# Differential Object Marking in Paraguayan Guaraní

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation *with research distinction* in Linguistics in the undergraduate colleges of The Ohio State University

by

Cory Shain

The Ohio State University
December 2008

Project Advisors:
Dr. Peter Culicover, Department of Linguistics
Dr. Judith Tonhauser, Department of Linguistics

## Abstract

Differential Object Marking (DOM), the appearance of a direct object marker on some but not all direct objects in a language, has been the focus of a substantial body of literature. This study analyzes DOM in Paraguayan Guaraní, a language that has not yet been discussed in the DOM literature. I show that object-marking in Guaraní is differential. I then apply two analyses of the distribution of DOM to a corpus of textual Guaraní data, testing the effectiveness of their predictions against the actual distribution of DOM in Guaraní and comparing the success of each analysis. The first analysis is that of Aissen (2003b), who proposes prominence as a condition on DOM, where the prominence of an object is how "subject-like" that object is in terms of animacy and definiteness. This prominence-based analysis predicts that highly prominent (subject-like) objects will be object-marked, while low-prominence objects will be unmarked. The second analysis is similar to that of Gerner (2008), who argues that in Yongren Lolo object-marking occurs in clauses in which ambiguity exists as to which NPs fill which grammatical functions. Under this analysis, DOM serves the purpose of disambiguating object from subject. I put both of these analyses to the test against the corpus. I find that both are supported, in that neither of their predictions fail for Guaraní. However, I argue that the ambiguity-based analysis is superior to the prominence-based analysis for Guaraní in terms of coverage, testability, and simplicity, and that an ambiguity-based analysis of DOM is therefore preferable in the case of Guaraní.

## Acknowledgments

I'm very grateful for the time and effort invested in me and my research by both my advisors, Judith Tonhauser and Peter Culicover. Judith helped me immensely throughout my running-start introduction to Guaraní and Guaraní studies, giving painstakingly detailed feedback, encouraging me to press on when I was ready to switch topics altogether, and providing the data without which this study would not have been possible.

Peter has been a readily available coach in the process of doing research and writing about it. His insights were invaluable to this project, quickly isolating strengths and weaknesses of argument and presentation. I'm indebted to his investment of time and concern, as well as his tolerance for rapid-fire email exchanges.

I'm sincerely thankful for Terrell Morgan and Scott Schwenter and their time, feedback, encouragement, and fresh ideas. I owe the completion of this paper to Julie McGory, who, via meeting after meeting, helped me navigate the process of Honors Thesis writing from the administrative side.

Finally, I'd have gotten nowhere without my wife Rachel, who tolerated the long nights of hibernation just as easily as she did the frequent incoherent verbal onslaughts she was subjected to as I tried to work through these ideas with her, all while writing a thesis of her own! I depended on her support.

# Abbreviations

| | |
|---|---|
| 1 | 1st person |
| 2 | 2nd person |
| 3 | 3rd person |
| A | Set A verbal prefix |
| B | Set B verbal prefix |
| CAUSE | Causative verbal inflection |
| COMPL | Completive aspect |
| COND | Conditional marker |
| DES | Desiderative modal marker |
| DIM | Diminutive suffix |
| DO | Direct object |
| EMPH | Emphatic marker |
| Excl | Exclusive |
| FUT | The future suffix *-ta* |
| GUI | The postposition *-gui* |
| IMP | Imperative marker |
| IMPERS | Impersonal agent marker |
| Incl | Inclusive |
| IO | Indirect object |
| JE | The impersonal, reflexive, or passivizing prefix *-je* |
| NEG | Negation marker |
| NOM | The nominalizing suffix *-ha* |
| NOM.FUT | The nominal future marker *rã* |
| NOM.PAST | The nominal past marker *kue* |
| Obj | Object |
| OPT | Optative mood |
| PE | The spatiotemporal postposition and argument-marker *-pe* |
| PL | Plural |
| PROG | Progressive aspect |
| PRO | Pronominal |
| PURP | The purposive marker *-gua* |
| QU | Question marker |
| RC | Relativizer |
| REHE | The postposition *-rehe* |
| REP | Reportative |
| REQ | Requestative |
| Sg | Singular |
| (Sp) | Spanish loan word |
| Subj | Subject |
| VAERA | The necessity modal *va'erã* |

# Contents

# 0 Introduction

DIFFERENTIAL OBJECT MARKING (DOM), or the conditional appearance of a direct object marker on some, but not all, direct objects in a language, is a well-attested cross-linguistic phenomenon (Bossong, 1983-1984, 1991; Aissen, 2003b, 2003a; Comrie, 1989; Heusinger & Kaiser, 2003, 2005; Leonetti, 2004; Lima, 2006). While DOM exists in Paraguayan Guaraní (Guaraní), there is very little discussion of it in the literature. The morpheme *-pe*, which has been more thoroughly described with respect to its function as a spatiotemporal postposition (see Gregores & Suárez, 1967; Velázquez-Castillo, 2004; Jensen, 1998), also serves to mark some objects in transitive clauses, but not all, as seen in the following examples.

(1) Ka'i o-mby-tavý-rõ-guare **aguará-pe**.
monkey A3-CAUSE-fool-COND-when.past **fox-PE**
'When the monkey tricked **the fox**.'

(2) O-monga-kuaa karai **pe mitã**.
A3-raise-know gentleman **that child**
'The gentleman raised **the child**.'

In (1), the direct object is *aguarápe*, in which a noun *aguara* — "fox" — is marked with the morpheme *-pe*. In (2), the direct object is *pe mitã* — "that child," which has no *-pe*-marking. One DO is marked, the other is not, demonstrating that object-marking is differential.

The crucial question pursued in this study is which objects take this marking and which don't. What are the conditioning factors behind DOM in Guaraní? The purpose of this paper is to test two analyses of DOM already presented in the differential object marking literature against a corpus of Guaraní data and compare the success of their predictions.

The first analysis I explore, which I refer to as PROMINENCE, is based on Aissen (2003b), who proposes a hierarchy of what she refers to as the "prominence" of noun phrases, based on degrees of definiteness and animacy, as a standard by which to determine the markedness of an NP with respect to its grammatical function. High-prominence NPs are subject-like, while low-prominence NPs are object-like. Thus prominent NPs are unmarked as subjects and marked as objects, and vice versa. Aissen (2003b) claims that this markedness translates into actual morphological marking on objects once prominence surpasses a certain language-specific threshold. In this study, I put Aissen's analysis to the test against a body of Guaraní data.

The second analysis I explore, which I refer to as AMBIGUITY, is based on Gerner (2008), who argues that Yongren Lolo marks objects in clauses which are otherwise ambiguous as to which NPs fill which grammatical functions. According to Aissen (2003b), there may be some languages for which ambiguity is the sole factor underlying DOM, but in many languages object-marking occurs in both ambiguous and unambiguous clauses. Thus, there are three possible relationships between ambiguity and DOM:

(3) a. ambiguity $\Longleftrightarrow$ object-marking
b. ambiguity $\longrightarrow$ object-marking
c. object-marking is unrelated to ambiguity

In this study, I put the strongest ambiguity-based analysis (one that predicts (3a)) to the test against a Guaraní corpus: evaluating whether or not object-marking appears always and only when there is ambiguity as to which NP is the object. I find that, while the existence of counterexamples to (3a) requires a weakening of the analysis to predicting (3b), there is nevertheless too strong of a correlation between ambiguity and object-marking to conclude the validity of (3c). I discuss patterns within the counterexamples to AMBIGUITY and hypothesize about which other factors might interact with ambiguity in conditioning DOM.

I then present my findings from both analyses and compare them, discussing the success of their respective predictions and exploring the overlap between data they each predict and fail to predict. Based on the results of this comparison, I argue that AMBIGUITY is preferable to PROMINENCE for Guaraní on the basis of coverage, testability, and simplicity.

The data used in this study comes from a textual corpus of naturally-occurring Guaraní data collected in Paraguay by my advisor, Judith Tonhauser. The corpus consists of nine texts, totaling 6463 words in length. Table 1 presents their respective lengths, as well as a brief description of the content of each. Corpus research is advantageous for this distributional study, since it enables evaluation the frequency and context of the appearance of particular constructions (distribution).

My presentation of the study is structured as follows. Section 1 focuses on the features of Guaraní syntax relevant to this study. Section 2 reviews the relevant literature on DOM generally and DOM in Guaraní specifically. Section 3 presents both alternative analyses, discusses the results from the data, and compares their effectiveness at explaining DOM. Section 4 concludes, and presents some suggestions for future work.

| 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 |
| *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 unkown) | 375 |
| *Michi** | A personal narrative about growing up, as told by MM | 283 |
| *Kiri Kiri** | A fable about the adventures of a cricket (author unkown) | 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 |

Table 1: Guaraní Texts Examined in This Study
*Title, word count, and authorship information from Tonhauser and Colijn (to appear).

# 1 Relevant Aspects of Guaraní Grammar

In this section, I provide a brief overview of the aspects of Paraguayan Guaraní relevant to this study. Guaraní is an official language of Paraguay and is spoken by over 4 million people, many of whom are bilingual in both Spanish and Guaraní. It is a Tupí language of the Tupí-Guaraní family; its genetic relationship to Tupí and Cariban languages is discussed more fully in Rodrigues (1985), Derbyshire (1994), and Jensen (1999).

The discussion in § 1.1 concerns grammatical function (GF) assignment in Guaraní grammar — assignment of grammatical functions to particular NPs in a clause. This is particularly relevant to the ambiguity-based analysis of § 3.2, under which it is necessary to evaluate whether NPs are unambiguously linked to grammatical functions. Section 1.2 briefly discusses noun phrase structure and the precedence relations between components of the Guaraní NP, which is relevant to the prominence-based analysis of § 3.1, where the structure of the object NP provides clues as to its level of definiteness. Section 1.3 presents a synopsis of the overall distribution of the morpheme *-pe*, the appearance of which on DOs is the main focus of this study. Section 1.4 argues that the string *chupe* — the 3rd person object pronoun — is not a composite of the morphemes *chu* and *-pe* but has instead been lexicalized. Section 1.5 addresses the question of the grammatical category of *-pe*: postposition or case-marker?

## 1.1 Grammatical Function Assignment in Guaraní

There are generally three methods by which grammars indicate the grammatical functions of the noun phrases in a clause: dependent-marking, fixed word order, and verbal head-agreement, or some combination of the three (see Nichols, 1986; Sapir, 1921; Baker, 1996; Hawkins, 1986; de Vogelaer, 2007; van Everbroeck, 2003; Mallinson & Blake, 1981; Amberber & de Hoop, 2005). Under dependent-marking, functions are assigned by morphemes which attach to the noun phrases themselves. For instance, take example (4) from Japanese, a dependent-marking language (data from Kuno (1973, p. 129)).

(4) Boku ga tomodati ni hana o ageta.
1Sg SUBJ friend DATIVE flowers OBJ gave
'I gave my friend flowers.'

In this clause, the NPs *boku* — "I," *tomodati* — "friend," and *hana* — "flowers," are each concatenated with morphemes demarcating which grammatical functions they fill. The subject function is indicated by *ga*, which attaches to *boku*, the object function is indicated by *o*, which attaches to *hana*, and the indirect object function is indicated by the dative marker *ni*, which attaches to *tomodati*. Thus the only GF-assignment arrangement licensed by dependent-marking in this sentence is "I gave my friend flowers."

Under word order, particular functions are assigned by default to particular linear positions in the sentence. This is the case in English, for example, where the sentence in (5) cannot be taken to mean that *Jadyn* is the object and *Sydney* is the subject, since the pre-verbal position is assigned to the subject function and the post-verbal position is assigned to the object function.

(5) Jadyn admires Sydney.

Under head-agreement, functions are assigned via agreement morphology on the head verb, such that when an argument appears that fits the agreement properties a grammatical function is established for that argument. This is the case in Swahili, as seen in the following example from Deen (2006, p. 231).

(6) Juma a-na-m-pend-a Mariam.
Juma 3Sg.Subj-PRES-3SgObj-like-INDICATIVE Mariam
'Juma likes Mariam.'

In this sentence, the verb *pend* — "like" — co-occurs with two 3rd person NPs: *Juma* and *Mariam*. The prefixes on the verb serve to index which kinds of NPs fill which grammatical roles: the prefix *a* indicates that the subject is 3rd person

singular, and the prefix *m* indicates that the object is 3rd person singular. In this particular instance the agreement properties of both prefixes are the same (3rd person singular), and therefore match both co-ocurring NPs. Thus word order comes into play: since Swahili is SVO, this sentence is parsed in such a way that *Juma* fills the subject role and *Mariam* fills the object role.

I now discuss these three methods of argument linking in the case of Guaraní specifically. With respect to dependent-marking, Guaraní is a differential object marking language (§ 1.3.3). Whether this marking is driven by a GF-assignment function or by something else is the primary issue being explored by this study. Head-agreement and word order are discussed below. Section 1.1.1 describes the system of verbal head-agreement in Guaraní syntax, and § 1.1.2 evaluates the role of word order in GF-assignment for Guaraní. Section 1.1.3 points out some additional ways in which the grammar of Guaraní indicates which NPs fill which grammatical functions, and § 1.1.4 presents a synopsis of GF-assignment in Guaraní grammar.

### 1.1.1 Verb Agreement

Guaraní is an active-stative language with verbal agreement morphology divided into two classes (Gregores & Suárez, 1967; Velázquez-Castillo, 2004; Tonhauser, 2006; Tonhauser & Colijn, to appear). According to these descriptions, Set A inflections agree in person (and some in number) with the subject of both active transitive and active intransitive predicates, as shown in (7), while Set B inflections agree with the subject of stative intransitive predicates as well as with the object of active transitive predicates, as shown in (8).

**Set A Marking:**

(7) a. **a**-ñe-mbo-jere
**A1.Sg**-REFL-put-twist
'**I** spun around'

b. Upé-va niko **a**-hendu.
that-RC truly **A1.Sg**-hear
'**I** hear that.'

**Set B Marking:**

(8) a. Nda **che**-py'aguapy-ve-í-ko.
NEG **B1.Sg**-peace-more-NEG-EMPH.
'**I** couldn't be calmer.'

b. Ha Felípa **che**-r-echa.
CONJ Felípa **B1.Sg**-REL-see
'And Felípa saw **me**.'

In each of these examples, sentence (a) is intransitive and sentence (b) is transitive. Notice that each verbal agreement marker is glossed as "A" or "B," shorthand for "Set A verbal prefix" or "Set B verbal prefix." Both sentences in (7) exhibit Set-A-marked verbs because they they agree with the subjects of active predicates: "spin around" (intransitive) in (7a) and "hear" (transitive) in (7b). Meanwhile, both sentences in (8) are Set-B-marked. In (8a), this is because the verb agrees with the subject of a stative predicate, *py'aguapyve* — "be calmer." In 8b, this is because the verb agrees with the object rather than the subject of the active predicate *hecha* — "see."

Henceforth I label the three main clausal arguments according to the shorthand established for ergative languages in Dixon (1979), where the subject of an intransitive clause is the S-argument, the subject of a transitive clause is the A-argument, and the object of a transitive clause is the O-argument. Based on this notation, 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 (9).

(9)

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

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

Agreement Markers in Guaraní
As presented in Tonhauser and Colijn (to appear)

Predicates in Guaraní are inflected to agree with one of their S, A, or O arguments, namely, the highest one on both the PERSON and grammatical FUNCTION hierarchies laid out in (10a) and (10b) from Tonhauser (2006, p. 131), assuming the ranking of hierarchies in (10c):

(10) a. **PERSON HIERARCHY:** 1st person > 2nd person > 3rd person
b. **FUNCTION HIERARCHY:** A argument > O argument[1]
c. PERSON HIERARCHY > FUNCTION HIERARCHY

In other words, the verb agrees with the highest argument on the PERSON HIERARCHY, unless both A and O are on the same level, in which case the verb agrees with the highest argument on the FUNCTION HIERARCHY. The constraints in (10) are exemplified by the following sentences (data from Tonhauser (2006, p. 132)).

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

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

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

Since person outranks function in terms of agreement, the verb in (11) agrees with the 1st person subject while the verb in (12) agrees with the 1st person object. In both cases, verb agreement is determined by the highest argument on the PERSON HIERARCHY. When both arguments are on the same level in the PERSON HIERARCHY, as is the case in (13), the verb agrees with the subject over the object (takes Set A rather than Set B morphology), according to the FUNCTION HIERARCHY. The marker *-pe* in (13) distinguishes *Juan* as the object rather than the subject of *hecha* — "see," a phenomenon which is discussed in much greater detail later in this study.

The system in (9) assigns grammatical function in (i) intransitive clauses, in which the only co-occurring NP must be the S-argument, and (ii) transitive clauses with 1st or 2nd person arguments, since the verb must display head-marking to agree with them according to the hierarchies in (10), and the range of referents selected by this agreement is very small (only the speaker or hearer). However, when both A and O are in 3rd person, the verbal agreement properties always match both co-occurring NPs. This is the case in (14) from Tonhauser and Colijn (to appear, p. 29), where the 3rd person verb-marker *-o* does not discriminate between *Juan* and *Maria*.

[1] The *A* argument is the one crossreferenced by a Set A marker, while the *O* argument is the one cross-referenced by a Set B marker

(14) **Juan** o-hecha **Maria**.
**Juan** A3-see **Maria**
'**Juan** saw/sees **Maria**.'

This means that clauses with overt 3rd person subject and object always have two analyses consistent with the verbal agreement system.[2]

### 1.1.2 Word Order

All six possible orders between subject, object, and verb are grammatical in Guaraní (Tonhauser & Colijn, to appear). Gregores and Suárez (1967) tentatively claim that Guaraní is SVO. Velázquez-Castillo (1996, p. 11), on the other hand, claims that "the default order of elements seems to be VO [verb–object], with the subject appearing either before or after the verb." Tonhauser and Colijn (to appear) provide evidence for both VO and SO basic orderings, but argue against a basic placement of subjects with respect to verbs. Regarding VO ordering, they show that 94% of objects in their corpus are post-verbal (pp. 18-19). Regarding SO ordering they present two pieces of evidence. First, 78% of clauses in their corpus with overt subject and object realize SO order. Second, they claim, SO disambiguates in clauses which would otherwise exhibit A/O ambiguity (Tonhauser & Colijn, to appear, pp. 29-30). As evidence, they present the two examples in their corpus of ambiguous clauses with both NPs overt. One is the sentence in (2), reproduced below as (15). The other is shown in (16).

(15) O-monga-kuaa karai **pe mitã**.
A3-raise-know gentleman **that child**
'The gentleman raised **the child**.'

(16) Tuju-ry o-jagara-pa **la** **ij-ao**.
mud-juice A3-grab-COMPL **the(Sp)** **B3-cloth**
'Mud got all over (grabbed) **his clothing**.'

Tonhauser and Colijn (to appear) argue that these examples are ambiguous because neither grammar nor animacy/world knowledge clearly indicate which NPs should be linked to which grammatical functions: since both NPs in each clause are 3rd person, the agreement properties of the 3rd person verbal marker

[2] I should also mention that S/IO ambiguities are also possible; they are perhaps even more likely since IOs are usually recipients, which have many proto-agent properties (Dowty, 1991). Bossong (1991) and Aissen (2003b) point out that the particular accusative morpheme used for DOM cross-linguistically is usually identical to the dative marker, and that indirect objects and differentially-marked objects take the same morphology as a result. Aissen hypothesizes that this is due to the fact that both DOs and IOs are similar in their non-subjecthood and that IOs are more frequently human than DOs. As discussed in § 1.3.1 – § 1.3.3, *-pe* marks directional locatives, IOs, and DOs, conforming quite well to this pattern.

*o-* match both, and since both NPs in each clause have the same level of animacy (Human in (15) and Inanimate in (16)), neither object is more likely than the other to fill either grammatical function. However, I contend that in the case of (15), despite the equal animacy of the NPs, each of the NPs *is* much more likely to fill one role over the other according to world knowledge. Given the verb *mongakuaa* — "raise," it would be much less surprising thematically for *karai* — "gentleman" — to be the subject than *pe mitã* — "that child," since adults generally raise children and not vice versa. Thus Tonhauser and Colijn's corpus provides one clear example (16) in support of their hypothesis that SO is a disambiguating word ordering. In addition to corpus data, they elicited responses from speakers to sentences like those in (17) (Tonhauser & Colijn, to appear, pp. 29-30).

(17) 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**.'

They presented Guaraní speakers with all six possible permutations of $NP_1$, $NP_2$, and verb for both (17a), which has a human NP and an inanimate NP and is therefore unambiguous, and (17b), which has two 3rd person human NPs and is therefore ambiguous. They found that in all six variations of (17a), *Juan* is judged to be the subject and *peteĩ ita* the object. However, in (17b), only SO permutations resulted in a judgment of *Juan* as subject, providing evidence that SO is a disambiguating order. The fact that judgments about the unambiguous (17a) are unaffected by word order while judgments about the ambiguous (17b) *are* affected by word order is compelling evidence that SO is indeed a default ordering, at least in otherwise ambiguous clauses. However, Tonhauser and Colijn do not address the question of whether OS parsings of ambiguous clauses are ungrammatical or simply dispreferred. In other words, it might be the case that the default parsing of an ambiguous clause with two overt NPs is SO, but that an OS reading is nevertheless possible.

Various scholars have identified other factors that affect word order in Guaraní. Velázquez-Castillo (1995) finds topicality to be relevant to object placement, demonstrating that "if a thematic participant is coded as an object it must follow the verb (but the opposite implication does not hold) (Velázquez-Castillo, 1995, p. 572)." Thus, pre-verbal objects must not be topical, while post-verbal objects may be either topical or not. In addition to grammatical function and topicality,

Tonhauser and Colijn (to appear) provide evidence that animacy and discourse status affect word order, such that animate or discourse-new noun phrases are more likely to be pre-verbal than inanimate or discourse-old ones. All of the above criteria have been shown to affect word order in other languages as well (Brody, 1984; Enrico, 1986; Choi, 1999; King, 1995; Kiss, 2002; Aissen, 1992).

In sum, all possible orderings of S, O, and V are grammatical in Guaraní. Furthermore, word order seems to reflect a variety of semantic and pragmatic factors besides grammatical function. The data presented in the literature thus far does not conclusively show whether SO order biases GF-assignment or completely eliminates grammatical function ambiguity, such that an OS reading is not available in an otherwise ambiguous clause. While leaving this question open to future research, for the purposes of this study, I do not assume word order to be a means of disambiguation when assessing the ambiguity of the clauses in the corpus.

### 1.1.3 Other Forms of GF-assignment in Guaraní Grammar

Besides head-marking, other, less productive forms of grammatical disambiguation appear as well. Examples of this include constructions containing a necessarily transitive verb and only one NP. Since 3rd person object omission is disallowed (Tonhauser, personal communication) while pro-drop is permitted, the NP is unambiguously a direct object.

(18) 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**.'

In (18), the necessarily transitive verb *guereko* — "have" — co-occurs with only one NP: *peteĩ jagua piru'i peteĩ* — "a dog, a little skinny one." Thus it must be the case that the dog is the object of "have," and this clause exhibits no A/O ambiguity.

Similarly, when one of the NPs in a transitive clause is a subject-only proform such as *ha'e* – "3rd Sg Subj" – or *hikuái* – "3rd Pl Subj," the subject slot in the clause is by definition filled, rendering the remaining NP an object. An example of this is given in (19).

(19) Pero o-topa *hikuái* **la sirujúano por kasualidad**
but(Sp) A3-find *3Pl.Subj.PRO* **the(Sp) surgeon(Sp) for(Sp) chance(Sp)**
**o-ĩ-va** **avei**.
**A3-be-RC too**.
'But they found **the surgeon who happened to be there, too**.'

Here, there are two overt NPs: *hikuái* — "they (subject)" — and *la sirujuáno por kasualidad oĩva avei* — "the surgeon who just happened to be there, too." Since *hikuái* is necessarily a subject pronoun, the object function is the only one available to "the surgeon," and this sentence also exhibits no A/O ambiguity.

### 1.1.4 Summary

To summarize the tools of GF-assignment in Guaraní grammar: there exists a marker *-pe* which demarcates NPs as objects when it attaches to them. Additionally, the verb agreement system shown in (9) allows A/O ambiguity always and only when both NPs are in 3rd person. The preferred word order is SO, which I assume to bias GF-assignment but not to disambiguate clauses outright, since all orders of S, O, and V are grammatical. Finally, single NPs co-occurring with a necessarily transitive verb are objects, since object omission is not grammatical, and the presence of subject-only pronouns such as *hikuái* relegates any co-occurring NPs to object position. Thus grammar-related subject/object ambiguities appear in Guaraní when both arguments are in 3rd person and no additional grammatical disambiguating information is available.

## 1.2 Noun Phrase Structure

Given that the discussion in this study centers on NPs in Guaraní and their marking as objects, it is important to provide a cursory overview of noun phrase structure in Guaraní. Determiners, demonstratives, numerals, and possessive markers precede the noun (examples (20) and (21) are courtesy of Tonhauser (2006)):

(20) ko kyse
this knife
'this knife'

(21) peteĩ kyse
one knife
'a/one knife'

(22) i-ryguasu
3-chicken
'Their/her/his chicken'

Determiners and possessive markers may co-occur, in which case the demonstrative precedes:

(23) la i-po
the(Sp) 3-hand
'Their/her/his hand'

Attributive noun clusters in Guaraní take the form of a string of nouns in which the final N is the head and the preceding N the modifier. Thus attributive nouns precede the head. These attributive constructions can be possessive, as in (24), but aren't necessarily so, as in (25).

(24) che-sy sy róga
my-mother mother house
'My grandmother's house (mother's mother's house)'

(25) tetã mba'e mombe'u
country thing story
'A country story'

Adjectives follow the head noun, as does the plural marker *-kuéra*:

(26) che lapi pytã
my pencil(Sp) red
'my red pencil'

(27) la i-pypore-kuéra
the(Sp) 3-footprint-PL
'Their/her/his footprints'

Relative clauses follow the head noun in Guaraní, and are demarcated by a relative clause marker *-va*, which attaches as a suffix to the main verb in the relative clause (example (28) courtesy of Tonhauser and Colijn (to appear)).

(28) peteĩ kane'õ-va
one tired-RC
'one who is tired'

(29) i-personal-kuéra o-mba'apó-va kokúe-pe
3-personal.servant(Sp)-PL A3-work-RC chacra-PE
'her personal servants who worked on the *chacra*'

Guaraní has two nominal temporal markers (*-kue*, past, roughly translated "former" and *-rã*, future, roughly translated "prospective" (see Tonhauser, 2006)), as well as a nominalizing suffix *-ha*, which attaches to verb phrases and allows them to be used as noun phrases within a sentence (for more on Guaraní nominalization, see Gildea (1994); Derbyshire (1994)). These morphemes are phrase final:

(30) che lapi pytã-rã
my pencil(Sp) red-NOM.FUT
'my (prospective) red pencil'

(31) ja-je-descuida-ha
A1.Incl-REFL-neglect(Sp)-NOM
'The neglecting of ourselves'

## 1.3 The Morpheme *-pe*

Since this study of DOM focuses on the morpheme *-pe*, it is important to describe the range of *-pe*'s distribution generally. In § 1.3.1 I cover the use of *-pe* in oblique phrases. In § 1.3.2 I point out that *-pe* marks indirect objects in addition to oblique phrases. Lastly, I provide evidence in § 1.3.3 that *-pe* does indeed function as a direct object marker and that this object-marking is differential: not all DOs in Guaraní are marked by *-pe*.

### 1.3.1 *-pe* as a Spatiotemporal Postposition

In oblique phrases, *-pe* denotes a wide variety of thematic relationships, mostly spatiotemporal (see Tonhauser, 2006). The phrases in parentheses are my approximate translations.

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

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

(34) kuehe o-ho-rõ Lebonia **Paraguaý-pe** a-ncarga
yesterday A3-go-when Lebonia **Asunción-PE** A1.Sg-charge(Sp)
ichu-pe o-gueru haĝua chéve che lapi pytã-rã ...
3.OBJ-PE A3-bring PURP 1Sg.IO B1.Sg pencil(Sp) red-NOM.FUT ...
'Yesterday when Lebonia went **(to Asunción)** I urged her to bring me some red lipstick ...'

(35) A-po **pe tape yké-pe**.
A1.Sg-jump **that path side-PE**
'I jumped **(to the side of the path)**.'

(36) 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 ...'

(37) **Peteĩ mes Máyo-pe** o-ho la escúela-pe, cáda
**one month(Sp) May-PE** A3-go the(Sp) school(Sp)-PE every(Sp)
dia o-ho-há-icha voi.
day(Sp) A3-go-NOM-like later
'**(One May)** she went to school, like she went like every day.'

(38) Ou ambue jagua-kuéra ha o-ñarõ avei, **o-joyvý-pe**.
A3.come other dog-PL and A3-bark also **A3-combined-PE**
'Other dogs came and barked too, **(in a chorus)**.'

(39) Ha **kyhyje-pó-pe** ro-hasa ha che-kyvy haimete ho'a
and **fear-very-PE** A1.Excl-pass and B1.Sg-brother almost A3.fall
mbokaja rati-ári.
small.coconut.plant spine-on
'And we passed **(in fear)** and my brother almost fell onto the spines of a coconut plant.'

(40) **Karai-ñe'ẽ-me** che gana.
**gentleman-language-PE** B1.Sg gain(Sp)
'He wins out over me **(in Spanish)**.'[3]

### 1.3.2 *-pe* as an Indirect Object Marker

In addition to oblique phrases like those given above, *-pe* can also occur on indirect objects.[4] This marking is restricted to 3rd person and 2nd person plural objects. The other 1st and 2nd person pronouns have distinct IO forms, as seen below in (41).

[3] The initial obstruent in *-pe* is nasalized because of the nasal vowel that precedes it.

[4] The postposition *-gui* can also mark IO's on occasion, most often those that are in a "maleficiary" relationship to the event.

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

In oblique phrases, *-gui* has been analyzed as an ablative case-marker, meaning "away from" (see Tonhauser, 2006).

(41)

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

1st & 2nd Personal Pronouns in Guaraní
As presented in Gregores and Suárez (1967)

Indirect object *-pe* appears on both referential NPs and pronouns, as demonstrated by (42) and (43), where the same verbs take a *-pe*-marked referential NP in the (a) sentences and a *-pe*-marked pronoun in the (b) sentences.

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

b. Mba'e angá-pa o-jehu **ichu-pe**?
thing poor-totally 3A-happen **3.OBJ-PE**
'Poor thing! What happened **to him**?'

(43) a. Ñande **sý-pe-ko** ña-ombe'ú-va o-jehú-va
B1.Incl **mother-PE-EMPH** A1.Incl-narrate-RC A3-happen-RC
ñandéve!
1Incl.IO
'We should tell **our mothers** what happens to us.'

b. Aní-ke-na re-mombe'u **ichu-pe**, mamá!
NEG-REQ-IMP A2.Sg-narrate **3.OBJ-PE** mom
'Don't tell **him**, Mom!'

Examples (43) – (47) all exhibit *-pe*-marked recipients.

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

(45) E-ñe'ẽ porã-na **ichu-pe**.
A2.Sg-say good-IMP **3.OBJ-PE**
'Tell **him** nicely.'

(46) t-a-me'ẽ-mo **ichu-pe** avati-mi.
REL-1A.Sg-give-put **3.OBJ-PE** corn-a.little
'I'll get ready to give **him** a little corn.'

(47) Memoria-ke-na ko compadre **Juan Ramón-pe**,
memory(Sp)-REQ-IMP this godfather(Sp) **Juan Ramón-PE**
'Regards to compadre **Juan Ramón**,'

Examples (48) and (49) show the 2nd person plural pronoun *peẽ* taking *-pe* when functioning as an IO.[5]

(48) Oi-me-ku he'i-sé-va **peẽ-me** hína.
A3-EXIST-such 3.say-DES-RC **2Pl.PRO-PE** PROG
'There's something he wants to say **to you**.'

(49) o-sẽ porã-mba vaerã-ku **peẽ-me** ãga!
A3-leave good-totally VAERA-such **2Pl.PRO-PE** now
'Things should turn out totally fine **for you both**!'

### 1.3.3 *-pe* as a Differential Object Marker

I now argue that *-pe* is a differential object marker, a morpheme that attaches to the O argument in some, but not all, transitive clauses. Take (50) for example.

(50) a. Iñ-akã-ári-rupi, o-hecha ha o-japysaká-re,
3-head-above-through A3-see and A3-give.attention-REHE
o-hendu **kuimba'e-kuéra-pe** o-ñe'ẽ ha o-ñe'ẽ-api.
A3-hear **man-PL-PE** 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**
day.before.yesterday 3A.hear **that Remigia chatterbox Mrs. Leona**
**memby**, o-mbojá-ramo hese o-mbo-liga-ha ndéve
**child** 3A-come.close-when 3.IO 3A-put-league(Sp)-NOM 2B.Sg
Don Albérto-ndive!
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!'

[5] The initial obstruent in *-pe* is nasalized because of the nasal vowel that precedes it.

In this case, the same verb *hendu* — "hear" — selects a *-pe*-marked object in (50a) (*kuimba'ekuéra* — "men") and an unmarked object in (50b) (*ku Remigia ñe'ẽngatu, Ña Leona memby* — "that Remigia, Mrs. Leona's daughter"). This is evidence that object-marking in Guaran'ıis differential, since marking varies across two transitive clauses that contain the same verb.

Further evidence of DOM is presented in (51).

(51) a. Ha upéi upépe o-ĩ-jave hína o-hecha **Juán-chi ha**
and then there A3-be-while PROG A3-see **Juan-DIM and**
**Pirúlo-pe** o-ñe-moĩ o-hupi i-po ichu-pe.
**Pirúlo-PE** A3-JE-put A3-raise 3POSS-hand 3-PE
'And then, while he was there, he saw **Juanito and Pirulo** getting ready to say good-bye to him.'

b. Ha upéi o-maña h-enonde gotyo amó-ite yvyra ñeno-ári **ju'i**
and then A3-see 3-front towards there-very wood lie-on **frog**
o-puka hese-kuéra.
A3-laugh 3.OBJ-PL
'And then he saw in front of him, right there on a tree trunk, **the frog** laughing at them.'

In both (51a) and (51b), the verb takes a null subject, 3rd person inflection, and one 3rd person object NP, which also serves as the subject for an embedded clause. However, in (51a) this NP is marked by *-pe*, while in (51b) it is not. Thus the objects in this pair are also differentially marked.

Example (52) provides a pair of sentences with two verbs, *jura* (52a) and *jagara* (52b), with similar semantics (both are translated into English as "grab"). In (52a), the object, *ju'i* — "frog," is *-pe*-marked. In (52b), the object, *la ijao* — "his clothing," is not *-pe*-marked.

(52) a. Pirúlo n-o-guerovia-i mba'é-icha-pa Juán-chi
Pirúlo NEG-A3-believe-NEG thing-like-QU Juan-DIM
nd-o-jura-i **ju'í-pe**.
NEG-A3-grab-NEG **frog-PE**
'Pirulo couldn't believe how Juanito couldn't grab **the frog**.'

b. I-ky'a-pá-ite Juán-chi porque tuju-ry
B3-dirty-COMPL-very Juan-DIM because mud-juice
o-jagara-pa **la** **ij-ao**.
A3-grab-COMPL **the(Sp)** **B3-cloth**
'Juanito was extremely dirty because the mud completely got ahold of **his clothing**.'

Thus, object-marking in Guaraní is differential. The question of distribution raised by the above data is the driving force behind this study, and my analyses and results are discussed in § 3.

## 1.4 The Lexicalization of *chupe*

The above examples have included several instances of the pronominal *chupe* (allomorphic with *ichupe*), which has been treated as a combination of a 3rd person root *chu-* and the argument-marker *-pe*. I now examine the lexical status and distribution of *chupe* and discuss its similarities to and differences from NP-*pe* phrases, arguing that *chupe* has been "partially lexicalized" — acting as a single lexical entry but retaining many of the syntactic properties of its component parts.[6] As evidence for the lexicalization of *chupe*, take (53).

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

b. O-puka o-hechá-vo **chu-pe-kuéra**.
A3-laugh A3-see-when **3-PE-PL**
'He laughed when he saw **them**.'

In this pair, the plural marker *-kuéra* both precedes and follows *-pe*, depending on whether the constituent it attaches to is a referential noun (53a) or the pronominal *chu* (53b). It turns out that *-kuéra* regularly falls between its nominal head and the following postposition; the only exceptions to this pattern are co-occurrences of *-kuéra* with *chu*, in which *-kuéra* always follows the postposition; the precedence relations between *-kuéra* and *-pe* are reversed in the case

[6] This hypothesis is along the same lines as Dietrich (1994), who claims that the word *mba'e* — "thing" — has been partially "grammaticalized" for all of Tupí-Guaraní. By this he means that *mba'e* has been elevated beyond referential to functional status as an affix and therefore has a much broader distribution than its referential meaning alone would allow.

of *chu-*. This is strong evidence for a combinatorial difference between NP-*pe* and *chu-pe*, such that *chu-pe* is more tightly bound together. Further evidence of this tighter adjacency is the fact that the morpheme *chu* never appears independently in the corpus; it is always augmented either by *-pe* or *-gui*, while bare NPs do occur (see Tonhauser & Colijn, to appear).

Even more evidence of the lexicalization of *chupe* comes from phonology. Stress in Guaraní generally falls on the ultimate syllable. However, some suffixes may be stressed when they are word-final, and some may not. In the case of the latter, stress instead falls on the last possible preceding syllable. Such is the case with *-pe*, an unstressed suffix, as can be seen in many of the examples of *-pe*-marked NPs given above, where stress falls on the penultimate syllable, the last one preceding *-pe*. *Chupe*, however, is stressed on the ultimate syllable, conforming to the stress pattern of individual words but not to the stress pattern of *-pe*-marked NPs, where *-pe* is morphemic. This suggests that *chu* is not simply a third person pronoun which can take an argument-marker, but that *chupe* is an indivisible lexical item. For these reasons, I take *chupe* to be lexicalized as a third person object pronoun for Guaraní.

All of the above provide evidence supporting the conclusion that *chupe* is one lexical item, and not simply the combinatorial product of *chu-* and *-pe*. However, I have included the notion of "partiality" into this account of lexicalization because of the existence of *chugui*, which appears to display alternation between two postpositions over the same *chu* root, as seen in example (54).

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

Just like *chupe*, however, *chugui* violates the usual precedence relations between postpositions and *kuéra*, as seen in (54), where *-gui* precedes rather than follows *-kuéra* when attaching to *chu*. This supports an analysis of *chugui* as a single item. Furthermore, while *-gui* is an unstressed suffix like *-pe*, *chugui* nevertheless has final stress, evidence that *chugui* has also been lexicalized.

The existence of both these forms is possibly a carryover from a time in which this *chu* + postposition construction was more productive. In any case, the distribution of *chupe* and *chugui* is similar to that of NP + $\left\{ \begin{array}{c} \textit{-pe} \\ \textit{-gui} \end{array} \right\}$, as shown in the examples below.

(55) a. *He'i* **Pirúlo-pe**...
*A3.say* **Pirúlo-PE**...

'He *said* **to Pirulo**...'

b. *He'i* **chu-pe** ype...
*A3.say* **3-PE** duck...
'The duck *said* **to him**...'

(56) a. I-*poi*-pá-ite **pe i-pó-pe-gúa-gui** ha ho'a
3-*drop*-COMPL-very **that 3-hand-PE-of-gui** and A3.fall
otĩ-vo ý-pe.
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-moĩ-vo pe i-vosa-'í-pe o-jagará-ha-guã-icha
and A3-JE-put-when that 3-bag-DIM-PE A3-grab-in.order.to-like
ju'í-pe, ju'i o-po o-je-*poi* **chu-gui** ý-pe.
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.'

Thus I take the position that *chupe* and *chugui* are each single lexical items that behave in some ways as if they were the concatenation of two morphemes.

## 1.5 Adposition or Case-Marker?

The question of whether *-pe* is a postposition or a case-marker is still open for Guaraní, in large part because of the fact that definitive distinctions between these categories are difficult to find. The difference between cases and adpositions is not functional but formal; as Zwicky (1992, p. 370) says, "anything you can do with cases you can also do with adpositions, and vice versa." Similarly, Haspelmath (2006, p. 2) states 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, in terms of their implications for the grammatical status of their noun phrases, cases and adpositions have significant differences. Case-markers do not cause category shift: NP + case is still an NP. Thus, cases are modifiers and nouns are heads. In contrast, adpositions are phrasal heads: NP + adposition is an adpositional phrase. NPs are complements of adpositions. The structural implications of the case/adposition distinction are therefore non-negligible.

As discussed in Zwicky (1992) and Haspelmath (2006), generally speaking, cases are morphological inflections on nouns or noun phrases, while adpositions

are independent words. There is nevertheless disagreement in the Guaraní literature as to the grammatical status of *-pe*. Gregores and Suárez (1967) make the claim that there is a class of NPs in Guaraní which serve as "the axis of the postposition *-pe*" when they are objects. They therefore take the position that *-pe* is a postposition which demotes the object NPs to which it attaches to oblique status. Velázquez-Castillo (2004) concurs in her analysis of *-pe*, explicitly stating that object nominal roots "are not inflected for case" (p. 1426). Instead, she claims, *-pe*-marked NPs are the objects of postpositional phrases. Others do not make this assumption. For example, Tonhauser and Colijn (to appear) assume *-pe* to be an affix, that is, a case-marker. Adelaar (1994) also makes the claim that *-pe* is a case-marker, not a postposition. Under this view, objects are not demoted when taking object-marking, and the structure of transitive clauses with marked objects is essentially the same as that of those with unmarked objects. There is a third possibility which has not (to my knowledge) been articulated in the Guaraní literature: that there could be two distinct *-pe* morphemes in the language. In the case of oblique phrases, *-pe* could be postpositional. In the case of indirect and direct objects, *-pe* could be inflectional.

The piece of data presented in (57) provides evidence that *-pe* is postpositional:

(57) 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**.'

In this example, the locative phrase *y ha yvýpe* — "in the water and on the ground" — consists of two conjoined NPs (*y* and *yvy*) followed by *-pe*. If *-pe* were a nominal case inflection, we would expect both conjoined NPs to be inflected: *ype ha yvýpe*. But this is not what happens: *-pe* only appears at the end of the conjunction. This serves as evidence that *-pe* is a postposition (i.e., not a daughter of the NP) when it occurs in oblique phrases, though it does not show whether it is postpositional on objects as well as obliques. However, the existence of (51a), reproduced below as (58), suggests that *-pe* is also postpositional on objects.

(58) Ha upéi upépe o-ĩ-jave hína o-hecha **Juán-chi ha Pirúlo-pe**
and then there A3-be-while PROG A3-see **Juan-DIM and Pirulo-PE**
o-ñe-moĩ o-hupi i-po ichupe.
A3-JE-put A3-raise 3-hand 3Obj.PRO
'And then, while he was there, he saw **Juanito and Pirulo** getting ready to say good-bye to him.'

In this example, the conjoined NPs, *Juánchi* and *Pirulo*, serve as the object of

*hecha* — "see." As is the case in (57), *-pe* appears only at the end of the conjunction, not on both NPs individually as we would expect if *-pe* were a case-marker. Thus coordination seems to favor a categorization of *-pe* as a postposition with respect to obliques and objects, which implies that *-pe*-marked objects are demoted to oblique status.

For the purposes of this study, the question of case vs. adposition is more or less irrelevant. It is possible that DOM in Guaraní involves demotion to oblique status, and it is possible that it does not. However, the issue at hand is simply the appearance of *-pe* juxtaposed to some objects and not others, and the question being addressed is which factors underly this distribution. Further questions about the resultant categories and clausal structures under object-marking do not directly bear on the results of this study. I therefore leave the question open to future research.

# 2 Literature Review

In this section I provide an overview of the literature relevant to the study of DOM in Guaraní. Section 2.1 covers the current literature on Differential Object Marking generally, with a heavy emphasis placed on Aissen (2003b), from whom I draw significantly in the analysis presented in § 3.1. Section 2.2 discusses the existing perspectives on *-pe*-marking in Guaraní.

## 2.1 Differential Object Marking

There is a large body of literature on DOM and the factors that condition it. One of the seminal works on the topic is Bossong (1985), upon which later publications have built (Aissen, 2003b, 2003a; Bossong, 1991; Heusinger & Kaiser, 2003, 2005; de Swart, 2007; Morimoto, 2002; Bak, 2004; Lima, 2006; Leonetti, 2004). These publications address a cross-section of languages and present an array of distributional analyses. Most prevalent among these theories of DOM are (i) that the markedness (subject-likeness) of the object triggers DOM and (ii) that DOM arises in response to a need to assign grammatical functions to NPs in clauses which are otherwise ambiguous with respect to GF-assignment. It appears that neither is sufficient for predicting object-marking universally, and that DOM in some languages is governed by the constraints in (i) and in others by those in (ii) (de Swart, 2007).[7]

[7] The literature is not universally supportive of markedness as a legitimate basis for DOM, however. For a counterpoint, see Næss (2004).

Regarding the first analysis, Aissen (2003b) advances a theory of DOM that encapsulates the ideas contained in much of the other literature on the topic (see Heusinger & Kaiser, 2003, 2005; Leonetti, 2004; Bossong, 1983-1984): that classes of nouns can be ranked according to how subject-like or object-like they are, and that object-like objects have a low level of markedness while subject-like objects have a high level of markedness. Aissen (2003b) refers to this subject-likeness as "prominence" and assumes that prominence corresponds to morphological marking in languages with DOM; that is, highly subject-like objects should be object-marked, while highly object-like objects should be unmarked. Aissen (2003b) appeals to two sets of properties in measuring prominence: animacy and definiteness. In doing so, she lays out two hierarchies, where $a > b$ iff $a$ is more prominent (subject-like) than $b$:

(59) **ANIMACY HIERARCHY:** Human > Animate[8] > Inanimate

(60) **DEFINITENESS HIERARCHY:** Pronoun > Proper Name > Definite > Indefinite Specific > Indefinite Nonspecific

Thus, the higher any object falls on either of these hierarchies, the stronger its preference for marking by DOM. Aissen (2003b) then takes the cross-product of the sets of categories contained in both these hierarchies to create a Prominence Lattice in which NPs may be ranked with respect to one another according to their position on both the ANIMACY and DEFINITENESS hierarchies together. Put in more formal terms:

(61) "if $a$ ranges over values on one scale, and $b$ over values on the other, then a pair $(a_1, b_1) \geq (a_2, b_2) \Longleftrightarrow a_1 \geq a_2$ and $b_1 \geq b_2$" (Aissen, 2003b, p. 458).

Thus "Human Definite" ≥ "Animate Nonspecific," but < "Human Pronoun," since it must rank as or more highly on both scales in order to be greater than or equal.

Aissen represents the Prominence Lattice with the illustration in Figure 1. Each node in this figure contains a pair of values ($a$,$b$), where $a$ is a value on the ANIMACY HIERARCHY and $b$ is a value on the DEFINITENESS HIERARCHY. Each value on one hierarchy is paired with each value on the other, with the rankings of the items in the hierarchies preserved from bottom (lowest) to top (highest). A line connecting two nodes indicates a ranking of those two nodes, with the higher node having a higher level of prominence. Any two nodes in a dominance relationship to one another in Figure 1 are ranked with respect to each other, with the mother node having a higher level of prominence than the daughter (for example: Human Pronoun > Animate Nonspecific)). Any two nodes that are not

[8] In Aissen (2003b), "animate" is shorthand for "animate and not human." The same terminology is used through this study as well: [+human] NPs are not considered animate.

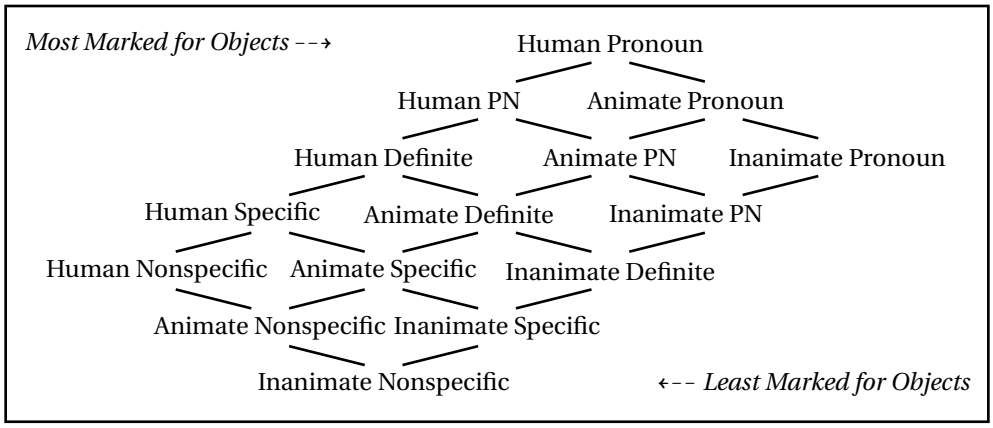


Figure 1: Prominence Lattice
As presented in Aissen (2003b).

in a dominance relationship (i.e. where neither dominates the other in Figure 1, such as Animate Pronoun and Human Nonspecific) are unranked with respect to each other. As a result, items which appear to be on the same level in Figure 1 are not equal but rather *incomparable* under this analysis. More formally, the set of pairs of nodes on the Prominence Lattice that are unranked with respect to each other consists of all pairs of nodes in which $a_1 \geq a_2$ while $b_1 < b_2$, or vice versa; i.e., pairs of nodes which outrank each other on different axes.

Aissen's theory predicts the following:

(62) a. If a node on the Prominence Lattice is never marked, then no nodes lower on the lattice will ever be marked.

b. If node on the Prominence Lattice is optionally marked, then no nodes lower on the lattice will be obligatorily marked and no nodes higher on the lattice will be obligatorily unmarked.

c. If a node on the Prominence Lattice is obligatorily marked, then no nodes higher on the lattice will ever be unmarked.

Aissen (2003b) therefore makes no predictions about DOM across nodes which cannot be ranked with respect to each other under this analysis. For example, this analysis makes no predictions about DOM on Animate Proper Names given object-marking on Inanimate Pronouns, since they each outrank one another on different axes. Rather than a universally-defined ranking in these situations, the ranking between such incomparable nodes is defined language-specifically. In addition to predicting how regions of obligatory marking, optional marking, and obligatory non-marking will be ranked, Aissen (2003b, p. 460-1) remarks

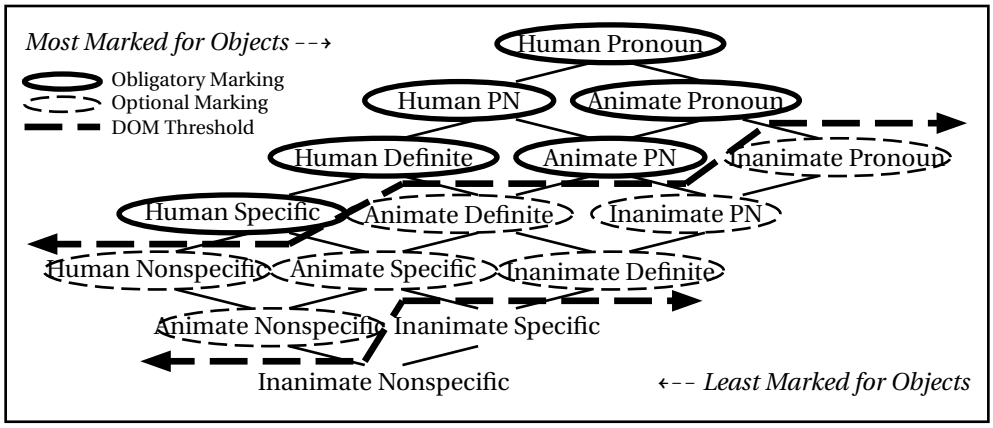


Figure 2: Prominence Lattice and DOM in Hindi
As discussed in Aissen (2003b)

that "in zones where DOM is optional, it is often the case that the probability of case-marking varies depending on the particular properties of the object... [H]igher (dominating) elements are more likely to be case-marked than lower (dominated) elements." This is what I refer to as a "suggestion" rather than a prediction of the analysis: to the extent that this turns out to be the case for a given language, the preferability of this prominence-based analysis of that language is reinforced. However, if the predictions presented in (62) hold for a language, then PROMINENCE correctly predicts DOM in that language regardless of the distribution of DOM in the optional region.

In other words, Aissen proposes that languages employ two thresholds of prominence in determining where to object-mark. Above the higher threshold, objects are obligatorily marked. Below the lower threshold, objects are obligatorily unmarked. Meanwhile, between the two thresholds, objects are optionally marked, with the additional suggestion that marking increases in likeliness as one moves up the lattice within the range of optionality. Thus Aissen makes universal predictions about the ranking of these thresholds for all DOM languages sensitive to prominence, but allows the location of these thresholds to be determined language-specifically: the set of categories which are obligatorily-marked, optionally-marked, and obligatorily unmarked varies from language to language.

As an example of the generalizations of this analysis applied the specifics of a language, I've presented Aissen's analysis of Hindi in Figure 2. In this figure, obligatorily-marked categories are circled with a solid line, optionally-marked categories are circled with a dotted line, and obligatorily-unmarked categories are uncircled. Between obligatory marking and optionality appears a dotted line

representing one threshold, and between optionality and obligatory non-marking appears a dotted line representing the other threshold. Thus we can see the specific ranges of optionality and non-optionality for DOM in Hindi, which conform quite well to the predictions of her theory: high-prominence objects are marked, low-prominence objects are not, middle-prominence objects are indeterminate. Additionally, while the Prominence Lattice represented by Figure 1 makes no claims about the ranking of Animate Proper Names and Inanimate Pronouns, for example, they are distinguished in Hindi specifically, since the former are obligatorily marked and the latter are not.

While the above supplies precedent for the prominence-based analysis presented in § 3.1, there is precedent in the literature for an ambiguity-driven analysis of DOM as well, as described in (ii) above, where DOM is conditioned by GF-assignment ambiguity. Gerner (2008) applies such an analysis to the Tibeto-Burman language Yongren Lolo. He claims that Yongren Lolo employs a disambiguating suffix *t$^h$ie$^{21}$* in cases of A/O ambiguity. As an example of this, he provides the following data (Gerner, 2008, p. 299):

(63) a. ŋo$^{33}$ ɕe$^{33}$mo$^{33}$ ʈʂ$^h$ ʒi$^{33}$.
1Sg snake follow go
'I follow the snake./The snake follows me.'

b. ŋo$^{33}$ ɕe$^{33}$mo$^{33}$ ʈ$^h$ie$^{21}$ ʈʂ$^h$ ʒi$^{33}$.
1Sg **snake** **OBJECT.MARKER** follow go
'I follow **the snake**.'

The clause in (63a) is ambiguous: either NP may be the subject or the object. In (63b), however, the NP "snake" takes the ambiguity-driven differential object marker *t$^h$ie$^{21}$*, demarcating "snake" as the object and eliminating A/O ambiguity. The object-marker is not necessary in example (64) from Gerner (2008, p. 301), since world knowledge renders the clause unambiguous: "you" would always want to plow "the earth" and "the earth" would never want to plow "you" under standard assumptions about the world.

(64) ni$^{33}$ mi$^{33}$ mo$^{21}$ ŋ$^{21}$me$^{33}$ ɛ$^{21}$.
2Sg earth plough want ALTERNATIVE.QUESTION.PARTICLE
'Do you want to plough the earth?'

The marker *t$^h$ie$^{21}$* disambiguates in cases of A/B[9] ambiguity in Yongren Lolo as well, as seen in (65).

[9] "B" is the label in Dixon (1979) for indirect object arguments

(65) a. ŋo$^{33}$ su$^{55}$bə$^{21}$ t$^{h}$i$^{21}$ bə$^{21}$ z̥ɔ$^{21}$ mo$^{55}$ (gə$^{21}$).
1Sg book one CLASSIFIER 3Sg show give
'I showed him a book/He showed me a book.'

b. ŋo$^{33}$ su$^{55}$bə$^{21}$ t$^{h}$i$^{21}$ bə$^{21}$ z̥ɔ$^{21}$ t$^{h}$ie$^{21}$ mo$^{55}$ (gə$^{21}$).
1Sg book one CLASSIFIER **3Sg OBJECT.MARKER** show give
'I showed **him** a book.'

In this example, while world knowledge rules out the possibility of "book" being anything but a direct object (books typically are not recipients or agents in the world), there is still indeterminacy in (65b) as to whether "I" or "he" is the recipient of the act of showing. This indeterminacy is resolved by the presence of the object-marker *t$^{h}$ie$^{21}$* in (65b), which attaches to "he" and marks it as the indirect object. As shown in (64) and (65), Gerner assumes world knowledge to disambiguate: in cases in which general familiarity with the way things happen in the world does not provide sufficient information to allow parsing of grammatical function for all NPs in the clause, *t$^{h}$ie$^{21}$* marks the NP which is intended to serve as the object.

Regarding a less distant language from Guaraní, Martins (2003, pp. 156-8) claims that the morpheme *-pe* performs a very similar function in Mbya Guaraní, a closely-related language to Paraguayan Guaraní, demarcating the object in instances of A/O ambiguity.

## 2.2 DOM in the Guaraní Literature

The appearance of *-pe* on DOs has been commented on already by several Guaraní scholars. Gregores and Suárez (1967, pp. 136-7) claim that *-pe* is an object marker which differentiates between two noun classes: nouns in one class "occur as the center or only constituent" in a clause and are not *-pe*-marked; nouns in the other class "occur... as the axis of the postposition *-pe*" and require *-pe*-marking when functioning as objects. While the distinction between these two classes is not said to have any inherent semantic or syntactic basis (other than +/– *pe*), Gregores and Suárez claim that *-pe* generally appears when "the object refers to a person, and sometimes to animals; this rule has some exceptions" (Gregores & Suárez, 1967, p. 156). In other words, according to Gregores and Suárez, object-marking in Guaraní is differential, conditioned by which noun class the object belongs to: when the object is human it is marked, and usually not otherwise. In a similar vein, Velázquez-Castillo (2004) claims that *-pe* is a non-inflectional postpositional marker of [+human] DOs.

Tonhauser (2006, p. 132) also observes that *-pe* marks O-arguments in Guaraní, in addition to spatiotemporal obliques. She also points out that this marking is differential; however, in contrast to Gregores and Suárez (1967), Tonhauser claims that DOM is conditioned not by the humanness of the referent of the DO, but instead by the properties of the A-argument: if A is overt and/or higher on the ANIMACY HIERARCHY than O, then *-pe*-marking is omissible. Thus, under Tonhauser's analysis, it is not the absolute animacy of the object but rather the relative animacy between subject and object, in addition to the locality of the subject, that conditions *-pe*. An object which is more animate than its subject and/or an object with a non-local subject necessarily takes *-pe*, which is otherwise optional. However, the approach to describing *-pe*-marking and DOM in Guaraní until now has been quite cursory in the literature. Many of the above points are gleaned from footnotes or single paragraphs. It is therefore the purpose of this paper to bring current theories of DOM to bear on the study of Guaraní, and to bring the data of Guaraní to bear on the conversation about DOM.

# 3 Analyzing the Distribution of *-pe*

This section presents two possible analyses of the range of objects that are differentially marked in Guaraní. Section 3.1 appropriates the prominence-based analysis put forward by Aissen (2003b), as discussed in § 2.1, and tests it against the Guaraní data. Section 3.2, on the other hand, examines whether DOM could be based on subject/object ambiguity in a clause. I hypothesize that *-pe* overtly marks objects in otherwise ambiguous clauses, and that DOM is therefore motivated by ambiguity. I then test this against the corpus as well, and present a comparison of the results from both analyses in § 3.3.

## 3.1 Prominence-Based DOM

As noted above, the concepts underlying PROMINENCE play a large role in current studies of DOM. Various studies have found animacy and/or definiteness to be relevant to object-marking in languages with inflectional case-marking systems such as Sinhalese (Gair, 1970), Hebrew (Givón, 1978), Romanian (Farkas, 1978), Turkish (Enç, 1991), Korean (Bak, 2004), Spanish (Leonetti, 2004; Kliffer, 1982; Weissenrieder, 1990), and Persian (Karimi, 1999; Lazard, 1982), as well as in languages with head- rather than dependent-marking, such as the Bantu family (Morimoto, 2002), which is of particular relevance to Guaraní, a head-marking

language. I now evaluate the degree to which PROMINENCE accounts for DOM in Guaraní. Section 3.1.1 outlines the methodology of the study and the predictions of the analysis, while § 3.1.2 tests these predictions against the corpus data.

### 3.1.1 The Theory and Methodology

To determine whether or not Aissen's characterization of prominence plays a role in DOM in Guaraní, I catalogued every transitive clause in the corpus, taking note of the level of animacy and definiteness for each direct object. I found 401 total transitive clauses.

There were four particular types of objects in the corpus that require additional discussion: (i) the 3rd person object pronoun *chupe*, (ii) objects expressed by a Set B verbal agreement morpheme rather than an independent NP, (iii) NPs consisting exclusively of demonstratives and their modifiers, and (iv) NPs consisting exclusively of the wh-questioner *mba'e* and its modifiers . Each of these exhibits no differentiality of object-marking in the corpus: type (i) objects (3rd person object pronouns) are always marked, while the other three types never are. There can be one of two reasons for this non-differentiality: first, that the conditioning factors on DOM are such that objects of these types show no variation in marking, or second, that objects of these types are necessarily marked/unmarked for some reason independent of the factors conditioning DOM. With respect to (i) and (ii), marking is not related to DOM. As I argue in § 1.4, *chupe* is one indivisible lexical item, and *chu* can never be unmarked. Verb agreement markers, on the other hand, are prefixes, which by definition can't be suffixed and therefore never take *-pe*. Thus *-pe* in these instances is conditioned by factors of the lexicon, not by the factors underlying DOM. It is counterproductive to include objects with pre-determined marking in this analysis, since they are not sensitive to the conditions on DOM being tested and would produce anomalous results that spread across the distinctions made by both PROMINENCE and AMBIGUITY. Thus I exclude (i) and (ii) from consideration. However, while objects of type (iii) and (iv) are never marked in the corpus, it is unclear whether this fact results from the system of DOM or from something else. If the former, then the analysis is responsible for predicting their non-marking; if the latter, then they too should not be considered when analyzing DOM. The data available is insufficient to decide between these two possibilities, since corpora cannot provide conclusive evidence about ungrammaticality; i.e., there is no way to determine from a corpus whether *mba'e-pe* or Demonstrative-*pe* as objects are ungrammatical or simply don't appear. If I were to assume that types (iii) and (iv) were irrelevant to DOM and exclude them, and this assumption were incor-

rect, the empirical consequence would be an insufficiently general description of the distribution of DOM, one that captures some but not all of the relevant data. However, if I were to assume the opposite (that a description of DOM is responsible for accounting for non-marking on types (iii) and (iv)) and this assumption were incorrect, the consequence would be significantly reduced statistical clarity for both analyses generally, since they would be shown as failing to predict a large amount of anomalous data, data which need not have been explained by an analysis of DOM in the first place. Since there is more to lose under the second assumption (the entire analysis could be compromised), for the purposes of this study, I assume objects of types (iii) and (iv) to be ungrammatical, and not sensitive to DOM, and therefore exclude all of the above types, (i) – (iv), from evaluation, while leaving the question of the status of types (iii) and (iv) open to future research. Removing these objects from the data set has two side-effects. First, the Pronoun category of the DEFINITENESS HIERARCHY is no longer relevant, since *chupe* is the only object pronoun available. Second, the number of relevant clauses is reduced to 250.

The criteria I used to determine an object's location on the DEFINITENESS HIERARCHY are as follows. Objects modified by the indefinite marker *peteĩ* counted as indefinite, while objects modified by the definite marker *la* or a demonstrative, such as *upe* — "that" — or *ko* — "this," counted as definite. When an object was bare (appeared with no discourse status morphology), it was classified as indefinite if it was discourse-new and definite if it was discourse-old. Indefinites were further subcategorized with respect to specificity: whether the object NP refers to a particular entity in the world or simply a generic member of the set of items described by the NP. Specificity was evaluated from the perspective of the agent of the clause: if the agent apparently knew which particular referent was the object of the action, the NP was judged to be specific. If the agent did not know which particular referent was the object of the action, the NP was judged to be nonspecific. As an example of an indefinite specific object, take (66), where the agent sees a frog on a leaf in the water.

(66) O-hasá-vo upéi o-hecha **ju'i-pe** o-guapy y mbyté-pe peteĩ
A3-pass-when then A3-see **frog-PE** A3-sit water middle-PE one
yrupẽ-ári.
sieve-on
'In passing he saw **a frog** sitting in the middle of the water on a water lily leaf.'

Here, *ju'i* — "frog" — is bare (unmarked for definiteness) and discourse-new, satisfying the criteria for classification as indefinite. It is specific rather than non-

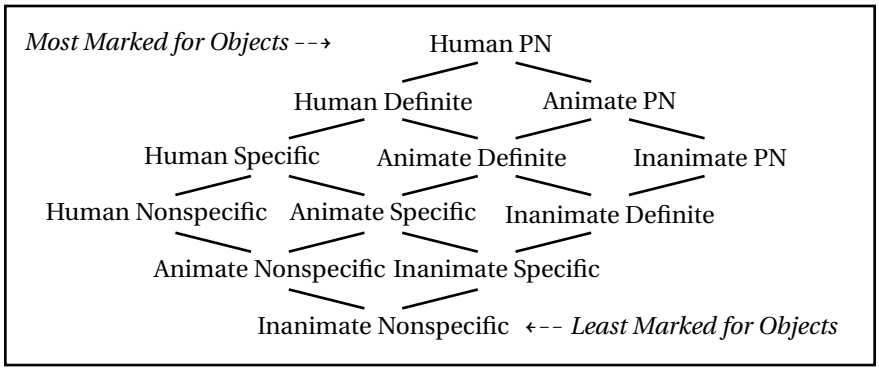


Figure 3: Revised Prominence Lattice
The "Pronoun" category of Figure 1 has been removed.

specific because the agent knows which particular member of the set of frogs he is seeing. Example (67) contains an indefinite nonspecific object: *overavante* — "that which shines."

(67) ha **o-verá-va-nte** pe-hecha-se, ta lata jepe!
CONJ **3A-shine-RC-only** 2A.Pl-see-DES be(Sp) tin(Sp) even
'and you only want to see **what shines**, even if it's tin!'

The object is a nominalization of the verb *overa* — "it shines," and has no discourse status marking morphemes; i.e., it is bare. Furthermore it is discourse-new. Thus it qualifies as indefinite. It is nonspecific because the agent in this sentence is said to love any member of the set of things that shine, not one particular shiny thing. Pronouns and proper names were categorized straightforwardly according to whether the object was (i) pronominal or (ii) a name rather than a description of an entity in the world.

I present a revision of Aissen's hierarchy in Figure 3; the only difference between Figure 1 and Figure 3 is the absence of any Pronoun categories, since these are not being considered in the analysis.

In order for the predictions of this theory to hold, there should be no obligatorily-marked objects lower in the Prominence Lattice than optionally- or obligatorily-unmarked ones, and no optionally-marked objects lower than obligatorily-unmarked ones. In addition to these predictions, the analysis strongly suggests that higher objects within the range of optionality should be more frequently marked than lower objects. Possible challenges to this suggestion include (i) an opposite trend, (ii) an even distribution of *-pe*-marking across optionally-marked categories, or (iii) a more-or-less random distribution of *-pe*-marking

across optionally-marked categories. While such challenges may not explicitly falsify this analysis, they would greatly reduce its appeal as an explanation of the conditions on DOM in Guaraní.

The cross-product of the items in the hierarchies should have equal or greater predictive power than single hierarchies; that is, objects which are low on both hierarchies should disprefer marking as or more strongly than objects that are low on only one, while objects that are high on both should prefer marking as or more strongly than objects that are high on only one. There is also no inherent weighting system between the ANIMACY and DEFINITENESS hierarchies; it is possible that one hierarchy is a much stronger predictor of DOM than the other.

### 3.1.2 The Data

Table 2 presents the PROMINENCE findings from the corpus survey, which are discussed in § 3.1.2. At the intersection of each category of the ANIMACY and DEFINITENESS hierarchies, three pieces of data are recorded: in the first row, the number of examples of that category that are *-pe*-marked is presented on the left, while the number of examples of that category that are not *-pe*-marked is presented on the right; in the second row, the percentage of the objects within that category that are *-pe*-marked is recorded in bold. Take the Human Nonspecific category, for example, shown as the upper left box in the table. The number 1 under "+*pe*" represents the fact that there was one Human Nonspecific object in the corpus that was *-pe*-marked. The number 8 under "–*pe*" represents the fact that there were 8 Human Nonspecific objects in the corpus that were unmarked. The bolded number underneath is the percentage of Human Nonspecific objects that are *-pe*-marked, which in this case is 1 out of 9, or about 11%. This percentage indicates the frequency of object-marking at that location on the Prominence Lattice. A higher percentage for a category indicates a preference for DOM in that category, while a lower percentage indicates a dispreference.

As shown in the lower right corner of Table 2, 10% of all objects in the corpus are *-pe*-marked. Therefore I assume 10% as the average likelihood of *-pe*-marking given objecthood. If the percentage of any particular category is significantly higher or lower than 10%, *-pe*-marking is preferred/dispreferred by DOM for that category. If, however, the percentage within any category is close to 10%, that category is not an informative predictor of DOM. If 100% of objects in a category are *-pe*-marked, that category falls into the obligatorily-marked range; i.e. it is above the threshold between obligatorily-marked and optionally-marked prominence categories. For any obligatorily-marked category, the analysis predicts that all higher categories will be obligatorily-marked (100%) as well. If 0% of

| Animacy | | Definiteness: 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 | 8 | 1 | 0 | 8 | 7 | 1 | 4 | 11 | 19 |
| | % +*pe* objects | **11%** | | **100%** | | **53%** | | **20%** | | **37%** | |
| Animate | # of objects | 1 | 3 | 2 | 4 | 9 | 33 | 1 | 0 | 13 | 40 |
| | % +*pe* objects | **25%** | | **33%** | | **21%** | | **100%** | | **25%** | |
| Inanimate | # of objects | 0 | 57 | 0 | 26 | 0 | 84 | 0 | 0 | 0 | 167 |
| | % +*pe* objects | **0%** | | **0%** | | **0%** | | **–** | | **0%** | |
| Total | # of objects | 2 | 68 | 3 | 30 | 17 | 124 | 2 | 4 | 24 | 226 |
| | % +*pe* objects | **3%** | | **9%** | | **12%** | | **33%** | | **10%** | |

Table 2: Object-Marking Across the Prominence Lattice

objects within a category are marked, that category falls into the obligatorily unmarked range and is therefore below the threshold between optionally-marked and obligatorily-unmarked prominence categories. For any category that is obligatorily unmarked (0%), the analysis predicts that all lower categories will be obligatorily unmarked as well. Any category with a percentage between 0% and 100% falls into the optionally-marked range between the two thresholds, above which there should be no obligatorily-unmarked categories and below which should be no obligatorily-marked categories.

The most important aspect of the data in Table 2 is where the prominence thresholds fall, since the ranking of these thresholds is what is predicted by the analysis. There appears to be only one threshold: a range of categories (Inanimate) is obligatorily unmarked (0% +*pe* objects) and a range of categories (Animate and Human) is optionally marked (100% > % +*pe* objects > 0%), but none are obligatorily marked (100% +*pe* objects). There seems to be a mutual implication between Inanimacy and non-marking: all 0% marked objects are Inanimate, and all Inanimate objects are 0% marked. While there are two categories that show 100% -*pe*-marking, Human Specific and Animate Proper Name, there is only example of each in the entire corpus. This is too little data to generalize that all Human Specifics or all Animate Proper Names in Guaraní are obligatorily marked. Thus the only prediction made by Aissen's analysis for Guaraní is this: that obligatorily-unmarked objects should rank lower on the Prominence Lattice than optionally-marked objects. This prediction holds: Inanimates are obligatorily unmarked and rank lower than Animate and Human objects, which are op-

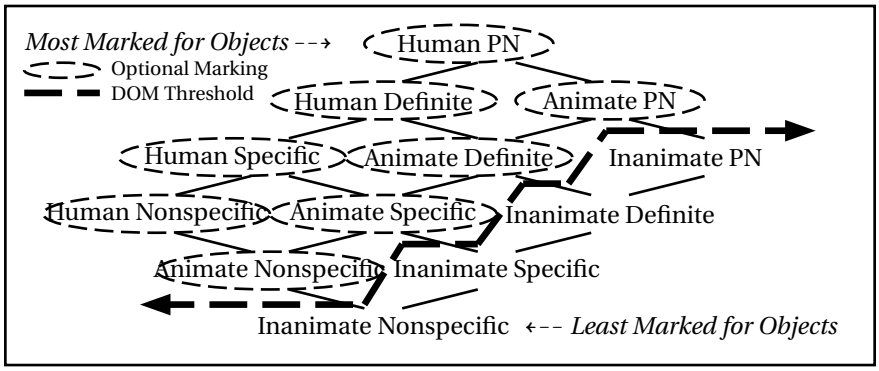


Figure 4: Prominence and the DOM Threshold in Guaraní
See Figure 2 for a comparison to DOM thresholds in Hindi.

tionally marked. This is captured by Figure 4, which shows all optionally-marked prominence categories in Guaraní circled by a dotted line and all obligatorily-unmarked prominence categories with no circle, along with a dotted line representing the threshold between the regions.

While the predictions of Aissen's prominence-based analysis are not contradicted by the Guaraní data, it is also important to analyze the degree to which Guaraní's optional range conforms to Aissen's suggestion about what should take place in that range, namely that more prominent objects should be marked with greater frequency than less prominent objects. I first discuss the ANIMACY and DEFINITENESS hierarchies individually, then look at the Prominence Lattice as a whole.

The data in Table 2 shows the suggested distribution of DOM along the individual hierarchies. This can be seen in the "Total" row for DEFINITENESS and the "Total" column for ANIMACY. The percentage of *-pe*-marked objects rises for NPs higher on both hierarchies, from 3% of Nonspecifics to 33% of Proper Names for DEFINITENESS and from 0% of Inanimates to 37% of Human objects for ANIMACY. The variation in *-pe*-marking from top to bottom of both hierarchies is similar in size; the highest category on the DEFINITENESS HIERARCHY exhibits about the same frequency of DOM as the highest category on the ANIMACY HIERARCHY, and likewise for the lowest categories of both hierarchies (33% vs. 37% and 3% vs. 0%). This means that high-animacy DOs prefer *-pe* about as strongly as high-definiteness ones, and low-animacy DOs disprefer *-pe* about as strongly as low-definiteness ones. The degree of variation in DOM frequency is slightly greater for ANIMACY, suggesting that animacy may be a more significant factor in determining the appearance of DOM in Guaraní. This is consistent with the

analysis of de Swart (2007), who argues that animacy is higher priority in general than definiteness with respect to object-marking, due to the fact that animacy is an inherent property of the referent while definiteness is contextually and pragmatically derived.

Along the DEFINITENESS HIERARCHY, Nonspecific objects disprefer DOM, since 3% is significantly less than the average (10%). At 33%, Proper Name objects prefer *-pe*-marking. Both Specific and Definite objects have percentages fairly close to the average (10%), which suggests that they do not significantly bias DOM. Meanwhile, along the ANIMACY HIERARCHY, Human objects are much more likely than chance to be *-pe*-marked (37%). Animate objects are less so, but still prefer *-pe*-marking (25%). This distribution of *-pe* along the ANIMACY HIERARCHY is evidence against the analyses of Gregores and Suárez (1967) and Velázquez-Castillo (2004), who claim that the conditioning factor behind DOM is humanness; Velázquez-Castillo (2004, p. 1426) says that *-pe* marks human objects, while Gregores and Suárez (1967, p. 156) say that *-pe* marks human objects and some animates. This does not at all hold for the data in the corpus: almost half of all Human objects are unmarked, while the ratio of *-pe*-marking on Animates is actually higher than chance, indicating a preference for *-pe*-marking on animate objects which is not predicted by either Gregores and Suárez or Velázquez-Castillo. Inanimate objects categorically disprefer object-marking: 0% of them are *-pe*-marked.

Turning now to optional marking on the cross-product of the categories on both axes (the Prominence Lattice), the story becomes less clear cut. In the Animate category, Proper Names have the highest rate of DOM (100%), followed by Specifics (33%), with Nonspecifics (25%) and Definites (21%) bringing up the rear, an ordering that does not reflect the ranking on the DEFINITENESS HIERARCHY. Thus the Animate category fails to exhibit a correlation between definiteness and the frequency of marking.

Similarly, in the Human category, the ranking of definiteness categories in terms of frequency of *-pe*-marking is Specific (100%) > Definite (53%) > Proper Name (20%) > Nonspecific (11%), which is almost the opposite of how these categories are ranked on the DEFINITENESS HIERARCHY. This is not the direct relationship between ranking on the hierarchy and frequency of marking suggested by Aissen (2003b).

Nevertheless, these do not constitute conclusive falsifications of the analysis, since it only definitively predicts the ranking of prominence thresholds, or the cutoff points between optionality and non-optionality in DOM. The above

is variation within the optional range.[10] Furthermore, just as sample size appears to have adversely affected the results for Human Specifics and Animate Proper Names, it could also be coloring the results in other categories (see Table 2), especially Animate Nonspecific (4 examples), Animate Specific (6 examples), and Human Proper Names (5 examples). Thus, some of the inconsistencies with PROMINENCE discussed above could simply be the result of insufficient data, and a larger corpus might bear out the suggestions of Aissen's analysis more evidently.

As a final note, in labeling particular prominence categories as permitting "optional" DOM, I am allowing for two possible, mutually exclusive explanations. The first is that DOM in these categories is truly optional for any given object, and that, for example, when a speaker is presented with a minimal pair of sentences that contain a Human Definite object and vary with respect to *-pe*-marking, both will be acceptable and have similar semantics and pragmatics. The second is that the analysis is overlooking one or more factors, and that while a particular prominence category is not sufficient to predict DOM, *-pe* is nevertheless not optional for any given object since other factors also determine the presence or absence of *-pe*. I leave the question open as to which of these explanations is correct.

In conclusion, the data from Guaraní does not contradict a prominence-based analysis like that of Aissen (2003b), but it does not clearly corroborate it either. While the patterns of DOM do not conclusively point to PROMINENCE as the correct analysis for Guaraní, they do suggest PROMINENCE as a possible analysis; a larger corpus study considering a larger number of factors would allow for a more revealing evaluation of the distribution of DOM within the optional range. Furthermore, certain generalizations are captured by this analysis that must be dealt with in any alternatives — most significantly, the obligatory non-marking of Inanimates, as well as the higher-than-chance likelihood of DOM on Human and Animate objects. Thus, PROMINENCE does not fail, per se, but it does leave something to be desired. In this regard, we now turn to another possible basis for DOM: ambiguity.

## 3.2 Ambiguity-based DOM

Both Gerner (2008) and de Swart (2007) discuss an analysis of DOM in which ambiguity as to which NPs fill which grammatical functions determines object-marking. In this theory, object-marking will appear in clauses that do not other-

[10] This assumes that the categories exhibiting 100% are still probably optional, since this percentage is based on only one instance of each, in which case optionality is impossible to determine since that one example will either be marked or not.

wise supply enough information to link NP to function. This might be the case in Guaraní, a possibility which I now put to the test. Section 3.2.1 outlines the theory of ambiguity-driven differential object marking and its predictions and discusses the methodology of the study. Section 3.2.2 tests these predictions against the actual distribution of *-pe* in the corpus.

### 3.2.1 The Theory and Methodology

In this study, I follow the precedent of Gerner (2008) and de Swart (2007) in adopting the term "ambiguity" to refer to a lack of sufficient information to link NPs to the grammatical functions in a clause. Section 1.1 discussed the tools of GF-assignment in the grammar of Guaraní, showing that A/O ambiguities regularly arise when both arguments are 3rd person. As I said in § 1.1, for the purposes of this study I do not assume word order to disambiguate when assessing ambiguity with respect to grammar, contrary to the claim of Tonhauser and Colijn (to appear).

Beyond grammatical cues, however, listeners make use of other sources of information in the process of parsing grammatical function. One of these sources is THEMATIC FIT, "event-specific world knowledge" (McRae, Spivey-Knowlton, & Tanenhaus, 1998, p. 283) or "general knowledge of how events typically occur in the world" (McKoon & Ratcliffe, 2007, p. 270). Listeners judge the likelihood of an NP serving as the subject of a particular verb based on their general knowledge of the event and its participants. For example, the sentence in (2), reproduced in (68), is grammatically ambiguous in Guaraní; both NPs — *karai* and *pe mitã* — match the 3rd person agreement morphology of the verb.

(68) O-monga-kuaa karai pe mitã.
A3-raise-know gentleman that child
'The gentleman raised the child.'

However, as discussed in § 1.1.2 for this example, "gentleman" fits much better into the agent role of *mongakuaa* — "raise" — than "child" does given world knowledge, since adults generally raise children. Therefore (68) would not be considered ambiguous in terms of THEMATIC FIT.

Another information source is the specific narrative or discourse CONTEXT in which a clause appears, which can influence the parsing of grammatical functions (Buvač, 1996; Christina et al., 2007). This is shown in example (69), where the agreement properties of the verb match both of the available noun phrases and thematic fit doesn't favor one NP over the other as the subject or object of *hecha* — "see."

(69) 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.'

However, in the narrative, the frog has been sneaking up behind Juanito and Pirulo, ergo the frog already sees them but not vice versa. Hence the context strongly favors Juanito and Pirulo as the agents of this sudden seeing event, and (69) would not be considered ambiguous in terms of CONTEXT. Thus, we observe three distinct domains which serve as sources of information in the parsing of grammatical function: (i) GRAMMAR, (ii) THEMATIC FIT, and (iii) CONTEXT.

We now turn to the question of whether DOM in Guaraní is driven by a need to resolve grammatical function ambiguities in one or more of these domains. For instance, is it the case that, in the event of a THEMATIC FIT ambiguity, the object is *-pe*-marked? Furthermore, if ambiguity is indeed relevant to DOM, which of the above domains, or which coordination of domains, conditions it? Thus the success of an ambiguity-based analysis is contingent on properly defining ambiguity with respect to which domains may contribute to it. If DOM in Guaraní is conditioned by ambiguity, there should be a strong correlation between ambiguity in one or more of these domains and *-pe*-marking. Similarly, the objects of non-ambiguous clauses should not realize *-pe*.

To determine whether or not ambiguity motivates DOM, I annotated the same catalogue of 250 clauses that were used when testing PROMINENCE,[11] this time with respect to ambiguity in the grammatical, thematic, and contextual domains. I was then able to evaluate correspondences between ambiguities in each of these domains and the appearance of *-pe*.

As discussed in the introduction, for the purposes of testing I assume a mutual implication:

(70) ambiguity $\Longleftrightarrow$ *-pe*-marking.

Thus all and only those clauses which exhibit opposite values (+/– or –/+) for these features constitute counterexamples to this implication. If counterexamples exist, it is possible that the implication is one way, and that other factors bear on DOM in addition to ambiguity:

(71) ambiguity $\longrightarrow$ *-pe*-marking.

I discuss this possibility in § 3.2.2.

[11] This number excludes clauses with objects that I assume to require or resist *-pe*-marking intrinsically, as discussed in § 3.1.1.

A clause was judged to be grammatically ambiguous when none of the methods of GF-assignment discussed in § 1.1 allowed for a clear assignment of grammatical function to the NPs in the clause,[12] as is the case in (72), which contains 3rd person verb agreement (*o-*) that matches both co-occurring 3rd person NPs (*peteĩ mitã tyre'ỹ* — "an orphan" — and *mymba-kuéra* — "wild animals"). Thus (72) is ambiguous in the domain of GRAMMAR.

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

A clause was judged to be ambiguous with respect to thematic fit if world knowledge about both referents and the event taking place did not significantly favor one thematic arrangement over another. For example, the sentence in (73) is not grammatically ambiguous, since only one overt NP (*hapichápe* — "his colleague") co-occurs with a necessarily transitive verb (*jora* — "untie"), and 3rd person object omission is disallowed in Guaraní. Thus the object function is the only one open for the NP to fill.

(73) O-je-po-kyty kyty ha o-jorá-ma-ne ra'e
A3-REFL-hand-dry dry and A3-untie-COMPL-OPT after.all
**h-apichá-pe**.
**B3-colleague-PE**
'He rubbed his hands and untied **his colleague**.'

However, it *is* ambiguous in the realm of thematic fit, since world knowledge doesn't favor either "his colleague" or "him" as the subject or object of "untie."

A clause was judged to be contextually ambiguous if an understanding of the specific actors, objects, and sequences of events in the narrative did not lend itself to a particular parsing; discourse-new predicates and NPs were thus strong indicators of contextual ambiguity. In this sense, "context" is primarily a residual category comprised of all non-syntactic information that is idiosyncratic to the narrative; i.e., not obtainable from an awareness of the general realities and patterns of behavior in the world. For example, while example (74) is not ambiguous with respect to grammar (2nd person verbal marking with a 3rd person NP) or thematic fit ("you" is much more likely the subject than "flour"), it is nevertheless contextually ambiguous, since both the event ("put") and the object ("flour") are discourse new, and no preceding discourse context clearly bears on their parsing.

[12] This of course does not include *-pe*-marking, the distribution of which this study investigates. Thus the sentence in (72) is determined to be grammatically ambiguous despite *-pe*-marking on the NP *mymbakuéra*, since the question at hand is whether the sentence would be ambiguous without *-pe*, and whether ambiguity is a condition on DOM.

(74) Ha he'i chéve: Kóva-pe re-hó-ta re-moĩ **arína.**
and A3.say B1Sg.IO this-PE A2sg-go-FUT A2-put **flour(Sp)**
'And he told me: "Go and put **flour** into this."'

Determining whether or not ambiguity is the motivating factor behind DOM requires an assessment of how A/O ambiguity would be defined in Guaraní if ambiguity were indeed relevant to DOM. Exactly which domains comprise the correct definition of ambiguity for Guaraní is not obvious *a priori*, which is why in § 3.2.2 I test the mutual implication given in (71) against a variety of definitions of ambiguity with respect to GRAMMAR, THEMATIC FIT, and CONTEXT, as well as the conjunction and disjunction of all or some of those domains. If any definition produces few or no counterexamples to the analysis, this serves as evidence that (i) DOM in Guaraní is ambiguity-driven and (ii) that particular definition of ambiguity is more or less correct. If no definition produces few or no counterexamples, this serves as evidence that ambiguity is not the conditioning factor behind DOM in Guaraní.

### 3.2.2 The Data

In this section I present three sets of tables, with each set corresponding to a particular way of defining ambiguity. Table 3 compares *-pe*-marking with the predictions made by my analysis for each domain individually. Tables 4 and 5 do the same for all possible disjunctions[13] and conjunctions[14] of domains.[15] The individual tables have two axes, each with binary values: the horizontal axis is *-pe*-marking, the vertical one is ambiguity. Each table is labeled with name of the domain(s) in which the feature [± ambiguous] is being evaluated: Table 3.1 presents ambiguity with respect to grammar only, Table 3.2 with respect to thematic fit only, etc. The quadrants of greatest interest are those in the lower left (↙) and upper right (↗) of each table, since each represents an infidelity to the prediction being tested: *-pe*-marking without ambiguity (↙) or ambiguity without *-pe*-marking (↗). The sum of the numbers in each of these quadrants equals the total number of counterexamples found in that domain, which is tallied for convenience in the last row of each table. The total number of clauses being looked at (250), as well as the number of *-pe*-marked (24) and non-*pe*-marked

[13] Disjunction = A or B or both, represented by ⋁.

[14] Conjunction = A and B, represented by ⋀.

[15] When I use the term "conjunction" I refer to the basis for positively rather than negatively identifying ambiguity. That is, when multiple domains are conjoined, a clause is only considered ambiguous if it is [+ ambiguous] with respect to all domains. This implies that [– ambiguous] clauses are [– ambiguous] on the *disjunct* of domains, since a non-ambiguity in any domain is sufficient to disambiguate. The converse holds for my use of the term "disjunction." Thus, when I refer to conjoined or disjoined domains, I am talking about the basis for labeling clauses as [+ ambiguous], not as [– ambiguous].

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 11 | 51 |
| – ambiguous | 13 | 175 |
| Counterexamples | 64 | |

3.1: Grammar (G)

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 22 | 25 |
| – ambiguous | 2 | 201 |
| Counterexamples | 27 | |

3.2: Thematic Fit (TF)

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 10 | 70 |
| – ambiguous | 14 | 156 |
| Counterexamples | 84 | |

3.3: Context (C)

Table 3: *-pe*-Marking and Individual Domains of Ambiguity

(226) objects, remains constant, while the definition of ambiguity varies. The domain(s) with the smallest number of inexplicable counterexamples is straightforwardly the closest to the right definition of ambiguity as it conditions DOM, if indeed it does so.

The data in Table 3 reveals that ambiguity in any single domain is not necessarily an adequate predictor of *-pe*-marking. The strongest candidate is THEMATIC FIT (Table 3.2) with only 27 counter-examples out of 250 total clauses, an 89% rate of success. The weakest is CONTEXT (Table 3.3) with 84 counterexamples (66% success).

The subpar performance of disjunction, as presented in Table 4, appears to take us further from a solution, with prediction rates only barely better than chance at the disjuncts of GRAMMAR $\vee$ CONTEXT and GRAMMAR $\vee$ THEMATIC FIT $\vee$ CONTEXT shown in Tables 4.2 and 4.4 (56% and 53% success, respectively), and with success rates only marginally better at the other two disjunctions. It is important to note the nature of the errors generated by disjoining domains. Tables 4.3 and 4.4 each exhibit 0 counterexamples in the $\swarrow$ quadrant, while Table 4.1 exhibits 1 and Table 4.2 exhibits 11, a small fraction of the 100 in the opposing quadrant ($\nearrow$). The result of defining ambiguity in these ways is that most *-pe*-marked objects appear in + ambiguous clauses. Meanwhile, all 4 tables show very large numbers of counterexamples in the upper right quadrant, meaning that a lack of *-pe*-marking on objects that are nevertheless ambiguous is quite common. In other words, disjunction greatly undergeneralizes (predicts *-pe* on too few objects).

This is not surprising, since disjunction is significantly broader than conjunction as a definition of ambiguity. By disjoining grammar and thematic fit, for example, we allow an ambiguity in *either* grammar *or* thematic fit *or* both to consti-

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 23 | 74 |
| – ambiguous | 1 | 152 |
| Counterexamples | 75 | |

4.1: G ⋁ TF

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 13 | 100 |
| – ambiguous | 11 | 126 |
| Counterexamples | 111 | |

4.2: G ⋁ C

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 24 | 89 |
| – ambiguous | 0 | 137 |
| Counterexamples | 89 | |

4.3: TF ⋁ C

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 24 | 118 |
| – ambiguous | 0 | 108 |
| Counterexamples | 118 | |

4.4: G ⋁ TF ⋁ C

Table 4: Disjunction of Domains:
+Ambiguous Clauses Are Ambiguous in at Least One Domain

tute a proper environment for *-pe*-marking. But this approach is rather counter-intuitive, since an explicitly non-ambiguous syntactic construction would most likely not be overridden by ambiguity or even opposing assignments on the thematic level. For example, the English sentence in (75) is patently non-ambiguous because of rigid word order constraints in the domain of GRAMMAR, despite the strong preference for an opposite reading in the THEMATIC FIT domain.

(75) The dog walked **the owner**.

We would therefore find it surprising if a definition of ambiguity that allowed sentences like (75) to count as ambiguous were a robust predictor of DOM, and the data does not show it to be so. Thus the relevant question, to which we now turn, is whether conjunction significantly *overgeneralizes* or whether it more or less effectively predicts DOM.

In fact, conjunction produces the best results by far, as seen in Table 5. GRAMMAR ⋀ THEMATIC FIT in Table 5.1 scores highest with only 16 counterexamples (94% success), while GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT in Table 5.4 comes in a close second with 18 counterexamples (93% success). Taking a look at which quadrants the counterexamples in Table 5 fall into, we see that conjunction produces a much higher number in the ↙ quadrant than disjunction (Table 4) and a slightly higher number in the ↙ quadrant than single domains (Table 3). Similarly, the ratio of ↙:↗ is much higher in Table 5 than in the other two, so much so that in 3 out of 4 subtables there are actually more items in the ↙ quadrant than in the ↗ quadrant, despite the much larger number of clauses available to appear in the ↗ quadrant. In other words, conjunction results in a greater frequency of non-ambiguous yet *-pe*-marked direct objects. Thus, we see that

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 10 | 2 |
| – ambiguous | 14 | 224 |
| Counterexamples | 16 | |

5.1: G ⋀ TF

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 8 | 21 |
| – ambiguous | 16 | 205 |
| Counterexamples | 37 | |

5.2: G ⋀ C

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 8 | 6 |
| – ambiguous | 16 | 220 |
| Counterexamples | 22 | |

5.3: TF ⋀ C

| | *+pe* | *–pe* |
|---|---|---|
| + ambiguous | 7 | 1 |
| – ambiguous | 17 | 225 |
| Counterexamples | 18 | |

5.4: G ⋀ TF ⋀ C

Table 5: Conjunction of Domains:
+Ambiguous Clauses Are Ambiguous in All Domains

conjunction does overgeneralize as suspected, but to a far lesser degree than disjunction undergeneralizes. Furthermore, conjunction significantly improves upon the outcomes from single domains given in Table 3.

While it may appear that GRAMMAR ⋀ THEMATIC FIT is the best definition of ambiguity since it results in the fewest counterexamples, I contend that GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT is a preferable definition for two reasons: (i) more of the counterexamples under this definition can be accounted for and (ii) it is a simpler definition: it allows all possible sources of disambiguating information to be available to DOM in Guaraní and relieves the analysis of the explanatory burden inherent in excluding only contextual information from consideration by DOM. Regarding (i), GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT has a higher ↙:↗ ratio than GRAMMAR ⋀ THEMATIC FIT, and there is a smaller number of underspecified objects (objects in the ↗ quadrant) for GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT than for GRAMMAR ⋀ THEMATIC FIT. If overspecification is more easily explicable than underspecification, then GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT is the preferable definition. I propose two hypotheses that might explain the redundant appearance of *-pe* in unambiguous clauses.

First, unlike properties of NPs like prominence, ambiguity is a property of clauses that must be calculated throughout the process of production. We would therefore expect some variability in the data, resulting from imperfect assessment by producers as to whether the structures of their sentences allow GF-assignment ambiguity. I hypothesize that speakers of Guaraní use *-pe* in a "cautionary" way, marking some objects in otherwise unambiguous clauses with *-pe* "just in case" when the ambiguity of a clause is more difficult to evaluate in real time. Such a hypothesis would therefore predict relatively few absences of *-pe*

on ambiguous objects but allow for occasional redundant *-pe*-marking on non-ambiguous ones. Furthermore, the latter should appear only on the fringes of ambiguity — i.e., in complex structures but not simple ones, when only one domain disambiguates but not when all do, etc.

Second, it is possible that the redundant appearance of *-pe* on unambiguous objects is the result of a construction analogous to that described by Gerner (2008, p. 301) for Yongren Lolo, in which the object-marker appears "as [a] focus particle to emphasize noun phrase participants." Gerner claims that it is possible for the disambiguating particle $t^hie^{21}$ to appear on objects in unambiguous predicational frames in discourse. When this happens, the purpose of $t^hie^{21}$ is no longer to assign grammatical function, but rather to place the O-argument in contrastive focus. Without the aid of prosody in the Guaraní corpus this hypothesis is difficult to evaluate, but it has the potential to explain redundant *-pe*-marking in Guaraní.

If one or both of these hypotheses is valid, then the ↙ examples can be accounted for. Since there are two ↗ counterexamples under GRAMMAR ⋀ THEMATIC FIT and only one under GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT, the latter definition leaves less data unexplained by the predictions of the analysis and the above two hypotheses. The single ↗ counterexample to GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT is presented in (76) below.

(76) I-ky'a-pá-ite Juán-chi porque tuju-ry
B3-dirty-COMPL-very Juan-DIM because(Sp) mud-juice
o-jagara-pa **la** **ij-ao**.
A3-grab-COMPL **the(Sp)** **B3-cloth**
'Juanito was extremely dirty because the mud completely got ahold of **his clothing**.'

(76) is ambiguous within the grammatical, thematic, and contextual domains. It has 3rd person agreement on the verb with two co-occurring 3rd person NPs, leaving grammatical indeterminacy as to which role each fills. There is no particular world knowledge that can supply information about whether mud grabbing clothes or clothes grabbing mud is more likely; both are anthropomorphisms of inanimate objects equally incapable of "grabbing." Finally, this event, as well as its participants, is discourse-new, rendering very little disambiguating power to the narrative context alone. Under an AMBIGUITY account, it should be marked, yet it is not. However, as discussed above, a single counterexample such as this is not defeating for the analysis, since a small amount of variation is expected. It could simply be that this sentence is ambiguous and the speaker didn't notice it.

Thus, for the purposes of this study I assume GRAMMAR ⋀ THEMATIC FIT ⋀

CONTEXT to be the correct definition of ambiguity to serve as the basis for DOM in Guaraní. The analysis does not hang crucially on this position, however, since the data bears out the predictions of GRAMMAR ⋀ THEMATIC FIT similarly well. Since there is only a difference of two counterexamples between them, the empirical impact of choosing GRAMMAR ⋀ THEMATIC FIT ⋀ CONTEXT over GRAMMAR ⋀ THEMATIC FIT is small. Regardless of the definition chosen, the strong version of AMBIGUITY is falsified by this data: even the best results include counterexamples. However, having settled on the best definition of DOM-related ambiguity, it appears that ambiguity does play a significant role in the distribution DOM in Guaraní: 93% of the data is correctly predicted.

In sum, a weakened version of AMBIGUITY, one which allows non-ambiguity-related factors to affect DOM, has the potential to predict the Guaraní data well. As to what the other factors are, I leave the question open. However, I suggest two possibilities for future research: that there is a tendency to overspecify in Guaraní when computing ambiguity in complex clauses, and that *-pe* is an object focus marker.

## 3.3 Prominence and Ambiguity Analyses Compared

We have seen that neither PROMINENCE nor AMBIGUITY are blatantly falsified by the data. This does not imply, however, that they are equal in explanatory power. In this section, I outline the similarities and differences between the the above two analyses and compare the sets of data they cover or fail to cover.

As the reader may have noticed, there are fundamentally similar criteria underlying both AMBIGUITY and PROMINENCE as analyses of DOM. The former marks objects which are easily construed to be the subject of the clause, while the latter marks highly *subject-like* objects. Thus, an object's prominence contributes to the likelihood of ambiguity (specifically, THEMATIC FIT ambiguity), since prominent objects are subject-like in nature. It is therefore unsurprising that PROMINENCE and AMBIGUITY both predict relatively similar distributions of *-pe*, and that these predictions more or less hold when tested against the corpus. We might even hypothesize that prominence-driven DOM systems are derivative from the need to disambiguate between easily confusable subjects and objects.

In terms of the data they predict, there is significant overlap between PROMINENCE and AMBIGUITY. Inanimate objects strongly disprefer *-pe*-marking, which fits the predictions of PROMINENCE. However, this is predicted by AMBIGUITY as well, since Inanimates have very few, if any, proto-agent properties (Dowty, 1991) and are therefore thematically unambiguous. Human objects prefer *-pe*-marking, which conforms to the predictions of both PROMINENCE and AMBIGU-

ITY, since Human objects are highly prominent and their many agentive qualities result in S/O ambiguities. Animates are relatively indeterminate in and of themselves as predictors of DOM, which is unsurprising from a PROMINENCE perspective because of their mid-level prominence and from an AMBIGUITY perspective because they produce thematic ambiguities with moderate frequency, depending on the specifics of the clause. Thus, both analyses harmonize well with each other.

The difference between the two perspectives lies in which *kind* of subject-likeness is assumed to be relevant. PROMINENCE looks at the object itself, while AMBIGUITY looks at the clause as a whole. Thus while high-prominence objects are *frequently* ambiguous and low-prominence objects are *frequently* unambiguous, they are not always so. It is the performance of each of these analyses at the margins which can set one apart from the other: are low-prominence ambiguous objects marked or unmarked? Likewise, what happens on high-prominence unambiguous objects? In order to explore these questions, I assess ambiguity within the categories of the Prominence Lattice. This allows me to answer two questions: first, within the categories appealed to by PROMINENCE, how effectively does AMBIGUITY predict *-pe*-marking; and second, can the counterexamples to AMBIGUITY be explained according to PROMINENCE? The results are presented in Table 6 on page 52.

This table looks similar to Table 2, with the Prominence Lattice represented by definiteness on the horizontal axis and animacy on the vertical axis. In the first row of each box, I record the number of *-pe*-marked objects on the left and the number of unmarked objects on the right. In the second row of each box, I record the number of ambiguous objects: those with *-pe*-marking on the left and those without it on the right. In order for the predictions of the AMBIGUITY analysis to hold, the number of +*pe* objects in the first row should equal the number of +ambiguous objects below it in the second row, while the number of +ambiguous/−*pe* objects in the second row should be zero. In the third row, I present the percentage of object-marking correctly-predicted by AMBIGUITY, where anything less than 100% represents a contradiction of the AMBIGUITY analysis. All percentages which represent an *incorrect* prediction by AMBIGUITY are shown in boxes in the table.

| Animacy | | | Definiteness: Nonspecific | | | | Definiteness: Specific | | | | Definite | | | | Proper Name | | | | Total | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | +pe | | −pe | | +pe | | −pe | | +pe | | −pe | | +pe | | −pe | | +pe | | −pe | |
| Human | # of objects | | 1 | | 8 | | 1 | | 0 | | 8 | | 7 | | 1 | | 4 | | 11 | | 19 | |
| | # of +ambiguous objects | | 1 | | 0 | | 1 | | 0 | | 1 | | 0 | | 0 | | 0 | | 3 | | 0 | |
| | % predicted by AMBIGUITY | | 100% | | 100% | | 100% | | – | | 13% | | 100% | | 0% | | 100% | | 27% | | 100% | |
| Animate | # of objects | | 1 | | 3 | | 2 | | 4 | | 9 | | 33 | | 1 | | 0 | | 13 | | 40 | |
| | # of +ambiguous objects | | 1 | | 0 | | 1 | | 0 | | 2 | | 0 | | 0 | | 0 | | 4 | | 0 | |
| | % predicted by AMBIGUITY | | 100% | | 100% | | 50% | | 100% | | 22% | | 100% | | 0% | | 100% | | 23% | | 100% | |
| Inanimate | # of objects | | 0 | | 57 | | 0 | | 26 | | 0 | | 84 | | 0 | | 0 | | 0 | | 167 | |
| | # of +ambiguous objects | | 0 | | 0 | | 0 | | 0 | | 0 | | 1 | | 0 | | 0 | | 0 | | 1 | |
| | % predicted by AMBIGUITY | | – | | 100% | | – | | 100% | | – | | 99% | | – | | 100% | | – | | 99% | |
| Total | # of objects | | 2 | | 68 | | 3 | | 30 | | 17 | | 124 | | 2 | | 4 | | 24 | | 226 | |
| | # of +ambiguous objects | | 2 | | 0 | | 2 | | 0 | | 3 | | 1 | | 0 | | 0 | | 7 | | 1 | |
| | % predicted by AMBIGUITY | | 100% | | 100% | | 67% | | 100% | | 18% | | 99% | | 0% | | 100% | | 29% | | 99% | |

Table 6: Ambiguity within the Prominence Categories

Within each category: the number of marked and unmarked examples, the number of ambiguous examples, and the percentage correctly predicted by AMBIGUITY, with percentages in boxes representing incorrect predictions

The results in Table 6 corroborate an AMBIGUITY-based analysis nearly perfectly for all *–pe* DO's. This is predictable from the data in Table 5.4, which shows only one [+ambiguous]/*–pe* counterexample in the entire corpus. It is this item, discussed in § 3.2.2 as (76) and reproduced below as (77), that accounts for the slight imperfections in the predictions of AMBIGUITY for *–pe* in the Inanimate Definite cell, since the example belongs to that type.

(77) I-ky'a-pá-ite Juán-chi porque tuju-ry
B3-dirty-COMPL-very Juan-DIM because(Sp) mud-juice
o-jagara-pa **la** **ij-ao**.
dA3-grab-COMPL **the(Sp)** **B3-cloth**
'Juanito was extremely dirty because the mud completely got ahold of **his clothing**.'

In favor of a PROMINENCE view, this counterexample to the AMBIGUITY perspective is accounted for by PROMINENCE. As discussed in § 3.2.2, AMBIGUITY predicts that the object *la ijao* — "his clothes," should be marked, yet this is not the case. However, *la ijao* — "his clothes" — is an Inanimate DO, which disprefers *-pe*-marking from a PROMINENCE point of view. Thus the latter analysis is preferable in this particular case. However, the very high degree to which AMBIGUITY accounts for the unmarked objects across prominence categories, which PROMINENCE cannot do, offers support for the AMBIGUITY position.

Turning now to the *+pe* examples, as shown in Table 5.4, there are 17 [–ambiguous]/*+pe* counterexamples in the corpus, and Table 6 reveals their distribution across the Prominence Lattice. The ANIMACY HIERARCHY seems to have relatively little bearing on the distribution of these counterexamples: there are no *-pe*-marked Inanimates and they can therefore be neither ambiguous nor unambiguous, and the success of AMBIGUITY in predicting *-pe*-marking on Animate vs. Human objects is almost equal (23% and 27% respectively).

The DEFINITENESS HIERARCHY, on the other hand, displays significant asymmetries in the predictive power of AMBIGUITY. Nonspecific DOM is perfectly predicted by AMBIGUITY, with decreasing effectiveness across the subsequent definiteness categories, as shown by the increasingly frequent occurrence of percentages in boxes when moving from left to right across Table 6, culminating in very poor performance (0% correct prediction) among Proper Name objects.

Thus AMBIGUITY is very powerful in predicting which objects will be unmarked within each category and variably effective in predicting which objects will be marked. Furthermore, this variation appears to be strongly conditioned by definiteness, where there is an inverse relation between rank on the DEFINITENESS HIERARCHY and the predictive effectiveness of AMBIGUITY.

Since AMBIGUITY predicts about 93% of DOM in the corpus (232 out of 250 clauses), while PROMINENCE makes no predictions about any of the data that falls in the optional range (all Animate and Human objects, a total of 83 clauses), AMBIGUITY is preferable to PROMINENCE on the grounds of coverage. However, the importance of definiteness to DOM must be accounted for. One possible explanation is the hypothesis discussed in § 3.2.2: that *-pe*-marking in unambiguous clauses may be similar to the object focus construction described by Gerner (2008) for Yongren Lolo, where a disambiguating particle appears on unambiguous objects only when shifting focus to that object. Since levels on the DEFINITENESS HIERARCHY roughly correspond to levels of focus (see Gundel, Hedberg, & Zacharski, 1993), it is reasonable to suppose that this construction would be far more active among high-definiteness NPs than low-definiteness ones, and that +pe/–ambiguous counterexamples to AMBIGUITY would therefore be more frequent higher on the DEFINITENESS HIERARCHY. This is exactly what we find in Table 6.

AMBIGUITY is also preferable to PROMINENCE in terms of testability. It makes clear claims that any given example may corroborate or falsify: if it's marked, it should be ambiguous, and if it's not marked, it shouldn't. PROMINENCE, on the other, makes clear claims only about the ranking of DOM thresholds: that obligatorily-marked categories should be higher in the Prominence Lattice than optionally-marked ones, and that optionally-marked categories should be higher than obligatorily-unmarked ones. Since a large chunk of Guaraní's Prominence Lattice is optionally marked (83 out of 250 objects), the definitive predictions of PROMINENCE are essentially unfalsifiable within that range; we can only evaluate whether the patterns of DOM are suggestive of a prominence basis or not. Since AMBIGUITY is therefore both more predictive of the distribution of DOM and more testable, it is significantly preferable.

Finally, AMBIGUITY is preferable to PROMINENCE on the basis of simplicity. Rather than assuming two hierarchies and setting them as axes of the Prominence Lattice, then ranking DOM thresholds within this space, AMBIGUITY only assumes a definition of A/O ambiguity, which in Guaraní is quite narrow. Thus, AMBIGUITY makes more accurate, more testable predictions on the basis of fewer and less controversial assumptions than does PROMINENCE, solidifying its preferability as an analysis.

# 4 Conclusion and Suggested Future Work

In this paper, I have presented evidence that Differential Object Marking is robustly active in Guaraní, a fact that has received relatively little attention in the literature on Guaraní and no attention in the literature on DOM. I have attempted to explore the question of the distribution of object-marking in Guaraní, applying two analyses from earlier literature about DOM to a corpus of Guaraní data. The first analysis, similar to that of Aissen (2003b), is based on the notion of PROMINENCE and predicts a direct relation between the level of prominence (subject-likeness) of an object and the likelihood of DOM on that object. The second analysis, similar to that of Gerner (2008), is based on ambiguity resolution and predicts that DOM appears always and only in clauses that would otherwise exhibit an ambiguity as to which NPs fill which grammatical functions. Neither of these analyses were severely undermined by the data in the corpus; the predictions of PROMINENCE held, and AMBIGUITY faced few counterexamples, most of which were explicable. However, when the particular data accounted for by both analyses is compared, as in Table 6, it becomes clear that AMBIGUITY performs better than PROMINENCE in terms of coverage, testability, and simplicity; it accounts for more data through simpler and more easily falsifiable predictions. AMBIGUITY is therefore preferable in the case of Guaraní as an analysis of the distribution of DOM.

I suggest six questions for future research that have been raised by this study. The first is that of the grammatical status of *-pe*, and the resultant place in the structure of objects which are *-pe*-marked. As discussed in §1.5, there has yet been no consensus in the literature as to whether *-pe* is a case-marker or a postposition. I have shown coordination evidence that suggests *-pe* is a postposition, but a more detailed examination is certainly warranted.

The second question is the relationship of *-pe* to other postpositions in Guaraní, especially *-gui* and *-rehe*, which might also mark direct and indirect objects. Tonhauser and Colijn (to appear), in their study of word order in Guaraní, assume some *-rehe*-marked NPs, such as that in example (78), to be objects (see Tonhauser & Colijn, to appear, p. 5-6).

(78) Ha áva-rehe piko rei-pota o-menda? **Pe mitã rú-rehe**!
CONJ who-REHE QU A2.Sg-want A3-marry **that youth father-REHE**
'And who do you want her to have married? **The child's father**!'

For the purposes of this study, NPs of this type were not analyzed, since my focus was on *-pe*-marked objects and unmarked objects. However, it would be interesting to explore the status of such phrases more fully. Specifically, I suggest fu-

ture research that focuses on the distribution of *-rehe-* and *-gui-*marked phrases with respect to the types of verbs with which they co-occur and the types of noun phrases to which they attach, evaluating whether or not these noun phrases constitute objects of some kind.

The third question is whether the kinds of objects which, within the context of the corpus data, appear to require or resist *-pe*-marking inherently really do so.[16] For example, is *-pe*-marking on *mba'e*-only and demonstrative-only DOs ungrammatical/unacceptable, rather than simply rare enough to have never appeared in the corpus? Similarly, is it truly ungrammatical, rather than just uncommon, for the morpheme *chu* to appear independently, without marking of any kind? If so, the assumption of this study is justified, and it is not the responsibility of an analysis of the distribution of DOM to account for the presence/absence of *-pe* in these cases. However, if these constructions, though absent from the corpus, are nevertheless acceptable, it becomes the responsibility of the analysis to explain why they don't occur. Corpus data cannot provide conclusive support for claims of ungrammaticality, and research on this topic will therefore require consultants.

The fourth question is whether *-pe* can serve as an object focus construction, one of the main explanatory hypotheses I propose for the counterexamples to the AMBIGUITY analysis. If it is true that all or most non-ambiguous *-pe*-marked DOs take on special focus properties, such as contrast, this would provide strong evidence that ambiguity is in fact the determining factor behind DOM. If it is not true, this would remove one possible explanation for 18 of the counterexamples to AMBIGUITY, but it would not necessarily undermine the analysis outright.

The fifth question is whether *-pe*-marking is "cautionary" in Guaraní, an overspecification that occurs during the production of clauses in which ambiguity is difficult to assess. If this is the case, it should turn out that redundant *-pe*-marking on DOs becomes increasingly likely the more complex the construction being produced, or the fewer the domains of GF-assignment in which disambiguating information is supplied. Since there were only 25 *-pe*-marked DOs in this corpus, answering this question will require a much larger corpus.

The sixth topic that I pose for future inquiry is that of DOM in other Tupí-Guaraní languages, or Tupí languages in general. Is this a feature isolated to Paraguayan Guaraní? If so, why did it develop there. If not, is DOM in other closely-related languages conditioned by the same factors? For example, as I mentioned earlier, Martins (2003) has already made the claim that AMBIGUITY

[16] These include (i) the object pronoun *chupe*, (ii) Set-B object agreement morphology on the verb, (iii) demonstrative-only NPs, and (iv) NPs consisting only of the wh-word *mba'e*. The reasoning behind their exclusion from consideration in this study is discussed more fully in § 3.1.1.

is the conditioning factor behind *-pe*-marking on objects in the closely-related language of Mbya Guaraní. It would be very interesting to see whether this claim would be born out by testing against a corpus of Mbya Guaraí data as it was for Paraguayan Guaraní. This is by no means restricted to these languages, however, and a wide-angle survey of DOM in Tupí languages in general could yield some very useful results.

# 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.

Bak, J. (2004). *Optional case marking of the possessor in Korean: A study of double accusative marking of possessor/possessee in Korean based on corpus study and optimality computation*. Unpublished master's thesis, University of Manitoba, Winnipeg.

Baker, M. C. (1996). *The polysynthesis parameter*. New York: Oxford University Press.

Bossong, G. (1983-1984). Animacy and markedness in universal grammar. *Glossologia*, *39*(2-3), 7-20.

Bossong, G. (1985). *Empirische universalienforschung: Differentielle objektmarkierung in der neuiranischen sprachen*. Tübingen: Narr.

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* (p. 143-170). Amsterdam: John J. Benjamins Publishing Company.

Brody, J. (1984). Some problems with the concept of basic word order. *Linguistics*, *22*, 711-736.

Buvač, S. (1996). Resolving lexical ambiguity using a formal theory of context. In K. van Deemter & S. Peters (Eds.), *Semantic ambiguity and underspecification* (p. 101-124). Stanford: CSLI Publications.

Choi, H.-W. (1999). *Optimizing structure in context: Scrambling and information structure*. Stanford: CSLI Publications.

Christina, A., Tsapkini, K., Bozikas, V., Karavatos, A., Fokas, K., & Nimatoudis, I. (2007, November 28 - December 2). Differential effects of sentence context on lexical ambiguity resolution in different subgroups of patients with schizophrenia. In *Annals of general psychiatry*. 3rd International Congress on Brain and Behaviour, Thessaloniki, Greece.

Comrie, B. (1989). *Language universals and linguistic typology* (2nd ed.). Chicago: University of Chicago Press.

de Swart, P. (2007). *Cross-linguistic variation in object marking*. Unpublished doctoral dissertation, Radboud Universiteit Nijmegen.

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* (p. 225-233). Somerville: Cascadilla Proceedings Project.

Derbyshire, D. C. (1994). Clause subordination and nominalization in Tupi-Guaranian and Cariban languages. *Revista Latinoamericana de Estudios Etnolingüisticos*, *8*, 179-198.

Dietrich, W. (1994). Word formation, syntax, or noun classification? Tupi-Guarani MBA'E — between lexicon and grammar. *Revista Latinoamericana de Estudios Etnolingüisticos*, *8*, 109-124.

Dixon, R. M. W. (1979). Ergativity. *Language*, *55*, 59-138.
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.
Farkas, D. (1978). Direct and indirect object reduplication in Romanian. In *Papers from the 14th regional meeting of the Chicago Linguistic Society* (p. 88-97). Chicago: Chicago Linguistic Society.
Gair, J. (1970). *Colloquial Sinhalese clause structures.* The Hague: Mouton.
Gerner, M. (2008). Ambiguity-driven differential object marking in Yongren Lolo. *Lingua*, *118*, 296-331.
Gildea, S. (1994). The Proto-Cariban and Tupi-Guarani object nominalizing prefix. *Revista Latinoamericana de Estudios Etnolingüisticos*, *8*, 163-177.
Givón, T. (1978). Definiteness and referentiality. *Universals of Human Language*, *4*, 291-330.
Gregores, E., & Suárez, J. A. (1967). *A description of colloquial Guaraní.* The Hague: Mouton.
Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*, 274-307.
Haspelmath, M. (2006). *Terminology of case.* Oxford.
Hawkins, J. A. (1986). *A comparative typology of English and German: Unifying the constrasts.* London: Croom Helm.
Heusinger, K. von, & Kaiser, G. A. (2003). 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"* (p. 41-65). Konstanz.
Heusinger, K. von, & 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"* (p. 33-70). Konstanz.
Jensen, C. (1998). Comparative study: Tupi-Guaraní. In D. C. Derbyshire & G. K. Pullum (Eds.), *The handbook of Amazonian languages* (Vol. 4, p. 487-618). Berlin: Mouton de Gruyter.
Jensen, C. (1999). Tupí-Guaraní. In R. Dixon & A. Y. Aikhenvald (Eds.), *Amazonian languages.* Cambridge: Cambridge University Press.
Karimi, S. (1999). Specificity effects: Evidence from Persian. *The Linguistic Review*, *16*, 125-141.
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.
Kliffer, M. (1982). Personal *a*, kinesis and individuation. In P. Baldi (Ed.), *Papers from the XII Linguistic Symposium on Romance Languages* (p. 195-216). Amsterdam: John J. Benjamins Publishing Company.
Kuno, S. (1973). *The structure of the Japanese language.* Cambridge: MIT Press.
Lazard, G. (1982). Le morphème r$^{a}$ en Persan et les relations actancielles. *Bulletin de la Société de Linguistique de Paris*, *73*, 177-208.
Leonetti, M. (2004). Specificity and differential object marking in Spanish. *Catalan Journal of Linguistics*, *3*, 75-114.
Lima, A. (2006, October). A minimalist view on differential object marking for specificity. In *The 36th Annual Meeting of the Michigan Linguistics Society.* Rochester.
Mallinson, G., & Blake, B. J. (1981). *Language typology.* Amsterdam: North-Holland Publishing Company.
Martins, M. F. (2003). *Descrição e análise de aspectos da gramática do Guarani Mbyá.* São Paulo: UNICAMP.

McKoon, G., & Ratcliffe, R. (2007). Interactions of meaning and syntax: Implications for models of sentence comprehension. *Journal of Memory and Language*, *56*, 270-290.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283-312.

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* (p. 292-314). Stanford: CSLI Publications.

Næss Å. (2004). What markedness marks: The markedness problem with direct objects. *Lingua*, *114*, 1186-1212.

Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, *62*(1), 56-119.

Rodrigues, A. (1985). Evidence for Tupi-Carib relationships. In H. Klein & L. Stark (Eds.), *South American Indian languages* (p. 371-404). Austin: University of Texas Press.

Sapir, E. (1921). *Language: An introduction to the study of speech.* New York: Harcourt, Brace and Company.

Tonhauser, J. (2006). *The temporal semantics of noun phrases: Evidence from Guaraní.* Unpublished doctoral dissertation, Stanford University.

Tonhauser, J., & Colijn, E. (to appear). Word order in Paraguayan Guaraní. *International Journal of American Linguistics.*

van Everbroeck, E. (2003). Language type frequency and learnability from a connectionist perpective. *Linguistic Typology*, *7*, 1-50.

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 J. Benjamins Publishing Company.

Velázquez-Castillo, M. (1996). *The grammar of possession: Inalienability, incorporation, and possessor ascension in Guaraní.* Amsterdam: John J. Benjamins.

Velázquez-Castillo, M. (2004). Guaraní (Tupí-Guaraǹ). 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, p. 1421-1432). Berlin: Walter de Gruyter.

Weissenrieder, M. (1990). Variable uses of the direct object marker *a*. *Hispania*, *73*(1), 223-231.

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* (p. 369-383). Amsterdam: John J. Benjamins Publishing Company.