as three-tiered,

such as in (12), where ">" means "greater in animacy than":

(12)  **Animacy Hierarchy**       Human > Animate[12] > Inanimate

---

[11]As de Swart (2007) notes, and as the examples in 2.2.2 will show, Mandarin appears to be opposite in this regard from most languages in which definiteness is relevant to DOM: elsewhere, more definite DOs take overt marking and less definite DOs take null marking (see §2.2.2).

[12]Throughout this study, "animate" is generally used to refer to things that are [+animate] and [–human], despite the fact that humanness technically implicates animacy.

Thus people are more animate than iguanas, which in turn are more animate than rocks.[13] Some include intermediate categories. Bossong (1991), for example, posits a category "pers" between Human and Animal for personified animals. Others, such as von Heusinger and Kaiser (2003), put anthropomorphized animals in the [+human] category. The cross-linguistic pattern seems to be that the higher an object falls on the Animacy Hierarchy, the more likely it is to be marked (Aissen, 2003b).

A variety of languages exhibit DOM conditioned at least in part by divisions at various points along the Animacy Hierarchy, i.e. marking [+human] objects and not [-human], marking [+animate] objects and not [–animate] objects, marking [+human] objects more frequently than [–human] objects, etc. One of these languages is Hindi (Indo-European), which, for the most part, requires human objects to be marked with the postposition *-ko*, while non-human objects may be realized without *-ko* (Comrie, 1989; de Swart, 2007; Mohanan, 1990). In the first example below, the human object *bacce* 'child' may not appear without *-ko*-marking, while the inanimate object *patr* 'letters' in (13b) may.

(13) a. (Adapted from de Swart, 2007, p. 127, citing Mohanan, 1990, p. 103)

      Ilaa-ne  bacce-**ko**/*baccaa uthayaa

      Ila-ERG child-**OM**/*child   lift.PERF

      'Ila lifted the/a child.'

    b. (Comrie, 1989, p. 133)

      Ye    patr   paṛhie.

      these letters read.POL

      'Please read these letters.'

Sardinian (Indo-European) displays a similar animacy-related markedness contrast, where human objects must be case-marked and non-human objects cannot be (Mardale, 2008).

(14) a. An       furatu *(**a**)   Ercole.

       3Pl.have steal   *(**OM**) Hercules

       'They have stolen Hercules.' (Adapted from Mardale, 2008, p. 450)

---

[13] I should note that not all treatments of animacy look exclusively at the hierarchy in (12). Comrie (1989), for example, labels (12) "literal animacy" and goes on to put other properties under the umbrella of (extended) "animacy," namely person (1st, 2nd, and 3rd), pronoun vs. non-pronoun, and proper name vs. common noun. Although he notes that none of these distinctions intuitively have anything directly to do with being sentient/alive, he includes them in his definition of "animacy" because they, like "literal animacy," condition case marking alternations cross-linguistically. Given this motivation, one might wonder whether Comrie's extended definition of "animacy" is actually a cover term for "properties that condition case marking alternations," which would be something of a departure from the way the term is generally used. Regardless, since (12) is very common in the literature, (12) is what I refer to by the term "animacy."

    b.  Appo    vistu (\***a**)  su    cane.

        1Sg.have see    (\***OM**) 2.POSS dog

        'I saw your dog.' (Adapted from Mardale, 2008, p. 458)

In (14a), the object *Ercole* 'Hercules' is human, and cannot appear without the object marker *a*. In (14b), meanwhile, the object *su cane* 'your dog' is animate (and non-human), and object marking is disallowed. Thus animacy can be seen to play a role in determining the licensing of object marking in Sardinian.

    Animacy-based DOM appears in Romanian as well (Klein, 2007). Animate-referring proper names must be marked, while inanimate-referring ones cannot be:[14]

(15)    a.  (Adapted from Klein, 2007, p. 1)

        L=am        lovit **pe**  Mihai.

        ACC.M=1.have hit   **OM** Mihai

        'I/we have hit Mihai.'

    b.  \*L=am        lovit Mihai.

        ACC.M=1.have hit   Mihai

        Intended: 'I/we have hit Mihai.'

(16)    a.  (Adapted from Klein, 2007, p. 1)

        \*Udo (l=)a        cumpărat **pe**  Mercedes-ul.

        Udo  ACC.M=3.have buy     **OM** Mercedes-DEF.M

        Intended: 'Udo has bought the Mercedes.'

    b.  Udo a       cumpărat Mercedes-ul.

        Udo 3.have buy      Mercedes-DEF.M

        'Udo has bought the Mercedes.'

The (a) examples above contain a marked DO, and only the human-referring name in (15a) licenses object marking. Similarly, the (b) examples contain an unmarked DO, and only the inanimate-referring name in (16b) licenses non-marking. Animate-referring common NPs are optionally marked, as shown below.

---

[14]Clitic doubling co-occurs with object marking. The clitic is co-referential with the object and is realized on the matrix verb (see Klein, 2007).

(17)    a.    (Adapted from Klein, 2007, pp. 1-2)

        L=am            lovit **pe**   copil-ul          vecin-ului.

        ACC.M=1.have hit    **OM** child-DEF.M neighbor-DEF.M.GEN

        'I/we have hit the neighbor's child.'

    b.    Am      lovit copil-ul          vecin-ului.

        1.have hit    child-DEF.M neighbor-DEF.M.GEN

        'I/we have hit the neighbor's child.'

Inanimate-referring common NPs may not be marked, however:

(18)    (Adapted from Klein, 2007, p. 2)

        Am      reparat  (\***pe**)  dulap-ul.

        1.have repaired (\***OM**) wardrobe-DEF.M

        'I've repaired the wardrobe.'

In Russian, masculine singular objects belonging to the first declension take the accusative case marker *-a* only if they are animate; inanimates of this declension have null marking (Comrie, 1989). This is shown by the following:

(19)    (Adapted from Comrie, 1989, p. 132)

        Ja    videl mal'čik-**a**/begemot-**a**/dub/stol

        1Sg see    boy-**ACC**/hippopotamus-**ACC**/oak/table

        'I saw the boy/hippopotamus/oak/table.'

In the above example, the O-arguments referring to the animates "boy" and "hippopotamus" take overt case-marking, while the O-arguments referring to the inanimates "oak" and "table" do not.

The Makú language Hup also exhibits animacy-related DOM (Epps, 2008). Take the following by way of example. In (20a) the object is human-referring, and must take the object marker *ăn*. In (20b) the object is animate non-human, and is optionally marked. In (20c) the object is inanimate, and cannot be marked.[15]

---

[15]For a more nuanced analysis of DOM in Hup, which takes into account the intriguing interaction between plurality and object marking, see Epps (2008).

(20)  (Adapted from Epps, 2008, pp. 175-6)

    a.  cet-ham-tubud-yíʔ-ɨ́y,        hup=ʔɛ́y-**ǎn**

        carry.on.back-go-3.INTS-TEL-DYN person=woman-**OM**

        'He carried the girl a long way off!'

    b.  mɔ̌**(-ǎn)**=mah      cɛ́p  tɨh hitɛ̃́ʔ-ɛ̃́h

        inambu**(-OM)**=REP other 3Sg imitate-DYN

        'The inambu bird is another that he imitates.'

    c.  děh  hi-wǎy     hám=d'əh yúp,    hɔ̃pkɔ̌k**(\*-ǎn)**,  mɔm-b'ɔ́k**(\*-ǎn)** hɨd tɔn-hám-ɛ́h

        water FAC-go.out go=PL    that.ITG fish.pull**(\*-OM)** iron-pot**(\*-OM)** 3Pl hold-go-DECL

        'Those who go out in the flooded forest to fish, they take along fishhooks and pots.'

In her discussion of object coding in Bantu, Morimoto (2002) takes DOM to include the presence vs. absence of verbal head agreement with the object. Assuming this definition, there are several examples of Bantu languages with DOM. One is Kiswahili, which requires overt head agreement with animate objects but allows both overt and null agreement with inanimate objects (Morimoto). This is exemplified by the following (see (32) for an example of marking variation on inanimates):

(21)  (Morimoto, 2002, p. 296, citing Vitale, 1981, pp. 123-4)

    a.  Juma a-li-**m**-piga    risasi tembo  jana     usiku.

        Juma SM-PST-**OM**-hit bullet elephant yesterday night

        'Juma shot an/the elephant last night.'

    b.  risasi i-li-piga    mti karibu na sisi.

        bullet SM-PST-hit tree near   us

        'The bullet struck the tree near us.'

Animacy also bears on object marking in Spanish, possibly the most thoroughly documented language with respect to DOM (Rodriguez-Mondoñedo, 2007, p. 89). In general, inanimate-referring objects may not be marked, while human-referring objects must be.

(22)  (Adapted from von Heusinger & Kaiser, 2003, p. 41)

    a.  Vi        \*(a)    la/una    mujer.

        1Sg.PST.see \*(OM) DEF/INDEF woman.

        'I saw the/a woman.'

b. Vi      (\*a)    la/una     mesa.

     1Sg.PST.see (\*OM) DEF/INDEF table.

     'I saw the/a table.'

In (22a), which contains the human-referring direct object *mujer* 'woman,' omission of the object marker *a* is ungrammatical, while in (22b), which contains the inanimate-referring direct object *mesa* 'table,' the presence of *a* is ungrammatical.

### 2.2.2   Definiteness and Specificity

A number of studies appeal to the concepts of definiteness and/or specificity in accounting for the distribution of object marking (Comrie, 1989; Aissen, 2003b; Leonetti, 2004; Shokouhi & Kipka, 2003; Lazard, 1982; de Swart, 2007; Morimoto, 2002; de Hoop & Malkuchov, 2008, to name a few). To briefly exemplify the intuitions underlying these terms, there is a definiteness contrast between the objects of the following sentences, exhibited by the appearance of a definite vs. indefinite article.

(23)   I ate an apple.

(24)   I ate the apple.

What exactly the difference between (23) and (24) consists of is a subject of debate. Some might claim that the apple in (24), unlike its counterpart in (23), takes a definite article because it is unique and identifiable within some relevant discourse context (Russell, 1905; Lyons, 1999), while others might view the appearance of the definite article in (24) as a result of its being familiar to those involved in the discourse (Heim, 1982; Kamp, 1981) — which of these is the more correct characterization (or neither, see Fraurud, 1990) is not yet settled. With respect to DOM, in some languages the object of (24) would prefer marking more strongly than the object of (23).[16] Specificity has to do with the *de dicto* vs. *de re* distinction, exemplified by the following.

(25)   John wants to marry a Norwegian.

There are two interpretations of (25), a specific (*de re*) one in which John has a certain Norwegian in mind that he wants to marry, and a non-specific (*de dicto*) one in which John would be perfectly satisfied to marry anyone so long as that person is Norwegian. The two interpretations are identical with respect to definiteness, regardless of one's conception of it: in either case, the Norwegian is neither familiar nor

---

[16]With the exception of Mandarin, as I mentioned before, which exhibits the opposite pattern, as shown in (11). I know of no other DOM language with similar "inverted" DOM patterns with respect to definiteness.

uniquely identifiable; i.e., the expression is indefinite (as shown by the determiner). A specific reading of indefinites in English can be forced by the use of adjectives like *specific, particular*, or *certain* (Enç, 1991; von Heusinger, 2003). For example, in (26) only the specific (*de re*) reading discussed above is available, and the nonspecific (*de dicto*) one is ruled out.

(26)   John wants to marry a $\left\{ \begin{array}{c} \text{specific} \\ \text{particular} \\ \text{certain} \end{array} \right\}$ Norwegian.

In DOM systems sensitive to specificity, specific objects are in general more amenable to object marking than non-specific objects (Aissen, 2003b).

There is a vast literature on both of these topics (and their relationship), which dates back at least to Russell (1905) and includes Christopherson (1939); Strawson (1950); Kamp (1981); Heim (1982, 1983); Prince (1981); Lyons (1999); Kadmon (1990); Kartunnen (1976); Roberts (2003); C. L. Baker (1966); Enç (1991); Matthewson (1999); van Geenhoven (1998b); van Geenhoven and McNally (1998, 2005); Kratzer (1998); Partee (1972); Hintikka (1973); Geach (1962); Fodor (1970); Lewis (1979); Stockwell, Schachter, and Partee (1973); von Heusinger (2003); Bleam (1999); Dobrovie-Sorin (1997); Givón (1978), and the list goes on. As the various arguments in the above publications indicate, the question of what makes something definite/indefinite or specific/nonspecific has not yet been resolved.

For the purposes of this thesis, I assume a *familiarity*-based approach to definiteness. That is, a nominal expression is definite if refers to a discourse referent that is familiar to the discourse participants, and indefinite otherwise (Heim, 1983; Kamp, 1981). Following Roberts (2003), I assume that definite expressions can exhibit either "strong familiarity," such that the referent of an expression has been explicitly mentioned in previous discourse, or "weak familiarity," such that the referent of an expression is entailed by the discourse context (cf. "inferability" from Prince, 1981). Others studying DOM, such as von Heusinger and Kaiser (2003), make similar assumptions.

Regarding definiteness in the DOM literature, a number of studies simply take definiteness to be a binary contrast between "definite" and "indefinite" (Schwenter & Silva, 2002; Leonetti, 2004; Enç, 1991), while others take it to be a scalar phenomenon (Aissen, 2003b; von Heusinger & Kaiser, 2007; Mardale, 2008). Of the latter approaches, Aissen's Definiteness Hierarchy is one of the more frequently cited. I provide it below, where ">" means "more definite than":

(27)  Personal pronoun > proper name > definite descriptive NP > indefinite specific NP > nonspecific NP

*abbreviated:*

Pro > PN > Def > Spec > NSpec (Aissen, 2003b, pp. 437,444)

Aissen (2003b) presents (27) as an expansion on an earlier definiteness hierarchy assumed in studies like Comrie (1986) and Croft (1988) (definite > specific > nonspecific), empirically motivated by the fact that DOM in some languages is contingent on pronominality and/or the distinction between proper names and descriptive NPs, distinctions that cannot be captured by the previous definiteness hierarchy (cf. the Romanian examples in §2.2.1). De Swart (2007) notes that (27) can be seen as a conflation of multiple feature rankings: definiteness (+ > −), specificity (+ > −), and "NP type" (Pro > PN > descriptive NP). Proposing such a hierarchy of multiple discrete semantic, pragmatic, and morphosyntactic properties would ideally be accompanied by some justification for why all and only these diverse features are included. It may at first seem to run the risk of circularly defining "definiteness" as a simple list of the various properties that underly object marking alternations and justifying that definition on the grounds that said alternations are based on it. However, I believe Aissen's Definiteness Hierarchy survives these objections for at least two reasons.

First, (27) is not merely a list of relevant conditions, but a hierarchy: a ranking of what are claimed to be definiteness-related factors. Thus, even though *pronoun* is a formal category and *nonspecific NP* is a both a formal and semantic/pragmatic one, if it is true that pronouns are systematically selected for marking over nonspecific NPs typologically, this supports the treatment of both categories as belonging to the same scale (of "definiteness" or something like it).

Second, as Aissen, 2003b notes parenthetically, (27) corresponds closely (though not exactly) to the implicational hierarchy of "givenness" proposed by Gundel, Hedberg, and Zacharski (1993) under entirely independent motivations: explaining the distribution of various forms of referring expression. Their Givenness Hierarchy, provided below, is a ranking of *cognitive statuses* that a given discourse referent may have. A referent that meets the requirements of a cognitive status on the hierarchy meets the requirements of all lower statuses, though not necessarily vice versa. Thus the set of referents of a particular level of givenness is a subset of the set of referents of all lower levels of givenness:

(28)  in focus > activated > familiar > uniquely identifiable > referential > type identifiable

Without going into the details of what these categories mean, beyond their intuitive labels, I would like to note that the results from Gundel et al. (1993) lend credence to the assumptions in Aissen (2003b). Gundel et al.'s findings about the reflections in linguistic form of the above statuses cross-linguistically support Aissen's judgment that pronominals should be ranked over descriptive definites: pronouns were cross-lingusitically "in focus," while descriptive definites ranged from "uniquely identifiable" to "activated." Though Gundel et al. do not explore the distribution of proper names, the ranking in (28) intuitively co-incides with Aissen's ranking of pronouns and proper names, since pronouns seem to be preferable to proper names for referents that are highly activated/in focus.[17] It is not clear to me how definite descriptions and proper names might rank with respect to each other according to (28), and if givenness is to be taken as justification for (27) this would have to be spelled out. However, it appears plausible that the single feature being tapped into by (27) is givenness/cognitive status/discourse activation, which would further motivate it as a basis for a theory of DOM.[18] Given the robustness of (27) in accounting for object marking cross-linguistically and the fact that a variety of analyses of DOM have appealed to it, I make use of it in my own analysis as well (see §3.2.2.2).

Like definiteness, the nature of specificity has been characterized in many different ways in the literature. Fodor (1970) and Fodor and Sag (1982) attribute specificity distinctions to quantifier scope. On that view, specific readings arise when the existential quantifier included in the meaning of the indefinite expression takes wide scope with respect to other operators in the clause, and non-specific readings arise when the quantifier takes narrow scope. Reinhart (1997), Kratzer (1998), and Matthewson (1999) analyze specificity effects as arising from choice functions from the set denoted by the descriptive NP to a particular individual within that set. Donellan (1966), Partee (1972), Hintikka (1973), and Ioup (1977) take the position that specificity distinctions result from whether the speaker has some particular individual in mind (specific) or not (non-specific). Enç (1991) analyzes specifics as belonging to a [+definite] (in the sense of "familiar") superset. Von Heusinger (2003) views specificity as "referential anchoring," that is, being functionally dependent on the referent of another expression. There are other analyses as well. Each of these

---

[17]Note the significant difference in naturalness between the following two examples:

(1) Boaz loves toads. ??Boaz freaks out whenever he sees one.

(2) Boaz loves toads. He freaks out whenever he sees one.

The very recent mention of the referent of *Boaz* presumably results in a high level of activation for that referent. Given this, notice that the pronoun *he* is much more acceptable in the following clause than a PN, suggesting that PNs are lower in givenness than pronouns.

[18]This would essentially amount to grounding the hierarchy in the "real world" of cognition. Such a maneuver would diffuse the critique of Carnie (2006), who takes issue with theories of grammar that appeal to hierarchies as primitives which are not "properly grounded either in the real world or in theoretical constructs that themselves have some grammatical status" (Carnie, p. 9).

approaches has its advantages/disadvantages, and I cannot survey them here. For the purposes of this study, I follow Donellan and others in assuming that specific expressions are those for which the speaker has a particular referent in mind (see §3.2.2.2). Under this approach, (25) is specific if the speaker has a particular Norwegian in mind and is asserting that John wants to marry that person, and non-specific if the speaker has no particular Norwegian in mind.[19]

Before turning to the DOM data, let me make some final comments about the relationship between specificity and definiteness. Notice that the Definiteness Hierarchy in (27) only registers specificity contrasts within the class of indefinite NPs. This requires one of two assumptions: either specificity is undefined with respect to definite expressions, in which case it matters only for indefinite expressions (Heim, 1982), or all definite expressions are inherently specific, in which case specificity distinctions only arise in indefinite expressions (Enç, 1991). For this study, which of these two assumptions is adopted does not particularly matter. However, I should note that while adopt (27), it is not universally agreed that definites cannot also exhibit specificity distinctions, i.e. that non-specific definites do not exist. Lyons (1999, pp. 167-8), for example, provides data that he takes to be supportive of the existence of non-specific definites in English. In the DOM literature specifically, some assess definiteness and specificity as independent features (Schwenter & Silva, 2002; Schwenter & Silva, 2003; Tippets & Schwenter, 2007; von Heusinger & Kaiser, 2003). However, Aissen's Definiteness Hierarchy, or something close to it, is adopted by several other works on DOM (Comrie, 1986; Croft, 1988; de Swart, 2007; von Heusinger & Kaiser, 2007; Lima, 2003; Mardale, 2008; Jäger, 2007; Morimoto, 2002; Nazareth, 2007, and others). It is this tradition that I follow here.

---

[19]I should note that the discussion is more complex than what makes something is specific or not. Farkas (1994), for example, observes several different subclasses of specific expression (see also Fodor, 1970). Furthermore, a growing literature is investigating a subclass of indefinite non-specifics that actually do not function as arguments at all. McNally (1995), van Geenhoven (1998b), Farkas (2002), Leonetti (2004), and many others take these indefinites to undergo "semantic incorporation," where their denotation becomes absorbed into the semantics of the verb to create a complex predicate. Such a picture can be used to account for the surprising syntax of the following example from Persian, which exhibits both intransitive morphology on the verb and an overt object.

(1) (Lazard, 1998, p. 94, citing Cartier, 1989, p. 178)

Ali ber-tanam  kopi.
Ali INTR-plant coffee

'Ali is a coffee-grower.'

If *tanam* 'plant' is a semantically incorporating predicate, the construction is accounted for, since the VP as a whole denotes a single property being predicated of Ali, rather than a relation between two arguments, and the verb is semantically intransitive. This is consistent with the observation in Lazard (2003) that objects in such constructions cannot take a determiner and must receive a generic interpretation, as seen in the translation of (1). While interesting, these points are not of crucial relevance to this study of DOM in Guaraní, since object incorporation is not under consideration.

Definiteness is relevant to a number of DOM systems cross-linguistically. One example is that found in Hebrew (Afro-Asiatic). Aissen (2003b, p. 453) provides the following data points:

(29)　a.　Ha-seret　her'a　　*('et-)ha-milxam

　　　　　the-movie showed *(ACC-)the-war

　　　　　'The movie showed the war.'

　　　b.　Ha-seret　her'a　　(*'et-)milxam

　　　　　the-movie showed (*ACC-)war

　　　　　'The movie showed a war.'

The object in (29a) takes the definite determiner *ha-* and is therefore obligatorily marked by the object marker *'et*. Meanwhile, the object in (29b) is indefinite because it lacks a definite determiner, and it cannot be marked. Thus DOM in Hebrew is clearly related to definiteness.

Balochi (Indo-European) also exhibits a definiteness-based DOM alternation (Butt, Ahmed, & Poudel, 2008; Korn, 2009). In Balochi, indefinite DOs may not be marked by the oblique case *aa*, while definite objects may be (Korn). This is exemplified by the following:

(30)　(Butt et al., 2008, p. 3)

　　　a.　aa　　　　man-aa kitab-**aa**　d-aa

　　　　　he.NOM I-OBL　　book-**OBL** give-Sg

　　　　　'He gives me the book.'

　　　b.　aa　　　　man-aa kitab d-aa

　　　　　he.NOM I-OBL　　book give-Sg

　　　　　'He gives me book(s).'

In (30a), the object *kitab* 'book' is marked, and given a definite interpretation. Meanwhile, the object in (30b) is unmarked and is given an indefinite interpretation. Thus definiteness seems to play a role in the distribution of DOM in Balochi.

DOM in Corsican (Indo-European) is also sensitive to definiteness, as the following examples show (Neuberger, 2009):

(31)　(Adapted from Neuberger, 2009, p. 1)

　　　a.　Vigu　　**\*(à)**　　Pedru.

　　　　　1Sg.see **\*(OM)** Pedru

　　　　　'I see Pedru.'

b. Vigu **(*à)** lómu.

1Sg.see **(*OM)** the.man

'I see Pedru.'

In (31a), the PN object requires the object marker *à*. In (31b), on the other hand, the definite (non-PN) object prohibits marking, suggesting that DOM in Corsican is restricted to objects that are PN or higher on the Definiteness Hierarchy.

Kiswahili (Niger-Congo) exhibits definiteness-based DOM as well (Morimoto, 2002). In the following examples, the object with corresponding overt object agreement morphology on the verbal head takes the definite reading shown in (32a), while the object without such agreement takes the indefinite reading shown in (32b):

(32) (Adapted from Morimoto, 2002, p. 298)

a. U-me-**ki**-let-a kitabu?

you-PERF-**OM**-bought-IND book

'Have you bought the book?'

b. U-me-let-a kitabu?

you-PERF-bought-IND book

'Have you bought a book?'

Hindi also exhibits definiteness-related object marking contrasts (Aissen, 2003b).[20] Aissen (2003b, p. 466), citing Mohanan (1990, pp. 87-88), points to the data in (33) by way of example.

(33) a. Ravii-ne kaccaa kelaa kaaṭaa.

Ravi-ERG unripe banana cut

'Ravi cut the/an unripe banana.'

b. Ravii-ne kaccaa kele-**ko** kaaṭaa.

Ravi-ERG unripe banana-**ACC** cut.PERF

'Ravi cut the unripe banana.'

The object in (33a) has no object marker *ko*, and it can be given either a definite or an indefinite reading. The object in (33b), on the other hand, is *ko*-marked, and it only takes a definite reading. Thus, in these examples, object marking is compatible only with a definite interpretation.

---

[20]Although de Hoop and Malkuchov (2008) take this Hindi contrast to be one of specificity rather than definiteness.

Nazareth (2007) claims that Tigrinya (Afro-Asiatic) exhibits a definiteness-based contrast in DOM on the basis of the fact that definite objects take overt agreement morphology on the verb, while indefinite objects do not:

(34)  (Nazareth, 2007, p. 9)

    a.  lamɨ    biʕɨrayɨ  riʕiy-a
        Sg.F.cow Sg.M.bull PERF.see-SM.3Sg.F
        'A cow saw a bull.'

    b.  ʔit-a      lamɨ    n-ät-i        biʕɨrayɨ  riʕiy-a-**to**
        DET-3Sg.F Sg.F.cow OBJ-DET-3Sg.M Sg.M.bull PERF.see-SM.3Sg.F-**OM.3Sg.M**
        'The cow saw the bull.'

In (34a), the object is indefinite, and no agreeing verbal morphology appears. In (34b), the object is definite, and it co-occurs with the verbal object agreement suffix *to*. Thus definiteness plays some role in DOM in Tigrinya.

With respect to the relevance of specificity to DOM, it seems to be the case cross-linguistically that [+specific] objects are better candidates for object marking than [–specific] objects (Aissen, 2003b; Schwenter & Silva, 2002; Leonetti, 2004). An example of a language with specificity-based DOM is Turkish (Altaic; Enç, 1991). This is exemplified by the following:

(35)  a.  (Enç, 1991, pp. 4-5)

        Ali bir   piyano kiralamak istiyor.
        Ali one piano   to-rent     wants
        'Ali wants to rent a (non-specific) piano.'

    b.  Ali bir   piyano-**yu**  kiralamak istiyor.
        Ali one piano-**ACC** to-rent     wants
        'Ali wants to rent a **certain** piano.'

In the minimal pair above, the second example contains an accusative-marked object, which requires a specific reading, while the first is unmarked and prefers a non-specific reading.

Persian (Indo-European) exhibits a specificity-related object marking distinction as well (Lazard, 1982).

(36)  (Adapted from Lazard, 2003, p. 2)

    a.  ketăb-i        xănd-am

       book-INDEF read-1Sg

       'I read a book.'

    b.  ketăb-i-**ră**        xănd-am ke...

       book-INDEF-**OM** read-1Sg COMP...

       'I read a **(certain)** book which....'

A specific reading arises in (36) only in the object marked example (36b). Example (36a) is both unmarked and non-specific.

Under the assumption that DOM may involve antipassivization, De Swart (2007, p. 139) presents the following data from van Geenhoven (1998a, p. 232) exhibiting a specificity-based coding contrast in West Greenlandic (Eskimo-Aleut):

(37)  a.  Nuka-p   **iipili**      neri-v-a-a.

       Nuka-ERG **apple.ABS** eat-IND-3Sg-3Sg

       'Nuka ate **a particular apple**.'

    b.  Nuka     **iipili-mik**   neri-v-**u**-q.

       Nuka-ABS **apple-INSTR** eat-IND-**INTR**-3Sg

       'Nuka ate **an apple**.'

In (37), the reading of the syntactically transitive construction in (37a) is specific, while its antipassivized counterpart in (37b) is non-specific.

García García (2005) provides the following examples from Spanish, in which the *a*-marked DO takes a specific reading and the unmarked DO does not. Thus, (38a) is an appropriate description of one's actions in the search for a lost secretary, while (38b) would be appropriate text for a classified ad soliciting secretarial services.

(38)  (Adapted from García García, 2005, p. 17)

    a.  Busco        **a   una   secretaria**.

       1Sg.PRES.search **OM INDEF secretary**

       'I am looking for **a (certain) secretary**.'

b. Busco **una** **secretaria**.

1Sg.PRES.search **INDEF secretary**

'I am looking for **a secretary**.'

The examples presented in this section demonstrate the cross-linguistic basis for positing a relationship between definiteness/specificity and object marking.

### 2.2.3 Relative Animacy and Definiteness

In light of the above discussion, I must note that it is not uncommon in the literature to find the argument that it is actually the *relative* animacy and/or definiteness of the object with respect to the subject rather than the absolute animacy or definiteness of the object alone that underlies DOM, language-specifically for some authors and typologically for others. Comrie (1989) posits relative animacy/definitess as typological conditions on DOM, as expressed in the following passage:

> Although in principle either of A and P[21] 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. (Comrie, 1989, p. 128)

Unfortunately, it is unclear what exactly is meant by "information flow" and in what sense there is a "strong tendency" for this flow to correlate with relative animacy/definiteness. As Næss (2004) points out, this appears to be an argument from frequency, yet no frequency data are provided. It is nevertheless fairly clear that the claim being made about language generally is that configurations in which S is less animate or definite than O are marked cases, which favor object marking. Whether this holds as a universal is an open empirical question, yet there is some evidence that, for at least some languages, the relevant computation with respect to DOM is of the relative properties of S and O, not of O alone. Comrie (1989) provides no examples of languages in which such relative distinctions matter to DOM.

Others, however, do. For example, Tippets and Schwenter (2007) found relative animacy to be the most important factor in their corpus study of the conditions on DOM in Spanish. Although inanimate Os are unlikely to be marked at all in Spanish, they were significantly more likely to be marked if A was

---

[21] "A" is Comrie's notation for "agent," which overlaps significantly with the GF "subject," and "P" is Comrie's notation for "patient," which overlaps significantly with the GF "object."

also inanimate. Furthermore, if O was more animate than A, the frequency of marking was considerably higher than if O was equal or lower in animacy than A.

De Swart ([2007](#)) points out that relative animacy is central to DOM in Awtuw (Sepik-Ramu), evidenced by the following:

(39)   (Adapted from De Swart, [2007](#), p. 90, citing Feldman, [1986](#), p. 110)

    a. Tey   tale      yaw d-æl-i.

        3Sg.F woman pig  FAC-bite-PST

        'The woman bites the pig/*The pig bites the woman.'

    b. Tey   tale-re      yaw d-æl-i.

        3Sg.F woman-OBJ pig  FAC-bite-PST

        'The pig bites the woman.'

The sentence in (39a) involves a human argument and an animate argument, and it may only be interpreted in such a way that the human argument is the agent. Since the human argument in (39b) is explicitly object-marked, the interpretation of agency is reversed. Thus, O-arguments that are more animate than their corresponding A arguments must be object-marked in Awtuw, though object-marking of less animate Os is not necessary. The claim that this contrast is based on relative animacy rather than animacy alone is supported by (40). Here, *yaw* 'pig' cannot be interpreted as an object as in (39a), because both arguments are equal in animacy. Instead, the sentence is interpreted as having what De Swart ([2007](#)) calls a "conjoined subject."

(40)   (Adapted from De Swart, [2007](#), p. 90, citing Feldman, [1986](#), p. 110)

    Piyren yaw di-k-æl-iy.

    dog    pig  FAC-IMPERF-bite-IMPERF

    'The dog and the pig bite.' **not:** 'The dog bites the pig.' / 'The pig bites the dog.'

This is evidence that DOM in Awtuw is sensitive to relative animacy.

Assuming word order variation to be DOM, De Swart ([2007](#)) also looks at Haida (Na-Dene), which exhibits a relative-animacy-related word order contrast. He provides the following examples:

(41)  (Adapted from De Swart, 2007, p. 102, citing Enrico, 1986, p. 74)

7adàahl-.uu    Mary Bill qing-gan.

yesterday-FOC Mary Bill see-PST

'Mary saw Bill yesterday' **not:** 'Bill saw Mary yesterday.'

(42)  (Adapted from De Swart, 2007, p. 102, citing Enrico, 1986, p. 75)

  a.  Yaank'ii.an-.uu Bill x-aay      gu'laa-gang.

     truly-FOC        Bill dog-DEF like-PRES

     'Truly Bill likes the dog' **not:** 'Truly the dog likes Bill.'

  b.  Yaank'ii.an-.uu x-aay      Bill gu'laa-gang.

     truly-FOC        dog-DEF Bill like-PRES

     'Truly Bill likes the dog' / 'Truly the dog likes Bill.'

The basic order of O and S in Haida is SO, which cannot be violated when O is greater than or equal in animacy to A, as seen in the fact that the OS translations of (41) — in which S and O are equally animate — and (42a) — in which O would be more animate — are ruled out as possible interpretations. However, the OS reading of (42b) is allowed, since the O is human and the S is an animal, and O is therefore higher in animacy than S. Thus relative animacy in Haida bears on the linear realization of O.

Clear examples of systems of DOM conditioned by contrasts in relative definiteness are significantly harder to come by. One example of such a system is found in Dutch, again assuming word order variation to count as DOM. Dutch is canonically SVO, and many contexts do not permit an OVS reading (Bouma, 2008). However, Bouma shows via an extensive corpus study that objects that are more definite than their subjects, specifically objects that are pronominal while the subject is a full NP, are significantly more likely to appear in an OVS construction than objects of lesser or equal definiteness to their subjects.

### 2.2.4  Ambiguity/Recoverability of Roles

Another frequent analysis of the distribution of DOM in some languages is that object-marking is driven by the need to unambiguously clarify which expressions in a clause serve which grammatical functions, something that De Swart (2007) calls "role recovery" and that I label "grammatical function (GF) assignment" in C. Shain (2008), the terminology I adopt here. It is sometimes argued that in languages which, in certain contexts, allow GF-assignment ambiguities (or indicate an incorrect GF-assignment), DOM may be conditioned by the (non-)appearance of such ambiguity, since a marked object cannot be interpreted

as S, and ambiguity is therefore eliminated (de Swart; Gerner, 2008; Feldman, 1986; Scott, 1978). That is to say, if GF-assignment ambiguity (henceforth GF-ambiguity) conditions DOM, then the object marker functions as a means of unambiguously assigning one NP to a particular GF (O), leaving the other NP to be interpreted as S. If this is the case, we would expect at least a one-way implication to hold for such languages:

(43)   ambiguous → marked

As an example of ambiguity-based DOM, take Yongren Lolo (*aka* Yi; Sino-Tibetan). According to Gerner (2008), Yongren Lolo does not have case-marking (aside from DOM), head agreement, or fixed word order. Thus when an object appears in what Gerner calls an "ambiguous predicational frame," that is, there exists ambiguity as to which NP is S and which is O, the object-marker $t^h ie^{21}$ appears, resolving the ambiguity.

(44)   (Gerner, 2008, p. 299)

    a.  ŋo$^{33}$ ɕe$^{33}$mo$^{33}$ ʈʂ$^h$    ʐi$^{33}$.

        1Sg  snake    follow go

        'I follow the snake.' / 'The snake follows me.'

    b.  ŋo$^{33}$ ɕe$^{33}$mo$^{33}$ **t$^h$ie$^{21}$** ʈʂ$^h$    ʐi$^{33}$.

        1Sg  snake    **OM**   follow go

        'I follow the snake.'

In (44), no object marker is realized, and either NP may be interpreted as S or O. In (44b), $t^h ie^{21}$ marks "the snake," eliminating the possible interpretation "the snake follows me." Clear thematic fit bias can reduce the need for object marking on the verb, as shown in (45), where "you" are much more likely to want to plow "earth" according to way the world usually works than the other way around, and no object marker therefore appears.

(45)   (Gerner, 2008, p. 301)

    ni$^{33}$ mi$^{33}$  mo$^{21}$    ŋ$^{21}$me$^{33}$ ɛ$^{21}$.

    2Sg  earth plough want    ALT

    'Do you want to plow the earth?'

Malayalam (Dravidian) seems to constrain DOM in a similar way. De Swart (2007) claims that Malayalam resists marking of inanimate objects. This tendency, however, does not always hold, as De Swart (p. 88) shows by the following examples from Asher and Kumari (1997, p. 204):

(46)  a.  Kappal tiramaalaka‍l‍-**e** bheediccu

      ship    waves-**ACC**    split.PST

      'The ship broke through the waves.'

    b.  Tiramaalaka‍l‍ kappal-**ine** bheediccu

      waves        ship-**ACC**  split.PST

      'The waves split the ship.'

In both of the above examples, an inanimate O takes accusative case-marking. The reason De Swart (2007) provides for this marking is that both ships and waves may split each other, thus resulting in thematic fit ambiguity. Accusative marking is permitted to appear in order to resolve this GF-ambiguity.

## 2.2.5 Topicality

There is an enormous literature on topicality, of which I cannot provide an overview here (see Gundel, 1973; Prince, 1979, 1981; Halliday, 1967a, 1967b, 1968; van Valin and Foley, 1985; Givón, 1983; Bolinger, 1952; Dahl, 1976; Büring, 1999; Firbas, 1966; and, for an extensive bibliography on topicality, García Landa, 2009). While the definition of "topic" is a subject of debate, topicality has been proposed as a conditioning factor in some systems of DOM cross-linguistically. For example, Mardale (2008) and Leonetti (2004) argue that the object marker is (in part) a topicality marker in Spanish. Mardale provides the following example.

(47)  (Adapted from Mardale, 2008, p. 451)

    a.  Un buen sablazo de sol  traspasaba **(\*a) la sacristia**.

      A   good cut     of sun pierced    **(\*OM) the vestry**.

      'A good cut of sun pierced **the vestry**.'

    b.  **(A) la sacristia$_i$**  la$_i$ traspasaba un buen sablazo de sol.

      **(OM) the vestry** it pierced    a   good cut    of sun.

      'A good cut of sun pierced **the vestry**.'

Indefinites are generally unmarked in Spanish, as I mentioned in §2.2.1. However, as (47) shows, this tendency can be overridden in certain circumstances. While the inanimate object *la sacristia* 'the vestry' in (47a) may not be object marked, the same NP may be marked if it is topicalized by fronting, as shown in (47b).

I mentioned previously that Persian marks [+ specific] objects. However, Shokouhi and Kipka (2003) argue that topicality plays a role in Persian DOM as well.[22] Their corpus study reveals that non-specific objects may still be marked if they are given (topical), and that objects with given referents were much more likely to be marked than objects with referents that were new or accessible.[23] They provide the example below.

(48)    (Adapted from Shokouhi & Kipka, 2003, p. 959)

    a.   rasti   **shahrieh** chi    shod?

       really **tuition**    what became.3Sg

       'by the way, what happened to **the tuition**?'

    b.   **shahrieh râ**   goft-an  pardaxt kard-an,

       **tuition    OM** said-3Pl pay      did-3Pl

       '(they) said (they) have paid **the tuition**.'

The question in (48a) establishes *shahrieh* 'tuition' as the topic of this piece of discourse. The answer, in which *shahrieh* is in object position, is *râ*-marked, suggesting topicality as a possible basis for DOM variation. Given their corpus results discussed above, it seems to be the case that topicality (as well as specificity) plays a role in the distribution of object marking in Persian.

Catalan was previously mentioned as a differentially object marking language, as shown in example (3). Escandell-Vidal (2008) argues that one of the conditions on DOM in Catalan is topicality. She shows that topicalized objects (left- or right-dislocated) may be marked even if they exhibit properties not generally associated with object marking in Catalan, such as being non-human or indefinite. Guntsetseg (2008) also finds (a variant of) topicality, which he calls "discourse prominence," to be relevant to DOM in Mongolian (Altaic). Defining the "discourse prominence" of a referent as the acceptability of reference to that individual via an anaphoric expression, he finds a positive relationship between discourse prominence and object marking, suggesting that recentness of mention plays a role in DOM in Mongolian. Guntsetseg also looks at persistence of a topic into subsequent discourse, and argues that "the direct object of a transitive clause is more likely to be marked by an accusative suffix if the speaker wants to tell more about it in the following discourse" (Guntsetseg, p. 64).

---

[22]Shokouhi and Kipka (2003) assume a *given/new* topicality distinction, where the topic is taken to be information already supplied (given) in the discourse.

[23]Shokouhi and Kipka (2003) define an "accessible" referent as "either derivable from the situation of the discourse or... a given referent which has not been activated for some time in the discourse" (thus it is similar to "weakly familiar" of Roberts, 2003).

Assuming verbal head agreement variation to be a form of DOM, there is data in Dalrymple and Niko-laeva ([2005](#)) that shows topicality to be relevant to the overtness of verbal object agreement in Northern Ostyak (Uralic). Looking purely at objectival or clausal properties, verbal object agreement appears to be optional.

(49)   (Adapted from Dalrymple & Nikolaeva, [2005](#), p. 75)

   a.  ma tam kalaŋ      we:l-s-əm

      I    this reindeer kill-PST-1Sg.SUBJ

      'I killed this reindeer.'

   b.  ma tam kalaŋ      we:l-s-∅-e:m

      I    this reindeer kill-PST-**Sg.OBJ**-1Sg.SUBJ

      'I killed this reindeer.'

In terms of animacy, definiteness/specificity, relative definiteness and animacy, and GF-ambiguity, the sentences in ([49a](#)) and ([49b](#)) are identical. However, ([49a](#)) has no object agreement on the verb, while ([49b](#)) does. What conditions this difference? Dalrymple and Nikolaeva ([2005](#)) appeal to the discourse properties of the referent of the object, showing that clearly topical objects must co-occur with overt agreement, while clearly non-topical objects cannot.

(50)   (Adapted from Dalrymple & Nikolaeva, [2005](#), p. 78)

   Q: What did you do to this reindeer?

   a.  (ma) tam kalaŋ      we:l-s-∅-e:m/*we:l-s-əm

      I       this reindeer kill-PST-**Sg.OBJ**-1Sg.SUBJ/kill-PST-1Sg.SUBJ

      'I killed this reindeer.'

   b.  (ma) we:l-s-∅-e:m/*we:l-s-əm

      I       kill-PST-**Sg.OBJ**-1Sg.SUBJ/kill-PST-1Sg.SUBJ

      'I killed (this reindeer).'

The question in ([50](#)) establishes the referent of *tam kalaŋ* 'this reindeer' as topical: it is what the discourse is about. As the subsequent examples then show, the sentences in which that referent appears as an object obligatorily take overt object agreement on the verb, regardless of whether the reindeer is referred to by a full NP, as in ([50a](#)), or not, as in ([50b](#)).

Schwenter and Silva ([2003](#)) show on the basis of corpus research that Brazilian Portuguese exhibits a topicality-related DOM contrast, assuming that object overtness contrasts are a type of DOM. Overt

objects (particularly overt lexical NPs, as opposed to pronouns or null objects) serve two topicality-related functions: (i) restoring a distant referent to topical status, or (ii) marking exceptionally high topicality. When the object referent was a specific animate, overt lexical objects performed function (i). When the object referent was not a specific animate, overt lexical objects performed function (ii).

In discussing topicality and DOM, I should note the relevance of the related notion of "focus," which, like "topic," has a variety of definitions in the literature. Looking particularly at contrastive focus, it turns out that there is cross-linguistic variation in its relationship to DOM: Northern Ostyak forbids overt agreement with contrastively focused objects, while Yongren Lolo permits marking of contrastively focused objects but not of (non-GF-ambiguous) unfocused ones.

Dalrymple and Nikolaeva ([2005]) show that objects in Northern Ostyak which are contrastive foci may not take overt object agreement, as in ([51]), in which the expression *tup* 'only' marks contrastive focus (it was the big cup as opposed to, e.g., the small one) and object agreement is disallowed.

(51)  (Adapted from Dalrymple & Nikolaeva, [2005], p. 75)

ma tup  wul a:n  il       pa:jət-s-əm/*pa:jət-s-∅-e:m

I    only big  cup down drop-PST-1Sg.SUBJ/*drop-PST-**Sg.OBJ**-1Sg.SUBJ

'I only dropped the/a big cup.'

However, in Yongren Lolo contrastive focus is one of the conditions under which object marking is permitted, as Gerner ([2008]) shows.

(52)  (Adapted from Gerner, [2008], pp. 301-2)

Q: What did Bolu do this afternoon?

bɔ³³lu²¹ mɔ³³lu³³ (*$t^h$ie²¹) tsʰɨ²¹ ɔ³³

Bolu    trousers  (***OM***)  wash DYN.PERF

Intended: 'Bolu washed trousers.'

(53)  Which trousers did Bolu wash?

mɔ³³lu³³ kʰo³³      mo³³ **$t^h$ie²¹** bɔ³³lu²¹ tsʰɨ²¹ ɔ³³

trousers DEM.DIST CL    **OM**   Bolu      wash DYN.PERF

'Bolu washed THAT pair of trousers (not THIS one).'

In (52), the question asked does not set up a context in which $mɔ^{33}lu^{33}$ 'trousers' may be contrasted with anything, and the object marker $t^hie^{21}$ is ungrammatical. The context in (53), on the other hand, does provide a potential for contrast, and $t^hie^{21}$-marking is permitted. Thus, the object marker $t^hie^{21}$ can be seen in part as a contrastive focus marker in Yongren Lolo.

It is therefore unclear what the typological predictions would be for the relationship between contrastive focus and object marking (since Northern Ostyak and Yongren Lolo are opposites in this regard). Furthermore, no descriptions of focus encoding in Guaraní have yet appeared, in which case I am not sure how contrastive focus might be indicated in Guaraní. For these reasons, I do not consider contrastive focus (or any other kind) in this study.

### 2.2.6 Other Conditions

The above properties are quite frequently appealed to in the literature on DOM. There are others, however, some of which I briefly mention here. Some propose telicity as a factor, such that objects in clauses with an inherent end-point are marked. If we assume object incorporation variation to be a kind of DOM, in some languages it has to do with the telicity of the clause; Mithun (1984) provides the following example of incorporation from Yucatec Maya:

(54)   (Mithun, 1984, p. 857)

   a.   t-in-č'ak-∅-ah          **če'**
        COMP-I-chop-it-PERF **tree**
        'I chopped a tree.'

   b.   č'ak-**če'**-n-ah-en.
        chop-**tree**-ANTIPASS-PERF-I.ABS
        'I chopped wood/wood-chopped.'

The clause in (54a) is an accomplishment, and the object is realized independently of the verb, while the clause in (54b) is an activity, and the object is incorporated. Thus telicity in Yucatec Maya is related to differential morphosyntactic realization of O.[24]

García García (2005) claims telicity is relevant to DOM in Spanish. He provides the example in (55), in which the verb *insultó* '(s)he insulted,' which is lexically telic, does not permit omission of *a* on human-referring objects.

---

[24] Notice that specificity might also play a role here.

(55)    (Adapted from García García, 2005, p. 25, citing Torrego Salcedo, 1999, p. 1786)

Marta insultó    *(**a**)    un    compañero.

Marta 3Sg.insult *(**OM**) INDEF colleague

'Marta insulted a colleague.'

García García (2005) also shows that DOM can influence the intepretation of activity verbs, of which he claims *besar* 'kiss' is one.

(56)    (Adapted from García García, 2005, p. 26, citing Torrego Salcedo, 1999, p. 1789)

Besaron        (**a**)    un niño.

kiss.past-3.pl (**to**) a    child

'They kissed a child.'

The telicity of (56), García García (2005) claims, rests crucially on the presence or absence of *a*. When *a* is present, the reading is telic. When *a* is absent, the reading is atelic. As evidence for this, he subjects (56) to a standard test for telicity: *in*/*for* modification. If a proposition is telic, it should be able to be felicitously modified by phrase meaning "in [some time]," such as "in a second," but not by a phrase meaning "for [some time]." The opposite holds for atelic propositions. The following examples show that (56) may be felicitously modified by the phrase *en un segundo* 'in a second' only when *a* is present. That is to say, the only way in which (56) may receive a telic reading is if its object is marked.

(57)    (Adapted from García García, 2005, p. 26, citing Torrego Salcedo, 1999, p. 1789)

a.  Besaron    **a**    un    niño en un    segundo.

3Sg.PST.kiss **OM** INDEF child in  INDEF second

'They kissed a child in one second.'

b.  *Besaron    un    niño en un    segundo.

3Sg.PST.kiss INDEF child in  INDEF second

Intended: 'They kissed a child in one second.'

Assuming DOM to include alternation between overt cases, the affectedness of the object also seems to play a role in DOM in some languages. Næss (2004) provides the following example from Finnish:

(58)   (Adapted from Næss, 2004, p. 1203)

    a.  Hän  jo-i        maido-**a**

        (s)he drink-3Sg.PST milk-**PART**

        '(S)he drank (some of the) milk'

    b.  Hän  jo-i        maido-**n**

        (s)he drink-3Sg.PST milk-**ACC**

        '(S)he drank the milk'

The alternation between partitive and accusative case corresponds to an alternation in part/whole affect-edness of the object. In (58b), which takes accusative case, the milk is necessarily interpreted as having been entirely drunk, while only part of the milk need have been drunk in order for (58a) to be true. Næss (2004) also shows a similar case alternation in Icelandic (Indo-European).

(59)   (Næss, 2004, p. 1205, citing Barðdal, 2001)

    a.  Hann    klóraði    **mig**

        he.NOM scratched **me.ACC**

        'He scratched me.'

    b.  Hann    klóraði    **mér**

        he.NOM scratched **me.DAT**

        'He scratched me.'

Though (59a) and (59b) are translated into English identically, Næss (2004) claims that in (59a) the scratching was violent and malicious, while in (59b) the scratching was helpful to the speaker, perhaps getting a hard-to-reach itch. Thus the accusative case-marked object is interpreted as being more affected by the event than the dative case-marked object.[25]

    Affectedness plays a role in overt object case alternations in Hungarian (Uralic) as well (Moravscik, 1978). In the following examples, the accusative-marked object is interpreted as entirely affected, while the partitive-marked object is interpreted as partially affected.

---

[25]I should note that, although I have presented telicity and affectedness as more or less unrelated here, the relationship between the two may be quite close (for discussion, see Dowty, 1991).

(60)   (Adapted from Lyons, 1999, p. 200, citing Moravscik, 1978)

    a.   Ette         a    sütemény-**t**.
      ate.3Sg.OBJ the pastry-**ACC**
      'He/She ate the pastry.'

    b.   Evett   a    sütemény-**böl**.
      ate.3Sg the pastry-**PART**
      'He/She ate some of the pastry.'

Moravscik (1978) also points to Kabardian (North Caucasian) as a language with affectedness-based case-alternation.

(61)   (Adapted from Lyons, 1999, pp. 200-1, citing Moravscik, 1978)

    a.   ḥă-m       q°ə́pṣḥă-**r** yeʒàqe.
      dog-ERG bone-**ABS** chews
      'The dog chews up the bone.'

    b.   ḥă-r       q°ə́pṣḥă-**m** yóʒàqe.
      dog-ABS bone-**LOC**  chews
      'The dog is chewing (on) the bone.'

In (61a), the absolutive-marked object is completely affected (eaten), while the locative-marked object in (61b) is only partially affected. Thus, both Hungarian and Kabardian exhibit affectedness-related alternations in overt object case.

    Næss (2004) takes such affectedness-related alternations to be the basis for her proposal that affectedness is the fundamental property underlying all DOM in accusative systems cross-linguistically. While this is an interesting approach, it quite quickly runs into the need to do some terminological stretching. In order to accommodate to her affectedness-based account the fact that more animate objects are more marked typologically, Næss absorbs into her definition of affectedness a concept she labels *salience*,[26] an approach which she justifies in the following quote.

> From a human point of view, which presumably is what linguistic expressions encode, some types of effects are both more easily perceptible and of more interest than others. More specifically, effects on some types of entities are more salient to human perception and interests than others. An effect on a

---

[26]Others, such as Magier (1987), appeal to "salience" in describing DOM as well. Unfortunately, the term is rarely defined explicitly in these discussions, and the intended intuition behind this type of description is often unclear (by contrast, see the explicit definition of "prominence" in Aissen, 2003b).

human participant is more likely to impinge directly on the lives of both the human in question and those surrounding him than an effect on an inanimate object; the situation described by *Peter killed John* is far more likely to have profound effects on all parties involved than, say, *Peter broke the pot*. (Næss, 2004, p. 1202)

Thus "affectedness" now refers not only to how completely the object alone is affected, but also to the less quantifiable notion of how important/impactful/interesting the effect itself would be to those indirectly involved, under the assumption that effects upon humans have greater collateral impact than effects upon inanimates. While it could be the case that "salience" plays a role in DOM, I believe it needs a more rigorous definition than that provided by Næss (2004) in order to be predictive. Also, it is not entirely clear to me whether this is an independently motivated redefinition of "affectedness," as opposed to the simple placing of a single label on "what the conditioning factors underlying DOM cross-linguistically have in common" so as to give the appearance of unified account.

### 2.2.7 The Interaction of Multiple Conditions on DOM

As I hope is clear by now, DOM in many, if not most or all, languages that exhibit it is conditioned by the interaction of multiple factors. In Spanish, DOM appears to depend on some combination of animacy, definiteness/specificity, topicality, telicity, affectedness, and probably more (see Tippets & Schwenter, 2007; García García, 2005). In Hindi, DOM is conditioned at least by definiteness and animacy (Aissen, 2003b). Persian seems to exhibit DOM based on topicality and definiteness/specificity (Lazard, 1982; Shokouhi & Kipka, 2003). Romanian DOM is conditioned by animacy, definiteness, and topicality (Klein, 2007; Mardale, 2008; von Heusinger, Klein, & de Swart, 2008). In Malayalam, DOM depends both on animacy and GF-ambiguity (Asher & Kumari, 1997; de Swart, 2007). The list could become quite long.

It is this reality of multiple underlying factors that drives much of the research on DOM. The research on Spanish is a case in point, as the analysis of *a* becomes increasingly rich in its evolution from "personal *a*" — assuming it to be based on the [±human] distinction — to current research which is continually uncovering more and subtler factors influencing the distribution of DOM in Spanish. Not only does a multiplicity of factors provide a number of interesting questions for language-specific investigation, it also drives an increasingly large literature on the typology of DOM, a literature concerned with the intriguing challenge of synthesizing the various and disparate properties associated with DOM (e.g., Aissen, 2003b; Hopper & Thompson, 1980, discussed below). Why is it the case that objects which are animate, definite, etc. form a natural class (in the terms of Comrie, 1979)? Is there a single property that the above

are manifestations of? Is a unified account possible? The questions remain open, and it is not my purpose here to answer them. Nevertheless, I consider it important to survey two of the major theoretical veins that attempt an answer: what I call the Markedness Approach (MA) and the Transitivity Approach (TA). I discuss these proposals in the following section.

## 2.3  Theories of Differential Object Marking

There are several syntheses of the properties underlying DOM, some of which have been more influential than others. García García (2005) argues that the common thread is *informativeness*, where the feature values more likely to trigger object marking are exactly those which are more informative. Næss (2004) considers *affectedness* to be the common ground between the conditions on object marking, as I discussed in §2.2.6. Some researchers, such as Comrie (1989) and Aissen (2003b), argue that DOM serves to mark deviation from an object prototype, a position I call the "markedness approach (MA)." Others, such as Hopper and Thompson (1980) and Næss (2007), assume DOM to be a morphosyntactic indication of a clause's conformance to certain aspects of a semantic/pragmatic transitivity prototype, a position I call the "transitivity approach (TA)." I perceive the MA and the TA to be the two most influential explanations of the distribution of DOM typologically, and here I briefly survey the claims of both. I conclude by showing that they both make identical predictions with respect to the phenomenon of Guaraní which I am examining, and thus that the central claims of this work do not rest crucially on the particular theory I assume.

### 2.3.1  The Markedness Approach: Aissen (2003b)

Although the MA has been assumed by several authors, I focus this review on Aissen (2003b). Aissen, employing reasoning from markedness theory, assumes that the more morphologically marked case corresponds to the atypical case, and the less morphologically marked case corresponds to the typical case. Thus, given a contrast in a language between overt and null object case-marking, the properties of the marked object are atypical for an object, and the properties of the unmarked object are typical for an object. Since animacy and definiteness are marked properties for objects, animate/definite NPs are therefore atypical as objects. Aissen proposes a "markedness reversal" between subjects and objects, whereby exactly the properties that are marked for objects are unmarked for subjects, and exactly the properties that are marked for subjects are unmarked for objects. Assuming animacy and definiteness to be different facets of one property, which she labels "prominence" (reminiscent of the "salience" found in other

proposals), Aissen claims that the typical subject has high prominence, while the typical object has low prominence. Much like Comrie (1989) earlier, Aissen predicts that deviation from the typical will result in a more morphologically marked form.

Aissen (2003b) hypothesizes that this prominence reversal is derivative of the need to clearly distinguish between S and O: if S and O are prototypically quite different from one another, then an atypical O will be one that exhibits S-like properties. Thus object-marking serves the primary purpose of marking deviation from the typical, but it also, indirectly, clearly disambiguates the GFs of two expressions that are similarly suited to the S slot. This logical by-product of marking atypical objects has led some to misinterpret Aissen as proposing disambiguation to be the core property underlying DOM cross-linguistically. For example, von Heusinger and Kaiser (2007) include Aissen's position among what they call "the Ambiguity Thesis," presumably under the impression that she claims the main reason DOM exists is to indicate the GFs of NPs in languages which cannot otherwise clearly do so. Malchukov (2008) also attributes to Aissen the claim that the primary function of DOM is distinguishing between S and O. However, this is a position she explicitly does not take, as shown in the following quotation:

> An intuition which recurs in the literature on DOM is that it is those direct objects which are most in need of being distinguished from subjects that get overtly case-marked. This intuition is sometimes expressed as the idea that the function of DOM is to disambiguate subject from object. There may be cases in which DOM is motivated precisely by the need to disambiguate, but it is also clear that DOM is required in many instances where the absence of case-marking could not possibly lead to ambiguity. (Aissen, 2003b, p. 437)

As I made clear above, the fundamental claim of the MA is that DOM arises in response to *markedness differences* between different types of objects, and performs a distinguishing function only indirectly as a result of the fact that the marked object is similar to the unmarked subject. While distinguishing A and O may have driven the formation of GF prototypes historically,[27] as De Swart (2007) argues with respect to animacy properties, the argument in Aissen is formulated primarily in terms of markedness. Thus, instances of object marking without GF-ambiguity do not constitute counterexamples to her analysis.[28]

---

[27] For an interesting development in research along these lines, see Jäger (2007). Assuming a prominence-based distribution of DOM quite similar to that of Aissen (2003b), he attempts to model the evolution of case-marking strategies in game-theoretic terms, arguing that all and only the attested case-marking patterns are evolutionarily viable.

[28] For criticism of the MA, see Næss (2004), Næss (2007), Haspelmath (2006a), and Carnie (2006).

### 2.3.2 The Transitivity Approach: Hopper and Thompson (1980)

Hopper and Thompson (1980) re-envision the notion of transitivity in grammar, departing from its common conceptualization as a morphosyntactic phenomenon and claiming instead that "the defining properties of transitivity are discourse-determined" (p. 251). In place of transitivity as a binary clausal property determined by the presence or absence of a direct object, Hopper and Thompson see Transitivity[29] as a scalar phenomenon consisting of a cluster of ten semantic/pragmatic features, which I reproduce in Table 2.1. The more "high" values a clause exhibits, the more Transitive it is, and the more "low" properties a clause exhibits, the less Transitive it is. The prototypical transitive clause is "high" with respect to every property.

|  | HIGH | LOW |
|---|---|---|
| PARTICIPANTS | 2 or more participants, A and O | 1 participant |
| KINESIS | action | non-action |
| ASPECT | telic | atelic |
| PUNCTUALITY | punctual | non-punctual |
| VOLITIONALITY | volitional | non-volitional |
| AFFIRMATION | affirmative | negative |
| MODE | realis | irrealis |
| AGENCY | A high in potency | A low in potency |
| AFFECTEDNESS OF O | O totally affected | O not affected |
| INDIVIDUATION OF O | O highly individuated | O non-individuated |

Table 2.1: The Transitivity properties of Hopper and Thompson (1980, p. 252)

Hopper and Thompson (1980) further break down "individuation" into another scalar cluster of properties, given in Table 2.2.

| INDIVIDUATED | NON-INDIVIDUATED |
|---|---|
| proper | common |
| human, animate | inanimate |
| concrete | abstract |
| singular | plural |
| count | mass |
| referential, definite | non-referential |

Table 2.2: The Individuation properties of Hopper and Thompson (1980, p. 253)

I should note that the TA is not simply a theory of DOM, but rather a theory of transitivity with implications for the theory of DOM. Notice that in Table 2.2, two of the individuation properties — animacy and

---

[29]Henceforth I use the capitalized "Transitivity" when referring to the concept from Hopper and Thompson (1980), since this is their notation as well.

definiteness — are also two of the main conditions on DOM described above. Since high individuation of O is a high Transitivity feature, it appears that DOM might be able to be given a uniform treatment in terms of Transitivity: since Hopper and Thompson implicitly assume object marking to be a more morphosyntactically transitive construction than non-marking, they predict that individuated objects will be marked because the clauses that contain them are highly Transitive.[30] Thus, the claim of this approach with respect to DOM is that object marking indexes higher Transitivity than the non-marked counterpart.[31]

As I claimed in the introduction to this section, both the MA and the TA can be seen to predict DOM essentially identically, namely, that animate/definite objects are those that are marked, and inanimate/indefinite objects are those that are unmarked. The MA predicts this because it assumes that inanimate and indefinite objects are the typical case, and that object-marking appears on the atypical case. The TA predicts this because it assumes definite/animate O to be a property of highly transitive clauses, and it assumes object-marking to be a more transitive morphosyntax than non-marking. Thus, definiteness/animacy and object marking should be positively related according to both frameworks. With this in mind, the particular theory chosen does not crucially affect the main objective of this work, which is to analyze the distribution of DOM in Guaraní.

## 2.4 Synopsis

Let me quickly review where we have come so far. I showed that there is a lack of consensus in the literature as to the definition of DOM: some include a large number of morphosyntactic alternations, and others are more restrictive; some include distributional properties in the definition, and some do not. As I will show, the particular Guaraní alternation under discussion here qualifies as DOM with respect to all the definitions of which I am aware, and thus the particular choice of definition for this thesis is not important.

Researchers of DOM have primarily attended to three important questions: (i) what are the distributional properties of DOM for particular languages that exhibit it, (ii) what cross-linguistic patterns hold with respect to the distribution of DOM, and (iii) to the extent that such patterns exist, how can these distributional facts be explained? With respect to (i) and (ii), which are primarily questions of description,

---

[30]Although it is never made explicit in Hopper and Thompson (1980) that some types of constructions are considered to be more transitive than others, such an assumption is necessary in light of the fact that they justify some of the Transitivity properties they present by showing that they are related to "morphosyntactic signals of high Transitivity," including "absolute case-marking on O" (Hopper & Thompson, p. 257). Subsequent work in the TA tradition has made this assumption explicitly, such as De Swart (2007) and Nuckols (2007).

[31]For criticism from within the TA framework of Hopper and Thompson's particular implementation of the TA, see Lazard (2002), Tsunoda (1985), and de Swart (2007).

I surveyed literature indicating several different factor groups relevant to DOM in the languages of the world:

(62)  a.  Animacy

    b.  Definiteness/Specificity

    c.  Relative Animacy

    d.  Relative Definiteness

    e.  GF-Ambiguity

    f.  Topicality

    g.  Telicity

    h.  Affectedness

I provided examples demonstrating the importance of each of these factor groups to DOM in certain languages. With respect to (iii), which is a theoretical question, I briefly outlined what I perceive to be the two most important theories of DOM in the literature: the Markedness Approach (MA), exemplified by Aissen (2003b), and the Transitivity Approach, exemplified by Hopper and Thompson (1980). The MA assumes that indefinite/inanimate DOs are the most typical DOs, and that deviation from this type results in a marked (atypical) expression that is more amenable to morphosyntactic object marking. The TA assumes that object marking is a more Transitive morphosyntax than non-marking, and that animate/definite objects are more individuated than inanimate/indefinite objects. Since high individuation of O is a high Transitivity indicator, clauses with animate/definite objects are more Transitive than clauses with inanimate/indefinite objects, and hence more likely to appear with the more Transitive overt marker than the less Transitive null marker. I pointed out that, while the picture painted by these two approaches is conceptually very different, both predict essentially the same thing with respect to DOM: that animate/definite objects will be marked, and inanimate/indefinite objects will not be. Since it is not the purpose of this thesis to develop a theory of DOM generally, I do not take a position regarding the MA, TA, or any other approach.

<div align="center">

# CHAPTER 3

# METHODOLOGY

</div>

Before discussing the results of the corpus investigation in Chapter 4, I must provide some critical background information: a clear set of criteria for determining which constructions are relevant to this discussion of DOM in Guaraní, and replicable techniques for testing the various hypotheses of the distribution of DOM.[1] Regarding the former, what is known in the sociolinguistic literature as a *circumscription of the variable context* (Tagliamonte, 2006), in §3.1 I outline in detail my methods and motivations for including certain clauses from the corpus as containing a relevant DOM alternant and excluding others. I preface this with a discussion of some properties of Guaraní grammar upon which circumscribing the variable context hinges. Following this, in §3.2 I lay out what I believe to be clear and replicable criteria according to which the relevant clauses in the corpus were categorized with respect to several factor groups of potential relevance to the distribution of DOM in Guaraní. This exposition is known as *operationalizing hypotheses* in the sociolinguistic literature (Tagliamonte). Features of Guaraní grammar pertinent to this operationalization for a given factor group are discussed in the relevant section.

## 3.1 Circumscribing the Variable Context

In this section I present a working definition of the phenomenon in question for Guaraní, placing restrictions on (circumscribing) the linguistic contexts in Guaraní in which one of the alternants of DOM as a linguistic variable is assumed to occur. This circumscription is not to be taken as a *claim* about which constructions in Guaraní should be counted as DOM, to the exclusion of those that do not qualify. It is simply a drawing of boundary lines regarding which kinds of things will be considered in this study and which will not. I lay out these restrictions in detail in §3.1.4. However, in order for them to be intelligible,

---

[1] Much of the following description of Guaraní draws heavily on my previous discussion in C. Shain (2008).

I must first describe some basic aspects of Guaraní grammar on which they are based. In §3.1.1 I briefly outline the verb agreement system in Guaraní. In §3.1.2 I discuss the distribution of the morpheme *-pe* generally, in its functions as a spatiotemporal postposition (§3.1.2.1), as an IO marker (§3.1.2.2), and as a differential object marker (§3.1.2.3), and I explore the syntactic category of *-pe* in §3.1.2.4. In §3.1.3 I address the question of whether the 3rd person object pronoun *(i)chupe* is a polymorphemic combination of the object marker *-pe* and what would then be a 3rd person pronoun *(i)chu*, or whether it is instead a single lexical item. Given this preface, I proceed to circumscribe the variable context in §3.1.4. All examples are from the primary Guaraní corpus unless otherwise noted.

### 3.1.1 The Verb Agreement System of Guaraní

Guaraní is what is known as an active-stative language (Mithun, 1991). It has two sets of verbal agreement morphemes: Set A inflections agree with S-arguments of stative predicates as well as A-arguments, and Set B inflections agree with S-arguments of active predicates as well as O-arguments (Gregores & Suárez, 1961; Velázquez-Castillo, 2004; Tonhauser, 2006).[2] By way of example, the clauses in (63) exhibit Set A inflection, and the clauses in (64) exhibit Set B inflection.

**Set A:**

(63)　a.　*Active S-argument:*

　　　　**A**-ju.

　　　　**A1.Sg**-come

　　　　'**I** came.'

　　b.　*A-argument:*

　　　　Upé-va niko **a**-hendu.

　　　　that-RC truly **A1.Sg**-hear

　　　　'**I** hear that.'

---

[2] For more thorough discussion of the distributional conditions on the two sets of agreement morphemes, see Mithun (1991) and Velázquez-Castillo (2002).

**Set B:**

(64)  a. *Stative S-argument:*

Nda **che**-py'aguapy-vé-i-ko.

NEG **B1.Sg**-peace-more-NEG-EMPH.

'**I** couldn't be calmer.'

b. *O-argument:*

Ha    Felípa **che**-recha.

CONJ Felípa **B1.Sg**-see

'And Felípa saw **me**.'[3]

On the basis of the classification of arguments in Dixon (1979, discussed in §2.1), active-stative languages like Guaraní are also referred to as "split-S" languages, since the agreement morphology is split across intransitive subjects (S-arguments), with active S-arguments taking the same morphology as A-arguments and stative S-arguments taking the same morphology as O-arguments. The Set A and Set B agreement morphemes are laid out in Table 3.1.

| Person | Set A | Set B |
|--------|-------|-------|
| 1sg | a(i)- | che(r)- |
| 1incl | ja(i)- | ñande(r)- |
| 1excl | ro(i)- | ore(r)- |
| 2sg | re(i)- | nde(r)- |
| 2pl | pe(i)- | pende(r)- |
| 3 | o(i) | i-/h- |

ro(i)- '12sg' po(i)- '12pl'

Table 3.1: Verbal agreement prefixes in Guaraní
As presented in Tonhauser and Colijn (to appear, p. 3)

Only one of these agreement morphemes appears on any given verb. As I mentioned above, the alternation between Set A and Set B prefixes on intransitive S-arguments is based on an active/stative distinction, where Set A prefixes occur with active one-place predicates, and Set B prefixes occur with stative one-place predicates. With regard to transitive verbs, Set A prefixes mark A, and Set B prefixes mark O. Which argument is selected for agreement is determined on the basis of the Person (65a) and Function (65b) Hierarchies (Tonhauser, 2006, p. 131), with the hierarchies themselves ranked as in (65c).

---

[3]The verbal prefix *r-*, which precedes the verb *hecha* 'see' in this example, has been a subject of significant debate in Tupí-Guaraní linguistics (Jensen, 1998). For an analysis of it, see Payne (1994). Its meaning does not matter for our purposes here, and thus verbs with and without it (e.g. *hecha* and *r-echa*) are glossed identically in this thesis.

(65) a. **Person Hierarchy:** 1<sup>st</sup> person > 2<sup>nd</sup> person > 3<sup>rd</sup> person

b. **Function Hierarchy:** A-argument > O-argument

c. Person Hierarchy > Function Hierarchy

This means that the verb agrees with the highest argument on the Person Hierarchy, regardless of its GF: if it is the subject, the agreement prefix will be Set A, and if it is the object, the agreement prefix will be Set B. If both arguments are on the same level on the Person Hierarchy (only when both are 3$^{rd}$ person in the corpus), the verb agrees with the highest argument on the Function hierarchy: the subject. Thus transitive verbs with two 3$^{rd}$ person arguments are always Set-A-marked. These constraints are exemplified by the following sentences (data from Tonhauser, 2006, p. 132).

(66) **A**-hecha   Juan.
**A1.Sg**-see Juan
'**I** see Juan.'

(67) **Che**-recha Juan.
**B1.Sg**-see  Juan
'Juan sees **me**.'

(68) **O**-hecha   Juán-pe.
**A3.Sg**-see Juan-PE
'**He/she/it/they** see(s) Juan.'

In (66) and (67), the highest argument according to the Person Hierarchy is the 1$^{st}$ person partici-pant. In (66), this argument is the subject, and the 1$^{st}$ person subject agreement morpheme therefore comes from Set A. The 1$^{st}$ person argument in (67), on the other hand, is the object, and the agreement morpheme is therefore from Set B. In (68), both arguments are 3$^{rd}$ person and thus equal on the Person Hierarchy. Therefore, according to the Function Hierarchy, the subject is selected for agreement over the object.

### 3.1.2   The Distribution of the Morpheme *-pe* in Guaraní

The morpheme *-pe* does more work in Guaraní than just the marking of DOs. It can also mark spatiotem-poral dependents as well as 3$^{rd}$ person indirect objects. While the latter two functions are not under con-sideration in this study, for completeness I briefly exemplify them here. In §3.1.2.1 I provide some exam-ples of the spatiotemporal use of *-pe*, in §3.1.2.2 I give examples of the occurrence of *-pe* on IOs, and in

§3.1.2.3 I establish the existence of variation in the use of *-pe* as a direct object marker. In §3.1.2.4 I briefly discuss the question of the grammatical category of object-marker *-pe*, i.e. whether it is a case marker or a postposition. I do not address the question of whether and to what degree the various uses of *-pe* are distinct lexical items, or explore their historical relationship to each other, but leave these questions to future research.[4]

### 3.1.2.1 Spatiotemporal Postposition

When *-pe* occurs in oblique phrases, it encodes primarily spatiotemporal relationships (see Tonhauser, 2006). The distribution of *-pe* in this function is broad both with respect to frequency of occurrence and the kinds of relationships it can express: Gregores and Suárez (1961, p. 271) claim that *-pe* "is the most frequent of all postpositions" and gloss it as 'on, at, in, to, for, with.' Here are some examples of oblique *-pe*.[5]

(69)  Oi-me   o-pyta  **pe   tranquera-pe**.
      3A-exist 3A-stay **that corner(Sp)-PE**
      'He's there, staying **(on that corner)**.'

(70)  Ha  upépe avei o-jepokuaa    la     o-je-karu-pá-rire     o-ñe-ñeño-mba-ite
      and there  too A3-be.used.to the(Sp) A3-JE-eat-COMPL-after A3-JE-lie.down-COMPL-very
      **yvyra-guý-pe   kátre-pe**.
      **tree-under-PE cot(Sp)-PE**
      'And there too after having eaten it was customary to lie down, **(under a tree) (on a cot)**.'

(71)  Ha  **i-pahá-pe** che-mo-sẽ        pero    ...
      CONJ **B3-end-PE** B1.Sg-CAUSE-leave but(Sp) ...
      '**(At the end)** he kicked me out, but ...'

### 3.1.2.2 Indirect Object Marker

Not only does *-pe* occur in oblique phrases as shown above, but it also is the basic IO marker for 3rd person objects in Guaraní.[6] I assume 1st and 2nd person pronouns have lexicalized IO proforms, presented in

---

[4]For some discussion of the historical relationship between *-pe* on spatiotemporal adjuncts and *-pe* on objects, see Chapter 5.

[5]The translations of the *-pe* phrases are in parentheses to represent the fact that they are approximations, and not claims about the denotation of *-pe* in these constructions.

[6]Some dative-type objects in Guaraní are marked by *-gui* rather than *-pe*. In C. Shain (2008, p. 14) I suggested that these objects take something like a "maleficiary" thematic role, in that they usually are the recipients of negative consequences of the action. The following is an example.

Table 3.2 (the 3<sup>rd</sup> person forms do not fit cleanly into this paradigm, and I discuss the 3<sup>rd</sup> person object pronoun in §3.1.3).[7]

|  | Bare Form | IO Form |
|---|---|---|
| 1Sg | *che* | *chéve* |
| 1Pl.Incl | *ñande* | *ñandéve* |
| 1Pl.Excl | *ore* | *oréve* |
| 2Sg | *nde* | *ndéve* |
| 2Pl | *peẽ* | *peẽme* |

Table 3.2: 1<sup>st</sup> & 2<sup>nd</sup> personal pronouns in Guaraní
As presented in Gregores and Suárez (1961, pp. 233-4)

The following are some examples of *-pe*-marked IOs in the corpus. In addition to these, there are many more examples of the 3<sup>rd</sup> person object pronoun *(i)chupe* in IO position. I discuss *(i)chupe* in more detail in the following section, and so I withhold exemplification until then.

(72) Ña    Maria, ña    Maria, rei-kuaa    ko    mba'é-pa o-jehu    **Rossáni-pe**?
     Doña Maria, doña Maria, A2sg-know this thing-QU A3-happen **Rossani-PE**
     'Doña Maria, doña Maria, do you know what happened **to Rossani**?'

(73) He'i    **Pirúlo-pe**...
     A3.say **Pirúlo-PE**...
     'He said **to Pirulo**...'

### 3.1.2.3   Direct Object Marker

Given the fact that analyzing the distribution of DOM in Guaraní presupposes its existence in Guaraní, I should establish that DOM is indeed an attested phenomenon in the language. This was essentially already done in Chapter 1, where I provided the constructions in (1), repeated below:

(1) O-mbo-tyai-pa              **chu-gui** y.
    A3-CAUSE-impure-COMPL **3-GUI**   water
    'He made all her water dirty **(to her)**.'

Aside from the above usage, Tonhauser (2006) analyzes *-gui* as an ablative case marker in oblique phrases, with a meaning something like "away from."

[7] It is possible that these IO forms are not actually single lexical items but instead are composed of a subject pronoun and the IO marker *-pe*, which is phonologically realized as *-ve*. This is the position taken by Gregores and Suárez (1961), Bossong (1985b), and Tonhauser (2006), and I do not intend to rule it out (although I know of no other instance in which *-pe* is realized as *-ve*). I simply assume for the purposes of this study that the forms given in Table 3.2 are single lexical entries of IO pronouns.

(1) (Adapted from Gregores & Suárez, 1961, p. 223)

    a. A-hecha **ne-róga**.

       A1.Sg-see **2Sg.POSS-house**

       'I see **your house**.'

    b. A-hecha **ne-rú-pe**.

       A1.Sg-see **2Sg.POSS-father-PE**

       'I see **your father**.'

In (1a) the verb *hecha* 'see' takes an unmarked DO, while in (1b) the same verb takes a marked DO, demonstrating that object marking is differential. The following examples further support this conclusion. In both pairs, the same verb takes a marked and unmarked variant — (a) and (b), respectively.

(74) a. Iñ-akã-ári-rupi,       o-hecha ha   o-japysaká-re,       o-hendu **kuimba'e-kuéra-pe**

      3-head-above-through A3-see   and A3-give.attention-REHE A3-hear   **man-PL-PE**

      o-ñe'ẽ ha   o-ñe'ẽ-api.

      A3-talk and A3-talk-be.correct

      'Over his head, while he was watching and paying attention, he heard **men** talking and discussing.'

    b. Kokuehe           o-hendu **ku   Remigia ñe'ẽngatu, Ña   Leona memby**,

      day.before.yesterday 3A.hear   **that Remigia chatterbox Mrs. Leona child**

      o-mbojá-ramo       hese       o-mbo-liga-ha         ndéve Don Albérto-ndive!

      3A-come.close-when 3.OBJ.PRO 3A-put-league(Sp)-NMLZR 2B.Sg   Don Albérto-with

      'The day before yesterday he heard **that chatterbox Remigia**, **Mrs. Leona's daughter**, when she came close to him, saying he was pimping you to Don Alberto!'

(75) a. Ha  upéi upépe o-ĩ-jave        hína  o-hecha **Juán-chi  ha  Pirúlo-pe** o-ñe-moĩ o-hupi

      and then there   A3-be-while PROG A3-see   **Juan-DIM and Pirúlo-PE** A3-JE-put A3-raise

      i-po          ichupe.

      3.POSS-hand 3.OBJ.PRO

      'And then, while he was there, he saw **Juanito and Pirulo** getting ready to say good-bye to him.'

b. O-jahú-ta-vo          peteĩ mba'e-ry-rú-pe          y-potĩ-porã-me          o-hecha **ju'i**

A3-bathe-FUT-when one   thing-liquid-recipient-PE water-clean-pretty-PE A3-see **frog**

o-po-po          o-hó-vo          h-apykuéri-kuéra.

A3-jump-jump A3-go-when 3.POSS-behind-PL

'And as he prepared his bath with clean water, he saw **the frog** coming, jumping and jumping

behind them.'[8]

### 3.1.2.4 Is *-pe* a Case Marker or a Postposition?

As I mentioned in Chapter 1, the terms "case marker" and "adposition" are used fairly interchangeably
in the literature on DOM to refer to the object marker in certain languages. This is not surprising, given
the functional equivalence cross-linguistically between case markers and postpositions; Zwicky (1992, p.
370) observes that "anything you can do with cases you can also do with adpositions, and vice versa,"
and Haspelmath (2006b, p. 2) claims that "in practice, we find considerable overlap between adpositions
and case inflection.... Thus, linguists will have to live with some indeterminacy in this area." However, the
question of whether object markers, like *-pe* in Guaraní, are adpositions still matters syntactically, since
adpositions are phrasal heads and case markers are not (see Tallerman, 2005).

There seems to be difference of opinion as to the syntactic status of *-pe* in Guaraní. Gregores and
Suárez (1961) and Velázquez-Castillo (2004) assume that *-pe* is a postposition, while Tonhauser and Colijn
(to appear) and Adelaar (1994) assume *-pe* to be a case marker when it occurs on DOs. In C. Shain (2008)
I presented the following examples, which might suggest that the former position is correct:

(76)   Jakare    peteĩ mymba oi-kó-va    **y ha yvý-pe**.

crocodile one   animal  A3-live-RC **water and earth-PE**

'The crocodile is an animal that lives **in the water and on the ground**.'

(75)   a. Ha  upéi upépe o-ĩ-jave    hína  o-hecha **Juán-chi  ha  Pirúlo-pe** o-ñe-moĩ o-hupi

and then there   A3-be-while PROG A3-see   **Juan-DIM and Pirúlo-PE** A3-JE-put A3-raise

i-po          ichupe.

3.POSS-hand 3.OBJ.PRO

'And then, while he was there, he saw **Juanito and Pirulo** getting ready to say good-bye to him.'

In both (76) and (75a), the constituent that is *-pe*-marked is a coordination of two nouns: *y* 'water' and
*yvy* 'earth' in (76), and *Juan* and *Pirulo* in (75a). The fact that such constructions are possible might in-

---

[8]The suffix *-me* in *ypotĩporãme* is an allomorph of *-pe*.

dicate that *-pe* is a postposition, since it does not have to appear on both conjuncts as we would expect if it were a case marker. I do not take this to be conclusive evidence, since the possibility exists that the coordinated nominal items in (76) and (75a) are of category N̄ rather than NP, and that the case marker is simply appearing on the maximal projection. Nevertheless, the case-marking account is complicated by the examples in (76) and (75a). As I mentioned in Chapter 1, this question is not of crucial importance to this study, which focuses primarily on the distribution of *-pe*, not on the correct syntactic analysis of it. I therefore leave the question open to future research.

### 3.1.3 The Grammatical Status of *(i)chupe*

I mentioned two possible parses of *(i)chupe* in §3.1.2.2: as a composed of the object marker *-pe* and a 3rd person pro-form *(i)chu*, or as a single 3rd person object pronoun. Here I present evidence that the correct analysis is the latter.

(77)  a. O-ĩ-ndaje  raka'e    peteĩ mitã tyre'ỹ   o-hayhú-va **mymba-kuéra-pe**.
          A3-be-REP long.ago one   child orphan A3-love-RC **wild.animal-PL-PE**
          'There once was an orphan who loved **animals**.'

     b. O-puka   o-hechá-vo   **chupe-kuéra**.
          A3-laugh A3-see-when **3.OBJ.PRO-PL**
          'He laughed when he saw **them**.'

The important thing to note about (77) is the linear order of *-pe* and the plural marker *kuéra*. In (77a), which contains a descriptive NP object, *kuéra* precedes the object marker. In (77b), however, the "object marker" precedes *kuéra*. As I discuss later in (3.2.2.1), *kuéra* in fact always precedes postpositions, including *-pe*, except when it occurs with *(i)chu*. This suggests that *-pe* in (77b) is not the object marker, but is instead part of the phonology of *(i)chupe*, a single lexical entry, explaining why *kuéra* follows. Supporting this conclusion is the fact that *(i)chu* never appears independently in the corpus. Indeed, speakers found the use of *(i)chu* by itself to be unacceptable.

As Judith Tonhauser has pointed out to me, there is also phonological evidence that *(i)chupe* is a single lexical item, rather than composed of two morphemes. Usually the last syllable of a stem is the one that is stressed in Guaraní, unless a stress-attracting suffix is added, in which case the suffix takes primary stress (Gregores & Suárez, 1961, p. 94). The Guaraní writing convention (used in the examples here) reflects this, since no accents are transcribed unless they fall somewhere other than the ultimate syllable of the word. However, *-pe* is an unstressed suffix; if it occurs word-finally, primary stress falls somewhere before it (see

the accent markers on *-pe*-marked phrases throughout this thesis). Nevertheless, *(i)chupe* is stressed on the ultimate syllable, which would be disallowed if the "pe" in *(i)chupe* were actually the object marker *-pe*, but which would be expected if *(i)chupe* were a single lexical item. I take this to be evidence that the apparent similarity between the last syllable of *(i)chupe* and the object marker *-pe* is only superficial, and that *(i)chupe* is in reality a single indivisible lexical item, an object pronoun.[9]

### 3.1.4 Circumscribing the Variable [±*pe*]

Having discussed several relevant and basic points of Guaraní grammar, I can now lay out the boundaries according to which clauses were considered relevant/irrelevant to the current study. For our purposes here, I limit the items under consideration to the following:

(78)    a.  Clauses involving overt, non-incorporated direct objects without any case/postposition marking

      b.  Clauses involving overt, non-incorporated direct objects that are marked by *-pe*

This is in keeping with the Principle of Accountability for the study of variation, according to which one must provide "an exhaustive report for every case in which a variable element occurs out of the total number of environments where the element could have occurred, but did not" (Tagliamonte, 2006, p. 13).

---

[9]I should note the existence of the form *(i)chugui*, which appears to be evidence of alternation of the postpositions *-pe* and *-gui* over *(i)chu* as a root. However, I again believe that this is only superficial, given the fact that, as was the case with *(i)chupe*, the *-gui* in *(i)chugui* precedes rather than follows the plural suffix, as shown in (1).

(1)  A-kañy-ma      **chu-gui-kuéra**
    A1sg-hide-COMPL **3-GUI-PL**
    'I've hidden **from them**.'

And, again like *(i)chupe*, while *-gui* is an unstressed suffix in Guaraní, *(i)chugui* is stressed on the ultimate syllable. This is evidence that both *(i)chupe* and *(i)chuqui* are single lexical items, and not composed of *(i)chu* and a postposition.

Despite this, both pronouns are similar in distribution to descriptive NPs + *-pe*/*-gui*, as shown in the following examples:

(2)  a.  *He'i*    **Pirúlo-pe**...
      *A3.say* **Pirúlo-PE**...
      'He *said* **to Pirulo**...'

    b.  *He'i*    **chupe**      ype...
      *A3.say* **3.OBJ.PRO** duck...
      'The duck *said* **to him**...'

(3)  a.  I-*poi*-pá-ite        **pe i-pó-pe-gúa-gui** ha ho'a otī-vo             ý-pe.
      3-*drop*-COMPL-very **that 3-hand-PE-of-gui** and A3.fall embarrassingly-when water-PE
      'He *dropped* **all that he was holding in his hands** and fell face first into the water.'

    b.  Ha o-ñe-moī-vo      pe i-vosa-'í-pe      o-jagará-ha-guā-icha    ju'í-pe, ju'i o-po      o-je-*poi*    **chu-gui** ý-pe.
      and A3-JE-put-when that 3-bag-DIM-PE A3-grab-in.order.to-like frog-PE frog A3-jump A3-JE-*drop* **3-GUI** water-PE
      'And as he was getting ready to grab the frog in his little bag, the frog jumped and he *dropped* **him** in the water.'

In light of this data, I conclude that *(i)chupe* and *(i)chugui* are each single lexical items that, in distribution and function, act similarly to NPs marked by *-pe* and *-gui*, respectively.

Assuming *-pe* to be that variable element, I consider all cases in which it occurs or could have occurred, but not those in which it could not have. Example (79) shows two sentences from the extended corpus that would be relevant under these criteria.

(79)  a.  **Juán-pe** n-oi-kytĩ-ri-kuri        kysé-pe   pe   kuimba'e.

          **Juan-PE** NEG-A3-cut-NEG-KURI knife-PE that man

          'That man didn't cut **Juan** with the knife.'

    b.  **Pe   pastel**   n-oi-kytĩ-ri-kuri        kysé-pe   pe   kuimba'e.

          **That cake(Sp)** NEG-A3-cut-NEG-KURI knife-PE that man

          'That man didn't cut **that cake** with the knife.'

The object *Juánpe* in (79a) is overt, non-incorporated, and marked by *-pe*. The object *pe pastel* 'that cake' in (79b) is also overt and non-incorporated, yet unmarked. Thus they are both relevant constructions to this study.

The above criteria rule out a variety of other constructions from consideration. The most obvious of these are intransitive clauses, which do not involve a DO and thus cannot exhibit object marking, such as the following.

(80)  Heẽ, a-ju.

    yes   A1.Sg-come

    'Yes, I came.'

The criteria in (78) also rule out some constructions that might conceivably involve a direct object, such as incorporated object constructions like that in (81b):

(81)  (Adapted from Velázquez-Castillo, 1996, p. 107)

    a.  A-ñamí-ta        pe   **vaka** moroti

        A1.Sg-milk-FUT that **cow**   white

        'I'll milk that white **cow**.'

    b.  A-**vaka**-amí-ta        ko   pyhareve

        A1.Sg-**cow**-milk-FUT this morning

        'I'll do some milking this morning.'

In (81a) the object *pe vaka* 'that cow' is a DO, morphosyntactically independent from the verb. In (81b), however, *vaka* 'cow' has been incorporated into the verb. Objects of the type in (81b), if they even are objects, are not considered in this study.

Null objects, exemplified below, are also excluded from examination.

(82)  a.  Pero   o-topa  hikuái      **la     sirujuano   por    kasualidad o-ĩ-va     avei**

          but(Sp) A3-find 3Pl.SUBJ.PRO **the(Sp) surgeon(Sp) by(Sp) chance(Sp) A3-be-RC also**

          'But they found **the surgeon who happened to be there, too**.'

    b.  O-topa  hikuái         ∅      ha     o-mbyvy chupe.

          A3-find 3Pl.SUBJ.PRO **(OBJ)** CONJ A3-stitch 3OBJ.PRO

          'They found ∅ **(him)** and he stitched her.'

The above sentences appear consecutively in the corpus. In the first, the object of *topa* 'find' is overt (the NP in boldface). In the second, the object of *topa* 'find' is null (anaphoric with the object in (82a)). Constructions like that in (82b) were not considered in this study.

Additionally, there are instances of overt, non-incorporated NPs that are not -*pe*-marked that are nevertheless marked by a different postposition/case marker. In cases in which such NPs have a clear oblique status, such as the bolded phrase in (83), they were not considered here.

(83)  Ha  o-moi  **i-lomó-re**      ha   vokoike o-ho.

     And A3-put **3-back-REHE** and soon    A3-go.

     'And he put it **on his back** and soon went.'

The NP *ilomo* 'his back,' marked by the spatial postposition -*re*, is clearly not the object of *moi* 'put,' but rather occurs in an oblique phrase expressing location. However, there are similarly-marked phrases in the corpus that seem to be much more object-like, in the sense that they seem like more canonical patients:

(84)  a.  A-mendá-ta,    a-mendá-ta     **ne-memby     kuñá-re**

          A1.Sg-marry-FUT A1.Sg-marry-FUT **2Sg.POSS-child woman-REHE**

          'I will marry, I will marry **your daughter**!'

    b.  Aniké-na nde-resarai **pe   nde-pó-gui**

          NEG-IMP B2.Sg-forget **that 2Sg.POSS-hand-GUI**

          'Don't forget **your hand**!'

The -*re*-marked NP in (84a), *nememby kuñáre* 'your child-woman (daughter),' is a DO in the English translation, at least, and seems to be causally affected in a way that the -*re*-marked NP in (83) is not. Other studies, such as Tonhauser and Colijn (to appear), do in fact consider such phrases to be direct objects. Similarly, given the fact that "your hand" is a DO in the English translation of (84b), it is possible that the

-*gui*-marked phrase *ndepógui* 'your hand' is also a DO in Guaraní. While I do not wish to rule these possibilities out, for the purposes of this study such phrases are not considered, since they do not exhibit either -*pe*-marking or non-marking.

There were 406 clauses in the corpus that fit the criteria in (78). There were a few construction types within this set that I further excluded from consideration:

(85)   a.   Clauses with a 1st or 2nd person DO.

     b.   Clauses with *(i)chupe* '3OBJ.PRO' as a DO.

     c.   Clauses with a proposition- or event-denoting DO.

The exclusion in (85a) is motivated by the fact that 1st and 2nd person objects do not permit DOM variation for independent reasons. As I discussed in §3.1.1, 1st and 2nd person objects are high on the Person Hierarchy and are therefore usually realized as verbal agreement prefixes, not as morphosyntactically independent DOs. This is shown in example (86):

(86)   a.   Por      poco      na-**che**-nupã-i.
          for(Sp)   little(Sp)   NEG-**B1.Sg**-hit-NEG
          'He hit **me** for nothing.'

     b.   Ñandejara ta-**ne**-rendu,      che-ama!
          God        IMP-**B2.Sg**-hear 1Sg.POSS-dear.one
          'May God hear **you**, my dear!'

     c.   **Roi**-su'ú-ta.
          **12Sg**-bite-FUT
          'I will bite you.'

The 1st person object in (86a) and 2nd person object in (86b) are realized as verbal prefixes and therefore cannot be object-marked. The same is true for the implicit 2nd person patient expressed by the portmanteau prefix *roi* in (86c). Thus, in none of these cases is object-marking possible. In cases in which 1st or 2nd person objects are realized independently of the verb agreement system, they take the form of 1st/2nd person IO pronouns, as shown in (87). Like *(i)chupe*, these proforms do not permit DOM variation. Thus the only objects for which DOM is a relevant variable are those in 3rd person.

(87)  (Adapted from Bossong, 1985b, p. 18)

(Che)      ro-juhu      (**ndéve**)

(1Sg.PRO) 12Sg-find (**2Sg.OBJ.PRO**)

'I found you.'

There were some 3<sup>rd</sup> person objects excluded as well, however. According to (85b), sentences like this one were not considered:

(77)  b. O-puka   o-hechá-vo   **chupe-kuéra**.

A3-laugh A3-see-when **3.OBJ.PRO-PL**

'He laughed when he saw **them**.'

As I argued in §3.1.3, *(i)chupe* is a single lexical item. Such objects cannot exhibit DOM variation, and sentences like (77b) therefore do not factor into my analysis.

The exclusion in (85c) rules out propositional or event-denoting objects like the following:

(88)  a. Ña   Maria, ña   Maria, rei-kuaa   **ko mba'é-pa o-jehu      Rossáni-pe**?

doña Maria, doña Maria  A2.Sg-know **this thing-QU A3-happen Rossani-PE**

'Doña Maria, Doña Maria, do you know **what happened to Rossani**?'<sup>10</sup>

b. Ha   nd-oi-kuaa-i       ju'i **mba'é-pa la      ha'e-ku'éra      o-japo-sé-va**.

CONJ NEG-A3-know-NEG frog **thing-QU the(Sp) 3Sg.SUBJ.PRO-PL A3-do-DES-RC**

'And the frog still didn't know **what they wanted to do**.'

c. **Mba'e piko** he'i   va'erã ne-memby-kuéra?

**thing  QU**   3.say must  2Sg.POSS-child-PL

'**What** will your children say?'

In all of the above examples, the object phrase in boldface does not pick out an individual or a discourse referent, but rather a proposition that can be known or said. They are not considered here for two reasons: (i) I know of no examples from the literature in which objects like these are explicitly said to be evaluated with respect to DOM, and (ii) of the more than 100 examples of propositional objects in the corpus, none receives an object marker. Thus most clauses with the verb stems *e* 'say' and *kuaa* 'know' are not consid-

---

<sup>10</sup>The *-pe* morpheme following *Rossani* indicates that *Rossani* is the indirect object of *jehu* 'happen.' It is not an object marker for the propositional object of *kuaa* 'know.'

ered here.[11] This is not categorical, however, as there are some non-proposition-denoting objects of these verbs in the corpus, such as the following:

(89)  Tuicha-ku hyapu hína **pe  téra**  he'i  va'e-kue!

      big-EMPH lie    PROG **that name** 3.say RC-KUE

      '**That name** he said is a big lie!'

In (89), the object of *he'i* 'say' is not an event or proposition, but rather an entity: a name. Thus it was in fact considered in this study.

I must note that the exclusions in (85) differ somewhat from those in my previous study of DOM in Guaraní (C. Shain, 2008). There I also excluded objects consisting only of demonstratives and their modifiers (i.e. without overt nominal heads) and objects consisting of the inanimate-referring wh-word *mba'e*. Both of these exclusions were made under the assumption that, since neither of these classes of objects is ever marked in the corpus, marking them was ungrammatical for reasons independent of the underlying conditions on DOM. This assumption turns out to have been incorrect, in the light of speaker data demonstrating that, though somewhat odd, the marked forms of demonstratives like *upeva* 'that (one)' and wh-words like *mba'e* are acceptable in object position. Thus it becomes the responsibility of the analysis to account for the absence of these marked forms in actual discourse. In my previous study I also did not distinguish between proposition- and entity-denoting objects, as I have done here.[12]

Finally, it was somewhat problematic to determine whether certain verbs selected an IO or a DO, since the IO and DO marker in Guaraní is the same (at least on the surface), namely, *-pe*. Since IO marking is not variable and is not the phenomenon being explored in this study, IOs were not taken into consideration throughout corpus annotation. I generally considered verbs to take IOs rather than DOs if the thematic role of the argument in question was fairly clearly a Goal/Recipient. Thus the Goal/Recipient arguments of speech verb stems like *ñe'ẽ* 'speak,' *mombe'u* 'recount,' and *e,* 'say,' transaction verbs like *me'ẽ* 'give,' and motion verbs like *ha* 'go,' *mboja* 'approach,' and *jupi* 'climb' were not considered DOs and were excluded from consideration here. A further complication in this area arose regarding verbs which include the causative *mbo,* such as the following:

---

[11]The verb *kuaa* 'know' is ambiguous: it can mean factual knowledge, in which case the object is a proposition, or it can mean personal familiarity with someone, in which case the object is human-referring. Clauses with *kuaa* of the latter type would therefore be considered in this analysis. However, none appear in the corpus.

[12]Thus, though I examine the same corpus in this work and in C. Shain (2008), the numbers I provide look somewhat different, since I am here considering some examples that I excluded before, and excluding some examples that I considered before. There were also some examples that I had simply analyzed incorrectly, and these have been fixed for this study.

(90)  Oi-ko-ndaje ka'i      rasa-ité-ma           o-mbo-py'a-rasy      **kuña-karai-mi**

A3-be-REP   monkey intense-very-totally A3-CAUSE-stomach-sick **woman-gentleman-DIM**

**peteĩ-me**.

**one-PE**

'They say that the monkey was very bothersome **to a lady**.'

Is the bolded argument a DO or an IO of *mbopy'arasy* 'cause X's stomach to be sick'? Bossong (1985b) provides a very useful test for IO vs. DO that might be helpfully applied here. He points out that that the portmanteau prefixes (which denote a 1st person subject and a 2nd person object), may only be used with verbs that take DOs, and not with those that take IOs. Thus if a portmanteau prefix appears with a verb, the lower argument of that verb is necessarily a DO, not an IO. He provides the following examples:

(87)  (Adapted from Bossong, 1985b, p. 18)

(Che)      ro-juhu    (**ndéve**)
(1Sg.PRO) 12Sg-find (**2Sg.OBJ.PRO**)

'I found you.'

(91)  (Adapted from Bossong, 1985b, p. 19)

(Che)   **a**-me'ẽ    ndéve      ko  kuatia.
(1Sg.PRO) **A1.Sg**-give 2Sg.OBJ.PRO this book

'I gave you this book.'

In (87), the verb *juhu* 'find' is a DO-selecting verb, licensing the appearance of a portmanteau prefix when the subject is 1st person and the object is 2nd person. In (91), however, no portmanteau prefix appears, despite the fact that the verb takes a 1st person subject and a 2nd object, since that object is an IO, not a DO.

While this test would be tremendously useful with consultants, its usefulness was somewhat limited for this corpus study, since portmanteau prefixes are fairly rare, and not all verbs in question appear with them in the corpus. However, I was able to find one example in the extended corpus of a causative verb which took a portmanteau prefix rather than a 1st person prefix and an IO pronoun, thus suggesting that the object argument of such verbs is a DO rather than an IO:

(92)  Michĩ re-ku'e      ne       aña-memby ha   yrupe-icha **ro**-mbo-kuá-ta      rei-kuaa

small A2.Sg.move B2.Sg.PRO devil-child   CONJ sieve-like **12Sg**-CAUSE-hole-FUT A2.Sg.know

hagua.

PURP

'Move one inch, you son of a devil, and I'll perforate you like a sieve in order for you to learn.'

While this is admittedly only *prima facie* evidence for classifying the objects of causatives like (90) as DOs, I believe it to be a reasonable basis for this classification in the absence of counterexemplification.

Taking the restrictions in (78) and (85) into account, along with the criteria for determining DO vs. IO status, there remain 233 eligible clauses in the corpus. Having thus circumscribed the variable under consideration, I now turn to the question of distribution. In §3.2 I explicate the metrics by which I operationalize some reasonable hypotheses of the distribution of DOM in Guaraní.

## 3.2   Operationalizing Hypotheses

In §2.2, I pointed out several underlying conditions on DOM proposed in the literature:

(93)   a.  Animacy

b.  Definiteness/Specificity

c.  Relative Animacy

d.  Relative Definiteness/Specificity

e.  GF-Ambiguity

f.  Topicality

g.  Telicity of the Proposition

h.  Affectedness of the Object

Since the above factor groups have an attested impact on the distribution of DOM cross-linguistically, they are a good place to start in formulating hypotheses about the distribution of DOM in Guaraní specifically. For the purposes of this study, I assume each of the factors in (93) to be possibly relevant to the distribution of DOM in Guaraní. This requires operationalization: the development of clear, replicable criteria by which to determine the value of any object/clause with respect to the above features. In light of this, I do not examine telicity or affectedness here, since I was unable to find reliable criteria independent of object-marking according to which the telicity of any given clause or the affectedness of any given

object might be assessed.[13] While I leave the possibility open that telicity and/or affectedness influence the distribution of DOM, without such clear criteria or speaker intuitions of my own I am unable to test this possibility systematically. I was able to adopt what I believe to be clear metrics for the other features, however, which I present directly.

### 3.2.1 Animacy

For the most part, the classification of DOs in terms of the animacy hiearchy in (12) was quite straightforward. If the object NP was used to denote a human being, it was assigned the value 'Human.' If the NP denoted an animal, it was assigned the value 'Animate.' If the NP denoted an entity that was not alive/sentient, it was assigned the value 'Inanimate.'

There were some objects that seemed to blur these lines, however, for which judgments had to be made. For example, body parts pose something of a problem: are they human, animate, or inanimate? For this study, I classified all body parts as animate non-human, following similar classifications in Klamer and Kratochvíl (2006), Quinn (2001), and Bell, Hadj-Bouziane, Frihauf, Tootell, and Ungerleider (2009).[14]

The assessment of animacy proved difficult for anthropomorphized animals as well. Since such animals are being attributed with characteristics exclusively possessed by humans, I viewed this as a type of coercion and categorized them as human. Support for this comes from Bossong (1985b, p. 22), who claims there is no distributional difference between object marking of personified animals vs. humans. I take this to be supportive of the choice to code anthropomorphized animals as human. Practically speaking, an animal was classified as human if there existed evidence in the discourse (independent of object-marking) that that animal was being portrayed with exclusively human properties. For example, if an animal was said to think rationally or speak, it was considered human. If no such independent evidence existed, the animal was considered animate.[15] Under this classification scheme, the object in (94), *aguarápe* 'the fox,' is human, since just before this the fox and the monkey are having a conversation with each other about marrying a human girl, a clear case of anthropomorphization.

---

[13]Although in the specific case of telicity, it does seem to be possible to do principled, corpus-based analysis of lexical aspect of verbs. For discussion, see R. Shain (2008).

[14]I should note that, while this is the assumption of the current study, it isn't always the case that body parts are assumed to be animate in the literature. Scheibman (2002), for example, assumes them to be inanimate.

[15]I must note that this classification system is different than the one I assumed in C. Shain (2008). There I classified all animals as animate, regardless of anthropomorphization. Since there were many such animals in the corpus, the animacy figures differ significantly between this study and the previous one.

(94) Ka'i,      peteĩ tesapirĩ-me o-ñapytĩ-su'u **aguará-pe**.

monkey one   blink-PE    A3-tie-bite    **fox-PE**

'The monkey, in a blink, strongly tied up **the fox**.'

### 3.2.2 Definiteness/Specificity

The coding criteria for definiteness/specificity presented below depend in part on morphosyntactic indicators of definiteness in the noun phrase. Therefore, prior to discussing these criteria, I provide a cursory overview in §3.2.2.1 of noun phrase structure in Guaraní.[16]

#### 3.2.2.1 The Syntax of the Guaraní NP

Determiners, demonstratives, numerals, and possessive markers are prenominal:[17]

(95) *Determiner*:

peteĩ kyse

one   knife

'a/one knife' (Tonhauser, 2006, p. 148)

(96) *Demonstrative*:

ko   kyse

this knife

'this knife' (Tonhauser, 2006, p. 148)

(97) *Possessive Marker*:

i-ryguasu

3.POSS-chicken

'Their/her/his chicken'[18]

When determiners and possessive markers appear in the same NP, the determiner precedes the possessive marker:

---

[16]For a more complete discussion of NP syntax in Guaraní, the reader is referred to Tonhauser (2006).

[17]With occasional exceptions for the indefinite determiner *peteĩ*, which is (infrequently) realized postnominally in the corpus. I take postnominal *peteĩ* to be a partitive construction, as I discuss in §4.10 with respect to example (143).

[18]Notice that throughout this thesis, the prefixes glossed as possessive markers are homophonous with the prefixes glossed as Set B agreement prefixes for stative verbs (see Table 3.1). I do not attempt to analyze the relationship between stative verbs and possessive constructions here, and refer the reader to Nordhoff (2004) for discussion.

(98) la       i-po

the(Sp) 3.POSS-hand

'Their/her/his hand'

Noun-noun compounds in Guaraní are realized as a string of nouns in which the final N is the head and the preceding N the modifier. These constructions may be possessive (99) or not (100).[19]

(99) che-sy        sy    róga

1Sg.POSS-mother mother house

'My grandmother's house (mother's mother's house)'

(100) tetã      mba'e mombe'u

country thing  story

'A country story'

Adjectives and the plural marker *-kuéra* are postnominal:

(101) che-lapi         pytã

1Sg.POSS-pencil(Sp) red

'my red pencil'

(102) la       i-pypore-kuéra

the(Sp) 3.POSS-footprint-PL

'Their/her/his footprints'

Relative clauses are postnominal and marked by the relative clause marker *-va*, which is a verbal suffix.

(103) (Adapted from Tonhauser & Colijn, to appear, p. 148)

peteĩ kane'õ-va

one   tired-RC

'one who is tired'

(104) i-personal-kuéra           o-mba'apó-va kokúe-pe

3.POSS-personal.servant(Sp)-PL A3-work-RC   chacra-PE

'her personal servants who worked on the *chacra*'

---

[19]For fuller discussion of attributive constructions in Guaraní, see Velázquez-Castillo (1996).

### 3.2.2.2 Operationalizing Definiteness/Specificity

Having briefly introduced the relevant components of NP structure, I turn to the operationalization criteria I used for definiteness/specificity. I assume the Definiteness Hierarchy of Aissen (2003b) for this study in classifying objects with respect to level of definiteness. The hierarchy was given in (27), which I reproduce below:

(27) Personal pronoun > proper name > definite descriptive NP > indefinite specific NP > nonspecific NP

*abbreviated:*

Pro > PN > Def > Spec > NSpec

In order to remain objective, whenever possible I assessed definiteness/specificity values according to overt morphosyntactic indicators. As will become evident below, the two domains underspecified by morphosyntax were bare common nouns and indefinites, the latter of which contain no morphosyntactic indicators distinguishing between specific and non-specific readings. In these cases, more subjective judgments were required.

In light of the variable circumscription given in §3.1.4, the highest item in (27), Pro, is irrelevant to this study. The reason for this is that all 1st and 2nd person objects are excluded from consideration, along with the only 3rd person object pronoun *(i)chupe*.[20] Thus there are no object pronouns in Guaraní for which DOM is a relevant variable. The criteria for classifying object phrases into the remaining categories in (27) are as follows.

---

[20]There is another 3rd person referring pronoun that appears to occupy object or object-like positions in Guaraní: *hese*. However, it seems to be the case that *hese* is the pro-form of *-rehe*-marked phrases. This is seen in the following examples:

(1) a. Upéi i-ñakã     o-sē-vo      pe y-gui    Juan-chi **ju'i rova-ite-rehe**   o-*maña*.
       then 3.POSS-head A3-leave-when that water-GUI Juan-DIM **frog face-very-REHE** A3-*look*

   'Then, when his head left the water, Juan was *looking* **at the frog's face**.'

   b. O-*maña* **hese**   ju'i ñemi-há-me.
      A3-*look* **3.PRO** frog hide-NMLZR-PE

   'The frog *looked* **at him** from its hiding spot.'

(2) a. Nde      piko re-*puká*-ta    **che-rehe**?
       2Sg.PRO QU A2.Sg-*laugh*-FUT **B1.Sg-REHE**

   'Will you *laugh* **at me**?'

   b. Ju'i o-*puka*   **hese-kuéra**.
      frog A3-*laugh* **3.OBJ.PRO-PL**

   'The frog *laughed* **at them**.'

Since *-rehe*-marked objects are excluded from consideration here according to (78), the proform *hese* is as well.

All and only names of individuals were classified as PN. Objects were considered definite descriptions if they contained a definite determiner,[21] possessive marker, demonstrative, had been explicitly mentioned previously, or were inferable from the discourse context (see Prince, 1981; Gundel et al., 1993). Under these criteria, all of the following bolded objects were considered definite.

(105)   *NP with a Definite Determiner*:

    A-hendu    **la    "kue, kue"** he'i-va

    A1.Sg-hear **the blah   blah** 3.say-RC

    'I heard **the "blah, blah"** that they were saying.'

(106)   *NP with a Possessive Marker*:

    O-ñami piko **ñande-vaka**,         Liboria ra'e

    A3-milk QU   **B1.Incl.POSS-cow** Liboria RAE

    'Did Liboria milk **our cow** yet?'

(107)   *NP with a Demonstrative*:

    Kuehe    **pe   kamby** ja-'u       va'e-kue h-yakuã      porã-iterei.

    yesterday **that milk**    A1.Pl-drink RC-KUE 3-smell.good good-very

    '**That milk** we drank yesterday smelled wonderful.'

(108)   *Bare NP, Previous Mention*:

    Context: Juan just tried to grab the frog but failed, and the frog jumped far away.

    Juan-chi   nd-o-jura-i      **ju'í-pe**.

    Juan-DIM NEG-A3-grab-NEG **frog-PE**

    'Juan couldn't grab **the frog**.'

---

[21]There is one (non-demonstrative) definite determiner in Guaraní, *la*, which alternates with a plural form *lo* according to Velázquez-Castillo (2004), although no instances of *lo* were found in the corpus. Since *la* is borrowed from Spanish, there is no inherent reason to assume that it behaves as a definite determiner in Guaraní as it does in Spanish. In calling *la* a definite determiner, I simply follow Tonhauser and Colijn (to appear, p. 10) who argue that NPs determined by *la* are discourse-old and uniquely identifiable.

(109)  *Bare NP, Inferable from the Discourse Context*:

The existence of the world is inferable, even without explicit previous mention.

Ja-korre   **mundo**.

A1.Pl-run **world**

'We run **the world**.'

Objects were ranked lower than definite if they were marked by the indefinite determiner *peteĩ* 'one' or if they were discourse-new and were not clearly inferable from the discourse context. Thus the bolded objects in the following examples were considered indefinite.

(110)  *NP with an Indefinite Determiner*:

O-guereko avei **peteĩ jagua**, **piru-'i      peteĩ**.

A3-have    also **one  dog    skinny-DIM one**.

'He also had **a dog, a little skinny one**.'

(111)  *Bare NP, Discourse New and Non-Inferable*:

The context, a narrative about going to a grandmother's house at a young age, does not imply the existence of cicadas and birds in the relevant domain of discourse.

Ro-heka      **ñakyrã ha   oi-me-raẽ-va       vyra-'i**  ro-juga-haguã      hese.

A1.Excl-search **cicada  CONJ A3-exist-after.all-RC bird-DIM** A1.Excl-play-PURP 3.OBJ.PRO

'We looked for **cicadas and any birds there** in order to play with them.'

Within this class of indefinite expression the hierarchy in (27) makes a distinction between specific and non-specific. An expression was considered specific if the speaker could reasonably be assumed to have a particular individual in the universe in mind when uttering it. Otherwise the expression was considered non-specific. Framing specificity in this way allows for a direct assessment of certain predictions in Bossong (1985b), who, in his analysis of DOM in Guaraní, appeals in part to whether the referent of the object is identifiable to the speaker or not. Under this criterion, the object in (110) is considered specific, since there is presumably one particular dog that someone is being reported to have owned, and the object in (111) is non-specific, since there is not presumbly a particular cricket, bird, or set of crickets and birds that the speaker is claiming to have looked for.[22] I also consider indefinite generics to be

---

[22] Interestingly, under these specificity criteria, *peteĩ* appears much more frequently in the corpus with specific objects than with non-specific ones, as shown in the following table.

non-specific for the purposes of this study, since the speaker presumably does not have an individual in mind when uttering them.[23] Thus the bolded object in (112) is considered non-specific, since it refers to animals generally:

(112)   O-ĩ-ndaje   raka'e peteῖ mitã tyre'ỹ   o-hayhu-va **mymba-kuéra-pe**.
        A3-be-REP before one    child orphan A3-love-RC **animal-PL-PE**
        'There once was an orphan who loved **animals**.'

### 3.2.3   Relative Animacy and Definiteness

Relative animacy (Rel-A) and definiteness (Rel-D) were coded as features with three values each: "+," "–," and "=." An object was assigned a value "+" for the relevant feature if it was more animate/definite than the subject, "–" if it was less animate/definite, and "=" it was equally animate/definite. The same animacy/definiteness criteria laid out above were applied to subjects in determining these relative values. This was somewhat complicated by the fact that Guaraní is a pro-drop language, and subjects are quite frequently null, as shown in the following examples:

(113)   a. Ha     upéi o-hecha sapy'a   **Juan-chi   ha     Pirulo** ju'í-pe.
           CONJ then A3-see    suddenly **Juan-DIM CONJ Pirulo** frog-PE
           'And then **Juan and Pirulo** suddenly saw the frog.'

        b. O-hecha ∅        ju'i  i-po-po.
           A3-see    **(SUBJ)** frog 3-jump-jump
           '**(They)** saw the frog jumping.'

In (113a), the subject *Juanchi ha Pirulo* 'Juan and Pirulo' is overtly realized. In (113b) there is no overt subject, though the implicit subject referent is the same as that in (113a). How should relative animacy/definiteness be assessed for such null subject cases?

    Given that there is still an implicit subject referent in the case of null subjects, relative animacy was fairly easy to calculate: the animacy of the object was compared to the animacy of the implicit subject

|                     | + *peteῖ* | – *peteῖ* | % with *peteῖ* |
|---------------------|:---------:|:---------:|:--------------:|
| specific object     | 7         | 11        | 39%            |
| non-specific object | 4         | 49        | 8%             |

A Fisher's Exact Test for statistical significance applied to these data results in a two-tailed p-value of 0.0041, which, assuming the standard maximum p-value for significance of .05, is very significant. It is not my purpose here to provide a detailed analysis of the distribution of *peteῖ*, but this preliminary data seem to suggest that *peteῖ* might be a specificity marker as well as an indefinite determiner, a possibility which, to my knowledge, has not yet been mentioned in the literature.

[23] I concede that this is not a given, and studies such as Carlson (1977) consider *kinds* to be individuals that generics refer to. I do not pursue this possibility in this study.

referent, which it turns out was overwhelmingly human in the corpus. Relative definiteness was more difficult to determine for null subjects, since "null" is not a category on the Definiteness Hierarchy in (27). How might null phrases be ranked with respect to other definiteness types? In making this decision, I drew from Gundel et al. (1993), who propose the Givenness Hierarchy in (28). The highest item on that hierarchy is "In Focus," which they define as being "not only in short term memory, but also at the current center of attention" (p. 297). They show in their study of Chinese, Japanese, Russian, and Spanish that null NPs were near categorically In Focus.[24] This provides some support for ranking null NPs at the top of the Definiteness Hierarchy. In search of at least *prima facie* confirmation of this choice, I assessed the level of "in-focus-ness" of 20 null subjects randomly selected from the corpus. Since Gundel et al. (1993) claim In Focus constituents are the current center of attention, I used a common topicality test, Referential Distance (RD), to determine how In Focus these subjects were (Givón, 1983). RD (to be discussed further with respect to topicality) is a measurement of the number of clauses that intervene between the constituent in question and the previous mention of its referent, with a lowest value of 1 (when the referent was mentioned in the immediately preceding clause) and an arbitrarily-set highest value of 20 (given to referents that are either mentioned more than 20 clauses ago or discourse-new, i.e. never having been mentioned in previous discourse). My assumption was that referents that are at the current center of attention are most likely to have been mentioned very recently, while referents that are not at the current center of attention are more likely to have been mentioned some time ago or not at all. Thus, I assumed a low RD value to indicate being In Focus according to Gundel et al.'s definition. The average RD value for the null subjects tested was 1.2. Considering the fact that the lowest possible RD value is 1 and the highest possible RD value is 20, I take this figure to be indication that null subjects are very much In Focus in Guaraní, and ought to be ranked higher than the other forms on the Definiteness Hierarchy. Since null objects were not considered in this study, this means that all objects co-occurring with null subjects receive a relative definiteness value of "–," since they are always lower on the Definiteness Hierarchy than their subjects.

### 3.2.4 GF-Ambiguity

I argued in C. Shain (2008) that the most promising definition of DOM-related GF-ambiguity in Guaraní is one in which a phrase is considered ambiguous if and only if none of the available information sources clearly assign GFs to the NPs in a clause (see §2.2.4).[25] In that study I identify three major domains in

---

[24] They also examined English, but those results are not relevant to this discussion, since Standard English does not have null subjects.

[25] Much of the following description draws heavily on my previous discussion in C. Shain (2008).

which GFs may be assigned to NPs in language: grammar, thematic fit, and context. By "grammar" I mean the morphosyntactic systems by which the GFs of NPs are overtly indicated. By "thematic fit" I mean "general knowledge of how events typically occur in the world," a definition borrowed from McKoon and Ratcliffe (2007, p. 270), according to which one argument may be more naturally interpreted as S or O than the other. By "context" I mean information specific to the discourse which cannot be gleaned from general world knowledge or grammatical cues, which again may bias interpretation of one NP as S or O over the other.[26] Regarding grammar, there are generally three methods by which grammars indicate the grammatical functions of the noun phrases in a clause: head-marking, dependent-marking, fixed word order, or a combination of the the above (terminology from Nichols, 1986, see also Sapir, 1921; M. C. Baker, 1996; Hawkins, 1986; de Vogelaer, 2007; van Everbroeck, 2003; Mallinson and Blake, 1981; Amberber and de Hoop, 2005).

Head-marking languages indicate GF-assignment via agreement inflection on the verb (the "head" of the clause). This inflection indicates what kinds of properties the argument which takes a certain GF must have: singular vs. plural, 1st, 2nd, or 3rd person, etc. The co-occurring NP that exhibits those properties is then interpreted as having that GF. Kiswahili is a head-marking language, exemplified by the following:

(114)   (Adapted from Deen, 2006, p. 231)

Juma a-na-m-pend-a                    Mariam.
Juma 3Sg.SM-PRES-3Sg.OM-like-IND Mariam

'Juma likes Mariam.'

In (114). the verbal prefix *a* indicates 3rd person subject, and the verbal prefix *m* indicates 3rd person object. In this case, both co-occurring NPs, *Juma* and *Mariam*, exhibit properties that match the requirements of both agreement morphemes, and the sentence remains ambiguous with respect to head-marking. In Kiswahili, this ambiguity is resolved by word order, which I discuss below.

Dependent-marking languages indicate the GFs of the NPs in a clause by means of morphemes that attach to the noun phrases themselves. Japanese is one such language, as is shown in the following example.

---

[26]For examples from Guaraní of thematic fit and contextual GF-assignment, see §3.2.4.

(115)    (Nichols, 1986, p. 61, citing Kuno, 1973, p. 129)

Boku ga    tomodati ni    hana    o    ageta.

1Sg    SUBJ friend        DAT flowers OBJ gave

'I gave flowers to my friend.'

Example (115) contains three NPs: *boku* 'I,' *tomodati* 'friend,' and *hana* 'flowers.' Each of these is marked by morphemes indicating grammatical function: *ga* for S, *o* for O, and *ni* for dative. By means of these markers, (115) is unambiguously interpreted as "I gave flowers to my friend."

Fixed word order languages assign grammatical functions to certain linear/structural positions in the clause. This is the case for English, where, canonically, the NP that precedes the verb is the subject and the NP that follows the verb is the object. Because of this word order constraint in English, example (116) cannot be interpreted with *the expedition* as the subject and *the lion* as the object.

(116)    The lion hunted the expedition.

In order to determine whether a Guaraní clause is GF-ambiguous, we must first understand how the grammar of Guaraní indicates which NPs in the clause take which grammatical function. How does Guaraní grammar make use of the aforementioned grammatical GF-assignment tools: (i) head-marking, (ii) dependent-marking, and (iii) word order? It turns out that Guaraní GF-assignment depends, in part, on all three.

Regarding (i), head-marking, as I discussed in §3.1.1, Guaraní has a system of verb agreement prefixes that index one (or in some cases both) of the NP arguments in a transitive clause. I will not rehash the presentation of this system already given above, but I leave it to the reader to verify that the verbal prefixes unambiguously indicate GF in all cases except those in which both S and O are 3$^{\text{rd}}$ person, since either argument bears the features indicated by the prefix, namely, not being a local participant in the discourse.

Regarding (ii), dependent-marking, Guaraní is a differential object marking language. As such, there are some instances in which an NP dependent is marked by *-pe*. In these cases, that object is unambiguously O, leaving only the S slot open to the remaining NP. However, since object marking is the variable under discussion for this study, GF-ambiguity is of course assessed irrespective of it. The question at hand is whether the appearance of a marker is driven to any degree by GF-ambiguity, which requires evaluation of whether or not the clause *would* exhibit GF-ambiguity *in the absence of* the marker. Given this, dependent marking does not play a role in the coding of GF-ambiguity for this study.

Regarding (iii), word order, my position here differs significantly from the one I took in C. Shain (2008), because I now acknowledge that I was operating under the incorrect assumption that word order does not play a role in the grammar of GF-assignment in Guaraní, the result of an erroneous interpretation of Tonhauser and Colijn (to appear). They investigated the interaction between word order variation and GF-ambiguity by presenting speakers with all six permutations of both NPs and the verb in each of the two following examples:

(117)  (Tonhauser & Colijn, to appear, pp. 29-30)

  a. **Juan** o-topa **peteĩ ita**.

   **Juan** A3-find **one   stone**

   '**Juan** found **a stone**.'

  b. **Juan** o-hecha **Maria**.

   **Juan** A3-see   **Maria**

   '**Juan** saw/sees **Maria**.'

Note that in (117a), thematic fit strongly biases GF-assignment, since the animacy asymmetry between *Juan* and *peteĩ ita* 'a stone' means that *Juan* is the only argument with the properties entailed by *topa* 'find' for its agent (sentient, perhaps volitional), that is without significant coercion. In (117b) this is not the case, however, since both *Juan* and *Maria* are equally capable of both seeing and being seen. Tonhauser and Colijn (to appear) show that *Juan* is always interpreted as the subject in the permutations of (117a), regardless of order. However, in the permutations of (117b) the first NP was consistently judged to be the subject. Tonhauser and Colijn provide this as evidence that the basic ordering of S and O in Guaraní is SO rather than OS. While in C. Shain (2008) I assumed that the above data were *inconclusive* as to whether OS readings of sentences like (117b) *could* arise but were simply dispreferred to some degree, I revise this position here in light of the fact that an OS reading of (117b) *is* in fact not good for speakers, either ruled out by the grammar or so severely dispreferred as to render any claims that (117b) exhibits GF-ambiguity highly suspect (Judith Tonhauser, p.c.).

In light of this, for the current study I do not consider clauses to be ambiguous if they exhibit SO ordering, with two major exceptions: wh-questions and nominal heads of relative clauses. Wh-phrases are usually realized clause-initially (135/141 wh-phrases in the corpus are clause-initial). The question in (118) is one example, in which the wh-word *máva* precedes the verb and object.

(118)  **Mava** pio pe    amo tape o-mboty-va   hína? ...

       **who**   QU that that  path A3-cover-RC PROG ...

       '**Who** is that blocking the road? ...'

Relative clauses are postnominal, in which case the nominal head always precedes its modifying clause (no overt nominal heads in the corpus follow a modifying relative clause). The sentence in (82a) exemplifies this. The nominal head, *la sirujuano* 'the surgeon,' precedes the relative clause *por kasualidad oĩva avei* 'who happened to be there, too.'

(82)  a. Pero    o-topa *hikuái*    **la**    **sirujuano**  **por**  **kasualidad o-ĩ-va**    **avei**

         but(Sp) A3-find *3Pl.SUBJ.PRO* **the(Sp) surgeon(Sp) by(Sp) chance(Sp) A3-be-RC also**

         'But *they* found **the surgeon who happened to be there, too**.'

These constraints apply regardless of whether the wh-phrase or nominal head is O or S, resulting in the possibility of either OS or SO realization in these particular cases. Thus otherwise ambiguous clauses involving wh-words as objects or nominal heads of relatives as objects of the subordinated verb were also considered ambiguous with respect to grammar.

In addition to head-marking, dependent-marking, and word order, there is an additional grammatical GF-assignment disambiguation in Guaraní. When one of the NPs in a transitive clause is *ha'e* '3Sg.SUBJ.PRO' or *hikuái* '3Pl.SUBJ.PRO,' which are subject-only pronouns, only the object GF remains open to the remaining NP. This is exemplified in (82a).

(82)  a. Pero    o-topa *hikuái*    **la**    **sirujuano**  **por**  **kasualidad o-ĩ-va**    **avei**

         but(Sp) A3-find *3Pl.SUBJ.PRO* **the(Sp) surgeon(Sp) by(Sp) chance(Sp) A3-be-RC also**

         'But *they* found **the surgeon who happened to be there, too**.'

In (82a), *hikuái* co-occurs with *la sirujuáno por kasualidad oĩva avei* 'the surgeon who just happened to be there, too.' Since *hikuái* is a subject-only pronoun, the object function is the only one available to the other NP, and this sentence also exhibits no GF-ambiguity.

To summarize the tools of GF-assignment in Guaraní grammar: the verb agreement system shown in Table 3.1 allows GF-ambiguity always and only when both NPs are in 3rd person. Additionally, there exists an object marker *-pe* which demarcates NPs as objects when it attaches to them. The order SO becomes fixed for otherwise ambiguous clauses, and therefore clauses with SO order are not GF-ambiguous. In necessarily-transitive constructions with only one overt NP, that NP must be the object. Finally, when NPs co-occur with the subject-only pronouns *ha'e* or *hikuái* they are obligatorily objects.

72

As I mentioned above, I also considered thematic fit (TF) and context to disambiguate GF-assignment. I considered a clause to be non-GF-ambiguous with respect to TF if one NP was much more typical according to world knowledge as a subject or object than the other. For example, in (119), "gentleman" fits much better into the subject role of *mongakuaa* 'raise' than "child" does given world knowledge, since adults generally raise children, and it was therefore coded as unambiguous with respect to thematic fit.

(119)  O-mo-ngakuaa karai      **pe  mitã**.
       A3-CAUSE-grow gentleman **that child**
       'The gentleman raised **the child**.'

Similarly, I considered context to disambiguate GF-assignment if knowledge about information specific to the discourse the information structure of the discourse would indicate which NP took which grammatical function.[27] This is shown in example (120). In the narrative preceding this sentence, the frog has been sneaking up behind Juanito and Pirulo, and thus the frog already sees them but not the other way around. Given this information, *ju'i* is much more likely to be the object of a sudden/surprising "seeing" event, and this clause was coded as unambiguous with respect to context.

(120)  Ha  upéi o-hecha sapy'a     Juán-chi  ha  **Pirúlo ju'í-pe**.
       and then A3-see   suddenly Juan-DIM and **Pirúlo frog-PE**
       'And then suddenly Juanito and Pirulo saw **the frog**.'

Summing up, a clause was considered to be non-GF-ambiguous if the intended mapping of GFs to NPs in the clause was unambiguously indicated either by grammar, thematic fit, context, or some combination of the three. Otherwise, the clause was considered GF-ambiguous. Although I argued that GF-ambiguity is the best predictor of *-pe*-marking in the corpus in C. Shain (2008), I drew this conclusion on the basis of a premise that I now reject, namely that word order was of no consequence to GF-assignment. In this study, for reasons given above, I consider otherwise ambiguous clauses that exhibit SO order to be unambiguous (with the exception of wh-questions and heads of relative clauses).

### 3.2.5   Thematic Fit

As I mentioned in §2.2.4 and §3.2.4, I assume thematic fit (TF) to be a source of GF-assignment information. That is to say, an NP fits thematically with the subject or object GF in a clause if general world knowledge would indicate that the referent of that NP has properties consistent with the role in question.

---

[27]For a discussion of the interaction between information structure and GF-assignment, see the various studies in Li (1976).

TF biases GF-assignment if one assignment of S and O to the NPs in a clause results in a more typical rela-tion than the opposite assignment. Thus, as I discussed above, in (117a) TF provides such a bias, since one NP is a much more typical subject of the verb than the other, while in (117b) it does not, since both NPs are equally typical as either the subject or object of the verb. While it is true that no clauses in the corpus are ambiguous according to the definition of GF-ambiguity given in §3.2.4, there is nevertheless reason to hypothesize that DOM in Guaraní may be sensitive to thematic fit, which I discuss in §4.2. Because of this I test the relevance of TF to DOM in this study, despite the fact that GF-ambiguity itself no longer seems to be a factor. Since TF is essentially non-linguistic in that it is based on intuitions about the way the world works, operationalizing it was fairly straightforward. If the referent of either NP seemed equally typical in either grammatical function, a value of "+" was assigned to the clause, indicating that it was [+ambigu-ous] with respect to TF. If world knowledge clearly suggested one GF-assignment as more typical than the alternative, a value of "–" was assigned to the clause, indicating that it was [–ambiguous] with respect to TF.

### 3.2.6   Topicality

In assessing the topicality of a given object, I relied on two topicality measures presented in Givón (1983): (i) Referential Distance (RD) and (ii) Topic Persistence (TP), also known as "look-back" and "look-ahead," repectively (Schwenter & Silva, 2002). The first I defined in §3.2.3 as a count of the number of clauses in-tervening between the NP in question and the previous mention of its referent. The latter is a count of consecutive subsequent clauses in which a mention of the referent of the NP in question appears. Low RD is assumed to indicate high topicality, as is high TP. Note that according to these measures topicality is a graded property, rather than a binary contrast between [± topical]. With respect to TP, other researchers, such as Myhill (1992), have employed an alternative measure that counts the number of subsequent men-tions within some arbitrarily determined number of following clauses, regardless of consecutiveness. For the sake of completeness, I use both measurement types in this study, calling the first "contiguous TP" (TPc) and the second "non-contiguous TP" (TPnc). For this study, I counted TPnc over the next 20 clauses following the object in question. I limited the definition of "clause" to "matrix clause," such that embedded clauses or relative clauses were not separately counted in the topicality measures.

Some have additionally noted that it might be desirable to be able to perform some sort of combina-tion of RD and TP to produce a measurement of "total topicality" (Scott Schwenter, p.c.). This is challeng-ing for three reasons. First, RD and TP are counting different things. RD is a measure of intervening clauses

between mentions, while TP is a measure of clauses in which mentions occur. Thus there is an apples and oranges problem in trying to combine them. Second, RD and TP values by nature relate oppositely with topicality: RD value is inversely related to topicality, while TP value is directly related to it. There is thus no readily apparent operation by which the two values might be directly combined to indicate total topicality. Third, there is an asymmetry in measurement between "look-back" and "look-ahead": since RD is backward-looking and TP is forward-looking, the domains they range over are mutually exclusive, and since look-back and look-ahead are measured differently, the results obtained for each are incomparable. Thus, given an object of RD = 20 and TP = 20, should it be considered high or low in topicality? More importantly, is it higher or lower in topicality than an object of RD = 0 and TP = 0? Beyond gut instinct, the measures as they stand offer no answers to these questions.

For this study, I implemented a slight revision to the above that I believe alleviates the many of the above problems and provides a simple means of measuring total topicality: I included a Forward RD (F-RD) and a Backward TP (B-TP, names courtesy of Scott Schwenter).[28] Using such tools enables us to measure total topicality straightforwardly: the total topicality value is simply the sum either of RD and F-RD (T-RD) or of TP and B-TP (T-TP).[29] Thus every eligible object in the corpus was assessed with respect to all seven of these topicality measures:

(121)    a.  Referential Distance (RD)

b.  contiguous Topic Persistence (TPc)

c.  non-contiguous Topic Persistence (TPnc)

d.  Forward Referential Distance (F-RD)

e.  Backward Topic Persistence (B-TP)

 f.  Total Topicality in terms of Referential Distance (T-RD)

g.  Total Topicality in terms of Topic Persistence (T-TP)

A word must be said about what counts as a "mention" of the same referent. Following Givón (1983), I only considered the discourse referent of an NP to be mentioned elsewhere if that referent was explicitly denoted by another NP or served implicitly as a semantic argument of another proposition. Thus hyper-

---

[28] For this study, I arbitrarily assumed a non-contiguous version of B-TP.

[29] I was unable to overcome the incomparability between counting intervening clauses and counting mentions, with the result that I must propose two distinct measures of topicality: T(otal)-RD and T(otal)-TP. However, for reasons I explain later, since these are "total" measures I believe them to be more informative than the two previously-proposed measures, RD and TP. Furthermore, T-RD and T-TP can be seen as quantifying topicality in two different senses: T-RD measures *proximity* of mention in discourse, and T-TP measures *density* of mention in discourse.

nym/hyponym pairs were not considered mentions of the same referent. An example of this is provided below:

(122) O-ĩ-ndaje  raka'e peteĩ mitã tyre'ỹ  o-hayhu-va **mymba-kuéra-pe**.

A3-be-REP before one   child orphan A3-love-RC **animal-PL-PE**

'There once was an orphan who loved **animals**.'

(123) Ha'e         h-éra   Juan-chi ha   o-guereko **peteĩ jagua h-éra-va    Piruli**.

3Sg.SUBJ.PRO 3-name Juan-DIM CONJ A3-have **one   dog   3-name-RC Piruli**

'He was named Juanito and he had **a dog named Piruli**.'

The above examples appear consecutively in the corpus. In (122), the object is the generic phrase *mymbakuérape* 'animals.' In (123), a particular animal is mentioned. Though the dog *Piruli* is implicated in the proposition that the orphan loves animals, I did not consider the above object expressions to refer to the same entity, and they were thus not counted as mentions of each others' referents in my assessment of topicality.

In addition to providing the aforementioned topicality measures, Givón (1983, p. 9) proposes three types of topic, as shown below, where a "chain" is taken to be a "string of clauses whose *main/primary* topic remains the same":

(124)   a. *Chain-initial topic*:

    i.  Newly introduced, changed, or returned to

    ii. *Discontinuous* in terms of the preceding discourse context

    iii.  Potentially *persistent* in terms of the succeeding discourse context

  b. *Chain-medial topic*:

    i.  Continuing/continuous topic in terms of the preceding discourse context

    ii. Persistent in terms of the succeeding discourse context

  c. *Chain-final topic*:

    i.  Continuing/continuous topic in terms of the preceding discourse context

    ii. Non-persistent topic in terms of the succeeding discourse context

These topic types are illustrated by the following example:

(125)   I had to take **[1] Marlo** to the vet yesterday morning. **[2] The poor guy** had been retching all night. **[3] He** really looked terrible.

In the above mini-discourse, the single referent of the bolded phrases is a different type of topic in each clause. Expression [1], the proper name *Marlo*, introduces a chain-initial topic.[30] Expression [2], the descriptive NP *the poor guy*, refers to a chain-medial topic, since its referent is discourse old (it was mentioned in the previous clause) and persistent (it will be mentioned in the next clause). Finally, expression [3], the pronoun *he*, is chain-final, since there are no subsequent mentions of the referent of [3], but that referent has nevertheless been recently talked about.

Given this, note that an advantage of using total topicality measures like T-RD and T-TP is that they are both able to reflect the fact that chain-initial and chain-final topics are more topical than non-topics, thus distinguishing them from non-topics in a way that RD alone or TP alone cannot. Taking only look-back RD into account, for example, there is no way to tell whether a referent of RD 20 (discourse new or not recently mentioned) is being newly introduced as the topic of discourse or is simply a non-topical referent fleetingly entering and exiting the discourse. The same holds for look-ahead TP: given that a referent has a low TP, there is no way to tell (in terms of TP alone) whether this is because the referent is very non-topical or because the referent had been a topic of a chain that is ending. However, since T-RD and T-TP take both look-ahead and look-back into consideration, they will reflect measurable differences in value between (i) chain-initial topics and non-topics, as well as (ii) chain-final topics and non-topics. Regarding (i), both chain-initial topics and non-topics will exhibit an RD of 20 and a B-TP of 0. However, chain-initial topics will also exhibit a low F-RD and a high T-TP, while non-topics will exhibit a high F-RD and a low T-TP. Thus T-RD should be significantly lower and T-TP significantly higher for clause-initial topics than for non-topics. The same applies to (ii): although both chain-final topics and non-topics will exhibit a high F-RD and a low TP, they will show opposite relationships with respect to RD and B-TP. Thus, chain-final final topics should be significantly higher in T-TP and lower in T-RD than non-topics.

## 3.3   Statistical Methods

Throughout the remainder of this study, statistical significance regarding binary feature values is determined via Fisher's Exact Test for independence as computed by the QuickCalcs™ web application at http://www.graphpad.com, a choice made on the basis of the fact that Fisher's Test produces accurate (exact) significance results for 2x2 contingency tables with small sample sizes[31] and of all frequency

---

[30]Assuming this story to be taken in isolation, with no preceding reference to Marlo.

[31]The rule of thumb for the $\chi^2$ test is that the expected number of tokens in each cell should be greater than or equal to 5 (Daya, 2002). In my case, several tables fail this rule. However, Fisher's Exact Test works for tables in which one or more expected values are less than 5.

distributions, unlike the $\chi^2$ test, for example (Fisher, 1935; Daya, 2002; Motulsky, 1995; Bower, 2003). The QuickCalcs™ software executes Fisher's test by summing small P values. Statistical significance of the differences in scalar feature values, such as topicality measurements, was determined using the unpaired $t$ test calculator at http://www.graphpad.com. All P values presented here are two-tailed, and I assume a maximum P value for significance of 0.05.

# CHAPTER 4

# THE DISTRIBUTION OF DOM IN GUARANÍ

In the previous chapter, I circumscribed the DOM variable for this study and provided criteria according to which DOM variants were classified with respect to various factor groups of potential significance to DOM. In this chapter I discuss my findings. In §4.1 I briefly survey some of the claims about DOM in the descriptive literature on Guaraní. In §4.2 I provide speaker data that I believe challenges any account of DOM stated in terms of categorical distributional constraints. I dedicate §4.3 – §4.8 to laying out the results from the corpus with respect to each factor group. Then in §4.10 I present my analyses of the distribution of DOM in Guaraní on the basis of these data, which show *-pe* to be a marker of human topics. In §4.11 I bring these results to bear on a previous analysis of DOM in Guaraní, Bossong (1985b), and discuss some weaknesses in his approach that they reveal. I conclude in §4.12.

## 4.1   Some Previous Proposals

Gregores and Suárez (1961, pp. 222-3) point to the existence of DOM in Guaraní in their categorization of nouns. According to them: "Syntactically, there are two large classes of nouns, according to the function they fill in a constitute which is the object of a transitive verb: some nouns occur as the center or only constituent of that constitute, other nouns occur in it as the axis of the postposition pe." In other words, direct objects may be *-pe*-marked or not in Guaraní, a distinction which Gregores and Suárez, if I understand them correctly, take to be one of *noun class*. According to them, the distinction between these classes is essentially reducible to animacy. More specifically, they claim that object marking appears when "the object refers to a person, and sometimes to animals; this rule has some exceptions (Gregores & Suárez, p. 223)." However, there is some evidence from the corpus that analyzing DOM in terms of noun

class results in incorrect predictions. For example, the following sentences from the corpus are problematic for this approach, since the same noun, *ju'i* 'frog,' is marked in one case and unmarked in another.

(75)  b.  O-jahú-ta-vo         peteĩ mba'e-ry-rú-pe        y-potĩ-porã-me      o-hecha **ju'i**
          A3-bathe-FUT-when one   thing-liquid-recipient-PE water-clean-pretty-PE A3-see  **frog**

          o-po-po         o-hó-vo     h-apykuéri-kuéra.
          A3-jump-jump A3-go-when 3.POSS-behind-PL

          'And as he prepared his bath with clean water, he saw **the frog** coming, jumping and jumping

          behind them.'

(126)  O-hasá-vo        upéi o-hecha **ju'í-pe** o-guapy y      mbyté-pe   peteĩ yrupẽ-ari.
       A3-pass-when then A3-see  **frog-PE** A3-sit    water middle-PE one   sieve-on

       'In passing, he saw **a frog** sitting in the middle of the water on a water lily.'

If we analyze *ju'i* 'frog' as being of the unmarked class, we fail to predict (126), and if we analyze it as being
of the marked class, we fail to predict (75b). The existence of nouns which can be marked in one context
and unmarked in another suggests that an analysis in terms of noun class is incomplete.

   Other scholars have noted the variation in object marking in Guaraní as well. Velázquez-Castillo (2004,
p. 1426) analyzes *-pe* as a marker of human-referring objects. Tonhauser (2006, p. 132) claims that object
marking is optional if the subject is 1st or 2nd person or more animate than the object. While this is not
stated, I take this to imply that object marking is obligatory if the subject is 3rd person and less than or
equal to the object in animacy. Perhaps the most thorough examination in the literature of the distri-
butional constraints on DOM in Guaraní is found in Bossong (1985b), who argues that *-pe*-marking is
conditioned by both animacy and definiteness. I discuss his proposal in more detail in §4.11. A common
feature of these approaches is the appeal to animacy as an important factor group to the distribution of
DOM in Guaraní. I believe this intuition to be born out by the data discussed in the current chapter, as
shown in §4.3.

## 4.2   *-pe*-Marking and Acceptability Judgments

Before I turn to the corpus itself, I believe it will be helpful to outline the space of *possible* variation of DOM
in Guaraní. As I pointed out in C. Shain (2008), this is something a purely corpus-based analysis cannot
do. The only information available from a corpus is whether a particular phenomenon is attested and, if
so, how frequently and in what contexts it appears. However, non-attestation does not constitute proof, or

even strong evidence, that the phenomenon is ungrammatical. To determine ungrammaticality, acceptability judgments are required. I am fortunate enough to have access to some telling acceptability data recently gathered by Judith Tonhauser in Paraguay. Four Guaraní speakers were consulted concerning the acceptability of *-pe*-marking on objects of varying degrees of animacy, definiteness, relative animacy, relative definiteness, thematic fit bias, and affectedness.[1] I present the results in Table 4.1, and subsequently discuss its implications with respect to the above factor groups.

| | | | SPEAKERS | | | |
|---|---|---|---|---|---|---|
| | PROMPT | MARKING | 1 | 2 | 3 | 4 |
| 1. | Melvin o-hecha **pe óga** | +*pe* | ✓ | ✓ | + | * |
| | 'Melvin sees **that house**' | −*pe* | + | + | + | + |
| 2. | Anita o-hupi **pe kyse** | +*pe* | ✓ | ✓ | + | * |
| | 'Anita lifts **that knife**' | −*pe* | + | + | + | + |
| 3. | Juan o-nupã **pe ita** | +*pe* | + | + | + | * |
| | 'Juan hits **that stone**' | −*pe* | + | + | + | + |
| 4. | Juan o-jagara **ju'i** | +*pe* | + | ✓ | + | + |
| | 'Juan grabs **the/a frog**' | −*pe* | + | + | + | * |
| 5. | Pe kava o-jopi **pe vaka** | +*pe* | + | + | + | + |
| | 'That wasp stings **that cow**' | −*pe* | ✓ | ✓ | + | * |
| 6. | Klara o-pysyrõ **peteĩ jagua** | +*pe* | + | + | + | + |
| | 'Klara saves **a dog**' | −*pe* | ✓ | + | + | * |
| 7. | Marco o-su'u **pe jagua** | +*pe* | na | + | + | + |
| | 'Marco bites **that dog**' | −*pe* | na | ✓ | + | * |
| 8. | Anita o-hecha **peteĩ kuña** | +*pe* | + | + | + | + |
| | 'Anita sees **a woman**' | −*pe* | ✓ | + | + | + |
| 9. | Juan o-ñapytĩ **pe mitã** | +*pe* | + | + | + | + |
| | 'Juan ties up **that child**' | −*pe* | ✓ | ✓ | + | * |

Table 4.1: Acceptability of marking/non-marking for various DOs

Objects are in bold. "+" = preferred. "✓" = acceptable. "*" = unacceptable. "na" = no judgments available.

It seems the clearest generalization to be gleaned from these data is that they are messy. The judgments are not uniform across speakers: for speaker 1, 2, and 3, no instance of either marking or non-marking was ungrammatical. Speakers 1 and 2 expressed preferences for one alternant over the other in certain circumstances, while speaker 3 considered all marking options equally good. For speaker 4, on the other hand, there was a categorical contrast in acceptability in all but one case (line 8). However, the fact that *-pe*-marking is not ruled out in any case for three of four speakers is actually very interesting. Contrast this with Spanish, for example, in which there are many cases for which a certain variant is simply unacceptable, as shown in (22), (47), and (55). I think it is reasonable to conclude from Table 4.1 that *-pe*-marking on any DO is grammatical in Guaraní, and that preference for one alternant over another arises

---

[1]As a discourse property, topicality is not directly accessible from elicitation data.

from other factors than grammatical rules governing object marking. It is the reponsibility of this analysis to elucidate what those other factors are.

There is more to be learned, I believe, from the data in Table 4.1. Despite differences in degree, the speakers showed similar intuitions as to which alternant was preferable, except in the case of line 4, where speaker 2 preferred the unmarked object and 4 preferred the marked one. Assessing these preferences across object features may provide some valuable insights.

|  | PROMPT | ANIMACY | MARKING | SPEAKERS | | | |
|---|---|---|---|---|---|---|---|
|  |  |  |  | 1 | 2 | 3 | 4 |
| 1. | Melvin o-hecha **pe óga** | I | +*pe* | ✓ | ✓ | + | * |
|  | 'Melvin sees **that house**' |  | −*pe* | + | + | + | + |
| 2. | Marco o-su'u **pe jagua** | A | +*pe* | na | + | + | + |
|  | 'Marco bites **that dog**' |  | −*pe* | na | ✓ | + | * |
| 3. | Juan o-ñapytĩ **pe mitã** | H | +*pe* | + | + | + | + |
|  | 'Juan ties up **that child**' |  | −*pe* | ✓ | ✓ | + | * |

Table 4.2: Animacy and acceptability
Objects are in bold. "+" = preferred. "✓" = acceptable. "*" = unacceptable. "na" = no judgments available.

Take, for example, animacy, exemplified in Table 4.2. There it is clear that non-marking is preferred for the inanimate DO, while marking is preferred for the animate and human DOs. In the absence of data from speaker 1 as to the acceptability of marking on the animate DO, it appears that the preferences are identical between the animate and human DOs. Thus this data might indicated that a relevant distinction to DOM in Guaraní is [± animate]. Definiteness also seems to interact with marking preference. As Table 4.3 shows, non-marking is less preferable for some speakers when the object is definite than when the object is indefinite.

|  | PROMPT | DEFINITENESS | MARKING | SPEAKERS | | | |
|---|---|---|---|---|---|---|---|
|  |  |  |  | 1 | 2 | 3 | 4 |
| 1. | Anita o-hecha **peteĩ kuña** | indefinite | +*pe* | + | + | + | + |
|  | 'Anita sees **a woman**' |  | −*pe* | ✓ | + | + | + |
| 2. | Juan o-ñapytĩ **pe mitã** | definite | +*pe* | + | + | + | + |
|  | 'Juan ties up **that child**' |  | −*pe* | ✓ | ✓ | + | * |

Table 4.3: Definiteness and acceptability
Objects are in bold. "+" = preferred. "✓" = acceptable. "*" = unacceptable. "na" = no judgments available.

Interestingly, relative definiteness and animacy do not seem to have a great deal to do with marking preferences, as Table 4.4 demonstrates. In the first line of the table, the subject and object are equally definite and animate. In the second line, the object is less animate and less definite than the subject. However, the judgments about the acceptability of marking/non-marking are quite similar between the

82

| | PROMPT | REL-D | REL-A | MARKING | SPEAKERS 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|
| 1. | Pe kava o-jopi **pe vaka** | = | = | +*pe* | + | + | + | + |
| | 'That wasp stings **that cow**' | | | −*pe* | ✓ | ✓ | + | ∗ |
| 2. | Klara o-pysyrõ **peteĩ jagua** | − | − | +*pe* | + | + | + | + |
| | 'Klara saves **a dog**' | | | −*pe* | ✓ | + | + | ∗ |

Table 4.4: Relative definiteness/animacy and acceptability

Objects are in bold. A value of "=" means that the object is equal to the subject according to the relevant factor group, and a value of "−" means that the object is less than the subject according to the relevant factor group. Speaker judgments are represented as follows: "+" = preferred; "✓" = acceptable; "∗" = unacceptable; "na" = no judgments available.

two, with the exception of 2's slight difference in judgment. This might suggest that DOM in Guaraní is not conditioned by the properties of the object relative to the subject.

| | PROMPT | TF | MARKING | SPEAKERS 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|
| 1. | Melvin o-hecha **pe óga** | more biased | +*pe* | ✓ | ✓ | + | ∗ |
| | 'Melvin sees **that house**' | | −*pe* | + | + | + | + |
| 2. | Juan o-nupã **pe ita** | less biased | +*pe* | + | + | + | ∗ |
| | 'Juan hits **that stone**' | | −*pe* | + | + | + | + |

Table 4.5: Thematic fit and acceptability

Objects are in bold. "+" = preferred. "✓" = acceptable. "∗" = unacceptable. "na" = no judgments available.

Another important piece of information to be gleaned from Table 4.1 is the interaction of thematic fit (TF) and marking preference (for discussion of TF, see §2.2.4, §3.2.4, and §3.2.5). While none of the above examples is ambiguous according to the previously laid out ambiguity criteria, there is nevertheless a difference in how strongly TF biases GF-assignment. I have labeled line 1 more biased than line 2 according to TF, because in line 2 one GF-assignment is much more typical than its alternative (the house seeing Melvin would be atypical), while in line 1 both arrangements are similarly typical in terms of world knowledge (stones can hit people and people can hit stones). In light of this, it is interesting to note that both 1 and 2 found -*pe*-marking to be more acceptable when TF did not provide a strong bias than when it did.

Let me make one final comment about the speaker data, this time concerning affectedness. As I mentioned previously, due to difficulties of operationalization, I did not take affectedness into account during corpus annotation. However, there are some examples in Table 4.1 that I intuit to exhibit an affectedness contrast. I present two of them in Table 4.6. In the first line of Table 4.6, the object is completely unaffected: neither it nor any of its parts undergo any change of state. In the second line, the object is causally affected by Anita's lifting action and undergoes a change of location. However, the markedness judgments

for all speakers between the two examples are identical, suggesting that affectedness is not a property to which the system of DOM in Guaraní is sensitive.

| | Prompt | Affectedness | Marking | Speakers 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|
| 1. | Melvin o-hecha **pe óga** | less | +*pe* | ✓ | ✓ | + | * |
| | 'Melvin sees **that house**' | | −*pe* | + | + | + | + |
| 2. | Anita o-hupi **pe kyse** | more | +*pe* | ✓ | ✓ | + | * |
| | 'Anita lifts **that knife**' | | −*pe* | + | + | + | + |

Table 4.6: Affectedness and acceptability

Objects are in bold. "+" = preferred. "✓" = acceptable. "*" = unacceptable. "na" = no judgments available.

I must note that these data are far from conclusive. None of the variables is systematically tested or isolated. They are presented simply as a preliminary exploration of which hypotheses of the distribution of DOM in Guaraní might be promising, and which might not be. They are also, more importantly, intended to highlight the fact that no clear categorical constraints on the distribution of object marking seem to hold in Guaraní, unlike some other differentially object marking languages such as Spanish. Therefore, whatever patterns emerge from the corpus must be considered patterns of usage distinct from patterns of grammaticality. Having explored the range of objects that *can* be marked/unmarked in Guaraní (possibly all of them), I now turn to the question of which objects *are* in fact marked in actual discourse.

## 4.3   Animacy as a Condition on DOM

Here I present the distribution of marked and unmarked objects across the categories of the Animacy Hierarchy. Animacy and -*pe*-marking are cross-tabulated in Table 4.7.[2]

| | +*pe* | −*pe* | % Marked |
|---|---|---|---|
| Human | 24 | 19 | **56%** |
| Animate | 3 | 55 | **5%** |
| Inanimate | 0 | 132 | **0%** |
| Total | 27 | 206 | 12% |

Table 4.7: Animacy and -*pe*-marking

As Table 4.7 shows, -*pe*-marking in Guaraní is closely related to animacy. No inanimate objects in the entire corpus receive an object marker, a statistically significant deviation from the mean frequency of 12% (p < 0.0001). Among animate objects, 5% are -*pe*-marked, a statistically significant increase in marking

---

[2]Recall that the only data represented in the statistical analysis are those from the primary corpus. Appearing throughout this thesis are examples labeled as being from the "extended corpus," which I did not systematically analyze. Thus these examples do not factor into the figures presented here. For details on the corpora, see the Appendix.

frequency from that of inanimate objects (p= 0.0274). Finally, 56% of human objects are marked, a much greater frequency than that of both animate objects (p < 0.0001) and inanimate objects (p < 0.0001). Thus the relationship between rank on the Animacy Hierarchy and likelihood of object marking appears to be quite strong.

## 4.4 Definiteness as a Condition on DOM

In this section I discuss the definiteness-related corpus results. In Table 4.8[3] I provide a cross-tabulation of the Definiteness Hierarchy and *-pe*-marking.

|  | +*pe* | −*pe* | % MARKED |
|---|---|---|---|
| PROPER NAME | 3 | 4 | **43%** |
| DEFINITE | 18 | 135 | **12%** |
| INDEFINITE SPECIFIC | 4 | 16 | **20%** |
| INDEFINITE NON-SPECIFIC | 2 | 51 | **4%** |
| TOTAL | 27 | 206 | 12% |

Table 4.8: The Definiteness Hierarchy and *-pe*-marking

The definiteness data in Table 4.8 are not as illuminating as the animacy data in Table 4.7. Marking frequencies do not perfectly line up with position on the hierarchy, since indefinite specific objects (20%) are more frequently marked than definite ones (12%). These data further show the relationship between the Definiteness Hierarchy and the distribution of DOM to be fairly weak in Guaraní. Among all six possible pairings of values on the Definiteness Hierarchy, only three exhibit statistically significant differences in marking frequency from one another: proper names and definites (p = 0.0485), proper names and indefinite non-specifics (p = 0.0092), and indefinite specifics and indefinite non-specifics (p = 0.0327). Furthermore, only one of the categories on the Definiteness Hierarchy shows statistically significant departure in marking frequency from the total mean of 12%: proper names (p = 0.0438). Thus, while there is not a clear increase in frequency of object marking along the Definiteness Hierarchy, there is a significant difference in marking between the extremes (proper name and non-specific). However, 6/7 proper names in the corpus are human-referring. Given Table 4.7, in which human objects are shown to be much more frequently marked than non-human objects, I take this proper name effect to be reducible to a humanness effect, especially considering that the marking frequencies of humans (56%) and PNs (43%) are quite comparable. Thus it appears that location on the Definiteness Hierarchy itself is not a strong predictor of *-pe*-marking.

---

[3]As discussed in §3.1.4, no object pronouns were considered in this study for independent reasons. Therefore the "pronoun" category on the Definiteness Hierarchy in (27) is omitted for clarity in Table 4.8.

This of course could simply be a function of sample size, and a larger corpus might exhibit more significant distinctions between definiteness categories. Even if this were so, the lack of significance of the Definiteness Hierarchy in tandem with the strong significance of the Animacy Hierarchy with respect to DOM within the same sample suggests that animacy is a much more important underlying condition on DOM than definiteness in Guaraní.

Before moving on to the other factor groups, I would like to remark on some other potentially consequential definiteness-related data. While the four-way contrast of the Definiteness Hierarchy does not appear to be a promising predictor of DOM, it is possible that collapsing levels on the hierarchy will yield more significant results, since the tokens would not be spread out over so many factors. Two possible collapsing schema come to mind: it might be the case that DOM in Guaraní is conditioned by a binary definite/indefinite contrast, or, assuming along with Enç (1991) that [+definite] entails [+specific], it might be the case that DOM is based in part on a specific/non-specific contrast. In the first case, the "proper name" and "definite" categories on the Definiteness Hierarchy would be collapsed into "definite," and both indefinite categories would be collapsed into "indefinite." In the second case, the "proper name," "definite," and "indefinite specific" categories on the Definiteness Hierarchy would be collapsed into "specific," and would contrast with "indefinite non-specific." The results for a binary definite/indefinite contrast are presented in Table 4.9, and the results for a binary specific/nonspecific contrast are presented in Table 4.10.

|  | +*pe* | −*pe* | % Marked |
|---|---|---|---|
| Definite | 21 | 139 | **13%** |
| Indefinite | 6 | 67 | **8%** |
| Total | 27 | 206 | 12% |

Table 4.9: Definiteness and -*pe*-marking

|  | +*pe* | −*pe* | % Marked |
|---|---|---|---|
| Specific | 25 | 155 | **14%** |
| Non-specific | 2 | 51 | **4%** |
| Total | 27 | 206 | 12% |

Table 4.10: Specificity and -*pe*-marking

While definites (13%) are more likely to be marked than indefinites (8%), this difference in frequency is not statistically significant ($p = 0.3782$). Nor is there a statistically significant difference in marking frequency between definites/indefinites and the total mean of 12% ($p = 0.6421$ and $p = 0.5198$, respectively), which means that the definiteness contrast as defined for Table 4.9 is not a strong predictor of object-marking. The specificity data in Table 4.10 are more promising: there is a 10% difference in marking fre-

quency between specific (14%) and non-specific (4%) direct objects. This difference is not large, in the sense that knowing whether an object is specific or non-specific does not greatly influence our ability to predict (non-)marking for that object. However, the difference is still statistically significant (p = 0.0499). Nevertheless, as with definiteness above, specific/non-specific objects do not exhibit a statistically significant difference in marking frequency from the total mean (p = 0.7546 and p = 0.1277, respectively). Thus the corpus does not show a definiteness effect with respect to DOM, although there appears to be a small DOM-related specificity effect.

## 4.5 Relative Animacy and Definiteness as Conditions on DOM in Guaraní

I now assess the degree to which relative animacy (Rel-A) and/or relative definiteness (Rel-D) play a role in determining whether a given object will be *-pe*-marked or not. In Table 4.11 I cross-tabulate *-pe*-marking and relative animacy level, and in Table 4.12 I cross-tabulate *-pe*-maring and relative definiteness level.[4]

|        | +*pe* | −*pe* | % MARKED |
|--------|-------|-------|----------|
| O > S  | –     | –     | –        |
| O = S  | 25    | 20    | **56%**  |
| O < S  | 2     | 186   | **1%**   |
| TOTAL  | 27    | 206   | **12%**  |

Table 4.11: Relative animacy and *-pe*-marking

|        | +*pe* | −*pe* | % MARKED |
|--------|-------|-------|----------|
| O > S  | –     | –     | –        |
| O = S  | 2     | 10    | **17%**  |
| O < S  | 25    | 196   | **11%**  |
| TOTAL  | 27    | 206   | **12%**  |

Table 4.12: Relative definiteness and *-pe*-marking

As these tables show, there are no objects in the corpus greater either in Rel-A or Rel-D than their subjects. Thus the only contrast in relativeness available for scrutiny is that between O and S being equal on the relevant hierarchy and O being lower than S on the relevant hierarchy. Rel-A exhibits a large difference in marking frequency between O = S (56%) and O < S (1%). This difference is statistically significant

---

[4]As I mentioned in §3.2.3, the symbol "=" here means "greater than S in $\left\{\begin{array}{c}\text{animacy}\\\text{definiteness}\end{array}\right\}$," and "<" means "less than S in $\left\{\begin{array}{c}\text{animacy}\\\text{definiteness}\end{array}\right\}$." Thus those objects which are "O = S" are equal to their subjects in animacy or definiteness, and those objects which are "O < S" are less than their subjects in animacy or definiteness.

(p < 0.0001). Thus it appears to be the case that Rel-A is a very important contributing factor to DOM in Guaraní. However, this conclusion is complicated by the fact that there is a severe interaction between Rel-A and mere animacy of the object, as noted by Tippets and Schwenter (2007). In this case particularly, this is because there are no O > S tokens, which means that all human objects in the corpus belong to the category O = S. This is shown in Table 4.13, in which I cross-tabulate the features [±human][5] and the relevant Rel-A features.

|  | O = S | O < S |
|---|---|---|
| [+HUMAN] | 43 | 0 |
| [−HUMAN] | 2 | 188 |

Table 4.13: Humanness and relative animacy

As seen in Table 4.13, all human objects are O = S, and all but two O = S objects are human. Thus there is a near categorical mutual implication in the corpus: an object is equal in animacy to its subject if and only if that object is human-referring. In light of this, teasing apart animacy and relative animacy for this corpus is, unfortunately, virtually impossible. By way of illustration, compare the contrast in marking frequency between O = S and O < S in Table 4.11 to the contrast in marking frequency between [±human] in Table 4.14.

|  | +*pe* | −*pe* | % MARKED |
|---|---|---|---|
| [+HUMAN] | 24 | 19 | **56%** |
| [−HUMAN] | 3 | 187 | **2%** |
| TOTAL | 27 | 206 | **12%** |

Table 4.14: Humanness and -*pe*-marking

The numbers in both tables are almost identical. As shown in Table 4.13, there are only two examples in the entire corpus which are at the margins of Rel-A and humanness (that is, O = S and O is [−human]). If they were to show the same thing, these examples might be suggestive as to whether Rel-A or animacy is the feature that is relevant to DOM. However, they point in opposite directions from one another. I present them both below.

(127)  Tuju-ry    o-jagara-pa     **la**     **ij-ao**.
        mud-juice A3-grab-COMPL **the(Sp) 3.POSS-cloth**
        'Mud had gotten all over (had grabbed) **his clothes**.'

---

[5] [±human] refers to humanness as an object property, both here and throughout the remainder of this thesis.

(142) Ha **i-vaká-pe**          nipo     o-jopi   káva ra'e.

CONJ **3.POSS-cow(Sp)-PE** perhaps A3-sting wasp RAE

'And maybe a wasp stung **his cow**.'

In (127), both the subject, *tujury* 'mud,' and the object, *ijao* 'his clothes,' are equally inanimate. If Rel-A were the deciding factor, the object should be *-pe*-marked. If humanness were the deciding factor, it should not be. The fact that no marker appears seems to suggest that animacy of the object is crucial feature, rather than Rel-A. However, the existence of (142) presents a challenge to this approach. Here both the subject, *káva* 'a wasp,' and the object, *ivaka* 'his cow,' are equally animate. Again, a Rel-A approach would predict marking here, while a humanness approach would predict non-marking. The appearance of a marker therefore favors Rel-A as the relevant feature to DOM. Thus even at the margins the data are conflicting. Recall that the acceptability data presented in Table 4.4 seemed to exhibit little difference in acceptability judgments between an O = S example and an O < S example, perhaps indicating that it is animacy, and not relative animacy, that DOM in Guaraní is sensitive to. If this were the case, then the presence of *-pe* in (142) would need to be explained on other bases. Unfortunately, I have no choice but to leave the question open for now. It is nevertheless clear that animacy in some sense (either relative or absolute) is a very important condition on DOM in Guaraní. Since the results from object animacy vs. Rel-A are so similar, for simplicity's sake I take only object animacy into account for the remainder of this study. Making the opposite choice would not drastically alter the results.

The data in Table 4.12 do not seem to show Rel-D to be relevant to the distribution of DOM in the corpus. The frequency of marking does not differ in a statistically significant way between O = S and O < S (p = 0.6353). Nor is there a statistically significant difference in marking frequency between either Rel-D category and the total mean of 12% (p = 0.6394 and p = 0.8834, respectively). A further limitation on this measure is the relative rarity of any Rel-D value except O < S (all but 12 tokens in the corpus). There is thus quite little variation in Rel-D throughout the corpus, and what variation there is does not relate significantly to *-pe*-marking. I take this as evidence that Rel-D does not factor significantly into the system of DOM in Guaraní.

## 4.6   GF-Ambiguity as a Condition on DOM in Guaraní

Here I discuss the corpus results for GF-ambiguity as a factor group. It turns out that defining GF-ambiguity as I have done in §3.2.4 has the effect of rendering only one clause in the corpus ambiguous at all, which I

discuss as (122) below. Therefore, in contrast to my previous proposal in C. Shain (2008), ambiguity does not play a significant role in the current analysis, since it exhibits so little variation in the corpus. I nevertheless do not wish to rule out GF-ambiguity altogether as a contributing factor. According to the criteria in §3.2.4, there are actually only three scenarios in which a clause would be considered [+ambiguous] with respect to GF-assignment: when both S and O are overt and 3$^{rd}$ person, neither thematic fit nor the discourse context bias one GF-assignment over the other, and either (i) O precedes S, (ii) O is a fronted wh-word, or (iii) O is a nominal head which can be the object of a modifying relative clause. The single qualifying clause from the corpus is the following one:

(122)   O-ĩ-ndaje   raka'e peteĩ mitã tyre'ỹ    o-hayhú-va **mymba-kuéra-pe**.

A3-be-REP before one   child orphan A3-love-RC **animal-PL-PE**

'There once was an orphan who loved **animals**.'

The sentence in (122) is the first one in the discourse. Thus context provides no disambiguating information. Furthermore, either argument (*peteĩ mitã tyre'ỹ* 'an orphan' and *mymbakuérape* 'animals') could equally well be the subject or object of *hayhu* 'love' according to world knowledge. Additionally, both S and O are overt 3$^{rd}$ person, and one of the arguments is the nominal head of a relative clause, qualifying as ambiguous according to (iii) above. Notice that in this one instance of true ambiguity, the object is *-pe*-marked, supporting the hypothesis that GF-ambiguity favors object marking.

In addition to (122), I was able to find the following relevant examples in the extended corpus.

(128)   a.   **Máva-pe** piko nd-o-juka-mo'ã-i        Bill?

**who-PE**   QU   NEG-A3-kill-think-NEG Bill

'**Who** won't Bill kill?'

b.   **Mitã  nde-icha-gua    jepe-ve**         o-juhú-rõ       o-juka hikuái.

**child 2Sg.PRO-like-of including-more** A3-find-when A3-kill 3Pl.SUBJ.PRO

'When they found them they would kill **even children like you**.'

Example (128a) is a wh-question, a prime place to look for GF-assignment ambiguities since the wh-phrase is fronted in Guaraní, regardless of whether its GF is S or O (see §3.2.4). The sentence in (128a) is therefore ambiguous aside from the object marker: there are two overt 3$^{rd}$ person NPs (*mávape* 'who' and *Bill*), both are human and equally capable of killing each other according to world knowledge, there is no contextual information according to which the correct GF-assignment might be decided on, and O precedes (or rather *can* precede) S. Thus the object marker on *mávape* is the only available indicator of

which argument in the clause bears which grammatical function. It seems reasonable to suppose that it is for this reason that the object marker appears here.[6] However, we have already seen that human objects are much more likely to be marked than others. Thus the marker in (128a) could simply have arisen as a result of the fact that *máva* 'who' is [+human]. But compare (128a) with (128b), which is rendered unambiguous by the fact that the subject is encoded as the subject-only pronoun *hikuái*, leaving only the object slot open for the remaining NP. Notice that no marker appears in this unambiguous clause, despite OS realization and the fact that the object is also [+human]. Thus it could be the case that GF-ambiguity is a conditioning factor on DOM in Guaraní. Nevertheless, these results can be nothing more than suggestive, since the difference between (128a) and (128b) is not necessarily attributable to GF-ambiguity alone. Therefore it is simply interesting to note that one of the major differences between (128a) and (128b), ambiguity, corresponds to a difference in object marking as well.

Regardless of whether ambiguity influences the distribution of DOM in Guaraní, it is quite clear that this distribution is not *reducible* to ambiguity, or even largely dependent on it, contrary to my argument in C. Shain (2008). This more-or-less goes without saying, since DOM variation is attested in the corpus while ambiguity variation is not. That ambiguity alone does not suffice to account for DOM is shown by the following examples from Gregores and Suárez (1961):

(129)   (Adapted from Gregores & Suárez, 1961, p. 261)

       A-hayhu    **che-retã-me**.

       A1.Sg-love **1Sg.POSS-country-PE**

       'I love **my country**.'

(1)   (Adapted from Gregores & Suárez, 1961, p. 223)

     b.  A-hecha    **ne-ru-pe**.

         A1.Sg-see **2Sg.POSS-father PE**

         'I see **your father**.'

The sentence in (129) is unambiguous for at least two reasons. First, and most importantly, the verb is inflected for 1st person subject agreement, leaving only the object slot open for the 3rd person co-occurring NP. Second, as I argued in §3.2.4, the fact that only one NP co-occurs with a transitive verb indicates that the NP is the object. There might be a TF bias arising here as well, to the extent that the speaker loving

---

[6]As a side point, the existence of (128a) presents a problem for the claim in Gregores and Suárez (1961, p. 262) that interrogative pronouns cannot take an object marker.

the country is a more natural kind of relation according to world knowledge than the country loving the speaker.[7] Example (1b) is unambiguous for similar reasons: the 1st person agreement morphology rules out the co-occurring 3rd person NP as the subject. Yet the object in (1b) is nevertheless marked. The fact that an object marker still appears in (129) and (1b) is clearly inexplicable on the basis of GF-ambiguity, as is the marking of 26 objects in unambiguous clauses in the corpus.

## 4.7 Thematic Fit as a Condition on DOM in Guaraní

In §4.2 I provided some data that suggested marking might be more acceptable/preferable if TF did not bias GF-assignment (see Table 4.5). In this section, I more fully explore the possibility that *-pe* is used in Guaraní to mark objects that occur in clauses which are GF-ambiguous with respect to TF (TF-ambiguous), in which case we would predict a higher frequency of marking for clauses that exhibit TF-ambiguity than for clauses that do not. In other words, I test whether *-pe*-marking is conditioned by *unpredictability* of meaning given the typical patterns of the world. Table 4.15 provides a cross-tabulation of TF-ambiguity and marking.

| | +*pe* | −*pe* | % MARKED |
|---|---|---|---|
| TF-AMBIGUOUS | 25 | 27 | **47%** |
| NON-TF-AMBIGUOUS | 2 | 179 | **1%** |
| TOTAL | 27 | 206 | **12%** |

Table 4.15: Thematic fit ambiguity and *-pe*-marking

According to Table 4.15, the difference in marking frequency between TF-ambiguous and non-TF-ambiguous clauses in the corpus is quite large, and statistically significant ($p < 0.0001$). This seems to be evidence that DOM in Guaraní is sensitive to TF-ambiguity. However, there is a problem with drawing this conclusion: like Rel-A, TF-ambiguity is closely tied to the animacy of the object, since human objects have properties consistent with subjecthood (volition, sentience, causation) that animate and inanimate objects generally lack (see Dowty, 1991). Thus there is a severe interaction between animacy and TF-ambiguity as conditions on the distribution of DOM. This interaction is demonstrated by the data in Table 4.16.

The relationship between humanness and TF-ambiguity shown in Table 4.16 is statistically significant ($p < 0.0001$). Although the overlap between humanness and TF-ambiguity is not as complete as the over-

---

[7]To me, at least, the latter relation is not at all inconceivable or even odd. The fact that TF bias seems fairly weak in this instance might possibly explain why this inanimate object is marked, despite the very strong dispreference for marking exhibited by inanimates, as shown in Table 4.7. I return to this possibility in §4.7.

|                | TF-Ambiguous | Non-TF-Ambiguous |
|----------------|:------------:|:----------------:|
| [+HUMAN]       | 37           | 6                |
| [−HUMAN]       | 15           | 175              |

Table 4.16: Humanness and thematic fit

lap between humanness and Rel-A shown in Table 4.13, it is nevertheless large enough to present difficulties disentangling humanness and TF-ambiguity. In fact, 24/25 of the *-pe*-marked TF-ambiguous clauses shown in Table 4.15 also had [+human] objects. While TF-ambiguity therefore accounts for the marking of one more object in the corpus than humanness does (see Table 4.14), there are 8 fewer unmarked human object clauses than unmarked TF-ambiguous clauses, and humanness is therefore a more predictive factor group than TF-ambiguity. Thus, as I did for the similar situation with Rel-A, for simplicity's sake I continue to look primarily at object animacy rather than TF-ambiguity for the remainder of this study.

As a final note, while it is true that TF-ambiguity and object animacy cannot be teased apart statistically in the given corpus, there was an example from the extended corpus that might indicate that DOM in Guaraní is sensitive to TF-ambiguity independently of object animacy.

(141)  Pe   tahachi piko n-oi-nupã-i-ra'e      **peteĩ mba'yru-guatá-pe**?
       that police   QU  NEG-A3-hit-NEG-RAE **one    car-walk-PE**
       'Didn't the police officer hit **a car**?'

The object in (141) is inanimate and non-specific, and lower in both Rel-A and Rel-D, in which case we would expect a strong dispreference for *-pe*-marking according to our data so far (Tables 4.7, 4.10, and 4.11). Furthermore, since S precedes O, there is no GF-ambiguity, which might have increased preference for marking according to the discussion in §4.6. However, (141) exhibits an object marker nonetheless. While this is certainly not the only conceivable property of (141) that might be associated with object marking, note that the example exhibits a pronounced TF-ambiguity. That is, based on the nature of cars, police, and the "hitting" relation in the real world, it seems to me no more typical for police to hit cars than for cars to hit police. In fact, world knowledge might even bias interpretation toward the latter reading (the image of a car running over a police officer seems more typical than that of a police officer punching a car). Given this, compare (141) to (130):

(130)  Heta   ary  rire,  masificación tekombo'é-pe oi-nupã **UNA rokẽ**.
       many year after, growth        education-PE A3-hit    **UNA door**
       'After many years, growth in education hit the **door of UNA** (the National University of Asunción).'

Although both examples involve the same verb (*nupã* 'hit'), unlike (141), (130) does not seem to exhibit any TF-ambiguity, given that we assume it to be less typical for "the door of UNA" to be the hitter of some abstract concept ("growth in education") than the other way around, where "hitting" is assumably taken as a metaphor for something like "arriving at." The fact that no object marker appears on the object of (130) therefore further suggests that TF-ambiguity might be playing a role in (141). In light of this, as I did for Rel-A in §4.5, I leave open the possibility that TF-ambiguity is relevant to the distribution of DOM in Guaraní, although limitations of the data at hand prevent further exploration of this possibility.

## 4.8  Topicality as a Condition on DOM in Guaraní

As I discussed in §3.2.6, seven different measures of topicality were applied to the tokens in the corpus, which I listed in (121), all of which in some sense are derivative of the intuition that referents that are topical will be mentioned often and in close proximity to one another in discourse, while referents that are not topical will be mentioned infrequently and far apart from one another in discourse. In the terminology of Givón (1983) we expect more topical referents to exhibit discourse continuity and less topical referents to fail to do so. While I do not wish to claim that topicality is reducible to any one or more of these measures, it is my assumption that their combined results should leave us with a reliable sense of how topical the object under consideration is. Two important questions must be answered empirically: (i) do the topicality tests agree with one another, and (ii) is there a significant relationship between topicality and -*pe*-marking? I begin to attempt to answer them by presenting the mean values for topicality of marked vs. unmarked objects in Table 4.17. The number of degrees of freedom for each row is 231.

| | +*pe* | −*pe* | Ratio +/− | P VALUE | *t* STATISTIC |
|---|---|---|---|---|---|
| REFERENTIAL DISTANCE (RD) | 9.37 | 15.3 | 0.612 | **0.0003** | 4.0743 |
| FORWARD RD (F-RD) | 6.48 | 12.5 | 0.518 | **0.0009** | 3.3757 |
| CONTIGUOUS TOPIC PERSISTENCE (TPc) | 3.26 | 0.451 | 7.23 | **< 0.0001** | 5.7868 |
| NON-CONTIGUOUS TP (TPnc) | 7.19 | 1.39 | 5.17 | **< 0.0001** | 8.6625 |
| BACKWARD TP (B-TP) | 5.30 | 0.767 | 6.91 | **< 0.0001** | 9.2061 |
| TOTAL RD (T RD) | 15.9 | 27.8 | 0.572 | **0.0003** | 4.6168 |
| TOTAL TP (T TP) | 12.5 | 2.16 | 5.79 | **< 0.0001** | 11.1726 |

Table 4.17: Topicality and -*pe*-marking

The data in Table 4.17 show a quite pronounced relationship between high topicality and -*pe*-marking in the corpus. By every measure of topicality used here, the difference in average topicality between marked and unmarked objects is statistically significant, and by every measure the topicality of marked vs. unmarked objects differs by a large factor. The first fact to notice about Table 4.17 is that all the different

topicality measures produce very similar results. All the RD-related measures (RD, F-RD, and T-RD) show a ratio of a little over .5 for marked/unmarked objects. Similarly, all the TP-related measures (TPc, TPnc, B-TP, and T-TP) show a ratio of around 6 for marked/unmarked objects. Since the RD- and TP-related measures agree so nicely with each other, for reasons of perspicuity I look primarily at T-RD and T-TP when discussing topicality throughout the remainder of this thesis, since both of these measures allow us to assign a single numerical value to the total (forward-looking and backward-looking) topicality of an object's referent (see §3.2.6).

Looking now at T-RD in Table 4.17, we see that marked and unmarked objects differ by a statistically significant factor of 0.572. This number means that the average distance in number of clauses between the previous mention and the subsequent mention for marked objects is 0.572 times that for unmarked objects. That is to say, the surrounding mentions of marked objects are a little more than half as far away as the surrounding mentions of unmarked objects. Given that RD is assumed to relate inversely with topicality, this means that marked objects are almost twice as topical as unmarked objects in terms of T-RD. Turning now to T-TP, we see that marked and unmarked objects differ by a factor of 5.79. That is, marked objects on average exhibit 5.79 as many mentions in the surrounding 40 clauses as unmarked objects. Since T-TP is assumed to relate directly with topicality, this means that marked objects are almost *six times* more topical than unmarked objects in terms of T-TP. I take this to be very strong evidence that topicality plays a role in the distribution of DOM in Guaraní, such that more topical DOs prefer -*pe*-marking.

I mentioned in §3.2.6 that T-RD and T-TP should reflect differences between chain-initial/final topics and non-topics, showing the former to be more topical. Thus we may explore which kinds of topics -*pe* generally marks in Guaraní. If -*pe* is a topical object marker, does it mark chain-initial, chain-medial, or chain-final topics, or some combination of the three? In order to test this, I compare the topicality of marked and unmarked objects which are potential chain-initial, chain-medial, or chain-final topics. The results are presented in Tables 4.18, 4.19, and 4.20.

|       | +*pe* | −*pe* | Ratio +/− | P VALUE | *t* STATISTIC |
|-------|-------|-------|-----------|---------|---------------|
| T-RD  | 31.1  | 34.0  | 0.915     | 0.3115  | 0.9791        |
| T-TP  | 4.30  | 0.929 | 4.63      | **0.0001** | 3.9424     |

Table 4.18: Topicality and -*pe*-marking of DOs with RD = 20
(df = 150)

In Table 4.18 I give the average T-RD and T-TP values of objects which are potential chain-initial topics (i.e. with an RD of 20/B-TP of 0), comparing the results for marked and unmarked cases. In Table 4.19 I do

|       | +*pe* | −*pe* | Ratio +/− | P VALUE | *t* STATISTIC |
|-------|-------|-------|-----------|---------|---------------|
| T-RD  | 3.73  | 8.36  | 0.446     | **0.0344**   | 2.1698        |
| T-TP  | 19.5  | 6.50  | 3.00      | **< 0.0001** | 9.2823        |

Table 4.19: Topicality and -*pe*-marking of DOs with RD < 20 and TPnc > 0

(df = 54)

|       | +*pe* | −*pe* | Ratio +/− | P VALUE | *t* STATISTIC |
|-------|-------|-------|-----------|---------|---------------|
| T-RD  | 35.1  | 37.2  | 1.08      | 0.4059  | 0.8341        |
| T-TP  | 0.857 | 0.336 | 0.851     | 0.1875  | 1.3256        |

Table 4.20: Topicality and -*pe*-marking of DOs with TPnc = 0

(df = 118)

the same with potential chain-medial topics (those with RD < 20 and TPnc > 0), and in Table 4.20 I do the same with potential chain-final topics (i.e., with a TPnc of 0/F-RD of 20). As I discussed in §3.2.6, objects falling into each of these categories may be either topical or non-topical. If there is a significant difference in marking frequency for any one of these categories, this suggest that -*pe* is picking out the more topical from the less topical objects within that category.

Since chain-initial and chain-final marked objects are relatively rare in the corpus (10 tokens and 7 tokens, respectively), some of these data are inconclusive. Although marked objects are more topical than unmarked objects both chain-initially and chain-finally according to both T-RD and T-TP, only one of these differences proved statistically significant: the contrast in average T-TP between marked and unmarked chain-initial objects, highlighted in Table 4.18. With respect to chain-medial object referents, the large differences in topicality between the marked and unmarked cases are statistically significant, demonstrating an asymmetry in marking preferences for clause-medial topics. Although a topicality asymmetry is not clear with respect to T-RD in chain-initial object referents, the difference is clear with respect to T-TP in such referents. Thus I take the data in Tables 4.18 and 4.19 to indicate that -*pe* is a marker of both chain-initial and chain-medial topics in object position in Guaraní. There was too little data to determine whether the same is true for chain-final topics, although it seems strange to me to posit a system in which speakers necessarily have the prescience to forecast the impending doom of a topic chain and, on principle, leave the final mention unmarked. To the extent that the reader agrees with this intuition, we can conclude that -*pe* is probably also a marker of chain-final topics.

The fact that -*pe* may mark all kinds of topics means that the object marker may be used both to introduce a new referent as the topic of discourse and to reinforce the continuation of an old topic as a topic. The following examples from the corpus shows this:

(126)   O-hasá-vo          upéi o-hecha **ju'í-pe**   o-guapy y       mbyté-pe   peteĩ yrupē-ari.
        A3-pass-when then A3-see   **frog-PE** A3-sit    water middle-PE one   sieve-on
        'In passing, he saw **a frog** sitting in the middle of the water on a water lily.'

(131)   Pirulo n-o-guerovia-i          mba'e-icha-pa Juan-chi   nd-o-jura-i          **ju'í-pe**.
        Pirulo NEG-A3-believe-NEG thing-like-QU   Juan-DIM NEG-A3-grab-NEG **frog-PE**
        'Pirulo couldn't believe how Juanchi couldn't grab **the frog**.'

Example (126) is the first mention of the frog in the story, in which the frog goes on to become one of the main characters (an important *topic* of the narrative, TPnc = 14). Thus the *-pe*-marked object in (126) is fairly clearly the introduction of a new topic. Example (131) follows in the same story several clauses later, after the frog has been repeatedly mentioned (RD = 1, B-TP = 7). Thus the object referent in (131) is clearly a chain-medial topic, and it is marked as well as the chain-initial topic in (126). If *-pe* is indeed a topicality marker, which it appears to be according to the above data, then it serves the dual purpose of highlighting the speaker's intent to introduce a new topic to the discourse and indicating which object referents are continuous topics throughout the discourse.

A final topicality-related question that ought to be addressed is why topicality should be such a strong factor group and definiteness such a non-factor group with respect to DOM, given the apparent relationship between topicality and definiteness. Given the definitions of topicality and definiteness assumed here, we might expect definite things to be more topical and topical things to be more definite, in which case these results are somewhat surprising. Givón (1990), who proposed the basic framework for the topicality measures employed here, even says that "it seems that the primary grammaticalization of both 'definite' and 'indefinite' is motivated by topicality" (Givón, p. 474). Nevertheless, given the fact that the results do differ so significantly regarding these two factor groups, it should be fairly clear that the relationship between topicality and definiteness as I have defined them is not especially strong. As shown in §3.2.2.2, definiteness was assessed with respect to morphosyntactic indicators as well as consideration of whether the referent was discourse-old (previously mentioned) or inferable from context. Topicality, on the other hand, was assessed with respect to counts of mentions and distances between mentions in the discourse. While definiteness has a lot to do with discourse-oldness, it is not explicitly connected to recentness of mention, as measured by RD, and certainly not to density of mention in the surrounding context, as measured by TP. Thus, while highly topical objects may tend to be definite, it is not at all necessarily the case, given these definitions, that definite objects are topical. Furthermore, perhaps the most

topical of the definiteness categories in (27) is Pronoun (see Gundel et al., 1993), which has essentially been eliminated from consideration in this study for independent reasons (see §3.1.4). One possible result of this is a weakening of the link between definiteness and topicality. Additionally, definiteness as defined here is sensitive to weak familiarity, in which a discourse-new referring expression may be definite if its referent is inferable from context/world knowledge (Roberts, 2003). However, topicality, as defined here, is sensitive only to discourse-newness/-oldness and density of mention in the discourse, and inferability is not taken into consideration. Thus even a [+definite] object will have the highest possible RD value (20) if it is discourse-new. Finally, since our definiteness criteria only look at previous discourse while the topicality measures look additionally at subsequent discourse, an indefinite object can still be fairly topical if it is being introduced as a chain-initial topic and is thus frequently mentioned later. I believe these facts taken together show that definiteness and topicality can and do vary independently of one another, and provide insight as to how topicality can be a strong predictor of object marking when definiteness is not.

## 4.9   Interim Conclusion

To review what we have found so far: the data show that animacy, specificity, relative animacy, thematic fit ambiguity, and topicality exhibit statistically significant relationships with object-marking frequency. Of these, there is an unresolvable interaction between animacy, relative animacy, and thematic fit ambiguity, since the latter two properties depend crucially on the first. In light of this, I opted to deal primarily with animacy rather than relative animacy in the remainder of this study, given that the opposite assumption will change the results very little (see §4.5). I made a similar decision with respect to TF-ambiguity, choosing to look primarily at animacy, which appears to be a better predictor of object marking than TF-ambiguity throughout the corpus (see §4.7). Regarding animacy, no inanimate objects were marked, 5% of animate objects were marked, and 56% of human objects were marked. Regarding specificity, [+specific] objects were 10% more likely to be marked than [–specific] objects generally. Regarding topicality, *-pe*-marked objects were up to 6 times as topical as unmarked objects according to some topicality measures (see §4.8). As far as the other proposed factor groups are concerned, the Definiteness Hierarchy (as opposed to [±specific]), relative definiteness, and GF-ambiguity did not appear to be related significantly to object marking frequency.

## 4.10   Analysis

Having explored the (non-)relationships between the factor groups and *-pe*-marking in the corpus, I now attempt to synthesize this data into a predictive analysis of the distribution of DOM in Guaraní. Up to this point, it looks as if the one of the strongest predictors of object-marking is whether or not the object is human. However, simply adopting a hypothesis that humanness is the only relevant condition on DOM leaves two important things to be explained: (i) why the *-pe*-marked animate objects in the corpus are *-pe*-marked, and (ii) why only a little more than half of the human objects in the corpus are marked. Regarding these questions, certain insights can be gleaned from Table 4.21, where I present the average T-RD and T-TP values for marked and unmarked objects within each animacy category.

|  |  | *+pe* | *−pe* | Ratio +/− | P value | *t* statistic | df |
|---|---|---|---|---|---|---|---|
| Human | T-RD | 13.0 | 25.6 | 0.508 | **0.0031** | 3.1470 | 41 |
|  | T-TP | 14.2 | 3.36 | 3.91 | **0.0001** | 4.2579 | 41 |
| Animate | T-RD | 33.7 | 24.6 | 1.37 | – | – | – |
|  | T-TP | 0.333 | 3.16 | 0.105 | – | – | – |
| Inanimate | T-RD | – | 29.4 | – | – | – | – |
|  | T-TP | – | 1.53 | – | – | – | – |

Table 4.21: Animacy, topicality, and *-pe*-marking

Regarding (i) above, these data are not conclusive. The marked animates in the corpus are actually less topical than the unmarked animates according to both T-RD and T-TP, contrary to what would be expected if topicality were a condition on the marking of animates. However, since there were only three marked animates in the corpus, and since there is a large numerical imbalance between marked and unmarked animates (3/55), no reliable *t* test for significance is possible. Thus while the topicality values in Table 4.21 do not corroborate the hypothesis that more topical animates are marked more often, they do not pose a clear problem for it either. There simply isn't enough data to decide, and a larger corpus study would be required to arrive at a satisfactory conclusion about the relevance of topicality to the marking of animate objects. If, however, we should assume that topicality and *-pe*-marking are directly related, we should ideally also formulate an explanation for why these three quite non-topical animates in the corpus still take a marker. I return to this question later.

However, regarding (ii) above, the data in Table 4.21 are much more illuminating. There we see that marked human objects exhibit statistically significant differences from unmarked human objects with respect to both T-RD and T-TP. Marked objects have a little more than half the average T-RD value of unmarked objects, which means that they are almost twice as topical in terms of T-RD. Marked objects

also have almost four times the average T-TP value of unmarked objects, indicating that they are almost four times as topical in terms of T-TP.

The data in Table 4.21 give some indication of the way in which animacy and topicality condition DOM in Guaraní. We have not yet determined whether [±specific] must be included in the account. Table 4.10 might suggest that it ought to be, since specific objects are significantly more likely to be marked than non-specific objects. Before concluding this, however, it is important to ensure that the specificity effect is not reducible to other factors already under consideration, in this case topicality. To this end I cross-tabulate object marking and specificity in Table 4.22, presenting the average T-RD and T-TP values for each cell.

|  | | +*pe* | −*pe* | Ratio +/− | P VALUE | *t* STATISTIC | DF |
|---|---|---|---|---|---|---|---|
| SPECIFIC | T-RD | 13.3 | 26.0 | 0.512 | **< 0.0001** | 4.3670 | 178 |
|  | T-TP | 13.6 | 2.46 | 5.53 | **< 0.0001** | 10.8406 | 178 |
| NON-SPECIFIC | T-RD | 40 | 33.2 | 1.20 | – | – | – |
|  | T-TP | 0 | 1.24 | 0 | – | – | – |

Table 4.22: Specificity, topicality, and -*pe*-marking

According to Table 4.22, marked specifics are almost twice as topical according to T-RD and more than five times as topical according to T-TP as unmarked human specifics. There were only two marked non-specific objects in the corpus, and thus there was too little data to determine whether topicality plays a role in the marking of non-specifics. Nevertheless, it appears to be the case, according to Table 4.22, that the same topicality distinction among human objects in Table 4.21 underlies the specificity effect with respect to the marking of human objects. I take this to be evidence that much of this specificity effect is reducible to topicality, and for this reason I do not appeal to specificity in my analysis of the distribution of DOM in Guaraní. This is not to say that specificity is not a factor at all in this distribution. There are, in fact, some marked specific objects in the corpus with very low topicality figures, suggesting that specificity may still play a role. Nevertheless, the data in Table 4.22 indicate that excluding specificity from consideration and looking exclusively at animacy and topicality will simplify the analysis without incurring a large cost in terms of empirical coverage.

Simple cross-tabulation can produce a good approximation of the relative importance of these various factor groups to the distribution of DOM. However, a variable rule (Varbrul) analysis could provide much

more conclusive results.[8] To this end I ran a variable rule analysis with respect to (a simplification of) the three most promising factor groups given above:

(132)   a.  ± human

        b.  ± specific

        c.  ± topical

Regarding (132a), although there was a statistically significant difference in marking frequency between animate and inanimate DOs (Table 4.7) which cannot be captured by a [±human] distinction, a varbrul analysis will not work if there is a 0 in a cell, as is the case for the inanimate marked cell (i.e. there are no objects that fit this description in the corpus). Since inanimates and animates are similar in that -*pe*-marking is quite rare for both (unlike humans), I collapsed them into a [–human] category for the purposes of conducting this variable rule analysis.

Regarding (132c), certain varbrul-driven adjustments to the taxonomy had to be made. First, there were seven distinct measures of topicality proposed above. Since each of these measures was designed to capture certain aspects of the same underlying feature (topicality), severe interactions would result if more than one of them were included in the analysis. Thus I only looked at T-TP, which showed more pronounced differences in topicality than T-RD and would therefore potentially yield clearer results. Second, while (132c) is presented as a binary feature, topicality as defined here is not binary, but scalar. Varbrul treats all values on a scale as distinct factors, which means that, since T-TP ranges from 1-40, topicality would have 40 different factors! Not only would this spread the data out too thinly to reveal statistically significant patterns, but it would also test distinctions which do not matter for our purposes. For example, we are not particularly interested in whether or not there is a statistically significant difference in -*pe*-marking frequency between objects with T-TP = 38 and T-TP = 37. Intuitively, the contrast we are after is something more akin to "high" vs. "low," which requires the imposition of an artificial cutoff point in T-TP between [+topical] and [–topical]. I assign 7 as that cutoff, since it falls around halfway between the mean topicality values of marked (12.5) and unmarked (2.16) objects as shown in Table 4.17. The results of the varbrul analysis are presented in Table 4.23.

Briefly commenting on what the numbers mean, the "weight" columns show the *factor weight* of both the "+" and "−" factors for each factor group tested. A weight closer to 1 indicates that the factor favors

---

[8]Varbrul is a statistical package first implemented by David Sankoff to analyze the relationship between the distribution of a linguistic variable and its potentially relevant factor groups. By means of stepwise logistical regression, it determines the statistical (non-)significance of each factor group to the variable in question as well as the ranking of factor groups in terms of their importance to the variable in question.

|  | WEIGHT | | |
| FACTOR GROUP | + | − | RANGE |
| --- | --- | --- | --- |
| humanness | .96 | .32 | **64** |
| specificity | .55 | .32 | [23] |
| topicality | .77 | .44 | **33** |

Table 4.23: Variable rule analysis results for humanness, specificity, and topicality
Corrected mean = .03; Total N = 233
Range values in **bold** are statistically significant. Range values in [brackets] are not.

marking, and a weight closer to 0 indicates that the factor disfavors marking. The *range* is simply the difference between the highest and lowest factor weights for each factor group, multiplied by 100. The greater the range value, the more important that factor group is to the linguistic variable, which in our case is object marking. This enables *constraint ranking*, or the ordering of factor groups according to their importance to DOM. The bolded groups, humanness and topicality, were statistically significant. The group in brackets (specificity), was not statistically significant. This finding is consistent with the hypothesis put forward earlier on the basis of cross-tabulation: that the specificity effect observed in the corpus is essentially reducible to topicality. At least, this might be a reasonable explanation for the non-significance of specificity. Furthermore, the range of humanness is nearly twice as large as the range of topicality, indicating that humanness is the more important factor group. This again accords with conclusions drawn from the aforementioned finding that indefinites are categorically unmarked in the corpus, regardless of their topicality.

|  |  | TOPICAL | | NON-TOPICAL | | TOTAL | |
|  |  | *+pe* | *−pe* | *+pe* | *−pe* | *+pe* | *−pe* |
| --- | --- | --- | --- | --- | --- | --- | --- |
| HUMAN | # of objects | 16 | 4 | 8 | 15 | 24 | 19 |
|  | % marked | **80%** | | **35%** | | **56%** | |
| NON-HUMAN | # of objects | 0 | 23 | 3 | 164 | 11 | 179 |
|  | % marked | **0%** | | **2%** | | **6%** | |
| TOTAL | # of objects | 16 | 27 | 3 | 187 | 27 | 206 |
|  | % marked | **37%** | | **2%** | | **12%** | |

Table 4.24: Humanness, binary topicality, and *-pe*-marking

Before synthesizing these results into a predictive analysis of DOM in Guaraní, we must first determine whether either humanness or topicality is reducible to the other as the condition on DOM, in which case a simpler analysis in terms of one factor group would be preferable to the more complex analysis in

terms of two. The data in Tables 4.7 and 4.21 indicate that the topicality effect is not reducible to human-ness, since only half of human objects are marked and those that are marked are more topical. However, a reduction might be possible in the other direction (reducing humanness to topicality), suggested by exactly the aforementioned data. The (artificial) reduction of topicality to a binary contrast discussed above allows us to test this possibility, since *-pe*-marking frequency across topicality and humanness can now be compared using 2x2 contingency tables, as shown in Table 4.24.

Table 4.24 shows *-pe*-marking to be more frequent in human topical objects, as expected. However, the important feature of Table 4.24 for our purposes is the fact that there is a statistically significant difference in marking frequency between human topics and non-human topics ($p = 0.0001$). Thus stating the analysis strictly in terms of topicality alone would miss important generalizations, just as stating the analysis in terms of humanness alone would.

Therefore, on the basis of the data presented in this chapter, I propose the following analysis of the distribution of DOM in Guaraní.

(133)   In Guaraní:

    a.  If an object is [–animate], it will be unmarked with overwhelming frequency.

    b.  If an object is [–human] and [+animate], it will usually be unmarked, but will be marked more frequently than [–animate] objects.

    c.  If a [+human] object is high in topicality, it will usually be marked, and if a [+human] object is low in topicality, it will usually be unmarked.

In other words, *-pe* is roughly a marker of topical human objects, which is also slightly sensitive to the [± animate] distinction.

The predictions in (133) are stated in probabilistic terms. This is motivated for two reasons:

1.  As I discussed in some detail in §4.2, I was unable to discover any clear categorical constraints on the acceptability of *-pe*-marking. Thus the formulation of an analysis on the basis of such constraints faces immediate problems.

2.  Having assumed topicality to be a graded phenomenon, formulating categorical grammatical rules in terms of it is quite difficult without the imposition of an artificial binary contrast between "high" and "low" topicality relative to some cutoff point. While above I imposed a +/– cutoff point in topicality, this was motivated purely practically by the requirements of variable rule analysis. But

proposing a grammatical rule that speakers mark human objects with a T-TP ≥ 7 suggests that speakers are actually performing a T-TP calculation as they speak, which to me is counterintuitive, even impossible (how can speakers know how often a referent will be mentioned in future discourse?). I would be very hesitant to assume this cutoff point, or the measure on which it depends, to have some sort of ontological reality; it is simply a measurement device. Furthermore, the topicality data do not lend themselves well to categorical description. In reality, there are very topical and very un-topical DOs, according to these measures, in both the marked and unmarked categories, as the following examples show.

(113)   *Marked DO high in topicality*

T-RD = 2, T-TP = 19

a.  Ha      upéi o-hecha sapy'a    Juan-chi   ha      Pirulo **ju'í-pe**.
    CONJ then A3-see    suddenly Juan-DIM CONJ Pirulo **frog-PE**
    'And then Juan and Pirulo suddenly saw **the frog**.'

(77)   *Marked DO low in topicality*

T-RD = 40, T-TP = 0

a.  O-ĩ-ndaje   raka'e      peteĩ mitã tyre'ỹ   o-hayhú-va **mymba-kuéra-pe**.
    A3-be-REP long.ago one   child orphan A3-love-RC **wild.animal-PL-PE**
    'There once was an orphan who loved **animals**.'

(119)   *Unmarked DO high in topicality*

T-RD = 4, T-TP = 17

O-mo-ngakuaa  karai       **pe    mitã**.
A3-CAUSE-grow gentleman **that child**

'The gentleman raised **the child**.'

(134)   *Unmarked DO low in topicality*

T-RD = 40, T-TP = 0

O-heka     **tukumbo**.
A3-search **rope**

'She searched for **a rope**.'

Nevertheless, it is still the case that marked DOs are *on average* much more topical than unmarked DOs, as shown in Tables 4.17 and 4.21.

These facts are not amenable to a description of the distribution of DOM in terms of grammatical rules, necessary and sufficient conditions, and so on, as I believe my discussion in §4.11 of the problems faced by Bossong (1985b) will show. However, they fit well with a description stated in terms of preferences, such as the one I provide in (133).

How well does (133) handle the corpus data? The animacy component in (133a) accounts for the non-occurrence of marking on all inanimates. The humanness component in (133b) accounts for the infrequency of marking on non-human animates (95% unmarked). The topicality component in (133c) accounts for the split between marked and unmarked humans, since marked humans are much more topical according to Table 4.21. However, as I mentioned previously, it would be nice to have an account of the 5% of animates that are marked (3 total). All of them have been presented already and are reproduced below.

(142)  Ha    **i-vaká-pe**        nipo    o-jopi   kava ra'e.
           CONJ **3.POSS-cow(Sp)-PE** perhaps A3-sting wasp RAE
           'And maybe a wasp stung **his cow**.'

(122)  O-ĩ-ndaje   raka'e peteĩ mitã tyre'ỹ   o-hayhu-va **mymba-kuéra-pe**.
           A3-be-REP before one   child orphan A3-love-RC **animal-PL-PE**
           'There once was an orphan who loved **animals**.'

(143)  O-henói **h-ymba**        **jagua peteĩ-me**.
           A3-call   **3.POSS-animal dog**    **one-PE**
           'He called **one of his dogs**.'

Before discussing these, let me first reiterate that they are not counterexamples to (133), which is probabilistic and which differentiates between [–animate] objects (overwhelmingly unmarked) and [–human] objects (usually unmarked, but more frequently marked than [–animate]). A marking frequency of 5% for animates, compared to 0% for inanimates, fits well within these predictions. Nevertheless, ideally we would be able to provide some sort of explanation for why these particular animates took a marker over all others.

Example (142) is O = S in terms of relative animacy (both S and O are animate non-human), which was shown in §4.5 to be significantly related to *-pe*-marking (although the interaction between Rel-A and

absolute object animacy was impossibly to disentangle). It could be the case that the marker here appears because of this Rel-A value. Additionally, given the argument in Bossong (1985b) that "thematic fronting" constructions favor object marking, it might be the case that *ivakápe* is morphosyntactically topicalized.[9] Furthermore, (142) very nearly qualifies as GF-ambiguous: both NPs are 3$^{\text{rd}}$ person and overt, O precedes S, and there is no disambiguating contextual information in the discourse. However, TF provides a fairly clear bias, since a cow stinging a wasp does not even seem to be a possible interpretation if real world knowledge is taken into account. From a production perspective, perhaps the fact that (142) comes close to true ambiguity motivates the appearance of a "cautionary" object marker for the sake of clarity. Taking the above into account, (142) seems to exhibit several features which, it seems reasonable to assume, might have overcome the dispreference for marked animate non-humans exhibited by the corpus data.

As I argued in §4.6, example (122) is actually the only GF-ambiguous clause in the corpus. It thus seems likely that the reason the object is marked here is that object marking is the only source of disambiguating information as to which NP takes which GF. Another factor that might play a role is topicality. According to the topicality measures as I defined them in §3.2.6, (122) is as un-topical as possible (T-RD = 40, T-TP = 0). This is the first clause in a narrative about the orphan and his interaction with two animals, a dog and a frog. According to the restriction against counting hyponyms as having the same referent as a hypernym, the subsequent mentions of the dog and frog were not considered mentions of the referent of *mymbakuérape* 'animals.' However, assuming "topic" to be in essence what the discourse is about, it might be the case that (122) is in fact setting up the concept of "animals" a topic of the narrative, specifically instantiated by the dog and the frog. If references to the dog and frog are considered mentions of the same referent as that of *mymbakuérape* 'animals,' then the topicality of *mymbakuérape* increases drastically to T-RD = 21 and T-TP = 19. If (122) is reassessed as having a topical object, then the appearance of a marker is more predictable according to the general pattern in Table 4.17, despite the non-humanness of the object.

I am not entirely sure how the marking of the object in (143) might be accounted for. One possibility is that the object is incorrectly analyzed as a direct object, and that, given the semantics of the verb (a speech act directed toward the object), *jagua peteĩme* 'one of his dogs' is actually an IO rather than a DO. This would account for the marking of a non-human object, since IOs are always marked in Guaraní (see

---

[9]I must note that Bossong's analysis of thematic objects as being fronted is contradicted by other arguments in the literature. One example is Velázquez-Castillo (1995), who claims that thematic (topical) objects are obligatorily post-verbal. Another is Tonhauser and Colijn (to appear), who show that animate or discourse-old objects are more likely to take post-verbal realization than inanimate or discourse-new objects. Thus, the hypothesis that *ivakápe* in (142) is marked because it is topicalized may not hold much water. For a fuller discussion of topicality, animacy, GF, and linear realization, see Brody (1984); Enrico (1986); H.-W. Choi (1999); King (1995); Kiss (2002); Aissen (1992).

§3.1.2.2). This hypothesis falls through in light data from the extended corpus, however, which show that object marking of the object of *henoi* 'call' is variable.

(135)  a.  A-henoi-se-mi      ningo **che**     **embaháda-pe**.
         A1.Sg-call-DES-a.little EMPH **1Sg.POSS embassy-PE**
         'I want to call **my embassy**.'

      b.  Pe-henoi-mí-na     **peteĩ ambuláncia** i-katú-rõ!
         A2.Pl-call-a.little-IMP **one   ambulance** B3-possible-when
         'Please call **an ambulance** if it's possible!'

Both clauses in (135) involve the verb *henoi* 'call' and an object. In (135a), the object is -*pe*-marked, while in (135b) the object is unmarked, demonstrating variation in marking. Since IO marking is non-variable, this is strong evidence that *henoi* does in fact take a DO rather than an IO.[10] Another possible explanation lies in the object partitive construction in (143). As I mentioned in (2.2.2), Enç (1991) analyzes partitive constructions as having specific referents, since the referents belong to a familiar superset. If we should adopt this analysis, it would allow us to analyze the object of (143) as specific, perhaps to the effect that it takes object marking. While this is a possible way to go, I would like to avoid commitment to Enç's particular brand of specificity, since it was not the one I used in this study. In light of these facts, I leave the question open as to why (143) is selected for marking out of the remaining unmarked animates. Nevertheless, the distribution of -*pe*-marking throughout the corpus conforms quite well to (133). Furthermore, in light of the above discussion, (133) should be taken as an attempt to account for the vast majority of cases according to the most statistically important factor groups. It is not intended, however, to rule out the influence of apparently more minor factor groups, such as specificity and GF-ambiguity. I acknowledge that such factors may and probably do play a small role in the distribution of DOM in Guaraní. However, as I showed above, they are fairly weak predictors object marking as a whole. I therefore conclude that (133) handles the corpus data well.

---

[10]The difference in marking between (135a) and (135b) is worth exploring, since they both seem to be inanimate objects. If this is the case, why is *che embaháda* 'my embassy' marked at all? I believe an answer is supplied by the clause immediately following (135a) in the extended corpus:

(1)  Ere-mí-na      **chupe-kuéra** ha'e-ha     peteĩ ñe-henói-apuro-pe-guá-va.
    2.say-a.little-IMP **3.OBJ.PRO-PL** 1.say-NMLZR one   JE-call-urgent(Sp)-PE-of-RC
    'Tell **them** it's an urgent call.'

Throughout the corpus, the personal pronoun *chupe* is used exclusively with human referents. The fact that the singular inanimate antecedent *che embaháda* 'my embassy' is referred to with a human plural pronoun suggests that *che embaháda* is anthropomorphized in the first clause, and is used to represent the collection of people who work at the embassy. The context therefore provides independent evidence in support of analyzing *che embaháda* as [+human], accounting for the difference in -*pe*-marking between (135a) and (135b).

## 4.11 Implications of These Results for Bossong (1985b)

Here I summarize and respond to the analysis of DOM in Guaraní presented in Bossong (1985b). Bossong makes several claims about the distribution of DOM in Guaraní, most of which have to do with the animacy and definiteness of the object. In §4.11.1, I summarize Bossong's approach, highlighting several testable claims contained in it. Then, in §4.11.2, I put these claims to the test against the previously discussed corpus results, and highlight some considerable difficulties these results present for Bossong's account.

### 4.11.1 Bossong's Analysis

The first two of the aforementioned possible conditioning factors on DOM, animacy and definiteness/specificity, occupy a prominent place in the approach of Bossong (1985b), who proposes what I understand to be the only extensive analysis of DOM in Guaraní to date. For this reason, I will here look simultaneously at Bossong's proposal, the animacy and definiteness/specificity results from the corpus, and the implications of those results for Bossong's claims.[11]

Bossong (1985b, p. 3) appeals to two hierarchies in his discussion, according to which objects with higher properties are predicted to be more marked than objects with lower properties. The first of these is the hierarchy of Inhärencemerkmale, "inherence features," presented in (136):

(136)   [±deix] < [±propr] < [±hum] < [±pers] < [±anim] < [± discr] < [±concr][12]

This hierarchy can roughly be seen as a conflation of the Animacy and Definiteness Hierarchies used in the current study: whether an object is deictic (which I assume to mean "pronominal") or a proper name is a kind of definiteness feature, while the personification/animacy of an object is a kind of animacy feature. I do not take the discreteness or concreteness of the object into account here, but Bossong (1985b) does not make crucial use of these features either.

The other hierarchy to which Bossong appeals has to do with what he calls "Referenzmerkale," or "reference features." I present this hierarchy in (137):

(137)   [+ident ego ∧ +ident tu] < [+ident ego ∧ –ident tu] < [–ident ego ∧ –ident tu]

---

[11] All English quotations of Bossong (1985b) are translations from German by Judith Tonhauser.

[12] Although Bossong (1985b) does not explicitly say how these abbreviations should be interpreted, the discussions in other publications such as (Bossong, 1991) and Bossong (1983–1984) indicate that they stand for the following: "deix" = deictic (pronominal); "propr" = proper name; "pers" = personified; "anim" = animate; "discr" = discrete (count); and "concr" = concrete.

The feature [±ident ego] refers to whether the object's referent is identifiable to the speaker, and [±ident tu] to whether the object's referent is identifiable to the hearer. Notice that this "reference" property is strikingly similar to the definition of specificity assumed for this study, according to which an object is considered specific if the speaker presumably has a particular individual in mind (see §2.2.2 and §3.2.2). Bossong's criteria additionally consider whether the hearer can identify an individual as a referent.

Turning to Bossong's claims about where object marking can and must occur in Guaraní, there are several components. First, in his discussion of DOM in the related language of Mbyá Guaraní, he says: "The presence of the features [+def] ∧ [+anim] is a necessary condition for the use of *pe* with the direct object in current Mbya Guaraní. Is this condition also sufficient, as it is with current Paraguayan Guaraní?"[13] (Bossong, 1985b, p. 15) While this is not a direct claim about Paraguayan Guaraní, it nevertheless presupposes that being [+anim] and [+def] is a sufficient condition for object marking in Paraguayan Guaraní. Thus I take this to be a component of Bossong's analysis.

In addition to this, Bossong (1985a, pp. 17-8) makes three claims about the distribution of object marking in Guaraní:

> Im einzelnen beobachten wir bezüglich des Gebrauchs bzw. Nicht-Gebrauchs der… Postposition *pe*… im heutigen Avañe'ẽ[14] die folgenden Regularitäten:
>
> 1. Die entscheidende Trennlinie zwischen *pe* und ∅ verläuft in der Dimension der Inhärenzmerkmale, und zwar bei dem Merkmal [±pers]. Das bedeutet, daß Bezeichnungen für Tiere (also der Bereich [–hum ∧ +anim]) immer dann positiv markiert werden, wenn ihre Referenten identifizierbare Individuen, sozusagen mit einer ‚Persönlichkeit', sind. Durch den Gebrauch oder Nicht-Gebrauch der Postposition bei Tierbezeichnungen können im Avañe'ẽ subtile Nuancen zum Ausdruck gebracht werden: der Sprecher kann sein Verhältnis zu dem Tier dadurch präzise konnotieren, daß er es entweder vermenschlicht (→ *pe*) oder verdinglicht (→ ∅). Diese Erscheinung hat in vielen der zahlreichen Sprachen eine Parallele, welche Objekte nach der Inhärenzskala differenzieren.
>
> 2. Eine sekundäre Trennlinie verläuft auf der Skala der Referenzmerkmale durch die Zone [+ident ego ∧ –ident tu]: wenn die individuelle Identifizierbarkeit des Objektsreferenten night gegeben ist, wird die Postposition auch dann nicht gebraucht, wenn es sich um menschliche Wesen handelt. Allerdings handelt es sich hier um eine typische Übergangszone: gelegentlich, wenn auch nicht sehr häufig, kann ein formal indefinites Nominalsyntagma genau dann mit der Postposition versehen sein, wenn es die se-

---

[13]Original: "Das Vorhandensein der Merkmale [+def] ∧ [+anim] ist eine notwendige Bedingung für den Gebrauch von *pe* beim ‚direkten Objekt' im heutigen Mbyá. Ist diese Bedingung auch schon hinreichend, so wie sie es im heutigen Avañe'ẽ ist?"

[14]*Avañe'ẽ* is the Guaraní name for the Paraguayan Guaraní language. Throughout his article, Bossong (1985b) uses this term to refer to Paraguayan Guaraní (as opposed to Mbyá Guaraní and others). For the sake of clarity, I render *Avañe'ẽ* as *Paraguayan Guaraní* in the English translations.

mantischen Bedingungen [+ident ego] ∧ [+pers] erfüllt. In das hier verwendete dimensionale Modell lassen sich solche Übergangszonen mit schwankendem Gebrauch ohne Schwierigkeiten integrieren.

3. Schließlich ist unter bestimmten syntaktischen Bedingungen eine geringfügige Verschiebung der erstgenannten Trennlinie auf der Inhärenzskala nach unten möglich: bei thematisierender Voranstellung des Objekts können nicht-personale Tierbezeichnungen zur Verdeutlichung der Aktantenbeziehung mit *pe* versehen werden.

Translated into English, this reads:

In particular, we observe the following regularities in modern Paraguayan Guaraní with respect to the use or non-use of the . . . postposition *pe* . . . :

1. The line of division between *pe* and ∅ is in the dimension of Inherent properties, in particular at the property [± pers]. This means that expressions for animals (the area [–hum ∧ +anim]) are always then positively marked when their referents are identifiable individuals, so to say, with a "personality." Through the use or non-use of the postposition with expressions denoting animals, one can express subtle nuances in Guaraní: the speaker can precisely connote his relation to the animal by either making it human-like (→ *pe*) or making it thing-like (→ ∅). This phenomenon has a parallel in many of the languages that differentiate objects on the basis of the Inherence Hierarchy.

2. A secondary line of division runs on the scale of Reference properties through the zone [+ident ego ∧ –ident tu]: when the individual identifiability of the referent of the object is not given, the postposition is also then not used when it is a human. However, this is a typical transition zone: sometimes, although not very often, a formally indefinite noun phrase can be marked with the postposition when it fulfills the semantic condition [+ident ego] ∧ [+pers]. Such transition zones with variable use can be integrated into the here-used dimensional model without difficulties.

3. Finally, under particular syntactic conditions, a slight downward movement of the first-mentioned line of division on the Inherentness scale is possible: to clarify the relation between the arguments, *pe* can be used with non-personified animal-denoting expressions if the object is thematically fronted.

To summarize, Bossong (1985b) considers being animate and definite a sufficient condition for object marking in Guaraní. Beyond this, he predicts [+pers] objects (referring to personified animals), and presumably those higher on the Inherentness Hierarchy in (136) as well, to be always marked in Guaraní. This prediction is qualified by point 2: marked human objects must be at least [+ident ego ∧ –ident tu] on the Reference Hierarchy in (137), and if human objects exhibit this reference property they may be marked even if they are formally indefinite. Notice that this [+ident ego] criterion is essentially the same

as that by which I classified objects as specific in the corpus. Thus, for our purposes, we might reword this prediction to say that specific indefinite human objects may be marked.[15] Lastly, he claims that the restriction against marking of [–pers] and [+anim] objects (non-personified animals) may be overridden if those objects are fronted.

Bossong makes further additions to these predictions. On p. 20, he claims that proper names are always marked. And, although this follows from the earlier claim that [+anim] ∧ [+def] objects must be marked, he argues that marking on human definite objects is *zwingend vorgeschrieben* 'obligatory' (Bossong, 1985b, p. 20). He notes the following counterexample to the latter claim, in which the definite human-referring object *itajýra* 'his daughter' is unmarked:

(138)  (Adapted from Bossong, 1985b, p. 20)

Ore       rendota o-me'ẽ-ta   **i-tajýra**       pe   kuimba'é-pe.
1Excl.POSS chief    A3-give-FUT **3.POSS-daughter** that man-PE

'Our chief gave **his daughter** to the man.'

Bossong (1985b) labels this an isolated example and proposes two hypotheses as to the reason for its non-marking: (i) the human-referring DO is objectified in the sense that she is being treated like a thing and given away, and/or (ii) the speaker is trying to avoid having too many instances of the sound *pe* in the same clause (there would be three if the object were marked).

In addition to these points, Bossong (1985b, p. 21) also argues that there is DOM variation within noun phrases that are [+hum] ∧ [–def], but that most of them are unmarked. He provides examples of both. Finally, on pp. 24-5 Bossong argues that inanimate-referring objects are never marked, regardless of their discreteness, concreteness, or definiteness.

I believe Bossong's analysis can be distilled into the following basic claims (stated in the terminology used throughout this thesis for the sake of clarity):

(139)  In Guaraní:

    a.  Definite animate objects are always marked.

    b.  Human objects[16] are always marked.

    c.  Non-specific objects are never marked.

---

[15]This prediction holds only of indefinite specific humans because the previously mentioned prediction that [+anim] ∧ [+def] is a sufficient condition for object marking implies that human definite objects must be marked.

[16]Including personified animals ([+pers]), which Bossong (1985b) claims behave identically to humans with respect to object marking.

d. Specific human objects may be marked.

e. Topicalized [–human] objects may be marked.

f. Proper name objects are always marked.

g. Indefinite human objects may be marked but usually are not.

h. Inanimate objects are never marked.

### 4.11.2 Animacy/Definiteness Results from the Corpus, and the Problems They Pose for Bossong's Analysis

In light of the acceptability judgments in §4.2, it should be clear that many of the claims in (139) are too strong.[17] The acceptability of non-marking on definite animate objects like *pe mitā* 'that child' and *pe jagua* 'that dog' is incorrectly counterpredicted by (139a). Claims (139a) and (139b) are falsified by the fact that three speakers accepted the unmarked form of the human definite/specific object *pe mitā* 'that child.' Three of four speakers found marked inanimate objects to be acceptable, contrary to (139h). Furthermore, marking of inanimate objects was attested in the extended corpus, exemplified below:

(140)  Pe   tahachi      n-oi-pysyrõ-i-kuri      **pe   mohenda-há-pe**.
       That police.officer NEG-A3-save-NEG-KURI **that install-NMLZR-PE**
       'That police officer didn't save **the computer**.'

(141)  Pe   tahachi piko n-oi-nupã-i-ra'e      **peteĩ mba'yru-guatá-pe**?
       that police   QU  NEG-A3-hit-NEG-RAE **one   car-walk-PE**
       'Didn't the police officer hit **a car**?'

These are clear empirical problems for (139h). The remaining claims require examination of the corpus for testing, since none of the elicited examples in Table 4.1 were patently non-specific (139c,139d), thematically fronted (139e), or proper names (139f), and since (139g) is stated in terms of tendencies, not categorical rules of acceptability. For this reason we now turn to the corpus results.

The claims in (139d) and (139e) are not categorical constraints on DOM, but simply assert that certain constructions are licensed. I take no issue with these claims, since both marked specific objects and marked fronted non-human objects are attested in the corpus, such as the following:

---

[17] In fact, we can already see some internal discrepancies and redundancies within these claims. For example, (139b) and (139g) contradict one another, as do (139b) and (139c) under the assumption that human non-specifics exist. Meanwhile, (139d) is entailed by (139b). Nevertheless, I will deal with each claim individually here.

(126)   *Marked specific*:

O-hasá-vo    upéi o-hecha **ju'í-pe** o-guapy y    mbyté-pe peteĩ yrupẽ-ari.

A3-pass-when then A3-see **frog-PE** A3-sit water middle-PE one sieve-on

'In passing, he saw **a frog** sitting in the middle of the water on a water lily.'

(142)   *Marked fronted non-human*:

Ha    **i-vaká-pe**    nipo    o-jopi kava ra'e.

CONJ **3.POSS-cow(Sp)-PE** perhaps A3-sting wasp RAE

'And maybe a wasp stung **his cow**.'

The object in (126) is both specific and marked, corroborating Bossong's claim that marking of specific objects is possible. If I correctly understand Bossong's meaning by the term *thematisierender Voranstellung* 'thematically fronted,' the object *ivakápe* 'his cow' in (142) is thematically fronted, as well non-human and marked. It follows from this that such objects indeed can be marked in Guaraní.

The data in Table 4.7 indicate that in some sense Bossong's intuition underlying (139h) is correct. Although Table 4.1 clearly shows that animacy is not a prerequisite for the *acceptability* of object marking and that inanimate objects *can* be marked (for some speakers), it does appear to be the case that inanimates are very rarely marked in actual discourse. However, Table 4.7 presents a problem for (139b), since it is quite clear that not all human objects are marked. In fact, nearly half are not. Nevertheless if an object is human its chances of -*pe*-marking dramatically increase.

Bossong (1985b) also puts forward definiteness-related claims, to which we now turn. According to Table 4.8, only 3/7 proper names are -*pe*-marked. These data are in clear violation of Bossong's claim in (139f), that all proper names are marked.[18] Not only are they not all marked in the corpus, but they are marked fairly infrequently, relative to such a claim. Thus while the acceptability judgments in Table 4.1 render (139a), (139b), and (139h) problematic, the corpus data in Table 4.8 place (139f) in jeopardy as well. Furthermore, the claim in (139c) that non-specific objects are never marked is falsified by the existence of two marked non-specific objects in the corpus.

Since most of the claims in (139) are stated in terms of the intersection of animacy and definiteness values, fully testing them requires simultaneous examination of the dimensions of animacy, definiteness,

---

[18]Further counterevidence to (139f) comes from the fact that the permutations of (117b) were considered acceptable, all of which took an unmarked PN object.

and object-marking. For this reason, in Table 4.25 I cross-tabulate the Animacy and Definiteness Hierarchies and provide marking/non-marking counts for each cell.[19]

| | | NONSPECIFIC $+pe$ | NONSPECIFIC $-pe$ | SPECIFIC $+pe$ | SPECIFIC $-pe$ | DEFINITE $+pe$ | DEFINITE $-pe$ | PROPER NAME $+pe$ | PROPER NAME $-pe$ | TOTAL $+pe$ | TOTAL $-pe$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **HUMAN** | # of objects | 1 | 6 | 3 | 3 | 17 | 7 | 3 | 3 | 24 | 19 |
| | % marked | **14%** | | **50%** | | **71%** | | **50%** | | **56%** | |
| **ANIMATE** | # of objects | 1 | 5 | 1 | 0 | 1 | 50 | 0 | 0 | 3 | 55 |
| | % marked | **17%** | | **100%** | | **2%** | | **–** | | **8%** | |
| **INANIMATE** | # of objects | 0 | 40 | 0 | 13 | 0 | 78 | 0 | 1 | 0 | 132 |
| | % marked | **0%** | | **0%** | | **0%** | | **0%** | | **0%** | |
| **TOTAL** | # of objects | 2 | 51 | 4 | 16 | 18 | 135 | 3 | 4 | 27 | 206 |
| | % marked | **4%** | | **20%** | | **12%** | | **43%** | | **12%** | |

Table 4.25: Object-marking across the Animacy and Definiteness Hierarchies

In Table 4.25, limitations of sample size become unfortunately apparent. One cell (Animate PN) has no tokens. Another two (Animate Specific and Inanimate PN) have only one, and cannot therefore exhibit variation. Since there are only 3 animate marked objects in the corpus, and since they are evenly distributed across the Definiteness Hierarchy, the fact that there is no statistically significant difference in marking frequency between animate objects of different levels of definiteness tells us little. It could be the case that the Definiteness Hierarchy is truly irrelevant to the marking of animates, or it could be the case that there are real marking differences but the corpus is too small to reveal them. The same holds for the human object category. Of the six possible pairings of different levels of definiteness within the Human category, only one exhibits a statistically significant difference in marking frequency: human non-specific and human definite ($p = 0.0124$). However, considering the fact that each cell in the human category has relatively few tokens (with the exception of human definite), it could simply be the case that real differences in marking frequency are hidden behind the small sample size. These issues arise from the fact that in Table 4.25 the tokens in the corpus are being spread across three factor groups (object marking, animacy, and definiteness) with two, three, and four distinct factors, respectively. There simply isn't enough data to revealingly fill out such a large table, at least in some respects.

---

[19]The structure of Table 4.25 is identical to that presented in C. Shain (2008), in which I examined the predictive power of Prominence as discussed in Aissen (2003b) with respect to DOM in Guaraní. Since Aissen defines Prominence as a partial ordering on the Animacy and Definiteness Hierarchies, Table 4.25 can also be viewed as presenting the distribution of DOM in Guaraní across Prominence categories.

However, I still contend that there are very important things to be learned from Table 4.25, especially with respect to Bossong's analysis. As I discussed above, Bossong claims that animate definite objects are always marked. However, of the 51 animate definite objects in the corpus, only one is marked, which means that (139a) fails 98% of the time for this cell. However, considering the fact that (136) is an implicational hierarchy, which means that [+anim] ∧ [+def] includes objects with properties that are higher on the Inherentness hierarchy, we really should be taking into account human/definite, animate/PN, and human/PN objects when testing (139a). With these categories taken together, the results get somewhat better: 21/81 (26%) of such objects are marked. Nevertheless, this still means that (139a) has a 74% failure rate for the corpus as a whole, which I take to be strong counterevidence.

Furthermore, while I did not list this prediction in (139), point #3 from Bossong (1985b, p. 18) presupposes a prohibition against marking non-personified animals, which may be overridden if the object is fronted. However, two of the three marked animate non-humans in the corpus do not conform to this prediction:

(122)  O-ĩ-ndaje   raka'e peteĩ mitã tyre'ỹ   o-hayhú-va **mymba-kuéra-pe**.
       A3-be-REP before one   child orphan A3-love-RC  **animal-PL-PE**
       'There once was an orphan who loved **animals**.'

(143)  O-henói **h-ymba       jagua peteĩ-me**.
       A3-call  **3.POSS-animal dog    one-PE**
       'He called **one of his dogs**.'

In both (122) and (143), no independent evidence exists to suggest that these objects are anthropomorphized. Additionally, neither is fronted. Yet both are marked. This is inexplicable under Bossong's analysis.

There are other important facts shown in Table 4.25. Bossong argues that [+human] ∧ [+def] objects are obligatorily marked. Yet there are many unmarked human definites in Table 4.25. This is true regardless of whether [+def] is taken to include or exclude proper names. Excluding proper names, 17/24 human definites are marked, or 71%. Including proper names, 20/30 human definites are marked, or 67%. I take these data to be a significant problem for the aforementioned claim. Furthermore, they cannot be explained away as easily as the counterexample Bossong noted from his own corpus, which I presented in (138). Evidence of this is the following:

(119)  O-mo-ngakuaa karai     **pe   mitã**.

A3-CAUSE-grow gentleman **that child**

'The gentleman raised **the child**.'

Recall that Bossong (1985b) accounted for (138) by hypothesizing that (i) the apparently human object was being treated as an inanimate, and (ii) that there would otherwise be too many *pe* morphemes in the same sentence. Neither of these speculations can handle (119), however. The relation *mongakuaa* 'raise' does not seem to reduce its object to sub-human status; on the contrary, it seems closer to entailing humanness of the object. Furthermore, there are no occurrences of *-pe* at all in this clause, and supposing a strategy of avoiding excessive *-pe*-marking cannot account for the non-occurrence of object marking in (119). Thus (119) seems to me to be a compelling counterexample to the analysis in Bossong.

I showed earlier that (139f) incorrectly predicts the distribution of *-pe*-marking on proper names generally. However, the data in Table 4.25 also show that even if we were to revise (139f) to predict marking only on *human* PNs, this prediction would still fail. Even among human proper names, only half take *-pe*-marking.

Regarding (139g), Bossong (1985b) claims that human indefinites are infrequently marked. However, according to Table 4.25, 4/9 (44%) human indefinites are marked in the corpus. While Bossong is not specific as to what level of marking frequency would constitute a falsification of this approach, my intuition is that a marking rate of nearly half is not infrequent. Thus (139g) does not handle the corpus data well.

In sum, the distribution of *-pe*-marking with respect to values on the Animacy and Definiteness hierarchies highlights some significant weaknesses in the claims of Bossong (1985b). In fact, strong counterevidence is found to all of Bossong's claims about the distribution of *-pe* presented in (139) except (139d) and (139e), which the corpus data confirm. Animate definite objects can be and often are unmarked, contrary to (139a). Only half of human objects in the corpus are marked, contrary to (139b). Nonspecific objects can be marked, contrary to (139c). Proper name objects may be unmarked, contrary to (139f). Almost half of human indefinite objects are marked in the corpus, contrary to (139g). And indefinite objects can be and in fact are marked, contrary to (139h).

In light of this, I consider the analysis presented in §4 to be preferable. Especially given the data in Table 4.1, which demonstrate the grammaticality of (non-)marking a wide range of different object types in Guaraní, I believe the correct analysis of DOM in Guaraní must be stated in terms of marking *preferences*, and not in terms of categorical rules like many of those in (139).

## 4.12 Conclusion

In conclusion, the data show that an analysis of the distribution of DOM in Guaraní with very broad coverage is statable in terms of animacy and topicality. Human topics are the most frequently marked cases, inanimate non-topics are the least frequently marked, and some object marking occurs among [-human] or non-topical objects on the basis of the feature [±animate] and perhaps other marginally relevant factor groups, such as specificity and GF-ambiguity.

# CHAPTER 5

# THE EVOLUTION OF DOM IN GUARANÍ

An interesting follow-up question to this analysis is whether, and to what extent, the existence of DOM in modern Paraguayan Guaraní is due to contact with Spanish. I briefly address this question here. First, let me point out that there has been mutual influence between Spanish and Guaraní in other domains than DOM through the course of the nearly 500 year history of contact between the two languages (de Granda, 1996; Dietrich, 1993, 1995; Usher de Herreros, 1976). Giménez Caballero (1966, p. 125) estimates that up to half of the lexicon of certain dialects of Guaraní is of Spanish origin. J. K. Choi (2000), meanwhile, contends that the existence of null direct objects in Paraguayan Spanish is due in part to contact with Guaraní, which allows anaphoric null objects. Given the amount of contact between Spanish and Guaraní and the attested mutual influence between them in other linguistic domains, we might wonder whether DOM arose in Guaraní as a result of its prolonged contact with Spanish. Bossong (1985b, n.d.) advocates just such a conclusion, which I discuss below.

## 5.1 Bossong's Diachronic Analysis of DOM in Guaraní

Bossong (1985b, n.d.) argues that DOM was non-existent in Guaraní prior to Spanish contact, and that therefore "the formation of DOM with all its semantic properties, was actualized by language contact" (Bossong, n.d., p. 13). In support of this claim, Bossong makes reference to several Jesuit grammars of Guaraní written in the 16th and 17th centuries, including de Anchieta (1595), de Aragona (ca. 1625), and de Montoya (1640a). Since these texts were composed by Spanish-speaking missionaries, they of course do not represent an entirely pre-contact Guaraní language. The first Spanish settlement in Paraguay was founded in 1537, which places the publication of the aforementioned Jesuit texts around 50-100 years after contact began. Although this is plenty of time for contact-induced language change to take place, there

118

is some reason to believe that Spanish influence on Guaraní was somewhat more limited than on indigenous languages in other Spanish colonies. First, Paraguay's inaccessibility and lack of natural resources contributed in part to a very small amount of early Spanish immigration (Rubin, 1985). And second, the Jesuit interest in Guaraní afforded it a comparatively high social status; Guaraní became broadly used as the language of liturgy and education throughout the period of Jesuit involvement in the colony (Rubin; J. K. Choi, 2003). Regardless of the degree to which Spanish had already shaped Guaraní by the time these works were published, it has doubtless continued to exert influence since their publication. Thus these 16<sup>th</sup> and 17<sup>th</sup> century texts serve as time capsules, turning back the clock of Guaraní evolution nearly 400 years. To the extent that they reveal differences in DOM between modern Guaraní and its older form, they suggest Spanish contact as a source of change.

Regarding the details of Bossong's argument, he examines the portions of the above grammars in which the grammar of GF-assignment is discussed and points out that none of these linguists mentions object-marking as a means of GF-assignment, despite the claim by both de Anchieta (1595) and de Montoya (1640a) that GF-ambiguity exists when both S and O are 3<sup>rd</sup> person. De Montoya (p. 35) provides the example in (144) and claims that either GF-assignment is possible for it.

(144)  Peru  o-mbo'e Chua.
       Pedro A3-teach Juan
       'Pedro teaches Juan' or 'Juan teaches Pedro'

Bossong's reasoning is that a discussion of GF-assignment and ambiguity would be exactly the place to introduce DOM, since -*pe*-marking would disambiguate in this case. Thus the silence of all the early Guaraní grammarians on the subject of DOM is rather "loud," so to speak, and might constitute evidence that there was no object marker in Guaraní prior to Spanish contact. While this conclusion is reasonable, and it would certainly seem strange for no grammarian of Guaraní to discuss DOM if it indeed existed, this is still essentially an argument from *absence of description*, which is an unsatisfying basis for drawing conclusions about the non-existence of a phenomenon in a language. A more compelling case could be made on the basis of absence of *attestation*, that is, if DOM were non-attested in corpora written before substantial Spanish contact. Thankfully, de Montoya (1640b) provides just such a corpus: he translated the catechism into Guaraní and included it interlinearized with Spanish in his collection of works on the language. To my knowledge, this catechism has not been looked at with respect to DOM. Given Bossong's claims,

such a study would produce inherently interesting results, since they would either contradict Bossong's argument or provide significant support for it.

## 5.2  DOM in de Montoya's Catechism

For this reason, I exhaustively surveyed Montoya's Guaraní catechism in search of object marking. Let me make some preliminary remarks before discussing the results. An important difference between what Bossong calls "Missionary Guaraní (MG)" — the Guaraní of the Jesuit grammars — and modern Guaraní (Guaraní) is the non-homophony of the spatiotemporal postposition and the IO marker, both of which are realized as *-pe* in Guaraní. In MG, however, the postposition is realized as *-pe* and the IO marker is realized as *upe*. These facts are demonstrated by the following examples:[1]

(145)    a.  Yvá-**pe**, yvý-**pe**...

'In the sky, on the earth...' (de Montoya, 1640b, p. 64)

b.  Santissima Trinidad **upe** i-kuaave'ẽnga.

'He shows them to the Most Holy Trinity.' (de Montoya, 1640b, p. 202)

In example (145a), the morpheme *-pe* expresses a spatial relation to the sky and the earth. In (145b), on the other hand, *upe* attaches to the IO *Santissima Trinidad* 'Holy Trinity,' the Recipient/Goal of a "showing" action. Montoya's Guaraní-Spanish dictionary, *Tesoro de la Lengva Gvaraní*, confirms this generalization. He describes *upe* as a "dative particle, the same as *para*[2] (de Montoya, 1639, p. 406, my translation)." Meanwhile, he considers *-pe* to be a "posposicion de quietud, y movimiento (de Montoya, p. 263)," which I translate as "postposition of stillness, and movement." Perhaps this means that *-pe* as a spatiotemporal postposition can be both locative ("stillness") and directional ("movement"). Therefore, in Missionary Guaraní, *-pe* and the IO marker are formally and distributionally distinct.

Thus there is a question as to whether *-pe* as a DO marker in Guaraní evolved from the IO marker *upe* or the postposition *-pe* in MG. Bossong (1985b) takes the position that both modern object markers (IO and DO) evolved from *upe*. This seems reasonable in the light of claims in other literature that the differential object marker is often homophonous with the dative (IO) marker cross-linguistically (Bossong, 1985a; Aissen, 2003b).

---

[1]Since Guaraní has undergone significant changes since this translation of the catechism, no segment-by-segment glosses are provided in this chapter. The translations given are best approximations based on Montoya's Spanish-Guaraní dictionary (de Montoya, 1640a) and Guaraní-Spanish dictionary (de Montoya, 1639), as well as the original Spanish text. The orthography in the Guaraní examples has also been updated, since de Montoya used a transcription system that today is non-standard.

[2]The preposition *para* is a dative marker in Spanish, marking Recipients and Goals.

In the catechism, there were clearly no direct objects marked by the postposition *-pe*. The IO marker *upe* was somewhat less clear, since it is often uncertain (without speaker intuitions) whether a given object is an IO or a DO. For verbs which took a Recipient/Goal as an object, such as *me'ẽ* 'give,' *ñembo'e* 'pray,' *mombe'u* 'tell,' and *ñe'ẽ* 'speak,' that object was not considered a DO. This judgment is supported by the fact that, in the case of verbs which take both a Theme and a Goal, such as *mombe'u* 'tell,' the Theme argument takes no *upe*-marking and the Goal argument does. The following examples demonstrate this:

(146)   a.   **Tupã rerovia-ha**, e-*mombe'u*-ehéve

'*Tell* me **the Articles of the Faith**.'[3] (de Montoya, 1640b, p. 48)

b.   E-*mombe'u* **ore-ruva**

'*Say* **the Our Father**.' (de Montoya, 1640b, p. 62)

(147)   a.   A-ñe *nombe'u* **Tupã pãvẽ-rembiapia upe**.

'I *confess* **to God Almighty**.' (de Montoya, 1640b, p. 35)

b.   Pe-ñe *mombe'u* **Pai upe**.

'*Confess* **to the priest**.' (de Montoya, 1640b, p. 107)

The sentences in (146) and (147) all involve the verb *mombe'u* 'tell.' The examples in (146) contain Theme objects of this verb, namely, the thing that is told. In (146a) this object is *Tupã reroviaha*, literally 'the faith-in-God-things,' but, according to the Spanish counterpart, referring specifically to the "Articles of the Faith" laid out in the catechism. In (146b) this object is *oreruva*, the Our Father, the well-known prayer of Jesus also called "The Lord's Prayer." In contrast with (146), the sentences in (147) contain Goal objects of *mombe'u*: the person being spoken to. In (147a) this person is *Tupã pãvẽrembiapia* 'God Almighty,' and in (147b) this person is *Pai* '(the) priest.' Notice that the Theme objects in (146) have no *upe*-marking, while the Goal objects in (147) are *upe*-marked. This is consistent with an analysis of *upe* as an IO marker, as well as with the assumption that the Goal-type objects of these verbs are not actually DOs.

Some *upe*-marked objects, which did not involve a clear Recipient/Goal, were more difficult to judge with respect to whether they were DOs or IOs. The sentence in (148) is one such example.

(148)   Ñande arakuaa aivy pype **hemiapo upe** ja-vãhẽ-mō i-*kuaá*-pa.

'In our weak understanding may we come to *know* **his works**.' (de Montoya, 1640b, p. 266)

---

[3]I am not certain what the correct gloss of *ehéve* should be on the basis of Montoya's dictionary. My guess is that it is actually *chéve*, the 1st person singular IO pronoun, which would be consistent with the context.

Example (148) appears to contain an *upe*-marked DO of the verb *kuaa* 'know.'[4] To my mind, it would be fairly non-intuitive to conceive of *hembiapo* 'work' to be the Recipient of some "knowing" action; it rather seems like the Theme of a relation like "know/be familiar with." Since the thematic relation assigned to the object of *kuaa* 'know' is not the Goal/Recipient/Beneficiary commonly associated with IOs (see Jackendoff, 1990; Dowty, 1991), considering this object a DO seems to have more *prima facie* plausibility than doing so for the objects in (147), for example.

As I discussed at points above, I know of two linguistic tests for determining whether a certain object of a certain verb is an IO or a DO in Guaraní. The first is that pointed out by Bossong (1985b), namely, that DOs are encoded in portmanteau agreement prefixes when S is 1st person singular and O is 2nd person, while IOs in the same context are encoded as 2nd person IO pronouns (see §3.1.4). Thus, our task is to find out whether, when S is 1st person singular, we find the agreement prefix *ro* '12Sg' when O is 2nd person singular or the agreement prefix *po* '12Pl' when O is 2nd person plural (indicating that the object is a DO), or whether we find a 1st person agreement prefix and a 2nd person IO pronoun (indicating that the object is an IO). Unfortunately, I was unable to locate any instances of *kuaa* 'know' in the catechism with a 1st Sg subject and a 2nd person object, and this test yielded no results in this case.

The second test rests on the assumption that, as in modern Guaraní, IO marking was obligatory in MG. If this is true, then variation in the appearance of the IO marker *upe* on the object of a certain verb is evidence that that verb takes a DO, not an IO. Applying the latter test to the verb *kuaa* 'know,' we do find instances in which, unlike (148), it takes an unmarked object. Two are provided below.

(149)  **Ñembo'e opa-katu** oi-*kuaá*-ne i-mbo-po-potá-vo.

'He has to *know* **all the prayers**, with a desire to complete them.' (de Montoya, 1640b, pp. 115-6)

(150)  **Ore-tuva**, *kuaa* katu-ramo.

'To *know* **the Our Father**.' (de Montoya, 1640b, p. 44)

Both of the above examples contain an unmarked object of *kuaa* 'know.' Two related conclusions might be drawn from this: first, that *kuaa* 'know' takes a DO, not an IO, in which case that DO is marked in (148), and second, that the marking of the object of *kuaa* 'know' is differential. If these conclusions are valid, then these examples constitute counterevidence to Bossong's claim that "MG does not know DOM" (Bossong, n.d., p. 13). However, the story is somewhat more complex, since (148) is actually the *only* instance of

---

[4]The initial judgment that *hemiapo upe* is an object of *kuaa* rather than *vāhē* 'arrive' was made under the assumption that the "arrive" relation is intransitive (one-place) and on the basis of the fact that the original Spanish text reads "... conocer *sus obras maravillosas*," which translates into English as "... know *his marvelous works*." This assessment of *hemiapo upe* is revised later in light of certain data.

a marked object of *kuaa* 'know' in the entire catechism. Taking into account the fact that the verb *vāhē* 'arrive' appears in (148) in addition to *kuaa*, it is interesting to note that de Montoya (1639, p. 406) ascribes a directional meaning ("de movimiento") to the IO marker *upe* as well, which he exemplifies with the following:

(151)  **Tupã upe** o-*ho*-va'e-rã-ma

'Those who have yet to *go* **to God**' (de Montoya, 1639, p. 406)

In light of this, it is quite possible that *hembiapo upe* 'his works' in (148) is a Goal of *vāhē* 'arrive' instead of a Theme of *kuaa* 'know.' The following example is consistent with this hypothesis:

(152)  **Ko Sacramento upe** t-o-*vahē* yvykatu, o-javo.

'In order to *arrive* worthily **at this sacrament**.' (de Montoya, 1640b, pp. 272-4)

The marked phrase *ko Sacramento upe* '(to) this sacrament' seems to be a Goal of *vahē* 'arrive,' lending credibility to the hypothesis that *hemiapo upe* '(to) his works' in (148) is also a Goal of *vāhē* 'arrive.' This is of course speculative. However, in light of this possibility I do not think the existence of (148) constitutes sufficient evidence to reject outright Bossong's claim that DOM did not exist in MG, since the marked NP in (148) may not even be a DO at all. Furthermore, *hemiapo* 'his works' fits the profile of a patently unmarked DO according to both the patterns of DOM in modern Guaraní and the patterns of DOM typologically: it is inanimate, non-specific, and very low in topicality (T-RD = 40, T-TP = 0). These facts together call into question any analysis of *hemiapo upe* '(to) his works' as a marked DO.

Another potential challenge to Bossong's claim has to do with the verb *ñyrō* 'forgive.' Its objects exhibit *upe*-marking in the corpus. Although "forgive" is intuitively a three-place predicate involving a forgiver, a forgiven person, and an offense to be forgiven, objects of *ñyrō* 'forgive' can be *upe*-marked regardless of which of the latter two roles they fill. This is shown in the contrast between (153)/(154) and (155).

(153)  Nde-*ñyrō* **ore i-ñangaipá-va'e-upe**.

'You *forgive* **our sins**.' (de Montoya, 1640b, p. 3)

(154)  **Poromomarã-ha-rupe**, *ñyrõ*.

'*Forgive* **an insult**.' (de Montoya, 1640b, p. 23)

(155)  i-*ñyrō*-ha-re-y **guapicha-rupe**

'those who do not *forgive* **others**' (de Montoya, 1640b, p. 68)

In (153) and (154), the *upe*-marked object is the offense being forgiven: *ore inangaipáva'eupe* 'our sins' in (153) and *Poromomarãharupe* 'an insult' in (154). In (155), however, the *upe*-marked object is the person being forgiven: *guapicharupe* 'other people.' At least in English, both of these cases are realized as DOs. The object in (155), which is a person, could reasonably be conceived of as a Recipient of the transfer of forgiveness, and for that reason could be a candidate for IO status. The objects in (153) and (154), however, seem much less amenable to categorization as Recipients/Goals. They seem to be Themes of the forgiveness act. Thus it is possible that these are *upe*-marked DOs.

The question of whether *ñyrõ* 'forgive' takes an IO or a DO should be answered on the basis of the aforementioned tests. Unfortunately, as before, Bossong's test provides no information, because there are no uses of *ñyrõ* 'forgive' in the corpus with a 1st Sg subject and a 2nd person object. However, the variation test (according to which variation in *upe* marking is evidence against IO-hood), produces interesting results: I could find no instances of *ñyrõ* 'forgive' in the catechism without an *upe*-marked object. This absence of variation suggests that *ñyrõ* 'forgive' simply selects an IO object.[5]

Regardless of the existence of the above examples that might suggest, however weakly, that a system of DOM was already active in the early days of Spanish-Guaraní contact, the number of these examples is quite small. Counterbalancing them are many other examples in which clear candidates for object marking under the rules of modern Guaraní appear unmarked in the MG text. Some are provided below. They contain [+human] and highly topical DOs, the properties associated with high marking frequency in the modern Guaraní corpus:

(156)   Ere-*ayhu* **Tupã**-ne mba'e pãvẽ asoce.

'*Love* **God** more than all other things.' (de Montoya, 1640b, p. 10)

T-RD = 3, T-TP = 10

---

[5]There is an example from the corpus that does not fit cleanly with such a picture, presented as (1) below.

(1)   Nde-*ñyrõ* **anga** *chéve*.
'You *forgive* *me* **my soul**.' (de Montoya, 1640b, p. 214)

It looks to be the case that *ñyrõ* here actually takes two objects: a 1st person object realized as an IO, *chéve*, and a 3rd person object *anga* 'soul' which takes no IO marking. If this is true, it would suggest that object marking of the object of *ñyrõ* is indeed differential, since there is an unmarked object in (1). However, there are enough variables at play with this example that I am unable to reliably assess its status with respect to my conclusions above. First, since this is the only construction I found like this, it is potentially idiomatic. This would be consistent with the fact that the Spanish counterpart only has one 1st person object and makes no mention of 'soul': "Perdona**me**," or "Forgive **me**." Second, if there are in fact two objects here, the semantic relation between the objects and the verb seems to differ from those exhibited in (153) – (155), in that "soul" does not seem to be an offense to be forgiven, and it is unclear whether *anga*, *chéve*, or both are the Recipients of the forgiveness act. For these reasons I do not assess the impact of (1) on the above conclusions about the subcategorizational properties of *ñyrõ* 'forgive,' since without a more reliable analysis of (1) it is unclear what that impact would be.

(157)   A-*rovia* ave **Jesu Christo**.

      'I *believe* also **in Jesus Christ**.' (de Montoya, 1640b, p. 5)

      T-RD = 4, T-TP = 13

(158)   Hũmãme Tupã **aipo Angeles** i-*mõñãngi* ra'e?

      'Where did God *create* **those angels**?' (de Montoya, 1640b, p. 72)

      T-RD = 2, T-TP = 16

(159)   Ha'e-nia o-*hecha* **Tupã** ne.

      'He will *see* **God**.' (de Montoya, 1640b, p. 33)

      T-RD = 6, T-TP = 9

Examples (156) – (159) are exactly the kinds of clauses in which object marking would be predicted for modern Guaraní, yet none appears, suggesting that DOM was not active in MG, at least with distributional properties resembling those in the modern language. This taken in tandem with the general absence of variation in *upe*-marking across verbs (verbs seem to select either marked or unmarked objects), and the explicability in other terms of the single potentially compelling example of a differentially-marked DO, presented in (148), I believe Montoya's catechism provides substantial support for Bossong's claim that DOM did not exist in MG prior to Spanish contact.

    This analysis is consistent with the difference in distribution of the pronoun *ichupe* between modern Guaraní and MG. As I showed in §3.1.3, in modern Guaraní *(i)chupe* is a general 3rd person object pronoun, which can be either an IO or a DO. However, in the catechism, *ichupe* only co-occurs with verbs that take an *upe*-marked IO object. De Montoya (1639, p. 406) translates *ichupe* as *a él*, 'to him,' suggesting an exclusively dative usage of *ichupe*. This immediately raises the question of what the DO pro-form was in MG. I was unable to find such a proform in the catechism, although I did observe that anaphoric DOs were often given null realization. Perhaps the null object was the basic anaphoric DO pro-form in MG. It is outside the scope of this thesis to rigorously test this hypothesis here (looking for the absence of something requires a much more thorough survey of the corpus than looking for the presence of something), and I provide it as a suggestion for future research.

## 5.3   DOM in Guaraní and Spanish Today

Having shown that DOM is unattested in early Guaraní texts, consequently providing support for the hypothesis that DOM in Guaraní is borrowed from Spanish, an important follow-up question is whether

DOM is distributionally equivalent in both languages. That is, is it the case that a marked object in Guaraní will be marked in a Spanish translation, and vice versa, or has DOM been put to different uses in Guaraní since its appropriation from Spanish?

A cursory comparison between the analysis presented here and common analyses of Spanish shows that the two DOM systems are not fundamentally different from one another, in that they are sensitive to similar factor groups. Both are conditioned by animacy and topicality (see Laca, 1995; Leonetti, 2004), although Spanish DOM is also conditioned by definiteness/specificity (Aissen, 2003b; Tippets & Schwenter, 2007; von Heusinger & Kaiser, 2003), which was not shown to be significant for this Guaraní corpus. Nevertheless, it does not appear to be the case that dramatic distributional changes occurred in the transfer of DOM from Spanish to Guaraní.

However, the DOM in both languages is not truly distributionally equivalent. One important difference between them has to do with whether categorical distributional rules hold.

(22)   a.   Vi            *(a)      **la/una**       **mujer**.
            1Sg.PST.see *(ACC) **DEF/INDEF woman**.
            'I saw **the/a woman**.' (von Heusinger & Kaiser, 2007, p. 87)

(160)   Anita o-hecha **peteĩ kuña**.
        Anita A3-see   **one   woman**
        'Anita sees **a woman**.' (see Table 4.1)

(117)   b.   Juan o-hecha **Maria**.
            Juan A3-see   **Maria**
            'Juan saw/sees **Maria**.' (Tonhauser & Colijn, to appear, p. 30)

As (22a) shows, non-marking of the [+human] object *la/una mujer* is ungrammatical in Spanish. The equivalent restriction does not hold in Guaraní, as shown by the non-marking of the indefinite [+human] object in (160) and the definite (PN) [+human] object in (117b), which is presumably grammatical for both. There are in fact many examples in the literature on Spanish DOM of categorical constraints on (non-)marking (see §2.2). As I argued at length above, no such constraints appear to be active in Guaraní.

Further evidence for the non-equivalence of the distribution of DOM in Guaraní vs. Spanish comes from the extended corpus, which came interlinearized with a Spanish translation. In several instances, object marker *a* appears in the Spanish translation where *-pe* does not occur in the Guaraní text. Here are some examples:

(161) I-mandu'a'-ỹ-re        umi isrrael-guá-re umi taigua-nguá-re ha    opava-ve

B3-remember-without-REHE those Israel-of-REHE those Taiwan-of-REHE CONJ all-more

servísio    intelihénsia     rehe-gua o-mbotá-va **h-okẽ**.

service(Sp) intelligentsia(Sp) REHE-of A3-hit-RC   **3.POSS-door**

Spanish: '*Sin mencionar a los israelíes, los taiwaneses y a todo servicio de inteligencia que golpeara*

***a su puerta*.'

English: 'Not to mention those Israelis, Taiwanese, and every other intelligence service that knocked

**on his door**.'

(162) Noriega oi-porú-kuri **umi ahénsia   de    Intelihénsia**.

Noriega A3-use-KURI **those agency(Sp) of(Sp) intelligence(Sp)**

Spanish: '*Noriega usó **a las agencias de inteligencia***.'

English: 'Noriega used **those intelligence agencies**.'

(163) A-hecha   **Dr. Ojeda**.

A1.Sg-see **Dr. Ojeda**

Spanish: '*Yo veo **a Dr. Ojeda***.'

English: 'I see **Dr. Ojeda**.'

(164) Ha    Jesús o-hechá-vo **pe Jerusalén** o-ñe-pyrũ     h-asẽ.

CONJ Jesus A3-see-when **that Jerusalem** A3-JE-overcome 3.POSS-lament

Spanish: '*Y Jesús al ver **a Jerusalén** empezó a llorar*.'

English: 'And when Jesus saw **Jerusalem**, he was overcome with tears.'

(165) Anína   Nika re-juka **nde    rapicha**.

NEG.IMP Nika A2.Sg-kill **2Sg.POSS person**

Spanish: '*Nika... no le mates **a tu prójimo***.'

English: 'Nika, don't kill **your fellow man**!'

In each of the above examples, when an unmarked DO in Guaraní is translated into Spanish it takes the

Spanish object-marker *a*. These data points make it clear that, despite the similar kinds of features under-

lying DOM in Spanish and Guaraní (animacy and topicality), there are distributional differences between

the two languages with respect to DOM. The above examples give the impression that object marking is

simply more widespread in Spanish than in Guaraní, i.e. that perhaps for each marked object in Guaraní

the Spanish equivalent will be marked, but not necessarily vice versa. While this is a possibility worth exploring, I found this example in the extended corpus in which Guaraní marks an object that Spanish does not:

(140) Pe   tahachi      n-oi-pysyrõ-i-kuri      **pe   mohenda-há-pe**.
      That police.officer NEG-A3-save-NEG-KURI **that install-NMLZR-PE**

      Spanish: '*El policía no salvó **el ordenador**.*'

      English: 'That police officer didn't save **the computer**.'

The data presented in (22a), (117b), and (160) – (140) thus indicate that DOM in Guaraní and DOM in Spanish are not distributionally equivalent. The obvious next question is what exactly the distributional differences are between the two languages, and why these differences exist. I cannot treat this question here, and leave it to future research.

## 5.4   Conclusion

I believe the foregoing discussion provides significant support for the claim in Bossong (1985b, n.d.) that there was no system of DOM in Guaraní prior to contact with Spanish. It of course does not follow automatically from this that the existence of DOM in modern Guaraní is entirely due to Spanish contact, since many things can induce language change, but it seems like a reasonable follow-up hypothesis. This will need to be explored by much more detailed historical linguistic research in the future.

Finally, I must qualify the conclusions drawn here in light of the nature of the data. Unlike much of the modern Guaraní corpus used for the bulk of this thesis, de Montoya (1640b) is not naturally occurring speech data, but, as a catechism, was assumably composed in an elevated register. Neither is it the product of a native speaker of MG, which calls into question its reliability as a basis for drawing conclusions about the linguistic properties of MG. In spite of this, de Montoya seems to be regarded as a skilled linguist who produced an accurate representation of the language: Huonder (1912, p. 224) provides statements to this effect from Mulhall, Hervas, Von der Gabelentz, and Platzmann. Regardless of how well the properties of MG are preserved in Montoya's catechism, it is one of the only long texts in Guaraní from the period. In this sense it is much better than nothing, but the doubts surrounding its reliability as a representative sample of the Guaraní spoken prior to Spanish contact cannot be fully eliminated.

# CHAPTER 6

# CONCLUSION

To briefly summarize, I tested different potentially relevant factor groups against the actual distribution of DOM in the corpus.

(166)    a. Animacy

       b. Definiteness/Specificity

       c. Relative Animacy

       d. Relative Definiteness/Specificity

       e. GF-Ambiguity

       f. TF-Ambiguity

       g. Topicality

I found the combination of speaker data and corpus data to show that the previous analysis of DOM in Guaraní, Bossong (1985b), as well as my own approach in C. Shain (2008), faced some crucial problems. In my own case, correcting the definition of GF-ambiguity renders ambiguity a marginal consideration at best in the distribution of DOM in Guaraní, consistent with the claim to this effect in Bossong (1985b). In Bossong's case, the problems are traceable to the fact that his analysis is stated primarily in terms of categorical grammatical rules, none of which hold when tested against both the corpus and acceptability judgments from speakers.[1] However, his basic intuition that animacy is of crucial importance to the actual distribution of object marking in Guaraní is certainly born out by the corpus. In light of the problems with the previous analyses highlighted by this corpus study, I proposed the distributional analysis in (133) and showed it to be quite successful in predicting DOM variation in the corpus.

---

[1]The only predictions of his that are supported by the data are (139d) and (139e), which are not categorical rules.

In light of Bossong's (1985b) claim that DOM in Guaraní is the result of Spanish contact, I proceeded to survey Jesuit Guaraní literature published prior to substantial Spanish contact, in search of DOM. My findings supported Bossong's proposal, and no clear cases of DOM could be found. This suggests (though does not prove) that contact with Spanish is responsible for the appearance of DOM in Guaraní, and perhaps provides a hypothesis as to why the factors underlying DOM in Guaraní and Spanish are so similar.

The analysis presented here has implications for earlier work on the subject. For example, I showed in §4.11 that the analysis of the distribution of DOM in Guaraní presented in Bossong (1985b) runs into difficulties in predicting the corpus data. I showed that this is primarily because Bossong's analysis is formulated largely in terms of categorical distributional rules, which fail in the light of acceptability judgments and corpus data showing that both marking and non-marking are grammatical on apparently any kind of object. In order to capture these facts, I argued that a descriptively accurate analysis of DOM in Guaraní will be stated probabilistically rather than categorically.

The fact that topicality plays such a critical role in Guaraní cannot be straightforwardly accounted for either under the Markedness Approach (MA) of Aissen (2003b) or the Transitivity Approach (TA) of Hopper and Thompson (1980) as they stand, since neither of these theories makes explicit reference to topicality. I showed in C. Shain (2008), as well as the study above, that examining DOM in Guaraní with respect to animacy and definiteness/specificity alone leaves important facts unexplained. Since both the MA and TA look only at animacy and definiteness/specificity in accounting for DOM, neither can overcome this problem. If either of these approaches is to be general enough to handle the Guaraní data, topicality will need to be integrated into them. This integration seems to me to be potentially straightforward for either approach. Regarding the MA, assumptions about what kinds of objects are typical/atypical would need to be augmented with the assumption that objects are typically non-topical, and that highly topical objects are therefore the marked case.[2] This assumption follows well from the claim in the topicality literature that subjects are grammaticalized topics, and that predicates usually "comment on" the subject topic (Givón, 1976). Regarding the TA, an additional "Topicality of O" feature could be included as a Transitivity feature, with "low" corresponding to low Transitivity and "high" corresponding to high Transitivity. Whether or not this conflicts with the aforementioned association between Topic/Comment information structure and subject/predicate syntactic structure would need to be explored more completely.

In addition to the above theoretical implications, the above findings suggest that topicality may be a more fundamental condition on DOM cross-linguistically than has been generally assumed, insofar as

---

[2]Escandell-Vidal (2008) makes an argument to this effect.

topicality is related to the observed conditions on DOM such as animacy, specificity, and definiteness. It has been proposed in the literature that there is a significant relationship cross-linguistically between topicality and animacy, such that human objects are frequently topical, animate objects are less frequently topical, and inanimate objects are infrequently topical (Rosenbach, 2002; Leonetti, 2004; Schwenter, 2006; Laca, 1995; Bentivoglio, 1983). This pattern is born out by the Guaraní corpus data, demonstrated by Table 6.1, which shows the average T-RD and T-TP values for objects in each of the three animacy categories.

|      | Human | Animate | Inanimate |
|------|-------|---------|-----------|
| T-RD | 18.6  | 23.9    | 29.1      |
| T-TP | 9.51  | 3.51    | 1.61      |

Table 6.1: Topicality and the Animacy Hierarchy

In terms of both T-RD and T-TP, the average level of topicality is directly related to level on the Animacy Hierarchy in (12). There is a statistically significant difference in topicality between each pair of items on the Animacy Hierarchy for each topicality measure except for T-RD between human and animate objects ($p = 0.0982$). Thus I take the data in Table 6.1, to represent a real relationship between animacy and topicality.

That there is a positive relationship between specificity and topicality has also been proposed in the literature (Leonetti, 2004; Enç, 1991; Diesing, 1992; Portner & Yabushita, 2001). The corpus data corroborates this for Guaraní. Table 6.2 gives the average T-RD and T-TP for specific vs. non-specific objects in the corpus.

|      | SPECIFIC | NON-SPECIFIC | Ratio +/− | P VALUE    | $t$ STATISTIC | DF  |
|------|----------|--------------|-----------|------------|---------------|-----|
| T-RD | 24.2     | 33.5         | 0.722     | **< 0.0001** | 4.4711        | 231 |
| T-TP | 4.02     | 1.19         | 3.38      | **0.0013**   | 3.2569        | 231 |

Table 6.2: Topicality and specificity

As Table 6.2 demonstrates, there is a a statistically significant difference between specific and non-specific objects in the corpus in terms of both T-RD and T-TP that shows specific objects to be more topical than non-specific objects.

There also seems to be a cross-linguistic relationship between topicality and definiteness. Consistent with the results from Gundel et al. (1993), the difference between items on the Definiteness Hierarchy of Aissen (2003b) can be seen as a function of Referential Distance (a topicality measure), under the assumption that more recently mentioned individuals are more cognitively activated. Givón (1990) also sees definiteness and topicality as directly related to one another.

Although I know of no studies that demonstrate this, it is certainly possible that more affected objects or objects of telic propositions are more topical cross-linguistically than their alternants. In light of the above, a useful question to be explored by future research is the degree to which object marking alternations cross-linguistically are based on topicality or clearly topicality-related properties. Such an approach would need to allow room for the grammaticalization of object marking in a given language with respect to the topicality-related features mentioned above, such as animacy (as shown here for Guaraní, see Laca, 1995, for Spanish), specificity (Enç, 1991, Turkish), or definiteness (Danon, 2001, Hebrew). Nevertheless, hypothesizing topicality as the core property underlying object marking typologically might be a promising way of unifying the various features relevant to DOM, and capturing their fundamental relationship to one another (see Leonetti, 2004). If this hypothesis is born out, the oft-used and quite intuitive label of marked objects as "prominent" might be replaced by the more empirically measurable label of marked objects as "topical."

In addition to these broader theoretical questions, a number of other questions surfaced through the course of this work that I was unable to give a rigorous answer to, which I present as topics for future research. First, the significance of other factor groups explored here, such as the Definiteness Hierarchy and GF-ambiguity, was unable to be determined because of sample size limitations, although their relationships to DOM were demonstrably weaker than those of animacy and topicality. Furthermore, the interaction between animacy, relative animacy, and thematic fit ambiguity could not be teased apart. A larger corpus might supply enough examples at the margins of these factor groups to reveal which are more predictive of object marking in Guaraní. Additionally, I was unable to evaluate DOM with respect to object affectedness or telicity in any systematic way here. An important question for future exploration is the degree to which these properties influence the distribution of DOM in Guaraní, and to what extent they interact with each other (see Dowty, 1991) and other relevant factor groups like animacy and topicality.

Finally, although a distributional analysis of this was outside the scope of this thesis, Guaraní exhibits null/overt object and object incorporation alternations in addition to the null/overt case alternation discussed here. Both of these alternation types have ben described in the literature as DOM (Lima, 2006). An example of an object overtness alternation is given in (82).

(82) a. Pero    o-topa hikuái    **la      sirujuano   por    kasualidad o-ĩ-va    avei**
       but(Sp) A3-find 3Pl.SUBJ.PRO **the(Sp) surgeon(Sp) by(Sp) chance(Sp) A3-be-RC also**
       'But they found **the surgeon who happened to be there, too**.'

b. O-topa hikuái        ∅    ha    o-mbyvy chupe.

A3-find 3Pl.SUBJ.PRO **(OBJ)** CONJ A3-stitch 3OBJ.PRO

'They found ∅ **(him)** and he stitched her.'

In (82), the bolded objects have the same referent: the surgeon. In (82a), the object is a full descriptive NP. In (82b), there is no overt object at all.

The following contrast is an example of an object incorporation alternation:

(81)  (Adapted from Velázquez-Castillo, 1996, p. 107)

a. A-ñamí-ta        pe-**vaka**  moroti
A1.Sg-milk-FUT that-**cow** white
'I'll milk that white **cow**.'

b. A-**vaka**-amí-ta        ko-pyhareve
A1.Sg-**cow**-milk-FUT this-morning
'I'll do some milking this morning.'

In (81a), the object *pevaka moroti* 'that white cow' is realized morphologically independently from the verb. On the other hand, in (81b) the object *vaka* 'cow' is incorporated into the verb. An analysis of the distribution of object overtness and object incorporation alternations in Guaraní could yield informative results, and the relationship between the distributions of these two alternations and the *-pe*-marking alternation could have important ramifications for the study of DOM. Speaking entirely speculatively, perhaps object coding forms are distributed along a spectrum of prominence/topicality, such as the following, where ">" means "more prominent/topical than":

(167)   Null object > Independent marked object > Independent unmarked object > Incorporated object

I leave this question to future study.

# APPENDIX:
# CORPORA (AND OTHER SOURCES) USED IN THIS STUDY

There were four data sources used in this study:

(168)   a.  A corpus of naturally-occurring data collected by Judith Tonhauser

       b.  Acceptability judgments elicited from speakers by Judith Tonhauser

       c.  An "extended" corpus of Guaraní texts compiled from data made publicly available at `http://crl.nmsu.edu/say/guarani/`.

       d.  Data from published scholarly works

Of these, only (168a) was exhaustively analyzed with respect to the relevant features, and all of the statistical data presented in Chapter 4 come from this corpus alone. The other sources were referred to for illustrative purposes or to gather data relevant to non-statistical portions of the analysis.

Table A contains names, descriptions, and word counts for the various texts making up (168a), the primary corpus analyzed in this study. The sum total of the words in the corpus is 6463, and the sum total of the relevant tokens to this study is 233.

| Name | Content | Word Count |
|---|---|---|
| *José Mbegue* | A play about the struggles of a rural Paraguayan family, by Jaime Bestard | 3630 |
| *Rossani* | A mother's personal narrative about nursing her daughter through an accident | 1095 |
| *BDF-1\** | A folk tale about the adventures of a boy and a dog in pursuit of a frog, as told by SC | 412 |

Table A: Texts in the primary corpus, collected by Judith Tonhauser

*Title, word count, and authorship information from Tonhauser and Colijn (to appear).

Continued

Table A continued

| | | |
|---|---|---:|
| *BDF-2\** | A folk tale about the adventures of a boy and a dog in pursuit of a frog, as told by NC | 247 |
| *Ka'i\** | A fable about a mischievous monkey and his unlucky friend Fox (author unknown) | 375 |
| *Michi\** | A personal narrative about growing up, as told by MM | 283 |
| *Kiri Kiri\** | A fable about the adventures of a cricket (author unknown) | 196 |
| *Jakare\** | An explanation of crocodile behavioral patterns (author unkown) | 143 |
| *Ypei\** | A fable about a friendship between a duck and a frog (author unkown) | 82 |
| | | 6463 |

The "extended" corpus was comprised of a number of texts made available by the Computing Research Laboratory at New Mexico State University. The particular texts used in this study were taken from the "Parallel text — sentence aligned" corpus. A plurality of the words in the extended corpus come from published texts, but several thousand more come from elicitation sessions, and still more from transcribed spoken Guaraní. As I previously mentioned, this corpus was used as a subsidiary resource in this work. In Table B I provide filenames, descriptions, acquirer names, and word counts for each of the texts I included in the extended corpus. I refer the reader to the URL given in (168c) for access to these sources.

| File Name | Content Type | Acquirer | Word Count |
|---|---|---|---:|
| allsentences-rené.doc | Elicitation session | Juan Alberto | 3581 |
| phrasebookedited-René.doc | Elicitation session | Juan Alberto | 1927 |
| prasebookedited-Christina[1].doc | Elicitation session | Juan Alberto | 1795 |
| elicitation corpus1.doc | Elicitation session | Juan Alberto | 1775 |
| elicitationcorpus2.doc | Elicitation session | Juan Alberto | 3513 |
| OBJETIVOS 7° EDUC.MUSICAL.doc | Written instructions | Danito | 2137 |
| ARTES PLÁSTICAS 7°.doc | Written instructions | Danito | 1581 |
| La trayectoria de Noriega | Written narrative | Danito | 3250 |
| Kalaíto Vore (01).doc | Written narrative: Chapter 1 of *Kalaíto Pombero* | Danito | 822 |
| Kalaíto Vore (02).doc | Written narrative: Chapter 2 of *Kalaíto Pombero* | Danito | 455 |
| Kalaíto Vore (03).doc | Written narrative: Chapter 3 of *Kalaíto Pombero* | Danito | 874 |

Table B: Texts in the extended corpus

Made available on New Mexico State University's Computing Research Laboratory's website.

Continued

Table B continued

| File Name | Content Type | Acquirer | Word Count |
| --- | --- | --- | --- |
| Kalaíto Vore (04).doc | Written narrative: Chapter 4 of *Kalaíto Pombero* | Danito | 949 |
| Kalaíto Vore (05).doc | Written narrative: Chapter 5 of *Kalaíto Pombero* | Danito | 729 |
| Kalaíto Vore (06).doc | Written narrative: Chapter 6 of *Kalaíto Pombero* | Danito | 668 |
| Kalaíto Vore (07).doc | Written narrative: Chapter 7 of *Kalaíto Pombero* | chrenter | 1057 |
| Kalaíto Vore (08).doc | Written narrative: Chapter 8 of *Kalaíto Pombero* | chrenter | 962 |
| Kalaíto Vore (09).doc | Written narrative: Chapter 9 of *Kalaíto Pombero* | Danito | 1070 |
| Kalaíto Vore (10).doc | Written narrative: Chapter 10 of *Kalaíto Pombero* | Danito | 1258 |
| Kalaíto Vore (11).doc | Written narrative: Chapter 11 of *Kalaíto Pombero* | Danito | 1207 |
| Kalaíto Vore (12).doc | Written narrative: Chapter 12 of *Kalaíto Pombero* | Danito | 1397 |
| Kalaíto Vore (13).doc | Written narrative: Chapter 13 of *Kalaíto Pombero* | Danito | 1275 |
| Kalaíto Vore (14).doc | Written narrative: Chapter 14 of *Kalaíto Pombero* | Danito | 910 |
| Kalaíto Vore (15).doc | Written narrative: Chapter 15 of *Kalaíto Pombero* | Danito | 1212 |
| Kalaíto Vore (16).doc | Written narrative: Chapter 16 of *Kalaíto Pombero* | Danito | 1386 |
| Kalaíto Vore (17).doc | Written narrative: Chapter 17 of *Kalaíto Pombero* | Danito | 1172 |
| Kalaíto Vore (18).doc | Written narrative: Chapter 18 of *Kalaíto Pombero* | Danito | 1544 |
| grn-para-43-00985 English_ÑOMONGETA ANGEL VERA NDIVE.doc | Spoken dialogue | algez32 | 605 |
| grn-para-43-00609 English_ÑE'ËASÄI B1.doc | Radio broadcast | algez32 | 783 |
| grn-para-43-00611 English_ÑE'ËASÄI B2.doc | Radio broadcast | algez32 | 232 |
| grn-para-43-00612 English_ÑE'ËASÄI B3.doc | Radio broadcast | algez32 | 355 |
| grn-para-43-00614 English_ÑE'ËASÄI B5.doc | Radio broadcast | algez32 | 470 |
| grn-para-43-00997 English_PEHE PETEIHA.doc | Radio broadcast | algez32 | 213 |

Table B continued

| File Name | Content Type | Acquirer | Word Count |
|---|---|---|---|
| grn-para-43-00995 English_PEHË MOKÖIHA.doc | Radio broadcast | algez32 | 419 |
| grn-para-43-00994 English_PEHE MBOHAPYHA.doc | Radio broadcast | algez32 | 189 |
| grn-para-43-00993 English_PEHË IRUNDYHA.doc | Radio broadcast | algez32 | 567 |
| grn-para-43-00998 English_PEHË POHA.doc | Radio broadcast | algez32 | 568 |
| grn-para-43-00989 English_ÑOMONGUETA LUI NDIVE.doc | Radio broadcast | algez32 | 632 |
| grn-para-43-00165 English_IT HAPPENED IN A GUARANI AIRPORT | Radio broadcast | algez32 | 261 |
| ABC REMIANDU 31-03-06 | Written news | Danito | 98 |
| ABC REMIANDU 29-11-06.doc | Written news | Danito | 79 |
| ABC REMIANDU 28-11-06.doc | Written news | Danito | 79 |
| ABC REMIANDU 28-07-06.doc | Written news | Danito | 73 |
| ABC REMIANDU 27-07-06.doc | Written news | Danito | 80 |
| ABC REMIANDU 26-05-06.doc | Written news | renbaca | 93 |
| ABC REMIANDU 26 -07-06.doc | Written news | Danito | 87 |
| ABC REMIANDU 25-11-06.doc | Written news | chrenter | 74 |
| ABC REMIANDU 25-07-06.doc | Written news | Danito | 96 |
| ABC REMIANDU 24-07-06.doc | Written news | Danito | 78 |
| ABC REMIANDU 23-11-06.doc | Written news | Danito | 77 |
| ABC REMIANDU 23-05-06.doc | Written news | renbaca | 100 |
| ABC REMIANDU 22-11-06.doc | Written news | Danito | 90 |
| ABC REMIANDU 22-06-06.doc | Written news | Danito | 94 |
| ABC REMIANDU 21-11-06.doc | Written news | Danito | 77 |

Table B continued

| File Name | Content Type | Acquirer | Word Count |
|---|---|---|---|
| ABC REMIANDU 21-06-06.doc | Written news | Danito | 99 |
| ABC REMIANDU 20-11-06.doc | Written news | Danito | 68 |
| ABC REMIANDU 20-06-06.doc | Written news | Danito | 89 |
| ABC REMIANDU 19-06-06.doc | Written news | Danito | 81 |
| ABC REMIANDU 17-11-06.doc | Written news | Danito | 92 |
| ABC REMIANDU 16-11-06.doc | Written news | Danito | 111 |
| ABC REMIANDU 15-11-06.doc | Written news | Danito | 79 |
| ABC REMIANDU 15-06-06.doc | Written news | Danito | 107 |
| ABC REMIANDU 15-05-06.doc | Written news | Danito | 97 |
| ABC REMIANDU 14-11-06.doc | Written news | Danito | 76 |
| ABC REMIANDU 14-05-06.doc | Written news | Danito | 101 |
| ABC REMIANDU 13-11-06.doc | Written news | Danito | 93 |
| ABC REMIANDU 13-05-06.doc | Written news | Danito | 79 |
| ABC REMIANDU 12-05-06 | Written news | Danito | 91 |
| ABC REMIANDU 11-12-06.doc | Written news | Danito | 92 |
| ABC REMIANDU 11-11-06.doc | Written news | Danito | 100 |
| ABC REMIANDU 10-11-06.doc | Written news | Danito | 83 |
| ABC REMIANDU 09-11-06.doc | Written news | Danito | 85 |
| ABC REMIANDU 09-05-06 | Written news | Danito | 100 |
| ABC REMIANDU 08-11-06.doc | Written news | Danito | 72 |
| ABC REMIANDU 08-06-06.doc | Written news | Danito | 95 |
| ABC REMIANDU 07-11-06.doc | Written news | Danito | 87 |
| ABC REMIANDU 06-11-06.doc | Written news | Danito | 102 |
| ABC REMIANDU 04-11-06.doc | Written news | Danito | 99 |

Table B continued

| File Name | Content Type | Acquirer | Word Count |
|---|---|---|---|
| ABC REMIANDU 03-10-06.doc | Written news | Danito | 84 |
| ABC REMIANDU 02-10-06.doc | Written news | Danito | 83 |
| ABC REMIANDU 01-10-06.doc | Written news | Danito | 92 |
| ABC REMIANDU 16-06-06.doc | Written news | Danito | 91 |
| ABC REMIANDU 16-05-06.doc | Written news | Danito | 95 |
| ABC REMIANDU 13-06-06.doc | Written news | Danito | 83 |
| ABC REMIANDU 06-06-06.doc | Written news | renbaca | 97 |
| ABC REMIANDU 05-06-06.doc | Written news | Danito | 91 |
| ABC REMIANDU 02-06-06.doc | Written news | Danito | 99 |
| | | | 48068 |

# LIST OF REFERENCES

Adelaar, W. F. H. (1994). The nasal/oral distinction in Paraguayan Guaraní suffixes. *Revista Latinoamericana de Estudios Etnolingüísticos, 8*, 125–133.

Aissen, J. (1992). Topic and focus in Mayan. *Language, 68*(1), 43–80.

Aissen, J. (2003a). Differential coding, partial blocking, and bidirectional OT. *BLS, 29*, 1–16.

Aissen, J. (2003b). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory, 21*(3), 435–483.

Amberber, M. & de Hoop, H. (2005). Competition and variation in natural languages: The case for case. In M. Amberer & H. de Hoop (Eds.), *Competition and variation in natural languages: The case for case*. Amsterdam: Elsevier.

Asher, R. E. & Kumari, T. C. (1997). *Malayalam*. New York: Routledge.

Baker, C. L. (1966). Definiteness and indefiniteness in English (MA thesis, University of Illinois, 1966).

Baker, M. C. (1996). *The polysynthesis parameter*. New York: Oxford University Press.

Barðdal, J. (2001). Case in Icelandic — a synchronic, diachronic, and comparative approach. In *Lundastudier i nordisk språkvetenskap 57*. Lund University: Department of Scandinavian Languages.

Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B., & Ungerleider, L. G. (2009). Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *Journal of Neurophysiology, 101*(2), 688–700.

Bentivoglio, P. (1983). Topic continuity and discontinuity in discourse. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 255–312). Amsterdam: John Benjamins Publishing Company.

Bleam, T. (1999). Leísta Spanish and the syntax of clitic doubling (PhD thesis, University of Delaware, 1999).

Bolinger, D. (1952). Linear modification. *PMLA, 67*, 1117–1144.

Bossong, G. (1983–1984). Animacy and markedness in universal grammar. *Glossologia, 39*(2-3), 7–20.

Bossong, G. (1985a). *Empirische universalienforschung: Differentielle objektmarkierung in der neuiranischen sprachen*. Tübingen: Narr.

Bossong, G. (1985b). Markierung von aktantenfunktionen im Guaraní. In F. Plank (Ed.), *Relational typology* (pp. 1–29). Berlin: Mouton de Gruyter.

Bossong, G. (1991). Differential object marking in Romance and beyond. In D. Wanner & D. Kibbee (Eds.), *New analyses in Romance linguistics: Selected papers from the XVIII Linguistic Symposium on Romance Languages, Urbana-Champaign, April 7-9, 1988* (pp. 143–170). Amsterdam: John Benjamins Publishing Company.

Bossong, G. (1998). Le marquage différentiel de l'objet dans les langues d'Europe. In J. Feuillet (Ed.), *Actance et valence* (pp. 193–258). Berlin: Mouton de Gruyter.

Bossong, G. (n.d.). *The typology of Tupi-Guarani as reflected in the grammars of Jesuit missionaries: Anchieta (1595), Aragona (ca. 1625), Montoya (1640), and Restivo (1729).* Unpublished manuscript.

Bouma, G. J. (2008). Starting a sentence in Dutch: A corpus study of subject- and object-fronting (PhD thesis, Rijksuniversiteit Groningen, 2008).

Bower, K. M. (2003). When to use Fisher's exact test. *American Society for Quality, Six Sigma Forum Magazine, 2*(4), 35–37.

Bresnan, J. & Mchombo, S. A. (1987). Topic, pronoun, and agreement in Chichewa. *Language, 63*, 741–782.

Brody, J. (1984). Some problems with the concept of basic word order. *Linguistics, 22*, 711–736.

Büring, D. (1999). Topic. In P. Bosch & R. van der Sandt (Eds.), *Focus: Linguistic, cognitive, and computational perspectives* (pp. 142–165). Cambridge: Cambridge University Press.

Butt, M., Ahmed, T., & Poudel, T. (2008). Development of case in South Asian languages. Presented at the worksop "Sprachwandelvergleich," Bamberg.

Carlson, G. N. (1977). Reference to kinds in English (PhD thesis, University of Massachusetts, Amherst, 1977).

Carnie, A. (2006). Some remarks on markedness hierarchies: A reply to Aissen 1999 and 2003. In D. Siddiqi & B. Tucker (Eds.), *Coyote papers: Working papers in linguistics* (pp. 37–50). Tuscon: University of Arizona.

Cartier, A. (1989). Constructions en indonésien du type 'il casse la branche/la branche casse'. *Actances, 4*, 175–197.

Chesterman, A. (1991). *On definiteness: A study with special reference to English and Finnish.* Cambridge: Cambridge University Press.

Choi, H.-W. (1999). *Optimizing structure in context: Scrambling and information structure.* Stanford: CSLI Publications.

Choi, J. K. (2000). [–person] direct object drop: The genetic cause of a syntactic feature in Paraguayan Spanish. *Hispania, 83*, 531–543.

Choi, J. K. (2003). Language attitudes and the future of bilingualism: The case of Paraguay. *International Journal of Bilingual Education and Bilingualism, 6*(2), 81–94.

Christopherson, P. (1939). *The articles: A study of their theory and use in English.* Oxford: Oxford University Press.

Comrie, B. (1979). Definite and animate direct objects: A natural class. *Linguistica silesiana, 3,* 13–21.

Comrie, B. (1986). Markedness, grammar, people, and the world. In F. Eckman, E. Moravczik, & J. Wirth (Eds.), *Markedness* (pp. 85–106). New York: Plenum Press.

Comrie, B. (1989). *Language universals and linguistic typology* (2nd ed.). Chicago: University of Chicago Press.

Croft, W. (1988). Agreement vs. case-marking and direct objects. In M. Barlow & C. Ferguson (Eds.), *Agreement in natural language: Approaches, theories, descriptions* (pp. 159–179). Stanford: CSLI Publications.

Dahl, Ö. (1976). What is new information?: Approaches to word order. In N. E. Enkvist & V. Kohonen (Eds.), *Reports on text linguistics* (pp. 37–51). Åbo: Akademia.

Dalrymple, M. & Nikolaeva, I. (2005). Nonsubject agreement and discourse roles. In A. McNay (Ed.), *Oxford working papers in lingusitics, philology, and phonetics* (pp. 71–92). Oxford: The Centre for Linguistics and Philology.

Danon, G. (2001). Syntactic definiteness in the grammar of Modern Hebrew. *Linguistics, 39*(6), 1071–1116.

Daya, S. (2002). Fisher exact test. *Evidence-based Obstetrics and Gynaecology, 4*(1), 3–4.

de Anchieta, J. (1595). *Arte de grammatica da lingoa mais vsada na costa do Brasil.* Coimbra: Antonio de Mariz.

de Aragona, A. (ca. 1625). *Breve introducción para aprender la lengua guaraní.* New edition edited by Bartomeu Meliá and published in *Amerindia (2003), 4,* 23-61.

de Granda, G. (1996). Interferencia y convergencia sintáctica e isogramatismo amplio en el español paraguayo. *International Journal of the Sociology of Language, 117,* 63–80.

de Hoop, H. & Malkuchov, A. L. (2008). Case-marking strategies. *Linguistic Inquiry, 39*(4), 565–587.

de Montoya, A. R. (1639). *Tesoro de la lengva gvarani.* Reprinted in 1878 by Julio Platzman, Liepzig.

de Montoya, A. R. (1640a). *Arte, y bocabulario de la lengva gvarani.* Reprinted in 1878 by Julio Platzman, Liepzig.

de Montoya, A. R. (1640b). *Catecismo de la lengva gvarani.* Reprinted in 1878 by Julio Platzman, Liepzig.

de Swart, P. (2007). Cross-linguistic variation in object marking (PhD thesis, Radboud Universiteit Nijmegen, 2007).

de Vogelaer, G. (2007). Extending Hawkins' comparative typology: Case, word order, and verb agreement in Germanic languages. *University of Tromsø Working Papers on Language and Linguistics, 34*(1).

Deen, K. U. (2006). Subject agreement in Nairobi Swahili. In J. M. et al. (Ed.), *Selected proceedings of the 35th annual conference on African linguistics* (pp. 225–233). Somerville: Cascadilla Proceedings Project.

Diesing, M. (1992). *Indefinites*. Cambridge MA: MIT Press.

Diesing, M. (1997). Yiddish VP order and the typology of object movement in Germanic. *Natural Language and Linguistic Theory, 15*, 369–4273.

Dietrich, W. (1993). Mbyá, guaraní, criollo y castellano: El contacto de las tres lenguas estudiado en un grupo mbyá en Misiones. *Signo & Seña, 3*, 56–71.

Dietrich, W. (1995). El español del Paraguay en contacto con el guaraní. In K. Zimmerman (Ed.), *Lenguas en contacto en Hispanoamérica* (pp. 203–216). Frankfurt: Vervuert.

Dixon, R. M. W. (1979). Ergativity. *Language, 55*(1), 59–138.

Dobrovie-Sorin, C. (1997). Classes de prédicats, distribution des indéfinis et la distinction thétique-catégorique. *Le gré des langues, 20*, 58–97.

Donellan, K. S. (1966). Reference and definite descriptions. *Philosophical Review, 75*(3), 281–304.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*(3), 547–619.

Enç, M. (1991). The semantics of specificity. *Linguistic Inquiry, 22*, 1–26.

Enrico, J. (1986). Word order, focus, and topic in Haida. *International Journal of American Linguistics, 52*(2), 91–123.

Epps, P. (2008). Hup's typological treasures: Description and explanation in the study of an Amazon language. *Linguistic Typology, 12*, 169–193.

Escandell-Vidal, V. (2008). Topics from Ibiza: Differential object marking and clitic-dislocation. In G. A. Kaiser & M. Leonetti (Eds.), *Proceedings of the workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"* (pp. 23–43). Universität Konstanz.

Farkas, D. F. (1994). Specificity and scope. In L. Nash & G. Tsoulas (Eds.), *Langues et grammaires 1* (pp. 119–137). Paris.

Farkas, D. F. (2002). Specificity distinctions. *Journal of Semantics, 19*(3), 213–243.

Feldman, H. (1986). *A grammar of Awtuw*. Canberra: Research School of Pacific Studies, The Australian National University.

Finer, D. (1997). Contrasting A′-dependencies in Selayarese. *Linguistic Inquiry, 25*, 677–728.

Firbas, J. (1966). On defining the theme in functional sentence analysis. In J. Vachek (Ed.), *Traveaux linguistiques de Pragues* (Vol. 1, pp. 267–280). Tuscaloosa: University of Alabama Press.

Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society, 98*(1), 39–82.

Fodor, J. D. (1970). The linguistic description of opaque contexts (PhD thesis, MIT, 1970).

Fodor, J. D. & Sag, I. (1982). Referential and quantificational indefinites. *Language and Philosophy, 5*(3), 355–398.

Fraurud, K. (1990). Definiteness and the processing of NPs in natural discourse. *Journal of Semantics, 7,* 395–433.

García García, M. (2005). Differential object marking and informativeness. In K. von Heusinger, G. A. Kaiser, & E. Stark (Eds.), *Proceedings from the workshop "Specificity and the Evolution/Emergence of Nominal Determination in Romance"* (pp. 17–31). Fachbereich Sprachwissenschaft Universität Konstanz.

García Landa, J. Á. (2009). *A bibliography of literary theory, criticism, and philology.* Electronic resource: http://www.unizar.es/departamentos/filologia_inglesa/garciala/bibliography.html.

Geach, P. (1962). *Reference and generality.* Ithaca: Cornell University Press.

Gerner, M. (2008). Ambiguity-driven differential object marking in Yongren Lolo. *Lingua, 118,* 296–331.

Giménez Caballero, E. (1966). El yopará in Paraguay. In *Congreso de academias de la lengua española 4. Actas y labores (Buenos Aires, de 30 de nov. al 10 de dic. de 1964)* (pp. 124–136).

Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. Li (Ed.), *Subject and topic.* New York: Academic Press.

Givón, T. (1978). Definiteness and referentiality. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravscik (Eds.), *Universals of human language: Syntax* (Vol. 4, pp. 291–330). Stanford: Stanford University Press.

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1–41). 1983: John Benjamins Publishing Company.

Givón, T. (1990). *Syntax: A functional typological introduction.* Amsterdam: John Benjamins Publishing Company.

Gordon, R. G., Jr. (Ed.). (2005). *Ethnologue: Languages of the world (online version)* (15th ed.). Dallas: SIL International.

Gregores, E. & Suárez, J. A. (1961). A description of colloquial Guaraní (PhD thesis, Cornell University, 1961).

Gundel, J. K. (1973). The role of topic and comment in linguistic theory (PhD thesis, University of Texas, 1973).

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*(2), 274–307.

Guntsetseg, D. (2008). Differential object marking in Mongolian. In F. Schäfer (Ed.), *Working papers of the SFB 732 Incremental Specification in Context* (pp. 53–69). Universität Konstanz.

Halliday, M. A. K. (1967a). Notes on transitivity and theme in English: Part 1. *Journal of Linguistics, 3,* 37–82.

Halliday, M. A. K. (1967b). Notes on transitivity and theme in English: Part 2. *Journal of Linguistics, 3,* 199–244.

Halliday, M. A. K. (1968). Notes on transitivity and theme in English: Part 3. *Journal of Linguistics, 4,* 179–215.

Haspelmath, M. (2001). Non-canonical marking of core arguments in European languages. In A. Aikhenvald, R. Dixon, & M. Onishi (Eds.), *Non-canonical marking of subjects and objects* (pp. 53–85). Amsterdam: John Benjamins Publishing Company.

Haspelmath, M. (2006a). Against markedness (and what to replace it with). *Journal of Linguistics, 42,* 25–70.

Haspelmath, M. (2006b). *Terminology of case,* Oxford.

Hawkins, J. A. (1986). *A comparative typology of English and German: Unifying the constrasts.* London: Croom Helm.

Heim, I. (1982). The semantics of definite and indefinite noun phrases (PhD thesis, University of Massechusetts, Amherst, 1982).

Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In R. Bäuerle, C. Schwartze, & A. von Stechow (Eds.), *Meaning, use and interpretation of language* (pp. 164–189). Berlin: Mouton de Gruyter.

Hintikka, J. (1973). Grammar and logic: Some borderline problems. In J. Hintikka, P. Suppes, & J. Moravcsik (Eds.), *Approaches to natural languages* (pp. 197–214). Dordrecht: Reidel.

Hopper, P. J. & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56*(2), 251–299.

Huonder, A. (1912). Antonio Ruiz de Montoya. In *The Catholic encyclopedia.* New York: Robert Appleton Company.

Ioup, G. (1977). Specificity and the interpretation of quantifiers. *Linguistics and Philosophy, 1*(2), 233–245.

Jackendoff, R. (1990). *Semantic structures.* Cambridge, Massechusets: The MIT Press.

Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language, 83*(1), 74–109.

Jensen, C. (1998). Comparative study: Tupi-Guaraní. In D. C. Derbyshire & G. K. Pullum (Eds.), *The handbook of Amazonian languages* (Vol. 4, pp. 487–618). Berlin: Mouton de Gruyter.

Kadmon, N. (1990). Uniqueness. *Linguistics and Philosophy, 13,* 273–324.

Kamp, H. (1981). The theory of truth and semantic interpretation. In J. Groenendgijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277–322). Amsterdam: Amsterdam Center.

Kartunnen, L. (1976). Discourse referents. In J. McCawley (Ed.), *Syntax and semantics 7: Notes from the linguistic underground*. New York: Academic Press.

King, T. H. (1995). *Configuring topic and focus in Russian*. Stanford: CSLI Publications.

Kiss, K. E. (2002). *The syntax of Hungarian*. Cambridge: Cambridge University Press.

Klamer, M. & Kratochvíl, F. (2006). The role of animacy in Teiwa and Abui (Papuan). *BLS*, *32*.

Klein, U. (2007). *Clitic doubling and differential object marking in Romanian*. Presented at the workshop "Case marking in Bantu and Romance," University of Stuttgart.

Kliffer, M. (1982). Personal *a*, kinesis and individuation. In P. Baldi (Ed.), *Papers from the XII Linguistic Symposium on Romance Languages* (pp. 195–216). Amsterdam: John Benjamins Publishing Company.

Korn, A. (2009). Differential object marking in Balochi and beyond. Presented at the International SKY-Symposium "Case in and across languages," Helsinki.

Kratzer, A. (1998). Scope or pseudo-scope. In S. Rothstein (Ed.), *Events in grammar* (pp. 163–196). Dordrecht: Kluwer.

Kuno, S. (1973). *The structure of the Japanese language*. Cambridge: MIT Press.

Laca, B. (1995). Sobre el uso del acusativo preposicional en español. In C. Pensado (Ed.), *El complemento directo preposicional* (pp. 61–91). Madrid: Visor Libros.

Lazard, G. (1982). Le morphème ră en persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris*, *77*(1), 177–208.

Lazard, G. (1998). *Actancy*. Berlin: Mouton de Gruyter.

Lazard, G. (2002). Transitivity revisited as an example of a more strict approach in typological research. *Folia Linguistica*, *36*(3-4), 141–190.

Lazard, G. (2003). What is an object in a cross-linguistic perspective? In G. Fiorentino (Ed.), *Romance objects: Transitivity in Romance languages* (pp. 1–17). Berlin: Mouton de Gruyter.

Leonetti, M. (2004). Specificity and differential object marking in Spanish. *Catalan Journal of Linguistics*, *3*, 75–114.

Lewis, D. (1979). Score-keeping in a language game. In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view*. Berlin: Springer.

Li, C. (Ed.). (1976). *Subject and topic*. New York: Academic Press.

Lidz, J. (2006). The grammar of accusative case in Kannada. *Language*, *82*(1), 10–32.

Lima, A. (2003). Definiteness and differential object behavior (BA thesis, The University of Queensland, 2003).

Lima, A. (2006). A minimalist view on differential object marking for specificity. In *The 36<sup>th</sup> Annual Meeting of the Michigan Linguistics Society*. Oakland University. Rochester.

Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.

Magier, D. (1987). The transitivity prototype: Evidence from Hindi. *Word, 38*(3), 187–199.

Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua, 118*, 203–221.

Mallinson, G. & Blake, B. J. (1981). *Language typology*. Amsterdam: North-Holland Publishing Company.

Mardale, A. (2008). Microvariation within differential object marking: Data from Romance. *Revue Roumaine de Linguistique (RRL), 53*(4), 449–467.

Matthewson, L. (1999). On the interpretation of wide-scope indefinites. *Natural Language Semantics, 7*, 79–134.

McKoon, G. & Ratcliffe, R. (2007, 2). Interactions of meaning and syntax: Implications for models of sentence comprehension. *Journal of Memory and Language, 56*, 270–290.

McNally, L. (1995). Bare plurals in Spanish are interpreted as properties. In G. Morrill & R. Oehrle (Eds.), *Formal grammar* (pp. 197–222). Universitat Polytècnica de Catalunya.

Mithun, M. (1984). The evolution of noun incorporation. *Language, 60*(4), 847–894.

Mithun, M. (1991). Active/Agentive case marking and its motivations. *Language, 67*(3), 510–546.

Mohanan, T. (1990). Arguments in Hindi (PhD thesis, Stanford University, 1990).

Moravscik, E. A. (1978). On the case marking of objects. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravscik (Eds.), *Universals of human language: Syntax* (Vol. 4, pp. 249–289). Stanford: Stanford University Press.

Morimoto, Y. (2002). Prominence mismatches and differential object marking in Bantu. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG02 Conference, National Technical University of Athens, Athens* (pp. 292–314). Stanford: CSLI Publications.

Motulsky, H. (1995). *Intuitive biostatistics*. Oxford: Oxford University Press.

Myhill, J. (1992). *Typological discourse analysis: Quantitative approaches to the study of linguistic function*. Oxford: Blackwell.

Næss, Å. (2004). What markedness marks: The markedness problem with direct objects. *Lingua, 114*, 1186–1212.

Næss, Å. (2007). *Prototypical transitivity*. Amsterdam: John Benjamins Publishing Company.

Nazareth, A. K. (2007). Differential object marking and topicality in Tigrinya. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG07 Conference* (pp. 5–25). Stanford: CSLI Publications.

Neuberger, K. A. (2009). Why differential case marking in Corsican? Presented at the International SKY-Symposium "Case in and across languages," Helsinki.

Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, *62*(1), 56–119.

Nordhoff, S. (2004). Stative verbs and possession in Guaraní. Presented at the "Syntax of the Worlds Languages" Conference, Leipzig.

Nuckols, M. E. (2007). Case variation in Czech and Russian: Implications for the Transitivity Hypothesis (PhD thesis, The Ohio State University, 2007).

Partee, B. (1972). Opacity, coreference and pronouns. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 415–441). Dordrecht: Reidel.

Payne, D. (1994). The tupí-guaraní inverse. In B. Fox & P. J. Hopper (Eds.), *Voice: Form and function* (pp. 313–340). Amsterdam: John Benjamins Publishing Company.

Portner, P. & Yabushita, K. (2001). Specific indefinites and the information structure theory of topics. *Journal of Semantics*, *18*, 271–297.

Prince, E. F. (1979). On the given/new distinction. *Chicago Linguistic Society*, *15*, 267–278.

Prince, E. F. (1981). Toward a taxonomy of given/new information. In *Radical pragmatics* (pp. 223–255). New York: Academic Press.

Quinn, C. (2001). A preliminary survey of animacy categories in Penobscot. In J. D. Nichols (Ed.), *Papers of the 32nd Algonquin Conference* (pp. 395–426). Winnipeg: University of Manitoba.

Reinhart, T. (1997). Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy*, *20*, 335–397.

Roberts, C. (2003). Uniqueness in definite noun phrases. *Linguistics and Philosophy*, *26*(3), 287–350.

Rodriguez-Mondoñedo, M. (2007). The syntax of objects: Agree and differential object marking (PhD thesis, University of Connecticut, 2007).

Rosenbach, A. (2002). *Genitive variation in English*. Berlin: Mouton de Gruyter.

Rubin, J. (1985). The special relation of Guarani and Spanish in Paraguay. In N. Wolfson & J. Manes (Eds.), *Language of inequality* (pp. 111–120). Berlin: Mouton de Gruyter.

Russell, B. (1905). On denoting. *Mind*, *14*, 479–493.

Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Company.

Scheibman, J. (2002). *Point of view and grammar*. Amsterdam: John Benjamins Publishing Company.

Schwenter, S. A. (2006). Null objects across South America. In T. L. Face & C. A. Klee (Eds.), *Selected proceedings of the 8th Hispanic Linguistics Symposium* (pp. 23–36). Somerville: Cascadilla Proceedings Project.

Schwenter, S. A. & Silva, G. (2002). Overt vs. null direct objects in spoken Brazilian Portuguese. *Hispania, 85*(3), 577–586.

Schwenter, S. A. & Silva, G. (2003). Anaphoric direct objects in spoken Brazilian Portuguese: Semantics and pragmatics. *Revista Internacional de Linguistica Iberoamericana, 1*(2), 99–123.

Scott, G. (1978). *The Fore language of Papua New Guinea.* Canberra: Research School of Pacific Studies, The Australian National University.

Shain, C. (2008). Differential object marking in Paraguayan Guaraní (BA thesis, The Ohio State University, 2008).

Shain, R. (2008). The preverb *eis-* and Koine Greek aktionsart (MA thesis, The Ohio State Univeristy, 2008).

Shokouhi, H. & Kipka, P. (2003). A discourse study of Persian *râ. Lingua, 113,* 953–966.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages.* New Jersey: Humanities Press Inc.

Stockwell, R., Schachter, P., & Partee, B. (1973). *The major syntactic structures of English.* New York: Holt, Rinehart & Winston.

Strawson, P. (1950). On referring. *Mind, 59*(235), 320–344.

Tagliamonte, S. A. (2006). *Analysing sociolinguistic variation.* Cambridge: Cambridge University Press.

Tallerman, M. (2005). *Understanding syntax* (2nd ed.). London: Hodder Arnold.

Tippets, I. & Schwenter, S. (2007). *Relative animacy and differential object marking in Spanish.* Presented at NWAV 36, Philadelphia.

Tonhauser, J. (2006). The temporal semantics of noun phrases: Evidence from Guaraní (PhD thesis, Stanford University, 2006).

Tonhauser, J. & Colijn, E. (to appear). Word order in Paraguayan Guaraní. *International Journal of American Linguistics.*

Torrego Salcedo, E. (1999). El complement directo preposicional. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 2, pp. 1780–1805). Madrid: Espasa-Calpe.

Tsunoda, T. (1985). Remarks on transitivity. *Journal of Linguistics, 21*(2), 385–386.

Usher de Herreros, B. (1976). *Castellano-paraguayo. Notas para una gramática contrastiva castellano-guaraní.* Asunción: Centro de Estudios Antropológicos.

van Bergen, G. (2006). To *ba* or not to *ba* (MA thesis, Radboud University Nijmegen, 2006).

van Everbroeck, E. (2003). Language type frequency and learnability from a connectionist perpective. *Linguistic Typology, 7,* 1–50.

van Geenhoven, V. & McNally, L. (1998). *Redefining the weak/strong distinction*. Presented at the 1997 Paris Syntax and Semantics Colloquim.

van Geenhoven, V. & McNally, L. (2005). On the property analysis of opaque complements. *Lingua, 115,* 885–914.

van Geenhoven, V. (1998a). On the argument structure of some noun incorporating verbs in West Greenlandic. In M. Butt & W. Geuder (Eds.), *Lexical and compositional factors* (pp. 225–253). Stanford: CSLI Publications.

van Geenhoven, V. (1998b). *Semantic incorporation and indefinite descriptions*. Stanford: CSLI Publications.

van Valin, R. D. & Foley, A. (1985). Information packaging in the clause. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 1, pp. 299–347). Cambridge: Cambridge University Press.

Velázquez-Castillo, M. (1995). Noun incorporation and object placement in discourse: The case of Guaraní. In P. Downing & M. Noonan (Eds.), *Word order in discourse*. Amsterdam: John Benjamins Publishing Company.

Velázquez-Castillo, M. (1996). *The grammar of possession: Inalienability, incorporation, and possessor ascension in Guaraní*. Studies in Language Companion Series. Amsterdam: John Benjamins Publishing Company.

Velázquez-Castillo, M. (2002). Grammatical relations in active systems: The case of Guaraní. *Functions of Language, 9*(2), 133–167.

Velázquez-Castillo, M. (2004). Guaraní (Tupí-Guaraní). In G. E. Booij, C. Lehmann, J. Mugdan, & S. Skopeteas (Eds.), *Morphologie/Morphology: Ein internationales handbuch zur flexion und wortbildung/an international handbook on inflection and word-formation* (Vol. 2, pp. 1421–1432). Berlin: Mouton de Gruyter.

Vitale, A. J. (1981). *Swahili syntax*. Dordrecht: Foris Publications.

von Heusinger, K. (2003). Cross-linguistic implementations of specificity. In M. Jaszczolt & K. Turner (Eds.), *Meaning through language contrasts* (pp. 405–421). Amsterdam: John Benjamins Publishing Company.

von Heusinger, K. & Kaiser, G. A. (2003). The interaction of animacy, specificity, and definiteness in Spanish. In K. von Heusinger, G. A. Kaiser, & E. Stark (Eds.), *Proceedings of the workshop "Semantic and Syntactic Aspects of Specificity in Romance Languages"* (pp. 41–65). Universität Konstanz. Konstanz.

von Heusinger, K. & Kaiser, G. A. (2005). The evolution of differential object marking in Spanish. In K. von Heusinger, G. A. Kaiser, & E. Stark (Eds.), *Proceedings of the workshop "Specificity And The Evolution / Emergence of Nominal Determination Systems in Romance"* (pp. 33–70). Universität Konstanz. Konstanz.

von Heusinger, K. & Kaiser, G. A. (2007). Differential object marking and the lexical semantics of verbs in Spanish. In G. A. Kaiser & M. Leonetti (Eds.), *Proceedings of the workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages"* (pp. 85–110). Fachbereich Sprachwissenschaft, Universität Konstanz.

von Heusinger, K., Klein, U., & de Swart, P. (2008). *Variation in differential object marking*. Presented at the workshop "Case Variation," University of Stuttgart.

Weissenrieder, M. (1990). Variable uses of the direct object marker *a. Hispania, 73*(1), 223–231.

Yang, N. & van Bergen, G. (2007). Scrambled objects and case marking in Mandarin Chinese. *Lingua, 117*(9), 1617–1635.

Zwicky, A. (1992). Jottings on adpositions, case inflections, government and agreement. In D. Brentari, G. Larson, & L. MacLeod (Eds.), *The joy of grammar: A Festschrift for James D. McCawley* (pp. 369–383). Amsterdam: John Benjamins Publishing Company.

# LANGUAGE INDEX

# AUTHOR INDEX