Spanish-Origin Verbal and Grammatical Borrowing in Paraguayan Guaraní

Josefina Bittar Prieto

Follow this and additional works at: https://digitalrepository.unm.edu/ling_etds

Part of the Linguistics Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Linguistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
Josefina Bittar Prieto
Candidate

Linguistics
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Rosa Vallejos, Chairperson

Naomi Shin

Jill Morford

Maura Velázquez-Castillo
SPANISH-ORIGIN VERBAL AND GRAMMATICAL BORROWING IN PARAGUAYAN GUARANÍ

by

JOSEFINA BITTAR PRIETO

B.A., Spanish, Universidad Nacional de Asunción, 2011
M.A., Linguistics, University of New Mexico, 2016

DISSERTATION

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

Linguistics

The University of New Mexico
Albuquerque, New Mexico

July, 2021
DEDICATION

To Guille.
May you always ask a lot of questions.
ACKNOWLEDGMENTS

I have been very fortunate to count on the support of many people throughout my dissertation writing process. First, I could not have asked for a better advisor. Dr. Rosa Vallejos has been my toughest critic and my biggest advocate at the same time. She has taught me to double check my ideas while trusting myself. I will be forever grateful for her academic teachings and overall support. I also thank my dissertation committee members, Dr. Jill Morford, Dr. Naomi Shin, and Dr. Maura Velázquez-Castillo, for their insight on my research. Their comments not only improved my dissertation, but also inspired many ideas for future academic endeavors. I am also thankful for the UNM Linguistics faculty, especially Dr. William Croft and Dr. Caroline Smith, who have been very supportive Graduate Advisors. Also in the UNM Linguistics Department, I thank Jessica Slocum, the Graduate Coordinator, who very patiently answered all my questions (I had many) about the forms and paperwork I needed to complete to advance in the program.

My fellow graduate students were there whenever I needed to check an idea, to find a reference, or to have my work proofread. They were also there when I felt like asking random questions or talk about non-academic topics. Thank you, Martin Watkins, Meagan Vigus, Lindsay Morrone, Jens Van Gysel, Ricardo Napoleão de Souza, Keiko Beers, David Páez, Tim Zingler, Desirée Ramírez Urbaneja (I put you all together in a group here, but I recognize that you are all special and wonderful in a unique way). I would also like to thank Jet Saengngoen and Jairo Marshall at the Graduate Resource Center.
The work that this dissertation entailed (data collection, transcriptions, writing) was partially funded by the Latin American and Iberian Institute at the University of New Mexico (LAII). I would like to thank the LAII for this support, which was essential throughout my graduate student career.

I am also lucky to have a big supportive family. I don’t think they know how much they have supported me emotionally, but they have. With every meal and every conversation, they fill my life with joy. Abu, Mami, Ota, Tía Belén, Tío Rubén, Tía Bea, Tío Chilo, Rafa, Pau, Tai, Victoria, Gabriel, Miguel, and Constanza: thank you for our wonderful almuerzos, cenas and meriendas in Asunción! Tía María José, Tío Leandro, Sofía, and Clara, you had the tough task of making me feel like I was home, and you excelled at it, of course. Thank you for your hospitality in San Rafael, California!

Throughout the data collection process, I counted on the help of Israel Pedrozo and Antonio Zena. They both helped me record interviews and introduced me to great people. Antonio also worked with me on transcriptions and translations, and opened me the doors of his parents’ home in San Juan Nepomuceno. I would like to thank Antonio, Doña Rosa, Don Juan, and the entire Zena family for being so welcoming and nice to me.

Finally, I don’t think I can ever thank Tati and Guille (my husband and my son) enough for being by my side unconditionally. They got to see my not-so-pleasant side when working on my dissertation, and they didn’t judge me, nor did they complain. They gave me hugs and love instead. Tati and Guille, God only knows what I’d be without you.
SPANISH-ORIGIN VERBAL AND GRAMMATICAL BORROWING IN PARAGUAYAN GUARANÍ

by

JOSEFINA BITTAR PRIETO

B.A., Spanish, Universidad Nacional de Asunción, 2011
M.A., Linguistics, University of New Mexico, 2016
Ph.D., Linguistics, University of New Mexico, 2021

ABSTRACT

Previous work on language change in contact situations has treated lexical borrowing and grammatical borrowing as discrete phenomena (Appel and Muysken 1987, Thomason 2001, Campbell 2013). Claims that lexical borrowing needs to occur before grammatical borrowing (e.g. Thomason and Kaufmann 1988) or that grammar can be borrowed without any lexicon (e.g. Aikhenvald 2002) are often mentioned in the literature. This dissertation explores the interplay between grammatical constructions (in the sense of Goldberg 2006) and lexical elements in intense contact situations. More specifically, it addresses the following research question: Are grammatical constructions borrowed independently from the elements contained in said constructions? To answer this question, the present research project analyzes the influence of Spanish on Paraguayan Guarani, both languages spoken in South America, which have been in intense contact for five centuries.

The research question is addressed via three empirical studies using a 40-hour corpus of sociolinguistic interviews (Labov 2001). The first study focuses on the
motivations to borrow verbs and the effects that the adopted elements have in the internal organization of the recipient language (see, for instance, Haspelmath 2009). The main finding of this study is that loan verbs coexist with their native-origin counterparts in part because they occur in different grammatical constructions. The second study explores differential object marking, and the factors that favor the occurrence of -pe, a human direct object marker. Results show that younger, more Spanish-dominant speakers, have a higher usage rate of the feature, and that -pe suffixes to loans more frequently than to native-origin nouns. The third study compares the behavior of the polysemous verbal prefix je- among loan verbs and among native-origin verbs. It was found that the primary function of je- is as middle-marker (MM). Even more striking is that more than half of the hosts are loan verbs, and only a quarter of je-prefixed verbs are of native-origin.

Results from the DOM and MM studies suggest that the Guaraní locative postposition -pe and the reflexive marker je- underwent contact-induced grammaticalization (Heine and Kuteva 2003) to become an object marker (LOC > DOM) and a middle marker (REFL > MID), respectively. Both grammaticalization processes were triggered by loans, which entered the language as parts of larger constructions: in the case of –pe, of the [Verb + DOM + HumanOBJ] construction; and in the case of je-, of the [Middle Marker + Verb + Oblique] construction. These borrowed constructions, whose slots were initially favored by loans, became more productive and started recruiting native-origin elements.

In sum, the studies show that lexical and grammatical borrowing are not discrete phenomena. The overall findings support the hypothesis that changes occur in
constructions (Bybee 2015), that terms that are frequently used together in the donor language (constructions) will be borrowed together by the recipient language (Backus 2013), and that grammatical replication originates in everyday communication (Heine and Kuteva 2010). Finally, this dissertation also contributes to expanding our methods to examine contact phenomena involving a major colonial language and a language of indigenous origin.
# TABLE OF CONTENTS

List of Figures ................................................................................................................................xii
List of Tables ..................................................................................................................................xiii

1. **Introduction** .......................................................................................................................... 1
   1.0. Chapter overview.................................................................................................................. 1
   1.1. Theoretical motivations ....................................................................................................... 1
   1.2. Social motivations: Why Paraguay .................................................................................... 3
   1.3. Main research question and hypothesis .............................................................................. 5
   1.4. Structure of the dissertation ............................................................................................... 6
   1.5. Notation and glossing ......................................................................................................... 7

2. **Theoretical Background** ...................................................................................................... 10
   2.0. Chapter overview............................................................................................................... 10
   2.1. Typological characteristics of Guarani and Spanish .......................................................... 10
   2.2. Lexical borrowing .............................................................................................................. 13
   2.3. Grammatical borrowing .................................................................................................... 17
   2.4. Problems with the lexical vs. grammatical borrowing dichotomy ..................................... 21
      2.4.1. Idioms and function words .......................................................................................... 21
      2.4.2. Loan verbs .................................................................................................................. 22
   2.5. Usage-based approach to borrowing ................................................................................ 23
   2.6. Chapter summary ............................................................................................................. 27

3. **Methods** ............................................................................................................................... 29
   3.0. Chapter overview............................................................................................................... 29
   3.1. Data collection .................................................................................................................... 29
      3.1.1. Recruitment of participants ........................................................................................ 30
      3.1.2. The interviews .......................................................................................................... 31
   3.2. Transcription and coding .................................................................................................. 32
   3.3. Chapter summary ............................................................................................................. 35

4. **First Study: Loan verbs** ....................................................................................................... 36
   4.0. Chapter overview............................................................................................................... 36
   4.1. Introduction ....................................................................................................................... 36
   4.2. Research questions and hypothesis ................................................................................... 37
4.3. **Theoretical preliminaries** ........................................................................39
4.3.1. Verbal borrowing: motivations, strategies and outcomes .......................39
4.3.2. Spanish loan verbs in Paraguayan Guaraní .............................................42
4.4. **Methods** .................................................................................................48
4.4.1. Meanings of kuaa ..................................................................................49
4.4.2. Meanings of ko .......................................................................................53
4.4.3. Meanings of nupà ...................................................................................55
4.4.4. Meanings of heja ...................................................................................57
4.4.5. Meanings of rovia .................................................................................58
4.4.6. Meanings of pytyvõ ..............................................................................60
4.5. **Results** .....................................................................................................62
4.5.1. Kuaa/konose .........................................................................................62
4.5.2. Ko/vivi ...................................................................................................67
4.5.3. Nupà/golpea ..........................................................................................70
4.5.4. Heja/deha ..............................................................................................72
4.5.5. Rovia/kree .............................................................................................73
4.5.6. Pytyvõ/ajuda ..........................................................................................75
4.6. **Analysis** ...................................................................................................76
4.6.1. Conventionalization of loans ..................................................................76
4.6.2. Motivations of loans .............................................................................81
4.7. **Discussion** ..............................................................................................84
4.8. **Conclusion** ...............................................................................................85

5. **Second Study: Human Direct Object Marker** ........................................88
5.0. **Chapter overview** ..................................................................................88
5.1. **Introduction** ...........................................................................................88
5.2. **Theoretical background** ........................................................................91
5.2.1. Differential object marking: definition and universal tendencies ...........91
5.2.2. DOM in Spanish ...................................................................................93
5.2.3. DOM in Paraguayan Guaraní and other Tupi-Guaraní languages ...........95
5.2.4. Replica contact-induced grammaticalization .........................................97
5.3. **Research questions and hypotheses** .....................................................98
5.4. **Methods** ................................................................................................101
5.5. **Results** ..................................................................................................109
6. **Third Study: Middle Voice**.................................................................118

   6.0. **Chapter overview**.........................................................................118
   6.1. **Introduction** .................................................................................118
   6.2. **Theoretical background**.................................................................119
       6.2.1. From reflexive to middle ...............................................................119
       6.2.2. The functions of je- ..................................................................122
       6.2.3. The functions of Spanish se .........................................................127
   6.3. **Research question and hypothesis**.................................................130
   6.4. **Methods** ......................................................................................131
   6.5. **Results** .......................................................................................137
       6.5.1. Overall functions of je- (loans and non-loans combined) ..........137
       6.5.2. Functions of je- by verb type (loans vs. non-loans) .................137
       6.5.3. Verbs organized by semantic types ............................................138
       6.5.4. Summary of results ....................................................................153
   6.6. **Analysis** ......................................................................................153
   6.7. **Discussion** ..................................................................................155
   6.8. **Conclusion** ..................................................................................156

7. **Conclusion** .......................................................................................157

   7.0. **Chapter overview** .......................................................................157
   7.1. **Lexical and grammatical borrowing** .............................................157
   7.2. **Other findings** ............................................................................159
   7.3. **Limitations and future research** ....................................................160
   7.4. **Broader impact** ...........................................................................163

8. **Appendix** ........................................................................................164

   8.0. **Gloss** .........................................................................................164
   8.1. **Speakers** ....................................................................................165

9. **References** ......................................................................................167
LIST OF FIGURES

Figure 1. Transcription sample..........................................................33
Figure 2. Corpus spreadsheet.............................................................34
Figure 3. Kuaa vs. konose across meanings.......................................62
Figure 4. Ko vs. vivi across meanings................................................67
Figure 5. Nupã vs. golpea across construction meanings.....................72
Figure 6. Heja vs. deha across construction meanings..........................73
Figure 7. Rovia vs. kree across meanings ..........................................74
Figure 8. Pytyvõ vs. ajuda per token frequency ..................................76
Figure 9. Rate of DOM usage per speaker age, from older to younger........109
Figure 10. Semantic relations among middle and other situation types (Kemmer 1993, 202) .................................................................121
Figure 11. Distribution of je verbs across functions, loans and native-origin verbs combined ..........................................................137
Figure 12. Distribution of je-prefixed native-origin verbs across functions ..........138
Figure 13. Distribution of je-prefixed loan verbs across functions ...............138
Figure 14. Distribution of je verbs across middle semantic domains, loans and native-origin verbs combined ........................................139
Figure 15. Native-origin verbs' distribution of je verbs across middle semantic domains ........................................................................139
Figure 16. Loan verbs' distribution of je verbs across middle semantic domains ....140
LIST OF TABLES

Table 1. Effects of loans: Percentages for coexistence versus replacement..................42
Table 2. Meanings of kuaa in colonial texts and modern data ..................................53
Table 3. Meanings of ko in colonial texts and modern data........................................55
Table 4. Meanings of nupã in colonial texts and modern data .....................................57
Table 5. Meanings of heja in colonial texts and modern data ......................................58
Table 6. Meanings of rovia in colonial texts and spoken data .......................................60
Table 7. Meanings of pytyvõ in colonial texts and modern data ....................................61
Table 8. Percentage of DOM usage according to type of object referent ....................113
Table 9. Percentage of DOM usage with children referents .......................................115
Table 10. List of Guaraní speakers, from youngest to oldest .....................................165
Chapter 1

Introduction

1.0. Chapter overview

The present dissertation is motivated by theoretical debates in the field of contact linguistics, as well as social discussions that are specific to the Guaraní-Spanish bilingualism scenario. This chapter outlines these motivations: the theoretical ones are introduced in Section 1.1, and the social ones are explained in Section 1.2. One of the theoretical discussions in which this research engages is whether or not several items can be borrowed simultaneously as parts of a construction. Section 1.3 addresses this issue, as well as other main questions and hypotheses that have guided this dissertation. Then, Section 1.4 explains the structure of the manuscript; and Section 1.5, the notation and glossing practices that were used in this project, respectively.

1.1. Theoretical motivations

This dissertation deals with the linguistic outcomes of centuries of Guaraní-Spanish contact. While many studies have examined the influences of indigenous languages on Spanish—for instance, Paraguayan Spanish (Penner, Acosta, and Segovia 2012), Andean Spanish (Klee 1996), Amazonian Spanish (Vallejos 2014), Mayan Spanish (Michnowicz 2015), among many others—the impact of Spanish on indigenous languages has not been systematically explored. This dissertation aims to fill this gap. The focus of this study is on the influence that Spanish has had on the morphosyntax of Paraguayan Guaraní. These two languages have been in intense contact for the past five centuries, however, research dedicated to this contact scenario is scarce. Moreover, the
focus of the few studies on the Guarani-Spanish contact has been on Spanish. Thus, this
dissertation is motivated, in part, by the current asymmetry in the study of these two
languages.

Another motivation of this project is the birth of new models to approach contact-
induced change. For the past decades, corpus-based research and variationist methods
have gained traction in the field of contact linguistics, which has further promoted
quantitative explorations of language practices. This quantification helps identify the
nuances in patterns of use that arise from linguistic structure, social context, or both.
Identifying these patterns does better justice to the language users and can reduce
researchers’ bias, as generalizations about the language are posited inductively, that is,
from the tokens produced by the users.

Furthermore, in the past recent years, Cognitive Contact Linguistics has shed new
light on old contact-related debates. Mainly, it has proposed that categories can be seen as
points in a continuum (Backus 2013). For example, it is argued that the distinction
between code-switching and loans is a synchronic one. Conventionalized loans start as
code-switches; and code-switches can become integrated loans. During the process of
integration, loans or code-switches will exhibit different degrees of conventionalization.
Similarly, lexical borrowing and grammatical borrowing are also placed in a continuum,
as the borrowing process can sometimes involve both lexical-like and grammatical-like
elements, as in the case of idioms and other fixed constructions.

Finally, the integration of the tenets of cognitive linguistics with quantitative
variationist methods is a promising avenue in the study of contact phenomena, as it
combines the exploration of the mental mechanisms involved in language change and the
study of how this change spreads in society. It is, therefore, with this spirit of theoretical innovation, that this dissertation has been written.

1.2. Social motivations: Why Paraguay

In addition to the theoretical motivations listed above, the present work has also been motivated by linguistic ideologies surrounding Paraguay’s widespread bilingualism. This South American country with approximately 7 million inhabitants (Paraguay 2017) is a location of interest to study contact phenomena, as about 50% of its population is bilingual in two typologically-different languages, Spanish and Paraguayan Guaraní (Paraguay 2003).

Paraguayan Guaraní (henceforth, Guaraní) is a sociolinguistically atypical Tupi-Guaraní language, as it is not spoken only by indigenous people. It is spoken by more than 80% of Paraguayans, and it is predominant in rural areas (Paraguay 2003). Most Guaraní speakers also speak Spanish. In fact, Penner (2014) proposed that ‘true’ Guaraní monolinguals no longer exist in the country, as even those who do not speak Spanish can understand it, as it is ubiquitous in the media and in educational institutions.

Regarding status, Guaraní has lower overt prestige than Spanish, as it is not a language that facilitates social mobility (Gynan 2011). At the same time, it is the language of national identity (Choi 2003) and it was made co-official with Spanish in 1992, with the new (current) Constitution. Guaraní has also been taught as a second language or used as the language of instruction in all schools, public and private, from first to twelfth grade since the Educational Reform was in force, in 1994. One of the key questions of the implementation of this language in the educational system was the
“kind” of Guarani that was going to be taught. The educational reformers agreed that it should be a variety with the least Spanish influence possible; thus, even today, ‘school Guarani’ regards Spanish borrowings with negativity (Penner 2014).

These purist ideologies are not absent from academia. For example, scholars have described the Spanish loans in Guarani as erasers of native forms or as randomly-occurring across the language. Palacios Alcaine (1999), for instance, claims that “[t]here isn’t a normalization or a systematization that defines this linguistic variety [Present Day Guarani]. [I]t is fundamentally individual, influenced by the communicative situation, the type of addressee or the subject being discussed.” (30) [my translation]. Likewise, Kallfell (2016) proposes that Guarani only exists in its written form, and that the spoken variety is Jopara (also spelled Yopará), as proposed by who “conceive[s] Yopará as a way of speaking with two codes at the same time. […] Guarani [is] relegated almost exclusively to textbooks, manuals, poems […].” (334).

The scarcity of research on spoken Guarani has contributed to the spread of misconceptions about the language. Therefore, it is my wish that more research on spoken Guarani inform language and educational policies in Paraguay, as many of the current policies are based on how speakers should speak instead of how they actually speak. Thus, this dissertation attempts to place Guarani speakers’ linguistic knowledge at the center of the many debates on Paraguay’s linguistic reality.
1.3. Main research question and hypothesis

With Guarani-Spanish contact as a case study, the present work is interested in answering the following main theoretical question:

- Are grammatical constructions borrowed independently of the (lexical) elements contained in said constructions?

The main hypothesis of this dissertation is that grammatical constructions and the lexical elements that contain them can be borrowed simultaneously. To test this hypothesis, three (presumed) contact features in Guarani were selected based on existing previous literature. The first feature is the so-called ‘doublets’ that were described by Palacios Alcaine (1999) and Kalfell (2016): loan verbs with semantically similar native-origin counterparts. The second feature is the human object marking suffix -pe, which has been proposed to be a grammatical loan from Spanish by Bossong (2009) and Shain and Tonhauser (2010). The third feature is the middle marking prefix je-, which has been thoroughly described by Velázquez-Castillo (2007), and whose uses have been proposed as contact-induced by Gómez Rendón (2007). All of these features—verbs, human object markers, and middle markers—involve predicates and, in our view, they are intimately linked to clause-construction phenomena that involve the verbal phrase. Each one of these features is studied independently, but their studies complement each other. The verbs (first feature) are analyzed along with the frequent grammatical elements, such as prefixes, with which they occur; while the grammatical affixes -pe and je- (second and third features) are contrasted with the verbs and nouns (loans and non-loans) to which they are attached. Thus, the research question and hypothesis of this dissertation are
answered and analyzed through the results of three independent but interconnected studies. In turn, the main question and main hypothesis comprise three specific questions:

1. Compared to native-origin verbs, do verb loans appear in certain syntactic constructions more frequently than in other constructions?
2. Does the human object marker -pe, a presumed contact feature, correlate more often with lexical loans?
3. Does the middle marker je-, a presumed contact feature, correlate more often with verb loans?

With respect to the three specific questions, it is hypothesized that:

1. Loan verbs and native-origin verbs will not occur in the same syntactic constructions.
2. The human object marker -pe will occur more often with loan verbs and loan object nouns than with native-origin words.
3. The middle marker je- will occur more often with loan verbs than native-origin verbs.

1.4. Structure of the dissertation

As explained above, the present dissertation is a collection of three studies on the influence of Spanish on morphosyntactic traits of Guaraní. The three studies are based on the same data, and take a usage-based approach to its analysis. Chapter 2 contains an account of the general theoretical background which frames this dissertation, along with a description of contact-origin structural characteristics of Guaraní and Paraguayan
Spanish. Chapter 3 explains the methods for data collection, transcription, and general coding.

With respect to the studies, each one focuses on one aspect of the language, and is fully described in one chapter. The first study, described in Chapter 4, explores verb loans that have a native-origin translation equivalent. The second study, described in Chapter 5, analyzes human object marking as a presumed contact feature. The third study, explained in Chapter 6, compares middle voice marking among loans and native-origin words. All of the above-mentioned chapters (4-6) contain the following study-specific sections: introduction, theoretical background, methods, results, analysis, discussion, and conclusion.

Finally, Chapter 7 summarizes the findings of the three studies, discusses their limitations, and outlines ideas for future research.

1.5. Notation and glossing

The majority of examples provided in this dissertation come from the corpus of spoken Guaraní created for this project. Thus, throughout the dissertation, when I write about ‘the corpus,’ I refer to the data collected for this project, whose creation is described in Chapter 3.

When examples were elicited, that is, those that did not come from the corpus, they are signaled with “(ED)”, for elicited data, in the third line of the gloss, as in (b). All corpus examples are left unmarked, as in (a).
(a) *Nd-a-roviā-i chupe.*

NEG-1-believe-NEG him

‘I don’t believe him.’

(b) *A-rovia o-mano-ha.*

1-believe 3-die-NMLZ

‘I believe she/he/they died.’ (ED)

Some of the written examples, extracted from studies by other authors, had their own glosses, and different orthographic conventions. As each author had different glossing criteria, I have re-glossed some of the examples for consistency, as example (c) (original gloss) and example (d) (my gloss). In some cases, I have referred to the original gloss to discuss the diverse analyses of certain aspects of the language, as with example (c), which analyzes the prefix *je-* as a reflexive (“rf”), while I propose it functions as a middle marker (“MID”).

(c) *re-ye-řesivì*

thou=rf=receive

‘get your degree’

(Gregores and Suárez 1967, 154)
(d) re-je-resivi

2-MID-receive

‘get your degree’
Chapter 2

Theoretical Background

2.0. Chapter overview

The language contact literature has identified two mutually exclusive types of borrowing: lexical and grammatical, based on the premise that, in the language users’ minds, there is a dichotomy between the lexicon and the underlying grammar. This account limits our understanding of Spanish-Guaraní contact phenomena, as previous studies suggest that borrowing processes in this scenario are too complex to frame within binary categories. In Section 2.1 of this chapter, some typological differences between Guaraní and Spanish are described. Then, Sections 2.2 and 2.3 examine the main assumptions with respect to lexical borrowing and grammatical borrowing, and summarize some examples of these phenomena in the Guaraní-Spanish contact scenario. Then, Section 2.4 outlines the problems that the lexicon vs. grammar dichotomy pose for language contact studies. Finally, Section 2.5 explains why a usage-based construction grammar approach to language change can better account for contact phenomena in general, and in the specific contact scenario described in this dissertation.

2.1. Typological characteristics of Guaraní and Spanish

Paraguayan Guaraní is a Tupi-Guaraní language which belongs to the Subgroup 1 branch (Cabral and Rodrigues 2002) or the Guaranian branch (Michael et al. 2015). Both classifications place it closely related to Mbya and Ache Guaraní, which are spoken by indigenous communities in Paraguay. Paraguayan Guaraní is the name given to the Guaraní that is mainly spoken by people who do not identify as indigenous (in this
dissertation we use *Guaraní* for short). Guaraní is an agglutinative language (Estigarribia 2020) with an SVO word order (Gregores and Suarez 1967). However, the language exhibits typological characteristics of SOV languages: it has postpositions, possessor-possessed word order, and has a wider variety of suffixes than prefixes.

One of the most interesting characteristics of Guaraní is its active/inactive system and its cross-referentiality in transitive constructions. This means that certain constructions (in particular, body, emotion and cognition events like ‘cough’, ‘cry’, ‘remember’) are treated as inactive (example (e)); and others, as active (which include verbs like ‘walk, ‘arrive’, ‘break’) (example (f)) (Velázquez-Castillo 2002). In addition, the first person is usually marked in the transitive sentence regardless of whether it functions as a subject (g) or an object (h). Another key Guaraní feature is its lack of non-human third-person direct object pronouns, exemplified in (i).

(e)  
(Che)  *che-rasê*  
I  1-NACT-cry  
‘I cry.’

(Velázquez-Castillo 2002, 142)

(f)  
(Che)  *a-guata.*  
I  1-walk  
‘I walk.’

(Velázquez-Castillo 2002, 138)
As seen from the examples above, grammatical relations in Guaraní are expressed in a typologically unique way. Spanish does not have an active/inactive system nor does it allow for cross-referentiality, and it uses non-human third-person object pronouns. In terms of predicate argument structure, the two languages are quite distant. This concept of typological distance is key in contact studies because typologically different systems are expected to evolve in different ways (Aikhenvald 2002; Thomason 2008). Thus, if the unrelated contact languages show similar patterns—in particular with respect to the linguistic structures that differentiate them—, a contact hypothesis can be proposed. In
Chapter 5 and Chapter 6, I outline the main differences in specific syntactic characteristics in the two languages. I focus on object marking in the first, and on middle marking in the second.

Not surprisingly, the intense contact between Spanish and Guaraní—which are structurally quite different—has been the subject of several contact studies. However, the many inconsistencies in the definitions proposed by contact linguistics hinders even more the study of an already complex contact scenario. In the following sections, some of the definitions will be discussed in light of some studies on the Spanish-Guaraní contact.

2.2. Lexical borrowing

Loanwords are broadly defined as “word[s] that at some point in the history of a language entered its lexicon as a result of borrowing” (Haspelmath 2009, 36). Some key characteristics that have been proposed for loanwords are that 1. they are monomorphemic in the recipient language and, 2. they are morphologically and phonologically adapted into the recipient language (Haspelmath 2009; Campbell 2013).

This phonological criterion posits some problems to identify loans in Guaraní, as with increasing bilingualism, adaptation is not as evident. For example, old loans like Kirito from Spanish Cristo (‘Christ’) show vowel insertions that are not seen in what can be called ‘newer’ loans, such as ekrivi from Spanish escribir (‘write’). Overall, Spanish loans show different degrees of phonological adaptation (Pinta and Smith 2017).

The single morpheme criterion has also posed some problems, as certain words appear to be borrowed and treated as polymorphemic, such as the loan brasileros in example (j):
Brazilian talk many times.

(Gómez Rendón 2007, 528)

Gómez Rendón claims that the example above shows a case of double plural marking, and that it is the Spanish plural marker -s that triggers the use of the Guaraní optional plural marker *kuéra*. In other words, he does not analyze *brasileros* (‘Brazilian’ (PL)) as a monomorphemic stem, instead, he treats it as a bimorphemic loan.

Thus, phonological and morphological criteria alone cannot decide whether an item is a loan, a fully integrated member of the lexicon, or if it behaves more like a code-switch (the bilingual Guarani speaker switched to Spanish to produce the word *brasileros*). This code-switching vs. borrowing debate is, of course, not specific to the Guarani case. As Winford (2007) puts it: “there is still no agreement on what constitutes code-switching as opposed to borrowing” (24). In Section 2.5, an account on how usage-based approaches can help navigate this dichotomy is presented.

Despite the many issues in contact phenomena classification, thanks to quantitative and typological methods, crosslinguistic research on lexical borrowing is advancing in interesting directions. Recently, the most comprehensive study on lexical borrowing was released: the Loanword Typology Project (LWT) (Tadmor and Haspelmath 2009). This project compared data from 41 typologically different languages, and proposed: a Basic Vocabulary list, a data-driven word class borrowability hierarchy, and a semantic domains hierarchy. The Basic Vocabulary list (known as the Leipzig-Jakarta List of Basic Vocabulary) is a 100-item list of words that resist borrowing, similar
to the Swadesh list of basic vocabulary, that was created with statistical methods. Researchers found that the most borrowing-resistant words are ‘fire’, ‘nose’, ‘to go’, ‘water’, and ‘mouth’. With respect to borrowability, the Project provided further evidence for the claim that nouns are more likely to be borrowed than verbs, as proposed by Thomason and Kaufman (1988). Concerning semantics, the Project supported long-held views that cultural words (e.g. religion lexicon) are more likely to be borrowed than non-cultural words (e.g. perception lexicon) (Thomason and Kaufman 1988).

To my knowledge, there are no exhaustive studies on lexical borrowing in Guarani that could be contrasted with the results from the LWT Project. Some studies include lists of loanwords (e.g. Morínigo 1931; Estigarribia 2020), but they do not discuss their context nor frequency of use. Other studies (Matras 2007; Gómez Rendón 2020) provide some numbers with respect to loans but do not list the specific words. The studies on loanwords in Guarani are further discussed in Chapter 4, Section 3.

Regarding the motivation for lexical borrowing, most authors agree that there are two main reasons why language users borrow words: need and prestige (e.g. Campbell 2013; Tadmor, Haspelmath, and Taylor 2010). While need-motivated borrowing presupposes the lack of a term to convey a specific meaning in the recipient language, prestige-motivated borrowing presupposes the existence of an equivalent term in the recipient language. However, since different languages categorize events and objects differently, it is not easy to decide whether terms in two different languages should be considered equivalent. Otheguy and García (1993), for example, have found that English Easter and Spanish Pascua, although apparently semantically equivalent, were not used interchangeably by speakers of US Spanish, as the English loan Easter referred to the
American version of the holiday, and the native Pascua referred to the Latin American version. They also found similar results for other word pairs, such as comedor/lunchroom, and edificio/building.

Some scholars have argued that the effects a loan has on the lexicon of the recipient language depends on the motivations of its borrowing. The LWT Project proposes that when a loan is motivated by need, it becomes an addition to the recipient language lexicon (insertion). When the loan is not motivated by need, there are two possible outcomes: the loan will either replace the native-origin word or it will coexist with it (Tadmor and Haspelmath 2009). The first outcome, replacement, has also been described as lexical loss (Muysken 2001). The second outcome, coexistence, appears to be more common across languages (Tadmor and Haspelmath 2009) (English has many examples; for instance, native-origin deep and Latin-origin profound). Because borrowing effects are tied to borrowing motivations, determining a loan’s outcome—insertion, replacement, or coexistence—is a difficult task. As previously mentioned, even when the loan and the native-origin counterpart appear to be semantically identical, a close look at the data might indicate otherwise. Furthermore, from a diachronic perspective, it is expected that all words, loans and non-loans, will change and develop collocation preferences. With respect to the outcomes of loans from Spanish, Estigarribia (2020) proposes that Guaraní is undergoing relexification, similar to that in creole languages. Other researchers, like Gómez Rendón (2020) make a similar claim: the high number of Spanish loans in Guaraní has dramatically changed the lexicon of the language. However, exhaustive studies on motivation and outcomes on Spanish-origin
loans in Guarani (ones that account for frequency of use, context, and variation across speakers) are needed to provide evidence for such extreme processes.

### 2.3. Grammatical borrowing

Grammatical borrowing has been defined as the “change in which a foreign syntactic pattern […] is incorporated into the borrowing language through the influence of a donor pattern found in a contact language” (Harris and Campbell 1995, 122). Compared to lexical borrowing, grammatical borrowing is harder to prove, as languages can borrow the function of the grammatical feature but not its form (Matras and Sakel 2007a). When form is not borrowed, grammatical changes can be due to language-internal factors. Therefore, identifying and proving grammatical borrowing is a challenge for contact researchers, including those who focus on Guarani and Spanish.

One of the main discussions on grammatical borrowing has centered on which features can be borrowed and which ones cannot. On one end of the spectrum, Appel and Muysken (1987) claim that only very salient grammatical patterns can be imitated, and that contact-induced grammatical change is rather superficial. On the other end, Thomason and Kaufman (1988) state that contact-induced grammatical changes can be deep, that is, depending on the intensity of the contact, anything can be borrowed. Slight contact can lead to the borrowing of function words or syntactic features that do not cause typological change; moderate contact can lead to borrowing of adpositions. Intense and very intense contact can lead to greater structural changes, such as changes in word order, word structure rules, and borrowing of inflectional morphology.
Studies on Guaraní-Spanish contact seem to favor Thomason and Kauffman’s point of view, as both languages show similar structural traits that cannot be fully explained by internal factors alone. For example, Choi (2000) documents the omission of non-human direct object pronouns (at a 90-100% rate) among speakers of Paraguayan Spanish, which could be due—in part—to the substratum effect of Guaraní, which lacks these pronouns. Similarly, Velázquez-Castillo and Henderson (2013) show that the preposition *para* in Paraguayan Spanish has grammaticalized into an aspectual marker due to contact with Guaraní. With respect to Guaraní, there are very few studies that are solely dedicated to the thorough exploration of a specific grammatical feature. Shain and Tonhauser’s (2010) article is one of the few feature-specific studies, as it focuses exclusively on Guaraní differential object marking as a product of contact. Overall, however, research on grammatical borrowing has mainly focused on listing possible contact characteristics. Gómez Rendón (2007) has provided such a list, which includes, but is not limited to, the marking of possession, number, and case. Among the proposed contact features that are relevant to this dissertation are the marking of human direct objects (as studied by Shain and Tonhauser 2010) (example (k)), and the ‘Spanish-like’ use of the so-called reflexive *je-* as in (l):

\[(k)\quad Pe\quad aluno-pe\quad o-ñe-mbo’e-va’erã\quad pe\quad i-lengua-materna-pe\]

Those students-OBJ 3-PASS-teach-DEO that 3-tongue-mother-LOC

‘Students have to be taught in their own language.’

Gómez Rendón (2007, 533)
Then also farmers 3-REFL-CAUS-strong

‘Then, also, farmers got stronger.’

Gómez Rendón (2007, 537)

With respect to example (k), it is proposed that the object marker -pe replicates Spanish human object marker a, and with respect to example (l) it is claimed that ſe (oral allomorph of je-) replicates Spanish clitic se (Gómez Rendón 2007).

It is not surprising that both lexical and grammatical borrowing have been documented in the intense Spanish-Guaraní contact scenario, as many scholars agree that contact-induced grammatical change requires a higher degree of contact than lexical borrowing. Some researchers would argue that in this or any other contact situation, lexical borrowing happened first. For example, Thomason and Kaufman (1988) propose that “in a borrowing situation the first foreign elements to enter the borrowing language are words” (37). Moravcsik (1978) goes further and claims that [n]o non-lexical language property can be borrowed unless the borrowing language already includes borrowed lexical items from the same source language” (110). However, there are no studies on Guarani which prove that loanwords were introduced before contact-induced grammatical changes. Moreover, the literature offers many examples that contradict the “lexical before grammatical” claim, as some languages exhibit grammatical borrowing without lexical borrowing. In their study of languages in contact in India, Gumperz and Wilson (1971) find very similar grammatical systems across historically and typologically unrelated languages. They claim that “[w]hile language distinctions are
maintained, actual messages show word-for-word or morph-for-morph translatability, and
speakers can therefore switch from one code to another with a minimum additional
learning” (165). The authors call this process of two languages becoming grammatically
similar convergence. Johanson (2008) also claims that many cases of convergence, or
Selective Grammatical Copying, “serve isomorphism by creating convenient translation
equivalents in the interacting codes (77).” In South America, Aikhenvald (2002) has
found that, in the Amazon, due to linguistic exogamy, unrelated languages share many
grammatical characteristics (which could be the product of contact), but they do not show
lexical borrowing. Of course, grammatical borrowing without loanwords is not the case
of Guaraní, as lexical borrowing is amply documented in the language. However, to
understand the nature of grammatical change and to test the convergence hypothesis,
more research needs to be conducted. For example, in the US, Silva-Corvalan (1994) has
claimed that dialects of Spanish are converging with English to reduce the cognitive load
of the bilingual speakers. However, recent studies on bilingualism have shown that
bilingual linguistic practices in the US, like code-switching, do not lead to Spanish-
English convergence, which call in question the ‘cognitive load’ hypothesis (Torres
Cacoullos and Travis 2018). Thus, whether or not grammatical change in Paraguayan
Spanish and Guarani is facilitated by code-switching and caused by the cognitive load of
bilingual speakers is a topic yet to be studied.

Most importantly, claims that languages have to borrow loanwords before they
can borrow grammatical features, or that grammatical changes are independent from the
lexicon are based on the assumption that the lexicon and the grammar are discrete
entities. Because of the intense contact nature of Spanish and Guaraní and for the
problems of the lexicon-grammar dichotomy (explained in the following section), this
dissertation examines contact phenomena by exploring the lexicon and the grammar
simultaneously.

2.4. Problems with the lexical vs. grammatical borrowing dichotomy

2.4.1. Idioms and function words

Croft and Cruse (2004) claim that componential approaches to language structure
have trouble classifying idiomatic expressions: are they part of the grammar, the lexicon,
or both? This problem is particularly relevant for language contact studies because idioms
are highly borrowable across languages. Ross (2007) claims that “lexical collocations are
among the most dramatic witnesses to calquing [but, i]nterestingly, the grammar–lexicon
model has nowhere to put them, ignoring the huge role that they obviously play in a
language” (136).

The lexicon vs. grammar approach is also inconsistent in the classification of
function words. When borrowed, these words are sometimes considered lexical loans and
sometimes grammatical loans. For some scholars, “words with grammatical meaning”
(Tadmor, Haspelmath, and Taylor 2010, 231), like articles and pronouns are lexical
loans; for others, they are grammatical loans (e.g. Matras 2007). For example, depending
on the approach one follows, one could theoretically classify the definite article *la*,
borrowed from Spanish into Paraguayan Guaraní, as a lexical loan (because it is a free
morpheme, a word), or as a grammatical loan (because it is a word with a grammatical
function).
The inconsistency in the classification of idioms and function words has also impacted the borrowing hierarchies proposed by the literature, because many of them place form, rather than function, at its core. For example, Thomason and Kauffman (1988) and Bakker, Gómez Rendón, and Hekking (2008) claim that lexical elements are borrowed before grammatical elements, and that free forms are borrowed before bound forms. These hierarchies can be problematic, as they impede “straightforward comparability [because of] the vagueness with which category labels are used” (Matras 2007, 35). That is, even within a language, there is disagreement in what is considered a lexical item or a grammatical one, a bound form or a free form. Thus, when dealing with two languages, the comparison becomes even more complicated and inconsistent.

2.4.2. Loan verbs

Verbs are challenging for the lexical vs. grammatical borrowing dichotomy because, while they are traditionally considered content words, like nouns, when borrowed, they are not always directly inserted into utterances. Unlike loan nouns, languages use different strategies to 'accommodate' borrowed verbs (Wichmann and Wohlgemuth 2008). This syntactic rearrangement of utterances containing loan verbs should be expected, because, as Matras (2007) claims: when borrowed, verbs “serve as a referential lexical item—a content word, not dissimilar to a noun, adjective, or descriptive adverb […] and] initiate the predication […] which] serve as the principal anchor point for the entire proposition of the utterance (48).” The complexity of the verbal borrowing process suggests that the study of loan verbs does not benefit from a lexical-only perspective. However, not only are there very few studies dedicated to an in-
depth exploration of loan verbs (Tadmor, Haspelmath, and Taylor 2010), but there are also virtually no holistic analyses of constructions which include borrowed verbs.

To date, the only extensive study of verbal borrowing within a language is Mifsud's (1995) *Loan Verbs in Maltese*. Mifsud (1995) found that the incorporation of loans from morphologically-different languages—Arabic, Italo-Romance, and English—triggered changes in the morphology of Semitic Maltese. This effect of loans on morphology provides more evidence that loan verbs should be studied in the syntactic constructions in which they occur. Furthermore, recently, a study on French-origin psych verbs in English, like *please*, showed that while borrowing the verbs, Middle English also “copied the pronominal and nominal ‘dative’ argument of the foreign model code (Old French)” (Trips and Stein 2019).

### 2.5. Usage-based approach to borrowing

Contrary to hierarchies based on form (e.g. bound vs. unbound morphemes), Matras (2007) has placed function at the center of language contact studies. With respect to borrowing hierarchies, he has claimed, for example, that in the modality domain, obligation is borrowed more often than necessity, regardless of whether it is conveyed through free or bound morphemes, and of whether its borrowing counts as lexical or grammatical. In addition, Matras and Sakel (2007a) claim that bilingual speakers’ mental processes are at the sources of grammatical changes, because in some contexts they need to exploit their entire communicative repertoire, instead of adhering to the constraints of one of their languages.
To advance the concept of function at the center of the borrowing process, Matras and Sakel (2007a) propose the distinction between matter (MAT) and pattern (PAT) replication: A language can borrow form (matter) and structure or meaning (pattern), simultaneously or independently of each other. This distinction applies to all structural levels, that is, it comprises lexical and grammatical borrowing. In the syntactic level, a structural pattern can be replicated in the recipient language with matter (forms) or without it (function-only). In cases where only pattern is replicated, Matras and Sakel (2007b) claim that the recipient language users avoid the replication of matter, and “replicate the abstract organisational pattern of the model construction using suitable elements in the replica language” (830). Both of these types of replication can be seen in Paraguayan Spanish. Paraguayan Spanish has borrowed grammatical particles from Guaraní with both form (matter) and pattern. In example (m), the polite request particle -na is suffixing the verb. The language has also borrowed the pattern of the emphatic voi but not its matter, as it is the Spanish adverb luego which conveys the meaning, as seen in example (n). (however, borrowing of pattern and matter of voi has also been documented (Penner, Acosta, and Segovia 2012)).

Replication of pattern with matter:

(m) \textit{Tra-é-me-na eso.} \\
\hspace{1cm} \text{bring-IMPR-1OBJ-REQ that} \\
\hspace{1cm} \text{‘Bring me that, please.’ (ED)}
Replication of pattern without matter:

(n) \textit{Sab-ia luego.}

\textit{know-1.IMPF.PAST EMPH}

‘I knew it, really.’ / ‘I already knew it.’ (ED)

Matras and Sakel (2007b)’s model points at the \textit{whole-part} approach to language knowledge proposed by constructionist approaches to grammar, which defy the lexicon vs. grammar dichotomy (e.g. Croft and Cruse, 2004). These approaches propose that the construction (the ‘whole’, the utterance) is the basic unit, and the elements are its components (‘the parts’, words, affixes, etc.) (Croft and Cruse, 2004). According to Goldberg (2006), constructions have various degrees of schematization, that is, certain constructions allow for more items to fit their slots than others. The English Ditransitive is a highly schematized construction, as it has a fixed word order and it only allows for transfer verbs like ‘give’. In turn, Bybee (2007) argues that the more frequently used each construction is, the more entrenched it will be in the language users’ memory.

Summarizing, elements that are frequently used together—regardless of their function—form constructions and are stored together, as schemas, in the users’ minds.

In addition, this \textit{whole-part} view of the cognitive representation of language is opposed to the rule-based approach to grammar, which treats linguistic elements as building blocks that combine, following rules, to form larger units (\textit{part-whole} approach) (Croft and Cruse, 2004). In the latter approach, levels of linguistic structure—phonology, syntax, and semantics—are stored independently in the language users’ minds. For the constructionist approach, the users’ linguistic knowledge is constructional, and it
originates in linguistic usage. The usage-based approach places levels of linguistic structure in a continuum, based on evidence from language change (e.g. with frequent usage, words can become affixes, changing their semantics and phonology) (Bybee 2015). Furthermore, usage-based linguists challenge the notion of “parts of speech.” Langacker (2008), for instance, claims that “it is only in the context of a higher-level grammatical construction [...] that a lexeme takes on noun-like or verb-like properties, [challenging the notion that] particular lexemes are learned and stored specifically as nouns or as verbs” (96). Likewise, Croft (2001) has claimed that parts of speech, or word classes are language specific and, therefore, an approach based on this classification hinders cross-linguistic studies. This criticism is crucial for language contact studies, because contact phenomena require the comparison of at least two languages.

The application of functional cognitive approaches to the study of contact phenomena is relatively recent. According to Zenner, Backus, and Winter-Froemel (2019), a usage-based account of language contact, in particular a construction grammar approach, will help us understand contact phenomena as a continuum. In the case of borrowing, a constructionist approach can solve the problems that arise from the lexical vs. grammatical view, such as the place of function words or idiomatic expressions. As Backus (2013) summarizes: “one may hypothesize that the mechanisms of borrowing are the same for specific units (i.e. words and expressions), schematic units (i.e. grammatical patterns) and partially schematic units (i.e. constructions in the sense of Construction Grammar)” (34).

Usage-based accounts can also reconcile the concepts of code-switching vs. borrowing, as they are conceived as belonging to a conventionalization continuum.
Loanwords that are integrated in the language were first instances of code-switching; and along their process of integration, they exhibit different degrees of use in the community (Zenner, Backus, and Winter-Froemel 2019). Then, loanwords do not need to be defined in terms of phonological adaptation or morphological integration, but in terms of rate of use in the communities of the recipient language.

Finally, usage-based accounts of language change have recently been expanded by variationist methods. Exploring the different ways diverse people use the same language can help determine how conventionalized a specific linguistic item is, and disprove assumptions about language within a community. In contact scenarios, for example, Meyerhoff (2017) has shown that younger speakers of Nkep do not borrow more lexical items from Bislama than their older peers, which is a strong belief held by the community. Similarly, in a study of Paraguayan Guaraní, I showed that verbal borrowing was relatively stable across generations (Bittar 2016). Moreover, loans coexisted with the native-origin counterpart, and appeared to be developing specific uses.

### 2.6. Chapter summary

Traditional accounts of contact phenomena have classified borrowing into two mutually exclusive categories: lexical borrowing, and grammatical borrowing. While lexical borrowing concerns loanwords, grammatical borrowing concerns the replication of grammatical features. This classification is problematic mainly because the literature has difficulties deciding what counts as grammar and what counts as lexicon, and because fixed expressions, which are highly borrowable, usually are a combination of grammatical-like and lexical-like elements. Moreover, the borrowing of verbs—
traditionally considered lexical words—usually require syntactic accommodations from the recipient language, that is, they cannot be studied as independent lexical-only loans.

The lexicon-grammar dichotomy framework, and other binary distinctions, fall short to study an intense contact scenario like Guaraní-Spanish bilingualism in Paraguay. Previous work on the languages and preliminary explorations of the data suggest that speakers of Guaraní and Spanish have used (and continue to use) a wide range of strategies to incorporate elements from one language to the other, elements that are not limited to a lexical-only or grammatical-only level.

In the following chapter, I outline the data collection and coding methods that were used in this study, which aimed at reducing both the researcher’s and the participants’ internalized prejudice about the way Guaraní is spoken. Although Paraguay is an ideal scenario to explore contact phenomena, linguistic ideologies among speakers (which are sometimes reproduced by researchers) make it difficult to carry out an unbiased analysis of the language practices of bilingual, especially Guaraní-dominant, speakers.
Chapter 3

Methods

3.0. Chapter overview

This chapter goes over the data collection process of the dissertation, which consisted mainly of recording interviews with speakers from diverse social demographics (Section 3.1). Then, Section 3.2 describes the transcription of the interviews, and their organization and coding. These steps were completed before starting the three studies conducted for this dissertation. As previously mentioned, each study has a chapter dedicated to it. In each study chapter, there is a methods section, which contains the methodological details that are specific to that study.

3.1. Data collection

Following the tenets of usage-based approaches to language change, and informed by variationist methods, the data used in the present dissertation consists of sociolinguistic interviews. These interviews were recorded by the author in two sets of sessions: the first one, in June 2015, in urban areas of Paraguay; the second one, in October 2019 and January 2020, in rural areas of the country. All the speakers were recruited with the snowball sampling method and were self-assessed native speakers of Guaraní. Only two of the speakers identified as Guaraní monolinguals; the rest identified as Guaraní-Spanish bilinguals. Speakers were not offered any compensation for their participation in the study. A total of 55 interviews were collected.
3.1.1. Recruitment of participants

The rural and urban divide is a frequent topic in non-academic discussions on Guaraní. When I commented on my research project, both academic and non-academic people would tell me to go to the countryside, where “the real Guaraní” is spoken, as urban Guaraní is “contaminated” with Spanish. This divide has been questioned by Penner (2014), who has claimed that Spanish is so pervasive on the media that its influence extends beyond the city. This discrepancy between language attitudes and actual language use pointed at the importance of documenting the speech of both urban and rural Guaraní speakers. Thus, interviews were collected in two locations.

Most of the urban interviews were recorded in the Bañado Sur area of Asunción, the capital of Paraguay. Bañado Sur was chosen because it is a Guaraní-dominant, working-class area of the city. It is also where the main interviewer, Israel Pedrozo, grew up and worked at the time of the recordings. Israel is a self-assessed native Guaraní speaker. As he is a well-known person in the area, Israel recruited participants while we walked around the neighborhood. Prospective participants were told they were going to be interviewed for a project about how people use language in Paraguay, that the conversations were about daily life and topics of the interviewees’ interests, and that they could choose to speak in Guaraní, Spanish, or both. After talking to them for the first time, some people agreed to be interviewed on the same day, some people told us to come back the next day. Thus, participants were chosen solely on their time availability and willingness to talk to us.

Participants in rural areas were recruited in the same way as participants in urban areas. Rural speakers were recruited in the San Juan Nepomuceno area. San Juan is a
town located 200 kilometers from Asunción, it is a Guaraní-dominant area, and has an urban downtown but is mainly comprised of rural communities. The town was chosen because of my friendship and work relationship with a San Juan native, Antonio Zena. Antonio is a self-assessed native Guaraní speaker. About three months before recording the interviews, Antonio and I were working on transcribing the urban interviews. When I told him I had to collect similar interviews in rural areas, Antonio suggested to go to his home town and talk to his acquaintances. When in San Juan, Antonio would contact people over the phone. Just like urban speakers, people were told they were participating in a study about language use in Paraguay. If they agreed to be interviewed, we would meet participants in their houses. Sometimes, after finishing an interview, the participant would tell us to go knock on a specific neighbor’s door, because they thought they would like to be interviewed as well. Antonio knew most of the people we interviewed.

3.1.2. The interviews

The interviews aimed at eliciting the colloquial style of the speaker and thus followed the interviewing practices proposed by (Labov 2001). The interviewee was asked basic demographic questions at the beginning, then they spoke about a topic of their interest with little to no intervention from the interviewers. There were some common topics across many interviews: childhood memories, school life, love life, child rearing, and politics. In the urban interviews, Israel Pedrozo was the main interviewer; in the rural areas, it was Antonio Zena. I, a heritage speaker of Guaraní, was present in the interviews, handling equipment at the beginning of the interview and asking questions towards the end. In some interviews, family or community members were present, mainly
listening, but sometimes intervening from time to time: this aided in the prompting of the interviewee’s casual speech style.

The focus on natural speech stems from the usage-based claim that everyday communication is at the core of language change (Bybee 2015). Also, as prescriptivist ideologies are commonplace in the official discourse on Guaraní (Penner 2014), more restricted activities, like elicitation tasks, would likely prime a more careful style among participants.

3.2. Transcription and coding

Of the 55 interviews that were recorded in Guaraní, 40 were selected for the corpus of this dissertation: 20 urban interviews and 20 rural interviews. On one hand, the urban interviews were selected from a larger corpus of 35 interviews. The main criterion in the selection was interviewer/interviewee speaking time ratio (the longer the interviewee speaking time, the more likely to be selected). Demographic balance was the other criterion: the selection includes 10 female and 10 male speakers, from a wide range of ages: 18 to 75. On the other hand, all the 20 rural interviews were recorded for the purpose of this dissertation and were therefore included: 10 speakers were female, and 10 were male; from an age range of 18 to 85. (See Appendix for details).

Then, the 40 interviews were fully transcribed in the software Transcriber 1.5.2 by native Guaraní speaker Antonio Zena, and myself. This software is user-friendly, and allows for an audio file and text file pairing. The text file can be easily exported to spreadsheets and other types of files. An example of what our work in Transcriber
looked like can be seen in Figure 1. The interviews transcribed by me were verified by Mr. Zena. We used Guarani’s standardized orthography.

*Figure 1. Transcription sample*

![Transcription sample](image)

Once transcriptions were ready, they were combined into a single *Excel* spreadsheet, which included demographic information of the speaker: pseudonym, age, sex, and location, as can be seen in Figure 2. Some discrepancies in the orthography were unified as well.
Figure 2. Corpus spreadsheet

<table>
<thead>
<tr>
<th>speaker</th>
<th>other people</th>
<th>sex</th>
<th>age</th>
<th>location</th>
<th>sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sebastian</td>
<td>P</td>
<td>Male</td>
<td>72</td>
<td>R</td>
<td>umano, komo kristiano ajeande</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ha, che upéa la ajhui laa, entre familia upea laa, ko’anga laa, la hente idiferensia tshaitekere pe ymaguare oja, porke yma orrepetai la hapichape, ikakuaaguüme</td>
</tr>
<tr>
<td>Sebastian</td>
<td>P</td>
<td>Male</td>
<td>72</td>
<td>R</td>
<td>hie haa, micht he’i haa, ymaguarenteko umia, okhekanipamhe he’i, ootakaove chupekueria isykyuera, ymaguarema umia eheja</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>lei laa, ipohýí porke la ne familia, neenguepotiriaari, porke nderepojokári hese pea amoa,</td>
</tr>
<tr>
<td>Sebastian</td>
<td>P</td>
<td>Male</td>
<td>72</td>
<td>R</td>
<td>haa che upéa upe sentidope, ajhui vai porke, nde pe nde familla laa, eñiupata la ikausaro, ha repuranduaría chupe, mba’epa la ojapo nohenduire la ekastigha la nde familia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>haa che upéa upe sentidope, ajhui vai porke, nde pe nde familla laa, eñiupata la ikausaro, ha repuranduaría chupe, mba’epa la ojapo nohenduire la ekastigha la nde familia</td>
</tr>
<tr>
<td>Sebastian</td>
<td>P</td>
<td>Male</td>
<td>72</td>
<td>R</td>
<td>ore upeichaakue, ore mami yma ore mondo, rehota eri ñandee pe vaka anoitee fume of hina he’i orre</td>
</tr>
</tbody>
</table>

The combination of transcriptions resulted in a searchable corpus spreadsheet of approximately 250,000 words. This corpus was the starting point of the three studies in this dissertation. From this corpus, study-specific corpora were created. For example, for the first study, all tokens of twelve verbs were extracted. These were combined into a new spreadsheet. The social information was kept, but additional linguistic factors—like origin of the verb—were coded. Thus, because each of the three studies explored a different feature of the language, the codification of linguistic factors was different for each. The codification and the rationale for the inclusion of every factor and other methodological considerations are thoroughly explained in the methods section of each study. In the loan verbs study, methods are described in Section 4.4. In the object marker study, this topic is covered in Section 5.4. In the middle voice study, methods are explained in Section 6.4.

In addition to the dissertation corpus, the first study (Chapter 4) also explored a corpus of Colonial Guaraní: the LANGAS corpus (1630-1813), and three Guaraní
grammars: Anchieta (1595), Ruiz Montoya (1639, 1640), Restivo (1724). In the chapter, these texts are referred to as *Colonial texts*, and are compared to the *Modern data* (the corpus of this dissertation). These old texts were chosen based on digital availability. Furthermore, when there was not data available in the corpus to illustrate specific features, some sentences were created with the assistance of native speaker Antonio Zena. These drafted sentences are differentiated from the corpus sentences throughout the dissertation with the letters ED (elicited data).

### 3.3. Chapter summary

Before developing the three studies of this dissertation, a corpus of oral Guaraní was created. The building of this corpus began with the interviewing of Guaraní speakers of different ages, sexes, and locations. Then, forty interviews were selected, transcribed, and compiled into a single file. This corpus was the base for the three studies that are comprised in the project.

The next chapter describes the first one of these studies, which explores loan verbs in the corpus of the study, and also in written texts of colonial Guaraní.
Chapter 4

First Study: Loan verbs

4.0. Chapter overview

This chapter is entirely dedicated to Spanish-origin loan verbs in Guaraní. Section 4.1 introduces the topic and the theoretical motivations for the study. Section 4.2 explains the research questions and hypothesis. Section 4.3 describes previous literature on verbal borrowing in general and borrowing in Guaraní specifically. Then, Section 4.4 explains the methodological steps in the selection and coding of tokens of the studied verbs. Sections 4.5 and 4.6 describe the results and their interpretation. Finally, Sections 4.7 and 4.8 indicate the limitations and the conclusions of the study, respectively.

4.1. Introduction

Contact linguists have proposed that lexical borrowing that is not motivated by a lexical need can indicate language loss or attrition (Muysken 2001; Campbell 2013). However, more recent studies have found that coexistence of the two forms, the foreign and the native, is more common than replacement (Tadmor and Haspelmath 2009). In this paper, I explore coexistence and replacement scenarios of Spanish-origin loan verbs and their native-origin counterparts in Paraguayan Guaraní, through a usage-based approach. Tokens of highly-frequent native-origin verbs along with their broadly equivalent Spanish-origin loans were extracted from 40 interviews and were correlated with the semantic meanings of each verb. Results show that Spanish verb loans only partially replace their native “equivalent”. This partial replacement leads to coexistence of the two forms, which suggests that coexistence and replacement are not mutually
exclusive outcomes of borrowing, and that Spanish loans in Guaraní do not appear to indicate language attrition.

### 4.2. Research questions and hypothesis

The objective of this case study is to gain insight into the outcomes (changes in the lexicon) of borrowings that appear to have a semantically equivalent form in the recipient language. The study also aims to shed light on the behavior of loans in Guaraní. Although this language has a long history of intense contact with Spanish, the description of the outcomes of Spanish-origin loans in Guaraní—in particular, recent loans—has not been informed by usage-based analyses of the language. Furthermore, the purist ideologies embedded in the metalinguistic discourse surrounding the language (Penner 2014) have favored the spread of ideas that these loans are unsystematic (Palacios Alcaine 1999) and are online borrowings (i.e., they are not conventionalized) in the language (Kallfell 2016). However, a recent variationist account of verbal borrowing in Paraguayan Guaraní (Bittar 2016) has shown that the semantic and syntactic properties of loans are consistent across speakers. Thus, the study intends to answer the following questions:

- Do loan verbs occur in the same syntactic constructions as their native-origin counterparts?
- Are coexistence and replacement mutually exclusive outcomes of borrowing?
- Do loans verbs and their native-origin counterparts overlap semantically?
  How can the syntactic constructions in which these verbs occur inform on their specific meanings?

It is hypothesized that loans do not completely overlap with their native-origin counterpart. The semantic nuances of each verb will be evidenced by the syntactic constructions in which it occurs. As a result, it will be seen that loans might only partially replace their native-origin equivalent.

To test the hypothesis and answer the research questions, this study uses a usage-based, constructionist approach: the loans are studied in natural speech and within the construction where they occur, not in isolation.

The following sections provide an in-depth examination of the coexistence and partial replacement of loans and native forms. The study focuses on six of the most frequent native-loan pairs described in Bittar (2016). It specifically explores the usage of six Spanish-origin verbal loans in Paraguayan Guarani, and their native-origin broad equivalents: konose (Sp. conocer ‘know’/‘become acquainted with’) and kuaa; vivi (Sp. vivir ‘live’) and ko; golpea (Sp. golpear ‘hit’) and nupă; deha (Sp. dejar ‘leave’) and heja; kree (Sp. creer ‘believe’) and rovia; and ajuda (Sp. ayudar ‘help’) and pytyvõ. The data come from a corpus of 40 sociolinguistic interviews, balanced for age, sex, and location.
4.3. Theoretical preliminaries

4.3.1. Verbal borrowing: motivations, strategies and outcomes

4.3.1.1. Motivations

It is agreed on that lexical borrowing can be motivated by need or by other factors, such as prestige of the donor language (Campbell 2013). Likewise, the Loanword Typology Project (henceforth, LWT) (Tadmor and Haspelmath 2009), which compared data from 41 languages, distinguished cultural borrowings from core borrowings. While cultural borrowings are motivated by need and presuppose the lack of an equivalent word in the recipient language, core borrowings are motivated by factors other than need—usually prestige of the donor language—and presuppose the existence of an equivalent word in the recipient language. Another explanation the LWT Project gives for core borrowings is a high degree of bilingualism in a community. In this scenario, both native and borrowed words could be used interchangeably, as both will be understood.

Even in cases where this high degree of bilingualism favors lexical borrowing, does it mean the use of both words—native and borrowed—will be random? Many studies suggest the opposite, and show that the distinction between core and cultural borrowing is far from clear-cut, as words that are core borrowings in appearance (because the recipient language has a word for the borrowed item) are in fact cultural borrowings. For example, in his work on Norwegian speakers in the US, Haugen (1953) found that geography concepts, like ‘river’, ‘creek’, and ‘lake’, were conveyed by English loanwords; the Norwegian-origin counterparts could only describe the Norwegian landscape, as it was very different from the immigrants’ new home. Similarly, researchers have found that English loans make communication more efficient for
Spanish speakers in the US, either by designating a culture-specific item (e.g. Otheguy and Garcia, 1993) or by shortening word-length (Shin 2010).

Lastly, as verbal borrowing is hardly ever studied independently from nominal borrowing, the motivations that have been proposed for lexical borrowing apply to nouns, adjectives, and verbs. However, if the core vs. cultural borrowing is problematic for borrowed nouns, it is certainly more problematic for verbs. For instance, even in the presence of a core verbal borrowing in a highly bilingual community, the borrowed and the native forms are not usually directly interchangeable. This is evidenced by the speakers’ use of syntactic strategies to incorporate loan verbs, discussed in the following section.

4.3.1.2. Strategies

The most recent typological effort to classify the strategies that languages use to borrow verbs is Wichmann and Wohlgemuth’s (2008) study, which identifies four patterns: 1. The light verb strategy, in which a native verb like ‘do’ accompanies the loan verb. 2. Indirect insertion, in which a native affix accommodates the loan verb. 3. Direct insertion, in which the verb is borrowed as “root-like, infinitive-like, imperative-like, inflected for third person or nominalized […]” (99) 4. Paradigm transfer, through which the verb is borrowed with its stem and some of its inflection. These strategies show that verbal borrowing is morphosyntactically more complex than nominal borrowing. Thus, it follows that the borrowing of verbs would be better accounted for by looking at the syntactic constructions they are borrowed into and not just in isolation. It is true, however, that the study of borrowed nouns can also be better understood through
exploring the constructions in which they appear. Aaron (2015), for example, showed that when compared with Spanish nouns, English-origin nouns inserted into Spanish discourse were more likely to occur with an indefinite article or as bare than with a definite article.

4.3.1.3. Outcomes

Muysken (2001) claims that verbs whose borrowing is motivated by a factor other than need indicate language attrition and lead to lexical loss, as the borrowed verb replaces the native verb. This replacement outcome of what the LWT Project calls ‘core borrowings’, however, is difficult to measure. As previously mentioned, it is not easy to determine whether a borrowing is ‘cultural’ or ‘core’ in the first place. Second, even if the borrowings were not originally motivated by need—when they are ‘true’ core borrowings—there is no evidence to suggest that replacement will necessarily take place in all cases. In fact, the LWT Project points at two possible outcomes of core borrowings: replacement and coexistence (Tadmor and Haspelmath 2009). Even though the Project does not propose an explanation for each one of the outcomes, it offers crosslinguistic numbers—summarized in Table 1—that demonstrate these two effects. Of the total amount of borrowings that were studied (12,475), 4,823 (39%) resulted in insertions to the lexicon of the recipient language (because they were cultural borrowings); 2,542 (20%) resulted in coexistence, 1,667 (13%) in replacement, and 3,443 (28%) were not classified. Thus, loanwords that were reported to coexist with their native counterparts outnumbered those which replaced the native words, as coexistence accounts for 60% of the core borrowings effects.
Table 1. Effects of loans: Percentages for coexistence versus replacement

<table>
<thead>
<tr>
<th>type of borrowing</th>
<th>effect</th>
<th>number of loanwords</th>
<th>% of total</th>
<th>% core borrowings</th>
</tr>
</thead>
<tbody>
<tr>
<td>cultural</td>
<td>insertion</td>
<td>4,823</td>
<td>39</td>
<td>-</td>
</tr>
<tr>
<td>core</td>
<td>replacement</td>
<td>1,667</td>
<td>13</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>coexistence</td>
<td>2,542</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>unspecified</td>
<td>no information</td>
<td>3,443</td>
<td>28</td>
<td>-</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td>12,475</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

(Tadmor and Haspelmath 2009)

Furthermore, as all elements in a language, a core loan—and its semantic relationship to its native-counterpart—will change over time. That is, loans can overlap with their native-origin counterparts before they integrate or at the beginning of their integration. However, this overlap can disappear over time, which in turn leads to the coexistence of the two forms.

4.3.2. Spanish loan verbs in Paraguayan Guaraní

Spanish loans have been highly neglected in the literature on Paraguayan Guaraní. Almost a hundred years ago, Morinigo (1931) published a book *Hispanismos en el guaraní*, which to this day remains the only comprehensive inventory of these loans. Paraguayan Guaraní uses what Wichman and Wohlgemuth (2008) call *direct insertion strategy* to incorporate verbs from Spanish. Gregores and Suárez (1967) describe the process as the borrowing of the Spanish verb infinitive form minus the final –*r* (for instance, *leer*, ‘to read’, would be borrowed as *lee*) and the addition of Guaraní
inflectional morphemes (o-lee 3-read means ‘S/he reads’). In his study of a novel written in Jopara (a mix of Spanish and Paraguayan Guarani), Estigarribia (2015) corroborates the verbal borrowing pattern described by Gregores and Suárez (1967), and he adds that “the Spanish base almost invariably carries the native oxytone stress of Guarani, even when the Spanish verb is paroxytone in the correspondent TAM+person/number combination” (24). He also proposes that this consistency might indicate that verbal loans are integrated (i.e. conventionalized) in the Guarani lexicon, and that verbal loans that do not follow this pattern (he found two in his corpus) might be code-switches or nonce borrowings.

Recent studies have identified Spanish loans in Paraguayan Guarani as the key feature of the so-called Jopara, a variety of the language, or for Estigarribia (2015), a mixed lect. For example, Palacios Alcaine (1999) has pointed at Spanish loans as the key element that distinguishes Jopara from “Standard” Guarani, that is, if a Spanish loan is used in place of its Guarani equivalent then we are the presence of the Jopara variety and not Standard Guarani. Likewise, in his grammar of Jopara, in a discussion on Spanish loans, Kallfell (2016) mentions the existence of “doublets”: meanings that can be expressed with a Guarani or a Spanish form, like ‘to read’, which can be expressed with lee (Spanish) or moñe’ê (Guarani). Kallfell (2016) claims that since speakers of Jopara speak two languages at the same time, it is pointless to determine whether these borrowings are nonce borrowings (spontaneous switches) or integrated loans. He also suggests that, in the presence of these doublets, the variant chosen by the speaker will be conditioned by “the situation, their purpose, their knowledge, etc. [but that] in some cases there are semantic differences [between the variants]” (70) [my translation].
Similar to Kallfell (2016), Estigarribia (2017) proposes that Jopara has a low-degree of conventionalization, due to the speakers’ access to two languages. Estigarribia (2017), however, does not define Jopara as “spoken Guarani” but as a continuum of code-switching practices that arise in Paraguay’s bilingual scenario, that go from more Spanish-like to more Guarani-like. In this continuum, a low-degree of conventionalization seems natural, as two languages are involved. Nonetheless, as pointed out by Estigarribia (2015), in some cases what appears to be code-switching might actually be integrated loans, that is, words that were once borrowed from Spanish that are already part of the Guarani lexicon.

Because bilingualism is wide-spread in Paraguay, separating loans from code-switches is a difficult task. Some researchers, like Estigarribia (2015), have relied on the phonological adaptation criterion to define loans. However, Guarani speakers are so exposed to Spanish that their pronunciation of loanwords might be more Spanish-like. In this study, we follow Zenner, Backus, and Winter-Froemel’s (2019) proposition that contact phenomena should be studied as a continuum. As such, all foreign-origin elements were considered loanwords: a foreign-origin word which starts as a code-switch, with frequent usage, can become a conventionalized word in the borrowing language. As Schmid (2015) explains: all communicative conventions start as innovations, which are diffused in the speech communities, and then normalized. Then, because conventional linguistic practices are frequently used, they become entrenched in the language users’ memory. In turn, the more entrenched a linguistic unit is, also the more conventional.
With the goal of informing the literature on Spanish-origin elements in Guarani, Bittar (2016) conducted the first variationist study on loan verbs. The study used 35 interviews of the Corpus of Paraguayan Guarani and Spanish (CPGS), recorded in June of 2015 in Asunción, the capital of Paraguay. The interviews were recorded in Guarani-dominant neighborhoods, by me (a heritage Guarani speaker), accompanied by L1 Guarani speaker Israel Pedrozo, who is a member of the social network of most of the interviewees. The unscripted sociolinguistic interviews were about daily routines, childhood memories, and family life. Participants were told to speak Guarani, however, in some interviews, they would switch to Spanish.

The general objective of Bittar (2016) was to find the so-called verbal “doublets” and correlate each variant (the dependent variable) with both linguistic and social independent variables. Of 331 verbal meanings, only 13 had broadly equivalent Guarani- and Spanish-origin stems. Said differently, 96% of the verbal meanings were expressed with a single form: either Spanish- or Guarani-origin. This high percentage strongly indicates that Spanish verbal loans, contrary to what Kallfell (2016) and Estigarribia (2017) propose, are integrated, i.e. conventionalized, in the Guarani lexicon and their occurrence is not conditioned by external, non-linguistic, factors. Furthermore, when Bittar (2016) explored the constructions in which the thirteen ‘variable verbs’ appeared, she found that each variant—the Spanish-origin and the Guarani-origin stem—occurred in different constructions. For example, for the meaning ‘leave’, the broadly equivalent stems heja (Guarani-origin) and deha (from Spanish dejar) had different syntactic distributions: deha always occurred with the (middle) reflexive prefix je- and with the postposition –gui after the object (example (1)). The counterpart heja never occurred in
this Oblique Reflexive Construction but always in a Transitive Construction (example (2)).

(1) *Mama o-je-deha chu-gui.*

Mom 3-REFL-leave 3-OBL

‘Mom left him.’

Oblique Reflexive Construction: [S REFL-V OBL]

(2) *A-heja-pa la che memby-kuéra.*

1-leave-COMPL the my child-PL

‘I left all my children.’

Transitive Construction: [S V OBJ]

This complementary distribution of native-origin form and loan-origin form also occurred in the ‘hurt’ pair: *nupã* (Guaraní-origin) and *golpea* (from Spanish *golpear*). Bittar (2016) proposes that the loan *golpea* has a non-volitional, middle reflexive reading (as in example (3)) and, thus, occurs with the prefix *je*. When the volitional reading is expressed, the Guaraní-origin stem *nupã* is preferred (as in example (4)). This study also aims at corroborating these claims.

(3) *Ha’a che... che jyva ári ha a-je-golpea-vai.*

1.fall my my arm on and 1-REFL-hurt-bad

‘I fell on my… my arm and I hurt myself badly.’
Reflexive Construction: [S REFL-V]

(4) Che ha’e-akue la a-pillarô, a-hecha-rô,

I say-PST NMLZ 1-catch-if 1-see-if

‘I used to say: if I catch him, if I see him,

a-trosá-ta, ai-nupá-ta, a-juká-ta.

1-destroy-FUT 1-hurt-FUT 1-kill-FUT

I will destroy him, I will hit him, I will kill him.’

Transitive Construction: [S V OBJ]

Naturally, not all verb pairs in Bittar (2016) showed a complementary distribution like ‘leave’ and ‘hurt’. Other doublets, like rovia and kree (‘believe’), appeared in the same constructions; in this case, the Transitive Construction. In examples (5) and (6), the Guaraní-origin verb rovia and the Spanish-origin verb kree (from Spanish creer) are complemented with a human direct object (a full noun in (5), and a pronoun in (6)). However, more tokens are needed to see clearer patterns.

(5) Pediátra-pe che a-rovia.

Pediatrician-OBJ I 1-believe

I believe the pediatrician.

Transitive Construction: [S V OBJ]
Summarizing, the distribution of these broadly equivalent stems in Paraguayan Guaraní shows that: (1) The so-called “doublets” are very rare in the verbal meanings of the language (only 4%); (2) even the apparent “doublets” are not always interchangeable, suggesting they are not true “doublets”; (3) some verb loans in Guaraní occur in specific syntactic constructions to convey specific meanings.

4.4. Methods

This study analyzes the spontaneous usage of six of the most frequent verb pairs attested in Bittar (2016)’s study of loan verbs in Paraguayan Guaraní konose (Sp. conocer ‘know’/‘become acquainted with’) and kuaa; vivi (Sp. vivir ‘live’) and ko; golpea (Sp. golpear ‘hit’) and nupã; deha (Sp. dejar ‘leave’) and heja; kree (Sp. creer ‘believe’) and rovia; and ajuda (Sp. ayudar ‘help’) and pytyvô.

First, tokens of the native-origin verbs (kuaa, ko, nupã, heja, rovia, pytyvô) were explored in old Guarani grammars: Anchieta (1595), Ruiz de Montoya (1639, 1640), Restivo (1724), and in texts from the LANGAS corpus (1630-1813). These texts reflect Guaraní usage from colonial times, from a stage of early contact with Spanish to the beginning of the 19th century. Thus, the verb meanings extracted from these texts indicate how certain verbs were used prior to the incorporation of loans. However, colonial
written texts in Guaraní are heavily biased towards religious and political topics, thus, usages of words in everyday communication might not be portrayed in these documents.

Second, tokens of the above-mentioned native-origin verbs were extracted from the corpus. For each analyzed verb, meanings were extrapolated from the tokens. Third, the loan counterparts of the native-origin verbs were extracted from the spoken data corpus: *konose, vivi, golpea, deha, kree, ajuda*. The meanings of each loan verb were extracted using syntactic and contextual information.

Finally, to determine whether the loan is replacing the native-origin verb, the meanings of each loan were compared with the meaning of their native-origin counterparts. Is the loan *vivi*, for example, used to convey all the pre-contact meanings of its native counterpart *ko* or is it used to convey just one specific meaning in present-day Guaraní?

In the following sections, I present each of the verb meanings of the native-origin verbs.

4.4.1. Meanings of *kuaa*

Across old texts, *kuaa*—often spelled *quaa*—was mainly translated to Spanish with the verbs *saber* (‘know’): when it referred to KNOW something; and *conocer* (‘know’), when it meant BE FAMILIAR WITH something, BE ACQUAINTED WITH someone, or BECOME ACQUAINTED WITH someone. The four meanings are also present in the spoken data, as exemplified below.
1. KNOW

Colonial texts:

(7) *Oy-quaa oyeupe y-mombeu hague.*

3-know IOBJ 3-tell PST

Spa: Sabe lo que le dixeron á él.¹

‘He knows what was said to him.’

(Restivo 1724, 40)

Modern data:

(8) *O-moi pepe he’i nd-oi-kuia-i moõ-pa.*

3-put there 3-say NEG-3-know-NEG where-INT

‘He put it there, he says, he doesn’t know where.’

2. BE FAMILIAR WITH

Colonial texts:

---

¹ The Spanish or Portuguese translations to each of the examples extracted from Colonial texts are the original Spanish or Portuguese translations given by the authors. The original orthography is preserved in all languages.
(9)  *Nd-ai-quaá-bi.*

NEG-1-know-NEG

Spa: No conocer.

‘Not knowing.’

(Ruiz de Montoya 1639, 109)

Modern data:

(10)  *Ei-kuaa  hína  pe  Calle  Quinta?*

2-know  EMPH  that  Calle  Quinta

‘Do you know that *Calle Quinta* (Fifth Street)?’

3. BE ACQUAINTED WITH

Colonial texts:

(11)  *O-yo-quaá-ima*

3-RECP-know-COMP

Spa: Hanse conocido ya en buena y mala parte.

‘They have known each other, the good and the bad part.’

(Ruiz de Montoya 1640, 326/320)

Modern data:

(12)  *Che  nd-ai-kuua-i-mi  la  nde  ména-pe.*

I  NEG-1-know-NEG-EMPH  the  your  husband-OBJ

‘I don’t know your husband.’
4. BECOME ACQUAINTED WITH

Colonial texts:

(13) *Ay-quaá-bucá Tūpā ychupe*

1-know-CAUS God IOBJ

Spa: Hele hecho que conozca a Dios.

‘I had him meet God’

(Ruiz de Montoya 1640, 326/320)

Modern data:

(14) *Péa che la ai-kuua-akue.*

Like.this I NMLZ 1-meet-PST

‘This is how I met her.’

The four meanings are summarized in Table 2. In this and the following tables, a checkmark indicates the meaning is attested and a cross indicates the meaning was not attested.
Table 2. Meanings of kuua in colonial texts and modern data

<table>
<thead>
<tr>
<th>Meanings of kuua</th>
<th>Colonial texts</th>
<th>Modern data</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNOW</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BE FAMILIAR WITH</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BE ACQUAINTED WITH</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BECOME ACQUAINTED WITH</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

4.4.2. Meanings of ko

In old grammars, ko is mainly translated as Spanish vivir (‘live’), estar and andar (‘be’), and morar (‘reside’). The three meanings are also present in contemporary Guaraní, as exemplified below:

1. LIVE

Colonial texts:

(15) *Ai-có-catú*

1-live-well

Port: Viuo bem.

‘I live well.’

(Anchieta 1595, 11)
Modern data:

(16) \textit{Péicha piko nde rei-ko-se?}

Like this INT you 2-live-DES

‘Is this how you want to live?’

2. BE

Colonial texts:

(17) \textit{Ai-co-et.}

1-be-FRU

Spa: Ocioso estar.

‘Stay idle.’

(Ruiz de Montoya 1639, 616)

(18) \textit{Che año ay-có.}

I alone 1-be

Spa: Solo andar.

‘Be alone.’

(Ruiz de Montoya 1639, 714)

Modern data:

(19) \textit{Ai-ko porâ.}

1-be well

‘I am (doing) well.’
3. RESIDE

Colonial texts:

(20) *Ay-có.*

1-live

Spa: Morar, habitar.

‘Reside, inhabit.’

(Ruiz de Montoya 1639, 616)

Modern data:

(21) *H-óga-pe che ai-ko-akue.*

3-house-in I 1-live-PST

‘I resided in her house.’

<table>
<thead>
<tr>
<th>Meanings of <em>ko</em></th>
<th>Colonial texts</th>
<th>Modern data</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIVE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>RESIDE</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

*Table 3. Meanings of *ko* in colonial texts and modern data*

4.4.3. Meanings of *nupá*

In old texts, the verb *nupá* is mainly translated as Spanish *azotar* (‘whip’) and *castigar* (‘punish’); *pegar* and *golpear* (‘hit’) are also used to translate this verb. In the present-day Guaraní data, however, only the meaning ‘hit’ was attested for *nupá.*
1. WHIP

Colonial texts:

(22) *Ay-nupā uca.*

1-hit CAUS

Spa: lo hago azotar

‘I have him whipped.’

(Restivo 1724, 326)

2. PUNISH

Colonial texts:

(23) *Ay-nupā acĭ-catu.*

1-hit hard-well

Spa: Le castigue muy bien.

‘I punished him well.’

(Restivo 1724, 377)

3. HIT

Colonial texts:

(24) *Ay-nupā pota-biňa.*

1-hit want-?

Spa: Dar en vago el golpe.

‘I failed to hit him.’

(Restivo 1724, 283)
Modern data:

(25) *Heta-iteri ai-nupā chupe.*

a.lot-AUG 1-hit  him

‘I hit him a lot.’

<table>
<thead>
<tr>
<th>Meanings of <em>nupā</em></th>
<th>Colonial texts</th>
<th>Modern data</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHIP</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>PUNISH</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>HIT</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

4.4.4. **Meanings of *heja***

In both colonial data and present-day data, *heja* is used to convey LEAVE SOMETHING or LEAVE SOMEONE.

1. LEAVE SOMETHING

Colonial texts:

(26) *Che-po-pe o-heya.*

my-hand-LOC 3-leave

Spa: lo dejó en mi mano

‘He left it in my hand.’

(Restivo 1724, 245)
Modern data:

(27) $O$-heja $i$-pypore.

3-leave 3-footprint

‘He left his footprint.’

2. LEAVE SOMEONE

(28) $O$-heya guembireco cincuenta y dos

3-leave wife fifty and two

Spa: Dejan a sus mujeres, cincuenta y dos (son).

‘They leave their women, fifty-two (they are).’

LANGAS (2012[1770])

<table>
<thead>
<tr>
<th>Meanings of heja</th>
<th>Colonial texts</th>
<th>Modern data</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEAVE SOMETHING</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LEAVE SOMEONE</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

4.4.5. Meanings of rovia

In old texts the verb rovia (also guerovia) is used mainly to convey the following meanings: BELIEVE IN GOD, BELIEVE SOMETHING TO BE TRUE, and OBEY. In the modern data rovia is used to convey the first two meanings but not the latter.
1. BELIEVE IN GOD

Colonial texts:

(29) A-robia Tupa.

1-believe God

Spa: Creo a Dios

‘I believe in God.’

(Ruiz de Montoya 1640, 343/337)

Modern data:

(30) Katóiko péro nd-a-guerovia-pá-i.

Catholic but NEG-1-believe-COMPL-NEG

‘[I’m] Catholic, but I really don’t believe [in God]’

2. BELIEVE SOMETHING TO BE TRUE

Colonial texts:

(31) Ndo-ro-gueroviá-i h-emimboa

NEG-1EXCL-believe-NEG 3-will

Spa: No creemos que es su voluntad.

‘We don’t believe it is his will.’

(LANGAS 2012[1753])
Modern data:

(32) Nd-a-rovia-i-voi la i-túa-ha.

NEG-1-believe-NEG-EMPH the 3-father-NMLZ

‘I don’t even believe he is his father.’

3. OBEY SOMEONE

Colonial texts:

(33) A-robia che-ruba.

1-believe my-father

Spa: Obedezco a mi padre.

‘I obey my father.’

(Ruiz de Montoya 1640, 343/337)

Table 6. Meanings of rovia in colonial texts and spoken data

<table>
<thead>
<tr>
<th>Meanings of rovia</th>
<th>Colonial texts</th>
<th>Modern data</th>
</tr>
</thead>
<tbody>
<tr>
<td>BELIEVE IN GOD</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BELIEVE SOMETHING TO BE TRUE</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>OBEY SOMEONE</td>
<td>✓</td>
<td>X</td>
</tr>
</tbody>
</table>

4.4.6. Meanings of pytyvō

In the colonial texts and the spoken data the verb pytyvō is used mainly to convey the meaning HELP SOMEONE. No additional meanings were found.
1. HELP SOMEONE

Colonial texts:

(34) *Che-pitybo.*

1OBJ-help

Spa: Me ayuda.

‘He helps me.’

(Ruiz de Montoya 1639, 301)

Spoken data:

(35) *Roï-pityvô  la ore sósio-pe.*

1EXCL-help the our friends-OBJ

‘We help our friends.’

Table 7. Meanings of pytyvô in colonial texts and modern data

<table>
<thead>
<tr>
<th>Meanings of pytyvô</th>
<th>Colonial texts</th>
<th>Modern data</th>
</tr>
</thead>
<tbody>
<tr>
<td>HELP SOMEONE</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

To compare the native-origin verb meanings described above with the meanings conveyed by the loan verbs, token frequency was taken into account. That is, how many instances of the loan were used to convey which meaning, compared to instances of the native-origin counterpart to express the same meaning? Below, results are presented by verb pair.
4.5. Results

4.5.1. Kuaa/konose

Kuaa (‘know’) translates into Spanish as saber and conocer. While the borrowing of saber is not attested, the borrowing of conocer is. Then, it could be assumed that the borrowing konose is kuaa’s partial equivalent. However, when correlating kuaa and konose with their specific meanings, it was seen that kuaa was more frequently used to convey the BE ACQUAINTED WITH and BE FAMILIAR WITH meanings while konose was more frequently used to express the BECOME ACQUAINTED WITH meaning (Figure 3).

Figure 3. Kuaa vs. konose across meanings

The use of Spanish-origin konose was almost never used to express the BE ACQUAINTED WITH meaning. Example (36) was the only token of 95 where konose was used to convey this meaning:
(36) Pêa pe apódo a-ñe-konose-ve ke la che réra.

Like this this nickname 1-PASS-know-AUG than the my name

‘I am better known by this nickname than by my name.’

In the other 94 cases, Guarani-origin kuaa was used, as exemplified in (37)-(41):

(37) Ha óga-pe Julio mante o-je-kuaa chupe.

and house-in Julio only 3-PASS-know him

‘And at home he is only known as Julio.’

(38) Upépe o-je-kuaa-porã.

There 3-PASS-know-well

‘There, they are well-known.’

(39) Nde ko che-kuaa porã.

You EMPH OBJ-know well

‘You know me well.’

(40) Pedrito suégra nde rei-kuaa?

Pedrito mother-in-law you 2-know

‘Do you know Pedrito’s mother-in-law?’
(41) *Ni la che ermána ni che ndo-roí-kuaá-i la che ermáno*

Neither the my sister nor I **NEG-1EXCL-know-NEG** the my brother

‘Neither my sister nor I know my brother’

To convey the BE FAMILIAR WITH meaning, Spanish-origin *konose* was used in 3% the cases, that is, in only two tokens out of 70, as in (42) and (43):

(42) *Che réra eskáso o-ñe-konose.*

My name little **3-PASS-know**

‘My name is very little known.’

(43) *E-konose-se gua'u ofisio Ánte.*

2-know-DES supposedly job beforehand

‘You want to supposedly know the job beforehand.’

In the other 68 cases, the Guaraní-origin verb–*kuaa*–was used (examples (44)- (48) below):

(44) *Che ai-kuaa-pa-ite.*

1-know-COMPL-INTEN

‘I know it all.’
(45) Nde-rei-kuáá-i pe kamionéta-i?

NEG-2-know-NEG that truck-DIM

‘Don’t you know that little truck?’

(46) Ai-kuaa-pa-ite la múndo.

1-know-COMPL-AUG the world

‘I really know everything about the world.’

(47) Oi-kuaa traváho hikuái.

3-know work they

‘They know about working.’

(48) Oi-kuaa-voi la óga.

3-know-EMPH the house

‘In fact, he knows the house.’

Contrary to the BE FAMILIAR WITH and BE ACQUAINTED WITH meanings, the BECOME ACQUAINTED WITH meaning is expressed more often with *konose* than with *kuaa*: *Konose* represented 88% of the tokens, that is, 29 out of 33, as in the example (49)-(52) below:
(49) *Che tua che sy upevoi o-ño-konose ha’e-kuéra.*

My dad my mom like that 3-RECP-meet 3-PL

‘My dad and my mom, that is how they met.’

(50) *Péro nda-ha’e-vé-i-ma yma pe ro-ño-konose-hagué-icha.*

But NEG-be-AUG-NEG-COMPL past that 1EXCL-RECP-meet-PST-like

‘But things aren’t like when we met.’

(51) *Ha upéroguare a-konosé-ramo pe karia’y-pe.*

And that time 1-meet-just that guy-OBJ

‘And during that time I had just met that guy.’

(52) *Ha upéi a-konose Ña Teresa rovajá-pe.*

And then 1-meet Mrs. Teresa brother-in-law-OBJ

‘And then I met Mrs. Teresa’s brother in law.’

Although highly dispreferred, *kuaa* was also used to convey the BECOME ACQUAINTED WITH meaning, as in (53):

(53) *Péa che la ai-kuaa-akue.*

Like this I NMLZ 1-meet-PST

‘This is how I met her.’
Overall, the examination of *kuaa/konose* tokens indicate that *konose* is the preferred form to convey the BECOME ACQUAINTED WITH meaning, while *kuaa* is preferred to express the BE FAMILIAR WITH or BE ACQUAINTED WITH meanings.

4.5.2. *Ko/vivi*

*Ko* means LIVE, RESIDE, and BE. On one hand, the borrowing of *andar* and *estar*, equivalent of BE, is not attested in the spoken data. On the other hand, *vivi* is attested in the data, and it is used to express both LIVE and RESIDE. As seen in Figure 4, the loan occurs more often to express the LIVE meaning than the RESIDE meaning.

*Figure 4. Ko vs. vivi across meanings*

In the LIVE meaning, the use of *vivi* represented 7% of the tokens, that is, seven out of 94, as in (54) and (55).
(54) Péicha  ro-vivi.
Like.this  1EXCL-live
‘We live like this.’

(55) Ikatu-haguāicha  ore  pais-pe  mimo...
Can-so.that  our.EXCL  country-in  own
‘So that in our own country we could...

ro-vivi  porā  ro-karu  porā.
1EXCL.live  well  1EXCL.eat  well
live well, eat well.’

In the other 87 cases of LIVE, ko was used, as in (56) and (57):

(56) A-karu  porā  ha  ai-ko  porā.
1-eat  well  and  1-live  well
‘I eat well and I live well.’

(57) Péicha  ai-ko.
Like.this  1-live
‘I live like this.’

In the RESIDE meaning, the use of the loan vivi represented only 4% of the tokens, that is, nine out of 211. In the other 202 cases, ko was used. Even in cases where there
were other Spanish loans (in bold, below) in the sentence, \textit{ko} was used, as in the following examples:

(58) \textit{Kóa eskuēla frénte voi oi-ko.}

This school front EMPH 3-live

‘He lives in front of this school.’

(59) \textit{Kuátro año ai-ko pépe.}

Four years 1-live there

‘I lived there for four years.’

(60) \textit{A-ha peteĩ úna semána}

1-go one a week

‘I went, for a week...

\textit{ai-ko-akue che vesína róga-pe.}

1-live-PST my neighbor house-at
to live at my neighbor’s house.’

Although highly dispreferred, \textit{vivi} was used to express the RESIDE meaning as well, as in (61) and (62):
(61) Ha o-ho-kuri la pe
     And 3-go-PST the one
     ‘And he went to live...

che ermána róga-pe o-vivi.
my sister house-in 3-live
...in the house of that one of my sisters’.

(62) Óga-pe oi-ko ha'e o-nase.
     Home-at 3-live he 3-be.born
     ‘He lives at home (since) he was born.’

In summary, contrary to konose (‘become acquainted with’), vivi does not occur more often than its native counterpart to express a particular meaning. However, proportionally, the loan is used more often to convey LIVE than RESIDE.

4.5.3. Nupã/golpea

In present day Guaraní, nupã means ‘to hit something or someone’. The equivalent of nupã in Spanish is pegar or golpear. In the spoken data, while the borrowing of pegar is not attested, there are 11 tokens of the loan verb golpea. Eight out of these 11 tokens are used in a reflexive construction, that is, they indicate that the subject did not cause the action, as in (63):
3-REFL-hit-DIM that 3-head-in

‘He hurt his head a little.’

Of the 153 tokens of nupă, only one is used in a (indirect) reflexive construction, as in example (64) (the speaker is narrating how she accidentally hurt herself when working in a sunflower plantation):

3-REFL-hit-DIM that 3-head-in

‘He hurt his head a little.’

All other tokens of nupă are used in simple transitive constructions, as in (65).

3-REFL-hit-DIM that 3-head-in

‘He hurt his head a little.’

The data indicate that nupă is preferred in simple transitive constructions while golpea is favored in reflexive constructions, as presented in Figure 5.
4.5.4. Heja/deha

Heja is equivalent to Spanish dejar, as both mean ‘to leave something or someone’. However, the loan deha never overlapped with heja in the spoken data (Figure 6). Deha is always used in middle reflexive constructions, which conveys an ABANDON or QUIT meaning, as in (66). This type of construction is possibly a replication of the Spanish dejarse de + noun construction, as exemplified in the translation (66x). Heja is always used in simple transitive constructions, as in (67):

\[(66) \quad A\text{-je-deha} \quad la \quad sigarrillo-gui.\]

1-REFL-leave the cigarrette-from

‘I quit smoking’.
(66) \( Me \ dej\-é \ del \ cigarrillo \)

1.REFL leave-1.PST from.the cigarrillo

‘I quit smoking.’ (ED)

(67) A-\( heja \)-pa \( la \ che \ memby-kuéra. \)

1-leave-COMPL the my child-PL

‘I left all my children.’

Figure 6. Heja vs. deha across construction meanings

4.5.5. *Rovia/kree*

In present day Guaraní, *rovia* means BELIEVE IN GOD, BELIEVE SOMETHING TO BE TRUE, and BELIEVE SOMEONE. This latter meaning, however, was not found in colonial texts. In Spanish, *creer* is also used to convey the meanings listed for *rovia*. In
the modern data, however, the loan *kree* was never used to express religious or spiritual belief. *Kree* was the preferred form to express BELIEVE SOMETHING TO BE TRUE, 17 of 24 tokens (70%), as in (68), while *rovia* was favored to convey BELIEVE SOMEONE, 13 of 17 tokens (76%), as in (69). Figure 7 illustrates these findings.

(68)  *O-*kree-*terei-voi* pe *eré-a* *chupe.*

3-believe-AUG-EMPH that 3.say-NMLZ IOBJ

‘He really believes what is said to him.’

(69)  *Nd-a-rovıá-i* *chupe.*

NEG-1-believe-NEG IOBJ

‘I don’t believe him.’

*Figure 7. Rovia vs. kree across meanings*
4.5.6. Pytyvõ/ajuda

Contrary to the other verbs in this study, both pytyvõ and ajuda appear to have just one meaning: HELP SOMEONE. The examination of the spoken data showed that these two forms are interchangeable, as shown by examples (70) and (71). However, ajuda is more widely used (95 of 113) than pytyvõ, seen in Figure 8.

(70) Ai-pytyvõ chupe.

1-help OBJ

‘I help him.’

(71) A-ajuda chupe.

1-help OBJ

‘I help him.’
4.6. Analysis

4.6.1. Conventionalization of loans

As hypothesized, the results above show that loan verbs do not overlap completely with their native-origin counterparts. On the one hand, it was shown that five of the six loans in this study are not necessarily interchangeable with their native counterparts: konose (‘become acquainted with’), vivi (‘live’), golpea (‘hurt’), deha (‘leave’), and kree (‘believe’). The loan ajuda (‘help’), on the other hand, appears to overlap with its native-origin equivalent pytyvô in all cases.

The loan konose has conventionalized to convey a BECOME ACQUAINTED WITH meaning, leaving kuaa the meanings of BE ACQUAINTED/FAMILIAR WITH. Therefore, considering kuaa and konose “doublets” is an oversimplification of the linguistic practices of Guarani speakers. Also, as a conventionalized item, konose is neither a code-
switch nor an online borrowing (in Schmid 2015’s terms: it is no longer an innovation),
but rather a loan that is integrated in the Paraguayan Guaraní lexicon.

In addition, the conventionalization of konose’s meaning is made evident in two
excerpts from two interviews. In the first excerpt (72), when the interviewer (I) (whose
tokens were not quantified for this study) uses the Guaraní-origin form to convey the
BECOME ACQUAINTED WITH meaning, the interviewee (R) asks for clarification, using the
Spanish-origin form. This repair is then reinforced by the answer to the interviewer’s
question, where she uses the expected form to convey the BECOME ACQUAINTED WITH
meaning: konose. In the second excerpt (73), the participant also uses konose to convey
the ‘meet’ meaning, even when the interviewer used kuua in his question. It is unclear
why the interviewer, a native Paraguayan Guaraní speaker, used kuua instead of konose.

(72) I: Mba’êicha ei-kuua-raka’e chupe?
How 2-meet-PST OBJ
‘How did you meet him?’

R: A-konose-akue chupe pa?
1-meet-PST OBJ INT
‘[How] I met him, right?’

I: Hm.
‘Hm.’ (backchanneling)

R: Che a-konose chupe merkâdo-pe.
I 1-meet OBJ market-at
‘I met him at the market.’
While the loan *konose* showed clear patterns of conventional usage, the verb loan *vivi* did not. *Vivi*, however, is more favored by the meaning LIVE than by the meaning RESIDE (7% vs. 4%, respectively). This pattern might suggest that *vivi* is a nonce (online) borrowing and will remain one. It is still an innovation, and might not diffuse, as it is not frequently used.

Because *konose* refers to the specific event of ‘becoming acquainted with someone’ most of the times it is used, it can be argued that the pairing of this form and its meaning has conventionalized. The conventionalization that we see today might be due to
konose’ s status of an old loan (it was attested by Morínigo (1931)), which was either borrowed with a specific function, the BECOME ACQUAINTED WITH meaning, or it specialized along the way. Vivi, however, might be a newer loan (it was not attested by Morínigo (1931)), which might specialize to convey a specific meaning in the future or not.

Results indicate that the loans deha (‘leave’) and golpea (‘hurt’) are not used in the same constructions as their native-origin counterparts heja and nupã. While heja and nupã are mainly used in prototypical transitive constructions, deha and golpea are used in reflexive constructions. These meanings have also conventionalized, also evidenced by the fact that the following sentences are not equivalent. While añenupã means ‘I was hit’; ajegolpea means ‘I (accidentally) hurt myself’ (ñe and je are allomorphs).

(74) A-ñe-nupã.
    1-PASS-hit
    ‘I was hit.’ (ED)

(75) A-je-golpea.
    1-REFL-hit
    ‘I got hurt.’ (ED)

Similarly, the sentence ajeheja and ajedeha are not synonymous, as the first means ‘I was left’ and the second one means ‘I quit/abandoned’.
(76) *A-je-heja.*

1-PASS-leave

‘I was left.’ (ED)

(77) *A-je-deha.*

1-REFL-leave

‘I quit.’ (ED)

The loan *kree* is more frequently used to express belief in that something is true than to convey belief or trust in someone (70% vs. 30%, respectively). This pattern also shows item conventionalization, that is, the Spanish-origin verb is an integrated item in the speaker’s lexicon as opposed to a code-switch.

Finally, *ajuda* (‘help’) is the only loan in the study that does not have a different meaning or usage than its native-origin counterpart, *pytyvõ*. It could be argued that out of the six loans, *ajuda* is the only one that appears to be replacing the native-origin word, as it is used more frequently than *pytyvõ*. Ajuda is also a conventionalized item, as this is the speakers’ preferred verb to convey the HELP SOMEONE meaning.

This conventionalized usage of loans and native-origin verbs suggests that the first is not fully replacing the latter, as each verb has its specific meaning. This partial replacement, in turn, leads to coexistence of the two forms.
4.6.2. Motivations of loans

4.6.2.1. Konose and vivi: construction polysemy

The fact that konose and vivi are used to convey some meanings more often that others might inform us on their motivation. When we explore the constructions where konose and vivi occur, respectively, it can be noted they can be polysemous. Notice the sentence with kuaa, below:

(78) Ai-kuaa Joel.
    1-know/met Joel (ED)

Example (78) can be translated as ‘I know Joel’ (BE ACQUAINTED WITH) or ‘I met Joel’ (BECOME ACQUAINTED WITH), because tense and aspect suffixation is optional in Guaraní. Thus, I argue that konose entered the Guaraní lexicon through the polysemic construction to unambiguously convey the BECOME ACQUAINTED WITH meaning. Other factors, like widespread bilingualism might have contributed to the incorporation of the loan.

Likewise, ko also appears in constructions that can have at least two interpretations:

(79) Ai-ko porã.
    1-live/be well (ED)
The example above (79) can be translated as ‘I am well’—which has a more temporary meaning, like Spanish *Ando bien* (‘I am doing well’)—or as ‘I live well’—which has a more long-term implication, as in Spanish *Vivo bien* (‘I have a good life’). It can be argued that *vivi* is used to unambiguously convey the I HAVE A GOOD LIFE meaning.

Thus, it appears that speakers want to separate the two meanings and use one form for each. This disambiguation is facilitated by a loan. However, as polysemy is ubiquitous in language, it remains unclear why there is a need to disambiguate the different meanings of ‘know’ or ‘live’ and not other verbs. The explanation might be in that other elements of the construction (verb arguments, TAM markers) do not help the listener select the correct interpretation.

4.6.2.2. *Golpea* and *deha*: Reflexive constructions

The fact that *golpea* and *deha* almost always occur in reflexive constructions suggests that these loan verbs were borrowed as elements of larger linguistic units in the donor language. *Golpea* is borrowed as part of the Spanish Middle Reflexive Construction [REFL V]: *golpearse* (‘get hurt’), and *deha* is borrowed as part of the Spanish Oblique Middle Reflexive Construction [REFL V OBL]: *dejarse de + noun* (‘abandon + noun’). The colonial texts data indicate that neither *nupã* nor *heja* (the native-origin counterparts of these loans) were used in reflexive constructions in the past. Chapter 6 discusses more verbs like *jegolpea* and *jedeha*, which are borrowed with the reflexive and middle prefix *je*. Furthermore, the loans *golpea* and *deha* are only
“equivalent” to nupã and heja in isolation, but when explored in context, they have clearly distinct usages.

4.6.2.3. Kree and ajuda: unclear motivation

Both kree and rovia convey the BELIEVE SOMETHING meaning. Similarly, both ajuda and pytyvõ convey the HELP SOMEONE meaning. These native-origin verbs, however, do not occur in polysemous constructions, as kuaa and ko do. Examples (80) and (81) only have one interpretation, thus, the loans would not disambiguate any meanings, as shown in (82) and (83).

(80) A-rovia o-mano-ha.

1-believe 3-die-NMLZ

‘I believe she died.’ (ED)

(81) Ai-pytyvõ María-pe.

1-help María-OBJ

‘I help María.’ (ED)

(82) A-kree o-mano-ha.

1-believe 3-die-NMLZ

‘I believe she died.’ (ED)
The examples above also show that the loans and the native-origin verbs can occur in the same constructions, and thus, do not appear to be part of pattern-replication borrowing. The motivation of these loans, kree and ajuda, remains unclear for this study.

4.7. Discussion

The results of this study demonstrate the need to rethink traditional binary categories in the language contact literature, with respect to outcome of the loan (coexistence vs. replacement), and with respect to motivation (core vs. cultural borrowing).

Regarding outcomes of borrowings, five of the six loans in this study do not span the entire semantic complexity of their native-origin counterparts. It could be argued that this lack of complete overlap results in the coexistence of the native-origin verb and the loan. For example, in the cases of konose (‘become acquainted with’), vivi (‘live’), and kree (‘believe’), the loan appears to replace the native-origin form to convey specific meanings. The loans’ partial replacement of their native “equivalent” leads to coexistence of the two forms. Thus, coexistence and replacement are not mutually exclusive outcomes of borrowing. Furthermore, as absolute synonymity is virtually inexistent (Cruse 1986), these loan-native pairs might continue to coexist without completely overlapping.
Concerning motivation of borrowing, are the meanings attached to the loans explored in this study ‘core’ or ‘cultural’? The verb BECOME ACQUAINTED WITH, for example, conveyed by the loan konose, could be said to be ‘core’, as it is a relatively basic action. However, it can also be ‘cultural’, as becoming acquainted with a person can be a different experience across cultures. In a relatively small, tight-knit community, becoming acquainted with someone might be a rare event. When this event becomes more frequent, the social implications of said experience are new and different, and thus, a different term for it might surge. Furthermore, while it can be easy to see how the nouns in this list—like ‘fire’—are culture-free, actions are inherently more complex than objects, as a ‘basic’ action like ‘X ingests Y’ can be construed very differently from one culture to another. Consequently, the distinction between cultural and core borrowings cannot be applied to verbs to the same extent in which it can be applied to nouns.

Finally, the clear preference for the loan over the native-origin verb to express a specific meaning suggests that Guaraní speakers do not borrow words randomly (Palacios Alcaine, 1999), and that they do not speak two languages at the same time (Kallfell, 2016). Previous research on Spanish-origin loans in Guaraní has proposed that loans tend to be unsystematic (Palacios Alcaine, 1999), and are motivated by language-external factors (Kallfell, 2016).

4.8. Conclusion

This study explored naturally-occurring instances of six Spanish-origin verbal loans in Paraguayan Guaraní: konose (Sp. conocer ‘know’/‘become acquainted with’) and kuaa; vivi (Sp. vivir ‘live’) and ko; golpea (Sp. golpear ‘hit’) and nupã; deha (Sp.
dejar ‘leave’) and heja; kree (Sp. creer ‘believe’) and rovia; and ajuda (Sp. ayudar ‘help’) and pytynõ. Tokens of each verb were extracted from 40 unscripted interviews with native speakers of Paraguayan Guaraní of diverse backgrounds. These tokens were correlated with the specific meanings of the verbs. It was found that the loans were used more frequently to express one of the verbal meanings over others. For example, to convey the BECOME ACQUAINTED WITH, the loan konose was used 88% of the time. However, to convey BE FAMILIAR WITH or BE ACQUAINTED WITH, it was used only 3% and 1% of the time, respectively. To express a reflexive meaning, the loan golpea was used 89% of the time, but to convey a volitional meaning, it was only used 2% of the time.

The fact that the Spanish loans studied here are more frequently found to convey specific meanings or appear more often in some syntactic constructions over others points at the need to explore verbal borrowing from a holistic model, which includes—at least—lexical semantics and syntax, and not lexical semantics alone. Going beyond the binary lexical vs. grammatical borrowing distinction will improve our understanding of borrowing.

In terms of usage, it was seen that the loans and their native-origin counterparts are not always interchangeable, which suggests that each verb has or is developing specific semantic properties. The loans appear to be replacing the native-origin forms in some of its meanings, not in their entire lexical potential. This partial replacement allows for the coexistence of the two forms.

Finally, this study indicates that Spanish-origin words in Paraguayan Guaraní are not unsystematic, as previously proposed (Palacios Alcaine, 1999; Kallfell, 2016).
Speakers of the language use the loans to convey specific meanings, and thus, these Spanish-origin words constitute an integral part of the Paraguayan Guaraní lexicon.
Chapter 5

Second Study: Human Direct Object Marker

5.0. Chapter overview

This chapter describes the synchronic use of the Guarani human direct object marker, -pe, as a presumed grammatical loan from Spanish. Section 5.1 introduces the motivations of the study. Section 5.2 describes previous literature on differential object marking across the world’s languages, and in Spanish and in Guarani. Section 5.3 goes over the research questions and hypotheses. Then, Section 5.4 explains the methods that were followed to extract tokens from the corpus, and to code and quantify them. Sections 5.5 and 5.6 describe the results from the statistical analysis and their analysis. Section 5.7 describes future research on the topic, and, finally, Section 5.8 synthesizes the main findings of the study.

5.1. Introduction

This study explores present-day spoken Guarani data to assess the claim that the language’s differential object marking (henceforth, DOM) originated from contact with Spanish. Consider the following examples:

(84) *A-mongakuuaa-pa la che familia-kuéra.*

1-raise-COMPL the my child-PL

‘I raised all my children.’
While in example (84) the direct object, *familia-küera* (‘children’), does not get the object marker, in example (85) it is marked with this suffix. This Ø/-pe alternation is common in present-day Guaraní (Shain and Tonhauser, 2010). Nonetheless, studies suggest that the DOM is a new feature in the language, as it is absent in 16th- and 17th-century texts (Bossong 2009). Pointing at the recent diffusion of the feature, Bossong (2009) and Shain and Tonhauser (2010) claim that the DOM originated from contact with Spanish, which marks human direct objects with the preposition *a*. However, they do not explain the mechanisms through which the new pattern emerged.

This study is concerned with how Guaraní DOM originated. It proposes that the new feature arose from a contact-induced grammaticalization process, the replica type in Heine and Kuteva’s (2003) typology. In this process, speakers of the recipient language recruit internal items to replicate a grammatical category in the donor language. In the case of Guaraní, following the model of Spanish, speakers recruited the locative postposition and dative marker -pe to form a human direct object marker.

As early Jesuit grammars of Guaraní give examples of –(u)pe as a dative marker, it could be assumed that this use precedes colonial times, as seen in the following example:
Ruiz de Montoya (1639) and Anchieta (1595) associate this dative marker to the indirect object pronoun (*i*)ch*upe*. Ruiz de Montoya (1639) also relates it to the locative -(*u*)pe. Thus, as other researchers have proposed (e.g. Gimeno 2012) based on common paths of grammaticalization (Croft 2002), locative -pe grammaticalized into a dative marker. As -pe had these two functions (locative and dative) before colonial times, it is
difficult to distinguish which one was the source of the DOM. Thus, we propose the following grammaticalization process: LOC / DAT > DOM, and show evidence from variation to support the contact hypothesis.

5.2. Theoretical background

Determining whether contact with Spanish has resulted in differential object marking in present-day Guaraní requires an examination of universal tendencies of DOM, presence of DOM in Tupi-Guaraní languages, and shared characteristics of DOM in Spanish and Guaraní. This study takes a multiple causation approach to change, as internal and external factors can conspire together in the appearance and spread of a new feature (Thomason 2008). Furthermore, this study is concerned not only with if DOM is a contact feature but with how this change occurred. As we propose this is a case of contact-induced grammaticalization, we will describe Heine and Kuteva (2003)’s typology of this type of change.

5.2.1. Differential object marking: definition and universal tendencies

Differential object marking is the variation in overt marking in a language’s direct objects (Döhla 2014). Spanish, for instance, has DOM because—broadly speaking—direct objects are marked when they have human referents and unmarked when they are inanimate. Many scholars have tried to determine where languages draw the markedness line. Iemmolo and Klumpp (2014), for example, claim that semantic or pragmatic properties of the object—which include animacy, definiteness and specificity—are usually the determining factors. In his typological study of DOM,
however, Sinnemäki (2014) explored 744 languages and found no universal correlation between animacy or definiteness and DOM. He found that while, cross-linguistically, there is a statistically-significant preference for restricted case marking (such as DOM) versus non-restricted case marking, languages rely on diverse factors to draw the lines between which elements in the language get marked and which ones do not. These factors are not limited to animacy and definiteness, but also include “common/proper nouns, kin terms, tense/aspect, information structure, and disambiguation” (Sinnemäki 2014, 300).

Going beyond the tradition of examining semantic properties of the direct object as a motivation for DOM, some scholars have explored verb semantics and constructions. In the case of Spanish, Heusinger and Kaiser (2011) have correlated affectedness with overt marking: verbs like ‘kill’ and ‘hit’ highly affect the object, therefore, the object receives the marking. Delbecque (2002), on the other hand, has proposed conflict in force dynamics between the two event actors as a determining factor in the emergence of DOM in Spanish: the more agent-like role the object takes in relationship to the subject, the more likely it is the marked construal will be used.

In sum, contrary to traditional accounts of DOM, which put animacy and definiteness as the main influential factors of overt DO markedness, languages appear to treat DOM in diverse ways across the world’s languages (Sinnemäki 2014). Also, even when factors like animacy have been proposed as essential to DOM in certain languages like Spanish (Tippets 2011), a single factor does not account for the entire paradigm within a language. This cross-linguistic variation of DOM and the language-specific complexity of the feature point to the fact that when two typologically distant languages,
like Spanish and Guaraní, share similar DOM patterns, universal tendencies are an unlikely explanation.

With respect to DOM in contact, a recent study on Basque showed that Spanish loan verbs favored the use of the marker among bilingual speakers (Rodríguez-Ordóñez 2020). Rodríguez-Ordóñez followed Heine and Kuteva (2010), who claim that grammatical structure enters the language through the semantics of the loan verb.

5.2.2. DOM in Spanish

In broad terms, Spanish marks the direct object with the locative preposition a when its referent is human (Laca 2006), as in (89): *su amigo* (*‘her friend’*) takes the object marker. When the object is inanimate, it does not take a marker, as in (90), where *su abrigo* (*‘her coat’*) is not preceded by a.

(89)  
\[ \text{Llev-ó} \quad \text{a} \quad \text{su amigo} \quad \text{a} \quad \text{la escuela}. \]

Take-3.SG.PST OBJ her friend to the school

‘She took her friend to school.’ (ED)

(90)  
\[ \text{Llev-ó} \quad \text{su abrigo} \quad \text{a} \quad \text{la escuela}. \]

Take-3.SG.PST her coat to the school

‘She took her coat to school.’ (ED)

In Spanish, DOM is present in the language as early as the 12th century (Döhla 2014). Scholars have explored the role of referentiality, verb semantics, and constructions
in the usage rate of the human object marker across centuries. Overall, as seen by (Company Company 2003), the use of DOM with common human nouns went from 35% in the 15th century to 57% in the 21st century. However, when classifying these nouns into human definite and human indefinite, the advancement of DOM is even more visible. According to Laca’s (2006) corpus study, while in the 15th century 58% of human definite noun tokens were marked, in the 19th century, the percentage increased to 96%. Similarly, with indefinite human objects, the markedness rate went from 0% in the 15th century to 41% in the 19th century (Laca 2006). Thus, Laca proposes that DOM in Spanish has advanced following a referentiality path:

Human Proper Name > Human Def NP > Human Indef NP > Human Bare Nouns

(Laca 2006)

With respect to verb semantics, Heusinger and Kaiser proposed that the advancement of DOM in Spanish followed these paths:

- definite noun phrases: PERCEPTION , FEELING , ACTION >> PURSUIT , KNOWLEDGE

- indefinite noun phrases: PERCEPTION > FEELING , ACTION > KNOWLEDGE > PURSUIT Heusinger and Kaiser (2011)

Also, Delbecque (2002) has claimed that the a/Ø alternation depends on how the event is construed. If the patient-like participant is perceived as actively responsible for
the agent’s experience, then the $a$-construal will be used. This explains why a sentence like *Déjé a Madrid* (‘I left Madrid’) is possible even when *Madrid* is a nonhuman object.

### 5.2.3. DOM in Paraguayan Guaraní and other Tupi-Guaraní languages

In Guaraní, the object human referent is marked (Gregores and Suárez 1967), but it has been described as optional (Shain and Tonhauser 2010). Thus, both sentences below, (91) and (92) are acceptable in the language:

(91) $Ai$-ku$aa$ $nde$ $sy$.

1-know your mother

‘I know your mother.’ (ED)

(92) $Ai$-ku$aa$ $nde$ $sý-pe$.

1-know your mother-OBJ

‘I know your mother.’ (ED)

The Guaraní DOM has been attested since the 20th century (Shain and Tonhauser 2010). As Shain and Tonhauser (2010) noted, the available synchronic data points to the fact that when contact between Spanish and Guaraní began, in the 16th century, Spanish exhibited DOM but Guaraní did not. This discrepancy is essential to argue for any contact-induced change (Thomason 2001). While it is known that the frequency of DOM in Spanish increased with the centuries, there are no studies on Guaraní that explore an increasing or decreasing use of this feature.
With respect to Guaraní, Tonhauser and Shain’s (2010) study remains the only quantitative exploration of DOM in Paraguayan Guaraní or any other Tupi-Guaraní language. Of the data they analyzed (43 tokens), 56% of human direct objects were marked. Their findings also show that, unlike Spanish, definiteness does not appear to play a role in the developing of DOM. Instead, they propose animacy, and to a lesser extent, topicality, as the determining factors in the expression of the object marker. Bosson-Guérin (2009), however, claims that it is relative animacy that determines the occurrence of the marker: when subject and object nouns are equal in strength on the hierarchy, the marker is obligatory.

Regarding other Tupi-Guaraní languages, Bosson-Guérin (2009) states that Mbyá or Apopokuva show patterns of DOM but its usage is not as advanced as in Paraguayan Guaraní. Roessler (2019), however, finds DOM in four Tupi-Guaraní languages, which she groups in a cluster: Paraguayan Guaraní, Mbyá, Avá/Chiripá, and Paí-Tavãterã/Kaiowá. She claims that DOM in the four languages can be accounted for by animacy and specificity, but that the patterns are understudied: Paraguayan Guaraní is the only of the four languages whose DOM has been discussed in the literature. Roessler, Gasparani, and Danielsen (2014) are skeptical about DOM being a contact feature because it is present in all languages of the Tupi-Guaraní sub-group 1 (which includes the above-mentioned four languages and Aché, Chiriguano, and Ñandeva), and in Guarayo (sub-group 2).

Finally, indirect objects in Spanish and Guaraní are obligatorily marked with a and –pe, respectively, patterns whose similarity also suggest contact effects (Estigarriribia 2017).
In consequence, the parallel behavior of Spanish -a and Guaraní -pe across three constructions (locative expressions, indirect object marking, and differential object marking) is striking. The point to highlight here is that of the three constructions, the locative can be clearly reconstructed back to the Tupi-Guaraní family, whereas the last two are interrelated innovations.

5.2.4. **Replica contact-induced grammaticalization**

According to Heine and Kuteva (2003), there are two types of contact induced grammaticalization: ordinary grammaticalization and replica grammaticalization. In both types, speakers of one language identify a grammaticalization process in the donor language, and replicate it in the recipient language, by forming an analogy. The difference between the two types is that the ordinary type entails common grammaticalization paths documented cross-linguistically. The replica type refers to rare processes identified in two systems in contact; the only explanation for why these processes took place in both systems is language contact, that is, speakers of the recipient language assumed this process happened in the donor language and replicated it. Thus, the replica type requires more agency, noticing, and awareness on the part of the speakers. In the Guaraní-Spanish scenario, Guaraní speakers seem to have “replicated” the Spanish process of a locative adposition/dative marker grammaticalizing to a human direct object marker, but following a common path. In fact, as noted above, parallel diachronic processes have been identified in many Tupi-Guaraní languages, as well as several others which appear to have DOM as a result of contact: Italian Romanian in
contact with Standard Italian (Cohal 2014), Greek in contact with Turkish (Karatsareas 2020) and Basque in contact with Spanish (Rodríguez-Ordóñez 2020).

5.3. Research questions and hypotheses

This study aims to explore two questions: one specifically regarding the origin of DOM in Paraguayan Guaraní, and the other one concerning the mechanisms of language borrowing in general. Thus, the research questions are the following:

1. Is DOM in Guaraní a contact feature?
2. If DOM is a contact feature, how was it borrowed?

As for the first question, following Bossong (2009) and Shain and Tonhauser (2010), this study supports the hypothesis that DOM in Guaraní is a contact feature. It originated by replicating a Spanish pattern. Regarding the second question, while Shain and Tonhauser (2010) provide evidence of the absence of DOM in the early stages of contact with Spanish, the present study provides an account of how this pattern emerged and spread in Guaraní. It attempts to demonstrate that the clues of contact can be found in natural speech, as suggested by both usage-based approaches and variationist approaches to language change.

With respect to the mechanisms of borrowing, this study hypothesizes that the Guaraní DOM was borrowed as an element of a Spanish construction and not as an independent feature. This borrowed Spanish construction included a transitive verb, a human direct object, and the DOM: [V DOM NHuman]. It is hypothesized that the borrowing of the DOM Spanish construction can be attested synchronically in Guaraní,
by showing that the presence of loan verbs or loan nouns will predict the occurrence of the DOM.

The sentences (93) and (94) below—produced by the same speaker, a 31-year-old man from a rural area—exemplify our predictions. Both utterances include the phrase ‘to help our farmer(s)’ which, in turn, include the loan *agrikultor*, from Spanish *agricultor* (‘farmer’). However, in (93), the speaker uses the loan *ajuda*, from Spanish *ayudar* (‘to help’), while in (94) he uses the native-origin counterpart *pytyvõ*. The object marker –*pe* occurs in (93), the sentence that includes the loan verb, but not in (94), the sentence with the native-origin verb.

(93) *Nd-aipó-ri peteĩ politika en sí*

NEG-there.is-NEG one policy in itself

‘There isn’t a policy in itself...

{o-ajuda-haguã-icha ſande agrikultor-es-kuéra-pe}

3-help-to-as our farmer-PL-PL-OBJ

as to help our farmers.’

(94) *Sa’i [...] o-me’ẽ-a la govierno la*

Little 3-give-NMLZ the government NMLZ

‘It is little what the government gives...

{o-i-pytyvõ-haguã ſande rapicha agrikultor}

3-help-to our fellow farmer

to help our fellow farmer.’
Likewise, examples (95) and (96) below refer to the action of raising children, both include the native-origin verb *mongakuua* (‘raise’). However, (95)—produced by a 68-year-old woman from the city—refers to children using the native-origin noun *membykuéra*, while (96)—produced by a 33-year-old man from the same neighborhood—uses the loan *família*, from Spanish *familia* (‘child’). Of the two sentences, only (96) includes the marker –*pe*, which suffixes the loan.

(95) *E-mongakuuaa-porã la ne memby-kuéra.*

2-raise-well the your child-PL

‘You raise your children well.’

(96) *O-mongakuuaa-pa-ite la i-família-kuéra-pe.*

3-raise-COMPL-AUG the 3-child-PL-OBJ

‘He raised all his children.’

Thus, following examples (93) to (96) above, we predict a loan verb or an object noun, or the combination of both, will favor –*pe* (the dependent variable).

In sum, this study hypothesizes that:

1. DOM in Guaraní is a contact feature, that is, a feature that emerged from contact with Spanish.
2. Guaraní borrowed the Spanish DOM construction [V DOM OBJhuman], which led to the grammaticalization of the locative/dative marker into a human direct object marker.

5.4. Methods

This study explores DOM in the corpus of spoken Guaraní. All tokens with full noun human direct objects were manually selected and copied onto a new spreadsheet. Previous studies (Bossong 2009; Shain and Tonhauser 2010) and preliminary exploration of this study’s data showed that –pe is not used with inanimate objects and, thus, there is no variation with respect to its use. On the contrary, human direct objects show variation, as they can be suffixed with –pe or not suffixed at all. Thus, only full human object noun phrases were included in the statistical analysis. Also, as the purpose of this study is to test the influence of Spanish on Guaraní, in addition to having a human direct object, the selected tokens had to meet the following criteria: their translation to Spanish had to include a; and the use of –pe had to be variable. For the first criterion, I used my Spanish native speaker intuition to decide whether the translation required a DOM a or not. For example, sentence (97a) includes a human direct object without the marker. However, its translation to Spanish (97b) would not include the marker either, because the woman is talking about wanting a generic partner, not a specific one. Tokens like these one were not included. For the second criterion, I consulted with native speaker Mr. Zena to verify that both –pe and lack of –pe were allowed in each token.
(97a) Maerã-piko che ai-pota kompañero?
   Why-INT I 1-want partner
   Why would I want a partner?

(97b) ¿Por qué querría yo (*a) un compañero?

Interestingly, the marker –pe never occurred in a sentence that would not include the marker a in Spanish. The opposite, however, was true for some sentences, that is, –pe was not allowed in sentences whose Spanish translation would require a, per my consultant Mr. Zena. These tokens comprised sentences with nominalized transitive verbs, and sentences with three human participants. Both types of sentences did not allow the DOM marker, and thus, they were not included in the statistical analysis, as they are not variable. However, because they show interesting patterns, some examples are provided below.

Data show that –pe is not used with nominalized verbs which take direct objects, like in (98), where grupo de hóvenes (‘youth groups’) is the direct object of the nominalized verb omba’apóa (‘who work’). The DOM in this sentence cannot occur after the nominalized verb. However, –pe occurs, optionally, with nominalized verbs that do not take complements, like in (99) and (100). In (99), the marker suffixes the nominalized verb oka’úa (‘those who drink’) but in (100), the marker does not suffix the nominalized verb ouramóa (‘those who just came’).
The variable use of the marker in intransitive subordinate clauses and the
categorical unmarkedness of transitive subordinate clauses were confirmed by my
consultant, with whom I elicited sentences like (101)—with an intransitive subordinate
clause—and (101a)—with a transitive subordinate clause—, followed by its
ungrammatical versions (101b) and (101c).

(101) **Ei-kuaa pe mitarusu o-pitá-a(-pe).**

2-know that young.man 3-smoke-NMLZ(-OBJ)

‘You know that young man who smokes.’ (ED)
(101a) *Ei-kuaa pe mitarusu o-pítá-a mariuána.
2-know that young.man 3-smoke-NMLZ marihuana
‘You know that young man who smokes marihuana.’ (ED)

(101b) *Ei-kuaa pe mitarusu o-pítá-a-pe mariuána.
2-know that young.man 3-smoke-NMLZ-OBJ marihuana

(101c) *Ei-kuaa pe mitarusu o-pítá-a mariuána-pe.
2-know that young.man 3-smoke-NMLZ marihuana-OBJ

This Guaraní pattern differs from Spanish in that the Spanish DOM a is not
discouraged when the DO is a transitive subordinate clause. The marker is used with both
intransitive and transitive subordinate clauses, as in (102) and (103), respectively.

(102) Conoc-es a ese joven que fum-a.
know-2.PRES OBJ that young.man who smoke-3.PRES
‘You know that young man who smokes.’ (ED)

(103) Conoc-es a ese joven que fum-a marihuana.
know-2.PRES OBJ that young.man who smoke-3.PRES marihuana
‘You know that young man who smokes marihuana.’ (ED)
Thus, in general, Guaraní DOM is not used with transitive subordinate clauses. However, there is an exception to this pattern. When the DO in the subordinate clause has a human referent, the marker can be used as a disambiguation device. In examples (104) and (105) –pe indicates the object of the dependent clause: in (19) it is *tia* (‘aunt’); in (20) it is *mitarusu* (‘young man’). The disambiguation function is of the DOM is further evidenced by the ungrammatical example (106), in which both nouns take the object marker.

(104) *Ai-kuaa pe mitarusu o-hayhu-a che tiá-pe.*

1-know that young.man 3-love-NMLZ my aunt-OBJ

‘I know that guy who loves my aunt.’ (ED)

(105) *Ai-kuaa pe mitarusu-pe o-hayhu-a che tia.*

1-know that young.man-OBJ 3-love-NMLZ my aunt

‘I know know that guy whom my aunt loves.’ (ED)

(106) *Ai-kuaa pe mitarusu-pe o-hayhu-a che tiá-pe.*

1-know that young.man-OBJ 3-love-NMLZ my aunt-OBJ

The above-mentioned Guaraní pattern differs from Spanish in that a translation of (106) would require the use of the marker before both human direct objects (106S).
Another interesting event type where the marker is not used in Guaraní is in simple transitive constructions with three human participants. Likewise, in many dialects of Spanish, when human DO and human IO are present in a sentence, marking the human DO is discouraged (Moreno-Fernández, Penadés-Martínez, and Ureña-Tormo 2019), as seen in example (107a) and its ungrammatical version (107b). The Guaraní translations of these sentences also discourage the direct object marker, as seen in (108a) and (108b).

(107a)  *Le  d-i  a  mi  hij-o  a  mi  tía.

IOBJ give-1.PST OBJ my child-MASC IOBJ my aunt

I gave my son to my aunt

(107b)  *Le  d-i  a  mi  hij-o  a  mi  tía.

IOBJ give-1.PST OBJ my child-MASC IOBJ my aunt

(108a)  A-me’ê  che  memby  che  tiá-pe. (ED)

1-give my child my aunt-IOBJ

(108b)  *A-me’ê  che  memby-pe  che  tiá-pe.

1-give my child-OBJ my aunt-IOBJ
Corpus data also evidences the absence of the DOM in events with three human participants (examples (109), (110), and (111)).

(109) \[ O-u \ o-raha-pa \ ndehegui \ la \ mita. \]

3-come 3-take-COMPL from.you the child

‘They come and take all your children from you.’

(110) \[ O-heja \ la \ mita \ pe \ kuñatai-me. \]

3-leave the child that lady-IOBJ

‘She leaves the child to that lady.’

(111) \[ A-guerú-ta \ peē-me \ pende \ arpista-rā, \ petei \ mita-i. \]

1-bring-FUT 3.PL-IOBJ your.PL harpist-FUT a child-DIM

‘I will bring you your future harpist, a kid.’

Interestingly, the DOM is allowed in constructions where the third participant, the receiver, takes the –ndi postposition, equivalent to Spanish con, as seen in example (112).

(112) \[ E-heja \ la \ nde \ memby \ michi-a-pe \ la \ iñ-erman-o \ majór-ndi \]

2-leave the your child little-NMLZ-OBJ the 3-sibling-MASC older-with

‘You leave your little child with their older brother..."
Finally, the variable tokens (those where the DOM was optional) were coded. The dependent variable was the presence or absence of the DO marker. The linguistic independent variables included were: linguistic origin of the object noun: Guaraní or Spanish, and linguistic origin of the verb: Guaraní or Spanish. In addition to the mentioned linguistic factors, demographic information of the speaker—age (continuous), sex (female or male), and location (rural or urban)—were included. With respect to location, the variable rural versus urban was included because, traditionally, rural areas have been more Guaraní-dominant while urban areas have been more Spanish-dominant (Rubin 1968). Thus, location could have an effect on the production of any presumed contact-induced change: if the feature is more used in urban areas, where Spanish is more widely used, a contact effect argument could be made.

In summary, the variables that were accounted for in the study were:

- the dependent variable: absence vs. presence of DO marker –pe
- social independent variables: age (18-85), sex (male or female), location (rural or urban),
- linguistic independent variables: origin of the predicate’s object noun (Guaraní or Spanish), linguistic origin of the predicate’s verb (Guaraní or Spanish).

Finally, all interviewee-produced tokens (613) were correlated with the above-mentioned variables through a logistic regression in R-studio.
5.5. Results

Of the 613 tokens produced by the interviewees, 44% were marked with \(-pe\). The logistic regression showed age of the speaker \((p < 0.001)\), and language origin of the object noun \((p = .01)\) as the statistically significant factors favoring the overt marking of the direct object. Details from the regression are shown in Table 8. On the one hand, the younger the speaker, the higher their usage rate of \(-pe\), as seen in Figure 9. On the other hand, when the object noun is a loan from Spanish, instead of a native-origin noun, the marker is more likely to occur.

Figure 9. Rate of DOM usage per speaker age, from older to younger

Overall, the other social factors which were not hypothesized as significant—location and gender—did not have an effect on the dependent variable. The correlation of
age with the presence of the DOM was unexpected. However, this finding is compatible with our predictions, as it will be explained in the analysis section.

With respect to internal factors, it was predicted that the presence of loan verbs and Spanish-origin object nouns would favor the occurrence of –pe. This was true only for the object nouns, as the presence of loan verbs did not have an effect on the marker (at least, not synchronically).

A further exploration of the data showed that certain verbs, regardless of origin, correlate with a higher presence or absence of the marker. For example, ‘help’, ‘respect’, and ‘listen’ showed a very high rate of DOM presence (80-100%) while ‘take’, ‘correct’, and ‘put’ show a low presence of DOM (0%-7%). Likewise, it was found that direct objects whose referent was young people (e.g., ‘my children’) were marked less often than objects that referred to adults, 25% compared to 55%, respectively.

|                      | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 0.966050 | 0.354232   | 2.727   | 0.00639 ** |
| sexMale              | 0.239504 | 0.180809   | 1.325   | 0.18530   |
| locationU            | 0.030445 | 0.174068   | 0.175   | 0.86115   |
| age                  | -0.028253| 0.005376   | -5.255  | 1.48e-07 *** |
| ObjetoS              | 0.435645 | 0.172585   | 2.524   | 0.01160 *  |
| VerboS               | 0.106409 | 0.185932   | 0.572   | 0.56712   |

*Table 8. DOM logistic regression results*
5.6. Analysis

The logistic regression showed that –pe was more likely to occur in the presence of borrowed object nouns, but not in the presence of loan verbs. However, the most significant predictor in the occurrence of the marker was the age of the speaker. Both of these results provide further evidence for –pe as a contact feature in Guaraní, and support our contact hypothesis.

First, the fact that a loan object noun favors the occurrence of the Guaraní human object suffix suggests that the DOM is a contact feature. The borrowed lexical item proves the existence of a contact phenomenon in the sentence, that is, in an utterance like *oheja iprofesor* (‘leave his teacher’), which includes the loan *profesor* (‘teacher’), contact is self-evident. Results indicate that the presumed borrowed grammatical feature is used more frequently in sentences like these, where contact is evident, and thus, we argue the DOM emerges in this type of context. Put differently, if –pe was not a contact feature, it would not matter whether the suffixed noun was of native or Spanish-origin. This pattern is supported by usage-based approaches to contact-induced change, which posit that items that are frequently used together may be borrowed together (Backus 2013) regardless of whether the items are grammatical or lexical. Furthermore, it was the object loan and not the verbal loan that had an effect on the object marker, a pattern which can also be accounted for by co-occurrence. In Guaraní, the DOM is a suffix; the verb occurs further from it than the suffixed noun. Thus, when borrowing a noun that occurs with a case marker, the speaker might borrow the schematic unit [DOM OBJ]. Another explanation is that verbs played a role in the emergence of the pattern, as we hypothesized the DOM was part of a ‘larger’ loan, the DOM Spanish construction: [V
DOM OBJhuman]. However, over time, the DOM extended to verbs of native-origin, and thus, the verb effect was lost.

Second, in Paraguay, age correlates with degree of bilingualism, therefore, age being a highly significant predictor of the DOM’s presence is another piece of evidence that the feature is replicated from Spanish. The younger the person, the more likely it is they are self-assessed bilingual. In addition, according to the censuses of 2002 and 2012, the percentage of monolingual Spanish speakers increased by 5% (from 10% to 15%) (Paraguay 2003; Paraguay 2016). During fieldwork, this generational difference with respect to language preferences was highly evident. In rural areas, while older speakers would not speak Spanish to me, even knowing it is my dominant language; younger speakers would address me in Spanish when meeting me for the first time. In urban areas, it was the younger speakers who would code-switch the most before, during and after the interviews. If DOM was not a contact feature, the speakers’ language dominance (evidenced in their age) would not have an effect on their DOM usage rate.

As for the second question in our study—how does the DOM enter the Guaraní grammar?—, results suggest that Guaraní speakers did not borrow -pe alone, but as an item in the Spanish DOM construction [V DOM OBJ]. At the beginning, this Spanish construction included loanwords. In turn, the borrowing of this construction led to the replica grammaticalization of the Guaraní locative/dative marker to a human direct object marker. As part of its on-going conventionalization (and as bilingualism increases in Paraguay), the Guaraní DOM construction ‘accepts’ more native-origin verbs and object nouns. This pattern is expanded by younger speakers as not only they include more Spanish loanwords into Guaraní but also extend its use to native-origin nouns.
Another finding in our study is the effect of verb semantics and relative animacy in the occurrence of the DOM. These factors were not included in the logistic regression, as the tokens numbers vary widely across verbs. Nonetheless, the patterns found provide further explanations on the DOM in Paraguayan Guaraní but also give insight for future research. With respect to verb semantics, it was found that, across verbs (loans and non-loans) with five tokens or more, there was a wide range of DOM usage. On one end, three verbs categorically correlated with the absence of the marker: ‘correct’, ‘put’, and ‘gather’. On the other end, one verb categorically correlated with the presence of the marker: ‘help’. With respect to relative animacy, object nouns were more likely to occur with the DOM when they had adult referents than child referents, as summarized in Table 8.

Table 9. Percentage of DOM usage according to type of object referent

<table>
<thead>
<tr>
<th>object referent</th>
<th>Ø</th>
<th>-pe</th>
<th>total</th>
<th>% marked</th>
</tr>
</thead>
<tbody>
<tr>
<td>adult</td>
<td>170</td>
<td>211</td>
<td>381</td>
<td>55%</td>
</tr>
<tr>
<td>child</td>
<td>173</td>
<td>59</td>
<td>232</td>
<td>25%</td>
</tr>
</tbody>
</table>

We believe these two patterns, verb semantics and relative animacy, are better accounted for when explored together. For example, the loan korrehi (‘correct’) has ten tokens, in all of them the object refers to children: mitã (‘child’), família (‘offspring’), alúmeno (‘student’) and none of them gets the case marker. Thus, it could be argued that the phrase (and not the verb or the noun alone) ‘correct + child’ discourages the marker. In this type of event, it would be unexpected to have a child correct an adult. As proposed
by Comrie (1989) for universal patterns, Delbecque (2002) for Spanish, and Bossong (2009) for Guaraní, a marker would be expected when the object is higher in animacy than the subject. In this case, it could be argued that while children and adults are both equally animate, the prototypical correcting event would have an adult subject and a child object. If these roles are reversed, a less prototypical relation is construed and therefore the marker is needed, but tokens like ‘children correct their parents’ are not available in the data. However, other verbs can illustrate the occurrence of the marker when, in the event, both subject and object can perform the action. The sentences below include feeding events (the corpus has five tokens which include the verb ‘feed’). When an adult feeds another adult, as in example (113), the marker is used in the two tokens, but when an adult feeds a child, like in example (114), the marker is only used in one out of the three tokens.

(113) Ro-mongaru-haguã mamá-pe umía-pe merkádo-pe ro-mba'apo.

1EXCL-feed-to mom-OBJ those-OBJ market-in 1EXCL-work

‘We work at the market to feed mom and others.’

(114) A-mongaru che família.

1-feed my child

‘I feed my child.’

These patterns provide further evidence that DOM is favored in the presence of a loan. Even when the referent of the object is a child (or children)—and thus, the marker is
dispreferred—a loan with a child human referent increases the use of the marker. While only 10% of native-origin object nouns with child referents are used with the marker, loans with child referents are marked 38% of the time, as summarized in Table 9.

Table 10. Percentage of DOM usage with children referents

<table>
<thead>
<tr>
<th>object noun type</th>
<th>Ø</th>
<th>–pe</th>
<th>total</th>
<th>% marked</th>
</tr>
</thead>
<tbody>
<tr>
<td>native-origin</td>
<td>94</td>
<td>10</td>
<td>104</td>
<td>10%</td>
</tr>
<tr>
<td>loan</td>
<td>28</td>
<td>17</td>
<td>45</td>
<td>38%</td>
</tr>
</tbody>
</table>

Even though relative animacy has been described as significant in DOM patterns across the worlds’ languages (Comrie 1989) and also in Spanish specifically (Delbecque 2002; Tippets 2011), the child vs. adult distinction with respect to human object markers is not found in Spanish. Clearly, the donor and recipient language differ in this aspect of the DOM. However, this difference does not disprove the contact hypothesis, as borrowed items do not necessarily replicate the exact functions they had in the original language (Campbell 2013).

5.7. Discussion

This study has shown that the usage of DOM in Paraguayan Guaraní is favored by social and linguistic factors. With respect to social factors, the higher usage rate of –pe with human objects among younger speakers compared to older speakers shows a change in progress in Guaraní. This convergence of Guaraní and Spanish is not only accounted for by the increasing bilingual population, but also by positive attitudes towards Spanish.
Studies have shown that positive attitudes favor linguistic convergence. (Döhla 2014), for example, showed that the rate of DOM usage in Portuguese increased from the 13th until the 17th century but later declined. Döhla attributes this change in language structure to a shift in attitudes towards Spanish. Linguistic divergence between Spanish and Portuguese coincided with the period of national identity formation in Portugal, which included negative attitudes towards Spain. Thus, in Paraguay, the high value of Spanish, as it is the language for social mobility (Choi 2003), creates a favorable scenario for convergence between Spanish and Guaraní.

As to linguistic factors, overall, the marker is favored when it co-occurs with a Spanish-origin loan object noun and it is disfavored when the events shows the prototypical arrangement of subject having more agency than the object. This last factor explains why objects with child referents (who have less agency than adults) are less marked than nouns with adult referents. The effect that the presence of loans have on the occurrence of the DOM point to contact with Spanish as the origin of the feature.

For this study, 40 hours of recordings were analyzed. Some of these patterns, however, emerged from very few tokens or had many uncontrolled variables. For example, word order could have an effect on the realization of DOM. However, to better study the effect of word order, other factors would need to be somewhat controlled, like the type of verb, the type of object, etc. Thus, while naturally-occurring data provides essential instances of spontaneous linguistic usages that are at the core of language change, their vast variability can limit the analysis. We believe elicitation or experimental methods can be used in future research to verify the patterns that were learned from the data explored in this study.
5.8. Conclusion

This study aimed at exploring DOM as contact feature in Paraguayan Guaraní in a corpus of natural speech. It was hypothesized that: 1. DOM is a feature that emerged from contact with Spanish, and 2. that as a contact feature, the direct object marker would occur more often in the presence of Spanish loanwords. A logistic regression showed that –pe was more likely to occur in the presence of borrowed object nouns, but not in the presence of loan verbs. However, the most significant predictor in the occurrence of the marker was the age of the speaker. Both of these results provide evidence for –pe as a contact feature in Guaraní.

The data explored here indicated that in Guaraní, the object marker –pe grammaticalized from the Guaraní locative/dative marker after the Spanish DOM construction was borrowed. This Spanish construction included a loanword, which explains why DOM is (still) favored by a loan object noun. Thus, the findings of this study point at the need to explore borrowing from a lexical-grammatical continuum approach, instead of a lexical vs grammatical dichotomy, as words and particles that occur together might be borrowed together.
Chapter 6

Third Study: Middle Voice

6.0. Chapter overview

This chapter explores the distribution and functions of the marker \( je \) in loan verbs and native-origin verbs. The goal is to assess whether the functional versatility of \( je \)- is linked with that of its Spanish counterpart \( se \), and with the origin of the verb with which it combines. Section 6.1 summarizes the objectives of the study. Section 6.2 goes over the literature on middle voice in general, and middle voice in both Spanish and Guaraní. Section 6.3 states the research questions and hypotheses. Then, Section 6.4 describes the methodological steps to extract and classify the tokens. Sections 6.5 and 6.6 outline the results and how they are interpreted. Section 6.7 describes the limitations of the study and possible directions for future research. Finally, Section 6.8 summarizes the findings of the study.

6.1. Introduction

Middle voice constructions are a mid-point between prototypical two-participant events and one-participant events (Kemmer 1993). To form these constructions, many languages use an overt middle marker to decrease the valency of the verb, to leave a participant unspecified. In this study, we explore the interaction between two middle markers: Spanish clitic \( se \) and Guaraní prefix \( je \). More specifically, we explore potential effects of the Spanish middle voice construction in the use of Guaraní \( je \)- by comparing the behavior of \( je \)- in Spanish-origin loan verbs and Guaraní-origin verbs in 40 unscripted interviews with Guaraní speakers. The exploration of 2397 tokens of \( je \)-prefixed verbs
shows that *je*- can have several functions, including reflexive, passive, impersonal, and nominalizer. However, the origin of the verb can help predict the function of *je*-: the middle function is more frequent among loans than among native-origin items. Nevertheless, usage patterns suggest that the middle interpretation emerges from the interaction between *je*- and the causative marker *mo/-mbo-*, as almost half of *je*-prefixed native-origin verbs require a causative to have a middle reading. The results add to our previous findings (see Chapter 5) that grammatical categories are not borrowed in isolation but as part or larger constructions.

### 6.2. Theoretical background

#### 6.2.1. From reflexive to middle

Kemmer (1993) proposes that “the reflexive and the middle can be situated as semantic categories intermediate in transitivity between one-participant and two-participant events” (3). In reflexive constructions, there is an agent-like and a patient-like role, but both roles share a referent. In Croft’s (forthcoming) words: “there is one participant in two roles.” Similarly, middle constructions comprise events in which “the Initiator is also an Endpoint” (Kemmer 1993, 209), however, one of the participants is left unspecified (that is, unlike reflexive constructions, the event has a low degree of elaboration). Because reflexive and middle constructions are a mid-point between two-participant and one-participant events, reflexives are usually recruited for middle events (Kemmer 1993). Romance languages, for instance, extended the use of the Latin reflexive pronouns to middle uses. In these languages, the reflexive marker was first extended to non-translational motion verbs (‘turn’), then to grooming verbs (‘bathe’),
then to emotive speech act verbs (‘complain’) and then to emotion/cognition verbs (‘imagine’, ‘get mad’), and it has even extended to translational verbs (‘go’) to some extent (Kemmer 1993). Moreover, Kemmer (1993) found that this path from reflexive to middle in Romance languages is cross-linguistically common. She also hypothesizes that, diachronically, situation types which are closer to the prototypical reflexive, like non-translational motion events, will use a middle marker sooner than those that are less similar to the reflexive, like translational motion events. With respect to implications, she proposes that if a language exhibits middle marking in some translational motion events (e.g. ‘go’), one would expect to find the marking in non-translational motion events as well (e.g. ‘stand’). The diachronic and implicational paths for this study are the following:

reflexive > non-translation motion > grooming > emotive speech > emotion/cognition > change of state : translational motion

Although the change reflexive > reciprocal has also been attested in Guaraní, we believe this phenomenon deserves its own study, and thus, the topic will be only tangentially discussed.

The diachronic and implicational paths are summarized in Figure 10, where the prototypical reflexive is in the center, the starting point of the change; this change expands outwards, where the least reflexive-like situation types are located.
With respect to contact-induced change, Kemmer (1993) states that although secondary, influence from other languages plays a role in the development of middle marking systems, as languages that overtly mark the middle voice are areally clustered.
She adds, however, that the topic of middle systems in contact remains unexplored. Furthermore, it has been argued that morphemes are not borrowed as frequently nor easily as other linguistic items (Thomason and Kaufman 1988), this includes, of course, middle-marking morphemes. Mithun (2013) argues that morphemes’ resistance to borrowing might be due to their tight integration to other linguistic elements (that is, morphemes are not as salient as, for instance, discourse markers). In turn, this tight integration hinders speakers’ ability to establish morphological equivalence in the two languages: a necessary step to borrow items. However, Mithun (2013) claims that some linguistic factors might facilitate the consciousness of morphological structure, like transparency of meaning, degree of morphological fusion, and position of the morpheme. Social factors, like literacy, might also strengthen the language users’ ability to morphologically analyze the donor and the recipient language.

6.2.2. *The functions of je-*

In many descriptions of Guaraní, the verbal prefix *je-* has been described as a reflexive marker (115), a passive marker (116), an impersonal marker (117), and a nominalizer (118) (Gregores and Suárez 1967; Krivoshein de Canese 1983; Velázquez-Castillo 2004; Nordhoff 2004; Estigarribia 2020) (see section 6.3 for definitions of these functions). Nordhoff (2004), Velázquez-Castillo (2007), and Estigarribia (2020) have proposed *je-* can also form middle constructions (119). Estigarribia has grouped all the functions of the morpheme (nominalizer, reflexive, passive, impersonal, and middle) into a single category: agent-demoting voice. However, overall, authors agree in that *je-* (and
its nasal allomorph ňe-) is polysemic, as can be seen from the examples below (glosses were modified by me for consistency):

**Reflexive**

(115)  

*O-je-poi.*

3-REFL-let.go

‘They hurled themselves.’

(Velázquez-Castillo 2004, 1425)

**Passive**

(116)  

*Juan  o-je-juka.*

Juan 3-PASS-kill

‘Juan was killed.’ (It can also mean: ‘Juan killed himself.’)

(Nordhoff 2004, 32)

**Impersonal**

(117)  

*O-je-jeroky.*

3-IMP-dance

‘There is dancing.’

(Estigarrirbia 2020, 208)
Nominalizer

(118)  *Je-vy’a*

NMLZ-rejoice

‘Happiness.’

(Estigarribia 2020, 79)

Middle

(119)  *yvotryru o-je-joka*

vase 3-MID-break

‘The vase breaks.’

(Nordhoff 2004, 82)

The examples above show *je*-prefixed native-origin verbs. While most descriptions of Guaraní do not provide examples of *je*-prefixed loans, this morpheme is also used with Spanish-origin verbs. Gregores and Suárez (1967) provide the following examples:

(120)  *A-je-salva chu-gui*

1-REFL-save 3-OBL

‘I escaped from him.’

(Gregores and Suárez 1967, 214)
(121) *Ndē mbōy áño re-tūdia raka’ē la re-je-resivi haguā?*

you how many year 2-study PST the 2-MID-receive to

‘How many years have you studied in order to get your degree?’

(Gregores and Suárez 1967, 154)

In their original text, Gregores and Suárez (1967) gloss both instances of *je-* as reflexive markers. However, while (120) could have a reflexive reading (‘I saved myself from him’), (121) cannot. Clearly, the subject is not acting on herself (she is not the object of ‘receive’); that is, she is not ‘receiving herself’, she’s getting a degree. Thus, in (121), *je-* appears to be replicating the middle marker function of *se* in Spanish *recibir-se* (receive-MID, ‘graduate’). Example (122) below also shows a case where *je-* has a middle reading (originally glossed as a reflexive), which appears to be replicating the middle marker *se* in Spanish *perder-se* (lose-MID, ‘get lost’).

(122) *A-je-perde*

1-MID-lose

‘I get lost.’

(Gregores and Suárez 1967, 114)

Interestingly, when using the Guaraní-origin verb *kañy* (which also means ‘to hide’ and ‘to disappear’) to convey the GETTING LOST meaning, the middle marker is not allowed (123):
(123)  A-kañy

1-get.lost

‘I get lost.’ (ED)

(124)  * A-ñe-kañy

1-MID-lose

Examples of je-prefixed loans also include a reciprocal function, as seen below:

(125)  hama   na-ñe-komunika-mo'â-i

never  NEG-RECP-communicate-FRU-NEG

la  ña-ñe-komunika  háicha  la  Guaraní-me

NMLZ  1INCL-RECP-communicate  like  the  Guaraní-in

‘They would never communicate the way we communicate in Guaraní.’

(Gómez Rendón 2007, 538)

Gómez Rendón (2007) claims that the reciprocal function of je- might arise by influence of Spanish se, a clitic which conveys, among many, reflexive, passive and reciprocal meanings. Guaraní, however, has a dedicated reciprocal marker, jo-/ño-, as shown in (126).
Thus, examples (120) to (126) suggest that some new functions of je-, in particular middle and reciprocal, might have developed in the context of Spanish loan verbs. The literature on Guaraní, however, does not discuss how these functions of je- evolved, nor how frequent each function is in today’s Guaraní. The hypothesis developed in this chapter is that the reciprocal and medial functions were brought into Guaraní together with a specific type of Spanish loan verbs, the so-called pronominal verbs. However, as we saw in Chapter 4 with the DOM -pe, Guaraní did not borrow the morpheme se itself. In this case, an existing morpheme, je-, underwent analogical extensions diachronically. The use of this morpheme was extended to new situations modeled in the Spanish se patterns.

6.2.3. The functions of Spanish se

Unlike Guaraní je-, the Spanish clitic se has been widely described in the literature (e.g. Maldonado 1993; Sánchez López 2002; Mendikoetxea 2012; Clements 2015). Overall, authors propose that clitic se has the following functions: reflexive and reciprocal (anaphoric), and impersonal, passive, unnacusative, antipassive, middle, and aspectual (non-anaphoric). However, scholars disagree in where they draw the line between functions. For instance, while Clements (2015) proposes that antipassive and
middle *se* should be treated as a single function, Mendikoetxea (2002) argues that these should be treated as discrete functions. For the purposes of this study, we will follow Clements’ (2015) proposition that passive and unaccusative readings are indistinguishable, and that so are antipassive and middle functions. Examples of each function are provided below:

**Reflexive**

(127) *Juan se afeita.*

Juan REFLEX shave

‘Juan shaves (himself).’

(Sánchez López 2002, 15)

**Reciprocal**

(128) *Las vecinas se insultaron.*

The.F neighbors.F RECP insult

‘The neighbors insulted each other.’

(Sánchez López 2002, 15)

**Impersonal**

(129) *Se trabaja mucho en la fábrica.*

IMP works much in the factory

‘One works a lot in the factory.’

(Clements 2015, 248)
Passive and unaccusative

(130)  Se vend-en (los) mueble-s.

PASS sell-3PL (the) furniture-PL

‘(The) furniture is sold.’

(Clements 2015, 244)

Antipassive and middle

(131)  Luisa se enamora fácilmente.

Luisa MID falls-in-love easily

‘Luisa falls in love easily.’

(Clements 2015, 243)

Telic

(132)  Marta se durmió en casa de su amiga.

Luisa TEL sleep.PST at house of her friend.F

‘Marta fell asleep at her friend’s house.’

(Clements 2015, 252)

With respect to middle voice, Maldonado (2008) has followed Kemmer’s (1993) typology in classifying Spanish se into semantic groups. These semantic categories, “middle voice situations” (Maldonado 2008, 164), are the following: grooming (lavarse, ‘wash’), benefactive (conseguirse, ‘get’), non-translation motion (pararse, ‘stand up’),

Synchronously, authors agree that non-anaphoric functions of *se* developed from its function as a reflexive marker (e.g. Maldonado 1993, Clancy Clemens 2015). Kemmer (1993) argues that this path of change is cross-linguistically common, that is, the source of middle voice markers is often reflexive pronouns, as will be explained below.

### 6.3. Research question and hypothesis

This study aims at exploring the morphosyntactic distribution and functions of the verbal prefix *je-* in Guaraní. More specifically, it intends to answer the following question: Has *je-* replicated functions of Spanish clitic *se*, in particular, its middle voice functions? Cross-linguistically, it is common for reflexives to become middle markers. Thus, one possibility is that Guaraní *je-* could have developed middle-marking functions due to internal factors alone. However, we hypothesize that this process may have been supported by language contact. In other words, due to its contact with Spanish, Guaraní has extended the functions of reflexive/passive *je-* to include middle voice marking.

Evidence for this line of analysis is the following:

1. Middle markers are used more often with loans than with native-origin verbs.
2. The change from reflexive to middle has ‘skipped steps’: the middle is not used in all semantic domains (see Section 6.2), as would be expected from a language-internal process, as hypothesized by Kemmer (1993).

In the following sections we provide empirical evidence to demonstrate these two points.
6.4. Methods

To test the hypothesis that middle-marking je- occurs more often with loans than with native-origin verbs, all tokens of je verbs were extracted from the corpus. The 2397 tokens were coded for language origin of the verb (Spanish or Guaraní), function of je-: reflexive, reciprocal, impersonal, passive, nominalizer, and middle. These functions were classified following the definitions proposed by Kemmer 1993, and Croft (forthcoming).

Constructions were considered reflexive when one participant had, simultaneously, an agent role and a patient role (Croft forthcoming), and when a reflexive phrase (e.g. a sí misma ‘herself’) could be added to it (Kemmer 1993). This example in the corpus was classified as reflexive:

(133) ¿Mba'éicha nde e-je-hecha?

How you 2-REFL-see

‘How do you see yourself?’

Indirect reflexives, like the following, were also coded as reflexives:

(134) O-je-po-héi

3-REFL-hand-wash

‘He washed his hands.’
Constructions were classified as reciprocal when a pair of participants was involved in a symmetric relation (Kemmer 1993, Croft forthcoming), and a reciprocal phrase could be added (e.g. *una a otra*, ‘one another’) (Kemmer 1993). This example in the corpus was classified as reciprocal:

(135)  *Upéicha la ro-ñe-konose-akue.*

This NMLZ 1EXCL-RECP-meet-PST

‘This is how we met.’

Impersonal constructions were defined as event-oriented, that is, the event itself is given more salience than any of the participants. This focus on the event differentiates them from the passives, constructions in which the patient is given more salience, which makes passives more participant-oriented (Croft forthcoming).

(136)  ¿*Mba'ëicha o-je'-e?*

How 3-IMP-say

‘How do you say?’

(137)  *Pyà-e-arà e-je-gueru-arà la Pastoreo-pe*

Fast-DEO 2-PASS-bring-DEO the Pastoreo-to

‘You have to be brought to Pastoreo fast.’
Nominalizer constructions were those which used the marker to change an event-denoting word, to a reference-denoting word (Croft forthcoming).

(138) \textit{I-je-tu'u} \textit{la ſe-mba'apo, ¿ajéa?}
\begin{tabular}{ll}
\text{COP- NMLZ} & \text{get.stuck} \\
\text{the NMLZ} & \text{work right}
\end{tabular}

‘The work is difficult, right?’

Constructions were classified as middle when \textit{je-} was used to change the valency of the event, from a bivalent to a monovalent (Croft forthcoming), as in (139) and (140) below, and when they fell under the semantic domains, usually intransitive monovalent events, proposed by Kemmer (1993).

(139) \textit{Ha'ete ku a-ñe-ŋandu-porä-îtrei-a.}
\begin{tabular}{ll}
\text{It seems that 1-MID} & \text{feel-well AUG-NMLZ}
\end{tabular}

‘It seems that I feel very well.’

(140) \textit{Che upéa a-ŋandu-akue.}
\begin{tabular}{ll}
\text{I that 1-feel-PST}
\end{tabular}

‘I felt that.’

Then, to test the hypothesis that the middle marker has skipped stages in the path from reflexive to middle (as seen in Section 6.2), sentences containing middle readings were classified following Maldonado’s (2008) middle situation types for Spanish:
grooming, benefactive, non-translational motion, translational motion and body posture, emotional change, emotive speech, cognition, and spontaneous change of state.

**Grooming**

(141) O-ńe-monde ha o-u eskuéla-pe.

3-MID-dress and 3-come school-LOC

‘He gets dressed and he comes to school.’

**Benefactive**

(142) Priméra ves o-sē-rō-guare-kuri, o-ńe-venefisia ha'e-py

First time 3-leave-COND-when-PST 3-MID-benefit 3-DM

‘The first time it came out, well, he benefitted.’

**Non-translational motion**

(143) A-je-volea a-maña hese.

1-MID-turn 1-look him

‘I turned to look at him.’

**Translational motion**

(144) Ha pépe ha'e o-ho o-je-pasea.

and there he 3-go 3-MID-take.walk

‘And he goes there to take a walk.’
Emotional change

(145) *Upéi-ngo a-ñe-trankilisa-ma.*

then-EMPH 1-MID-calm-COMPL

‘Then I calmed down.’

Emotive speech

(146) *Nda-ikatu-i-ko ja-je-keha.*

NEG-can-NEG-EMPH 1INCL-MID-complain

‘We can’t really complain.’

Cognition

(147) *N-o-ñe-konfundi-mo’ā-i ha’e*

NEG-3-MID-confuse-FRU-NEG he

‘He won’t get confused.’

Spontaneous change

(148) *La pakova umía-ngo pya’e o-ñe-mbyai-py.*

The banana those-EMPH fast 3-MID-rot-DM

‘Bananas and other [fruit], well, they rot fast.’

The checklist for middle semantics in Kemmer (1993) was also used as a guide for classifying tokens. Following this checklist, two middle categories were added:
middle reflexives and middle reciprocals. Middle reflexives included behavior, self-protection and self-harm verbs, and middle reciprocals included group events.

*Middle reflexive*

(149) A-je-porta vai.

1-MID-behave bad

‘I misbehave.’

*Middle reciprocal*

(150) Ña-ñe-unì-arà.

1INCL-MID-unite-DEO

‘We have to unite.’

After all the tokens of *je* verbs were classified, the frequency of *je*- functions and middle situation types was counted for each verb, and for each verb class (Spanish-origin vs. Guaraní-origin).

In Guaraní, there are some *je*+stem verbs that appear to be lexicalized, like *jerovìa* (‘trust’), from *rovia* (‘believe’). These lexicalized *je*+stem verbs were not considered in the study because the function of *je*- is not transparent. However, they will be considered in future studies about this topic.
6.5. Results

A total of 2397 tokens of je verbs were extracted from the corpus, 1031 were of Guaraní origin and 1366 were of Spanish origin. 552 types were identified: 161 of Guaraní origin and 391 of Spanish origin.

6.5.1. Overall functions of je- (loans and non-loans combined)

As seen in Figure 11, based on type frequency, the functions of je- were distributed as follows: middle (42%), passive (27%), impersonal (18%), nominalizer (6%), reflexive (4%), and reciprocal (3%).

Figure 11. Distribution of je verbs across functions, loans and native-origin verbs combined

6.5.2. Functions of je- by verb type (loans vs. non-loans)

However, when the distribution of je- functions is explored separately by type of stem (loan or native-origin), there are three functions that show differences in their distribution: middles, impersonals, and nominalizers. As summarized in Figures 12 and
13, the middle marker is the most frequent function of *je*- in Spanish-origin verbs, that is, loans with middle readings account for 51% of all *je*-prefixed loans. Among Guaraní-origin verbs, middles account for 24% of all native-origin *je*-prefixed verbs. Impersonals also differ across verb types: they are the second most frequent function among native-origin verbs (26%), but the the third most frequent among loans (15%). Finally, nominalizers are the fourth most frequent function among native-origin verbs (16%), but one of the least frequent among loan verbs (2%). The remaining functions (passive, reflexive, and reciprocal) account for similar percentages for both native-origin and borrowed items. In inherited verbs, passives account for 27%; reflexives, for 4%; and reciprocals, for 4%. In loans, these functions account for 27%, 3%, and 2%, respectively.

Figure 12. Distribution of *je*-prefixed native-origin verbs across functions

Figure 13. Distribution of *je*-prefixed loan verbs across functions

6.5.3. Verbs organized by semantic types

With respect to middle marking, there were 1088 tokens of *je* that had this function: 220 came from prefixed inherited verbs, and 868, from prefixed loans. These tokens were distributed across 316 verb types: 66 of Guaraní origin, and 250 of Spanish
origin (verb types that occurred in more than one function were counted once for each function). As shown in Figure 14, based on type frequency, the types of middles for all verbs are, from most to least frequent: change of state (25%), middle reflexive (22%), emotional change (17%), cognition (11%), benefactive (7%), translational motion (4%), emotive speech (4%), non-translational motion (4%), reciprocal (3%), grooming (2%). As seen in Figures 15 and 16, the distribution is very similar when exploring Guaraní-origin verbs and Spanish-origin verbs separately.

*Figure 14. Distribution of je verbs across middle semantic domains, loans and native-origin verbs combined*

<table>
<thead>
<tr>
<th>Verb types distributed across middle semantic domains</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>benefactive</td>
<td>4%</td>
</tr>
<tr>
<td>change of state</td>
<td>22%</td>
</tr>
<tr>
<td>cognition</td>
<td>11%</td>
</tr>
<tr>
<td>emotive speech</td>
<td>17%</td>
</tr>
<tr>
<td>grooming</td>
<td>7%</td>
</tr>
<tr>
<td>emotional change</td>
<td>4%</td>
</tr>
<tr>
<td>non-translational motion</td>
<td>4%</td>
</tr>
<tr>
<td>reciprocal</td>
<td>3%</td>
</tr>
<tr>
<td>reflexive</td>
<td>2%</td>
</tr>
<tr>
<td>translational motion</td>
<td>25%</td>
</tr>
</tbody>
</table>

*Figure 15. Native-origin verbs' distribution of je verbs across middle semantic domains*
6.5.3.1. Middle reflexive

For Kemmer (1993), the middle reflexive category includes self-protection, involuntary self-harm and behavior verbs. The following verbs that belong to these categories were found in the data:

The Guaraní-origin verbs were classified into Kemmer’s (1993) and Maldonado’s (2008) middle semantic types. These verbs are described below in order of more reflexive-like to least reflexive-like.
**Involuntary self harm:**

- **akājoka** break one’s head
- **pysānga** trip
- **mbosarāi** fool oneself
- **juvy** hurt one’s throat

**Self-protection:**

- **mbopiro'y** refresh oneself
- **pohāno** medicate oneself
- **moï** put oneself in a situation
- **mbokoty** get into one’s room
- **mboty** lock oneself

**Behavior:**

- **ma'ēra** behave
- **hecha** meet (lit. see oneself with other people)
- **raha** get along

---

2 Per my consultant, *ma'ēra* is used by speakers when they are hesitant about which verb to use. Here it is glossed with the intended meaning.
At least two of the above-mentioned verbs appear to be calquing Spanish constructions: *jehecha* (verse, ‘meet’) and *jeraha* (*llevarse*, ‘get along’)

(151) *Che a-je-raha-porã mamá-ndi.*

I 1-MID-take-well mom-with

‘I get along with my mom.’

Spanish:

(152) *Yo me llev-o bien con mamá.*

I 1.MID take-1.PRES well with mom

I get along with my mom.’ (ED)

(153) *Luis katu a-je-hecha-ití hendie.*

Luis EMPH 1-MID-see-DUR 3.with

‘Luis, I still see him.’

Spanish:

(154) *Luis, me ve-o todavía con él.*

Luis 1.MID see-1.PRES still with him

‘Luis, I still see him.’ (ED)

6.5.3.2. Benefactive
Benefactive middles include verbs of coming into possession. Interestingly, many of the middle benefactive verbs in the data include the causative prefix *mbo-*/*mo-*. 

- **guakatey** appropriate
- **haihuka** get people to love you
- **hogapo** build a house for oneself
- **mohembireko** get oneself a wife
- **mohenda** get oneself a place to live
- **mokuña** get oneself a woman
- **me'ê** give oneself something

(155)  
\[Ro-ho \ péi\text{ch}a \ ro-\text{n}e-mo-henda.\]  
1EXCL-go like.this 1EXCL-MID-CAUS-place  
‘We went, like this we got ourselves a place.’

Other verbs of coming into possession do not take the prefix *je-*, like *jogua* (‘buy’). 

(156)  
\[A-jogua \ che \ r-\text{ó}ga-rā.\]  
1-buy my 1-house-FUT  
‘I bought myself a house.’ (ED)

(157)  
\[^{\ast}A-je-jogua \ che \ r-\text{ó}ga-rā.\]  

143
1-MID-buy  my 1-house-FUT

‘I bought myself a house.’

6.5.3.3. Middle reciprocal

Middle reciprocals include naturally reciprocal or group events (Kemmer 1993).

There are two Guaraní-origin verbs which express this type of meaning. They both appear to be replicating Spanish patterns.

_moĩ de akuérdo_  come to an agreement

_raha_  get along

(158)  _Ro-ñe-moĩ-mba-voi de akuérdo ha ro-sē-nte-ma_

1EXCL-put-COMPL-EMPH of agreement and 1EXCL-leave-just-COMPL

‘We all came to an agreement and we just left.’

(159)  _Ndo-ro-je-raha-porā-ve-i-ma-voi-kuri ore_

NEG-1EXCL-MID-take-well-DUR-NEG-COMPL-EMPH-PST we

‘We didn’t really get along anymore.’

Other reciprocal event verbs do not take the middle prefix, like _kovai_ (‘fight’)

(160)  _Ore-ngo ndo-roi-kovai-ri-akue._

We-EMPH  NEG-1EXCL-fight-NEG-PST
‘We didn’t fight.’

6.5.3.4. Grooming

One verb in the data belonged to the grooming domain:

*monde* get dressed

Other grooming verbs, like *jahu* (‘bathe’) do not take the prefix *je*-

(161) *Che a-jahu y ro'ysã-pe.*

I 1-bathe water cold-LOC

‘I bathe in cold water.’

(162) *A-je-jahu*

1-MID-bathe

‘I bathe.’

6.5.3.5. Non-translational motion

Non-translational motion and change in body posture also correlate with the occurrence of the causative *mbo-*/mo- in native-origin verbs. The tokens of *mombo* and *ty* (‘throw’) appear to be replicating Spanish tirarse, as seen in examples (45) to (48).

*mbohevipuku* making one’s tail longer

*mbovyvy* shake (oneself)
moï put oneself somewhere

mombo throw oneself

mongu'e move (oneself)

mopu'â get up

ty throw oneself

(163) Ro-jei-ty-arâ ápe.
1EXCL-MID-throw-FRU here

‘We would land here.’ (lit. throw ourselves)

Spanish:

(164) Nos tiramos acá.
1.PL.MID throw-1.PL.PRES here

‘We would land here.’ (lit. throw ourselves) (ED)

(165) A-ñe-mombo-nte-ma la píso-pe, mbararâmpe a-guejy.
1-MID-throw-just-COMPL the ground-in ONOMAT 1-get.down

‘I just jumped onto the ground, I got down rat-a-tat.’

Spanish:

(166) Solo me tir-é al piso, tararan me bajé.
Just 1.MID throw-1.PST to ground ONOMAT 1.MID get.down

‘I just jumped onto the ground, I got down rat-a-tat.’(ED)
Other verbs of body posture do not take the prefix je-, like *gwejy* (‘get down,’ seen in example above), *pu’ā* (‘get up’), *guapy* (‘sit’), *ñeno* (‘lie down’).

6.5.3.6. Translational motion

Translational motion events convey change of location of the subject, but there are very few Guaraní-origin verbs that allow the prefix je- to convey the middle meaning. In fact, the verb *(ro)va* listed below is also used without the prefix.

```
ma'ēra  get close
mboja  get close
mbotyriry  crawl
(ro)va  move
```

(167)  *A-pensa che la a-je-rova mba'ēro*

1-think I NMLZ 1-MID-move or.what

‘I thought what if I move or what.’

(168)  *Upēi-nte ro-va*

Afterwards-only 1EXCL-move

‘We only moved afterwards.’
6.5.3.7. Emotional change

There are some Guaraní-origin verbs that use je- to mark emotion middles. This is a typical semantic domain for middle verbs (Kemmer 1993). In our data, these verbs also tend to co-occur with the causative.

- **myasê**: start to cry
- **haru**: get spoiled
- **hasa**: go through
- **mbopochy**: pretend to be mad
- **mbotuicha**: get big
- **moakate'ŷ**: be petty
- **mokachiái**: pretend to be funny
- **mombarete**: get strong
- **mondýi**: get scared
- **mopu'â**: get strong (lit. get up)
- **ñandu**: feel
- **pokuua**: get used to
- **py'apy**: get worried

However, many Guaraní-origin verbs do not occur with je- to express emotion, like vy’a (‘get happy’), kyhyje (‘be afraid’), mbyasy (‘grieve’). Also, some emotion events are conveyed with a so-called inactive construction, like pochy (‘get angry’).
(169) A-je-py'apy-eteri che memby-kuéra-re.
1-MID-worry-AUG my child-PL-for
‘I worry a lot about my children.’

(170) Che-pochoy sapy'ánte.
1INACT-angry sometimes
‘Sometimes I get angry.’

(171) *A-je-pochoy
1-MID-angry
‘I get angry.’

6.5.3.8. Emotive speech

There was only one token of emotive speech being expressed with a je verb, ‘praise’. Other Guaraní-origin verbs of emotive speech, like japu (‘lie’) do not use the prefix.

mombe’uguasu praise

6.5.3.9. Cognition

Many of the Guaraní-origin verbs that express cognition middle are presumably replicating Spanish constructions, as can be seen from the examples.
hasa  cross a line
hesape'a  wake up
rovia  believe
mandu'a  remember
me'ê en kuénta  realize
 tavyporu  use one’s ignorance
 mbotavy  pretend to be a fool
hecha  see oneself (fix one’s problem on their own)

Many Guaraní-origin cognition verbs, like resarái (‘forget’), and mandu’a (‘remember’) belong to inactive constructions. Mandu’a, for instance, is more frequently used in inactive constructions, like the one below:

(172)  N-a-che-mandu'a-vé-i-ma
NEG-1-1.INACT-remember-DUR-NEG-COMPL
‘I don’t remember anymore.’

However, some speakers in the data have produced mandu’a with the prefix ſhe-. Some speakers would find this ungrammatical, and it could be argued this active construction of mandu’a is a calque from Spanish.
Perhaps one of the most interesting cognition events is ‘realize’, which is expressed with the phrase ňeme’ē en kuénta, which replicates Spanish darse cuenta but it adds en. This insertion is possibly influenced by a similar Spanish expression tener en cuenta (‘have in mind’).

(174) Kóa che n-a-ñe-me'ē-i en kuénta.
This I NEG-1-MID-give-NEG in account
‘I didn’t realize this.’

6.5.3.10. Change of state

Spontaneous change of state is the most frequent middle event expressed by Guaraní-origin verbs. It also co-occurs often with the causative.

ma'éra change
mboja'o divide
mbojeka break
mbojy cook
mboka'i rot
\textit{mboty} close
\textit{mbyai} rot
\textit{mokuñataĩ} become a woman
\textit{moñepyru} get something started
\textit{mongarai} become a man
\textit{poi} get loose
\textit{pyso} extend

However, the use of \textit{je-} with this type of verb is far from obligatory. In fact, many spontaneous change events do not take the \textit{je-} prefix, like ‘burn’, or ‘break’.

(175) \textit{Ro-gueru la rrosão, roi-ty ro-koĩ,}
1EXCL-bring the logs 1EXCL-throw 1EXCL-get.together
‘We bring the log, we throw them, we get together...
\textit{ro-hapy, upéi o-kái-pa-porā.}
1EXCL-burn then 3-burn-COMPL-well
we burn them, then they burn completely.’

(176) \textit{I-kadéra o-joka.}
3-hip 3-break
‘She broke her hip.’
6.5.4. Summary of results

On one hand, as predicted, results showed that the middle voice function is more frequent in je-prefixed Spanish-origin verbs than in Guaraní-origin verbs. On the other hand, it was predicted that the native-origin verbs were going to exhibit a middle function in certain semantic domains not in all of them. Contrary to this hypothesis, there was at least one je-prefixed Guaraní-origin verb in all middle semantic categories.

Another interesting finding was that of the 57 je-prefixed Guaraní-origin verbs with middle readings, 23 occurred with the causatives mbo/-mo- and -ka. Furthermore, eight verbs appear to be part of borrowed fixed Spanish expressions, like poner-se de acuerdo (‘come to an agreement’) and dar-se cuenta (‘realize’).

6.6. Analysis

The recruitment of reflexive markers for middle voice functions is cross-linguistically common. We had hypothesized, however, that the middle-marking function of “reflexive” je- in Guaraní was a result of contact with Spanish, which extensively uses “reflexive” se for middle functions. Important pieces of evidence come from the functions this marker has in the corpus and the verbs with which it combines.

The results in the study provide evidence that middle je in Guaraní is influenced by contact with Spanish. The main pieces of evidence are: 1. the fact that middle je-prefixes a Spanish loan twice as often as a Guaraní-origin verb; 2. the lack of the je-prefix in prototypical Guaraní-origin verbs of body posture, change of state, and other semantic domains; 3. the frequent co-occurrence of je- with causatives, and 4. je-including calques from Spanish. This evidence is explained below:
1. As the data show, the middle je- is more frequent with Spanish loans than native-origin verbs. This finding suggests that the middle function of je- entered Guaraní via Spanish loanwords.

2. Many prototypical (Guaraní-origin) verbs in each middle semantic domain do not take the prefix je, like the following body posture verbs: ‘sit’, ‘stay’, ‘lie down’, ‘stand’, ‘kneel’, and change of state verbs: ‘burn’, ‘infl ate’, ‘break’, ‘fall’. It is true, however, that there are examples of je-prefixed native-origin verbs in all middle semantic domains. This presence of native-origin verbs in all domains could indicate that reflexive je- extended to middle uses following the internal path of change proposed by Kemmer (1993), challenging the contact-induced hypothesis. Nonetheless, when comparing Guaraní-origin verbs within a semantic domain, it is the most prototypical verbs which do not take the prefix. Thus, because prototypical verbs, which are highly frequent, usually resist change (Bybee 2015), the lack of prefixation among highly frequent native-origin verbs suggest that the middle function of je- among Guaraní-origin verbs is a relatively recent feature.

3. Among Guaraní-origin items, je- occurs with causatives mbo-mo and -ka in 40% of the verbs. We interpret the frequency of this co-occurrence as evidence that it is the combination of je- and the causative, and not je- independently, what gives the verb a middle reading. However, this correlation needs further exploration, as well as the functions of je-, the functions of the causative markers also need to be explored.
4. Finally, 14% of the je-prefixed Guaraní-origin verbs are calques from Spanish fixed verbal phrases, as seen in some examples like (159), which include the phrase: je-raha-porã (MID-take-well) < llevar-se bien (take-MID-well).

When exploring the two last pieces of evidence described above, it can be seen that more than half of native-origin je-prefixed verbs, 35 of 66 verbs, exhibit a use of middle je- that is dependent on a causative and that appear to be a calque from Spanish.

6.7. Discussion

While this study provides evidence that the middle uses of je- in Guaraní have replicated middle uses of Spanish se, it is still unclear why this novel usage permeated some native-origin verbs and not others. It is possible that only low-frequency verbs allowed the prefixation while high-frequency verbs resisted the change. A comparative study on frequency of je-prefixed verbs and not prefixed verbs could shed light on this topic. Also, it would be interesting to compare semantically-similar verb pairs, as in the GET LOST example from the beginning of the chapter: When expressed with a native-origin form, it does not get the marker; when it is expressed with a loan, it does. Testing this hypothesis, however, can be challenging, as identifying semantically-similar pairs of Guaraní-origin and loan verb is not a straightforward task, as seen in Chapter 4.

An apparent-time study with elicitation tasks can also explore how middles in contact work, as it could test whether Spanish-dominant bilingual speakers are using je- as a middle marker in contexts where Guaraní-dominant speakers would not. Language-internal explanations can be further explored as well, especially by searching for middle
usages of the *je*- prefix in colonial Guaraní texts, and comparing them to present-day Guaraní, and paying special attention to the causatives.

6.8. Conclusion

This study aimed at exploring middle voice marking in Guaraní, by comparing the behavior of *je*- when prefixing loan verbs and when prefixing native-origin verbs in the speech of 40 speakers of Guaraní. Overall, results showed that middle *je*- occurs more often with loans, that Guaraní-origin *je*-prefixed verbs usually co-occur with causative affixes to form middle constructions, and that many *je*-prefixed verbal phrases are calques from Spanish. We have interpreted these results as evidence for a contact effect of Spanish *se* on Guaraní *je*-. That is, middle *je*- entered Guaraní with Spanish loan verbs; then, the middle function was extended to Guaraní-origin verbs. However, it is still unclear what determines which native-origin verb accepts the middle *je*- and which one does not. We believe that future research on frequency effects of each verb, and semantic tightness of verb networks will advance our understanding of how middles in contact arise and evolve.
Chapter 7

Conclusion

7.0. Chapter overview

This chapter provides an overview of the main findings of the three studies of this dissertation. Section 7.1 summarizes the results that show the need to study lexical borrowing and grammatical borrowing together, as part of a continuum. Section 7.2 describes other linguistic and social factors that favored the occurrence of the contact feature. Then, section 7.3 goes over the limitations of the present research, and the avenues it will take in the future. Finally, section 7.4 presents the social impact this dissertation can have for the speech communities that were featured in the studies.

7.1. Lexical and grammatical borrowing

Through three studies, this dissertation explored loan verbs, and two contact-induced grammaticalized verb affixes. The first study focused on loan verbs with broad native-origin equivalents. The second study explored the human direct object suffix -pe; and the third study analyzed the middle marker prefix je-. Thus, the first study could be classified as a lexical borrowing study; and the second and third studies, as grammatical borrowing studies. However, the approach taken throughout this dissertation is that as verbs and the said verbal affixes naturally occur together, the three studies explored the interplay between the loans and their morphosyntactic and semantic contexts.

Moreover, in the three analyses, linguistic variables were the predictors of the feature under scrutiny. There are several connections between the three studies. For example, the loan verb chapter (Chapter 4) showed that some verb loans are associated
with a specific verbal prefix. The loans *golpea* (‘hit’) and *deha* (‘leave’) almost always occur with the middle reflexive prefix *je-*; while their native-origin translation equivalents (*nupã* and *heja*) hardly ever do. The human object study (Chapter 5) revealed that the marker *-pe* occurred more often when suffixing a loan than when suffixing a native-origin noun. Likewise, the middle marker chapter (Chapter 6) showed that the polysemous prefix *je-* occurs more frequently as a middle marker when prefixing loans than when prefixing native-origin verbs.

The results from the verb loan study and the middle marker study suggest that many verbs and verbal prefixes are borrowed as components of a larger construction. This explains, for example, that loan verbs like *golpea* (‘hurt’) almost always occur with the middle reflexive marker *je-*.. Similar to *je-*; the human direct object marker *-pe* could have also originated with the borrowing of nouns. In Spanish, object-functioning nouns with human referents (preceded very often by the case marker *a*), like *a la maestra* (OBJ the teacher) could have been borrowed as *la maéstra-pe* (the teacher-OBJ). Then, the object suffix extended to non-loan nouns. This construction borrowing supports the hypothesis that items that are frequently used together in the donor language will be borrowed together by the recipient language (Backus 2013). In turn, the findings show that grammatical replication originates in pragmatics (Heine and Kuteva 2010), that is, in everyday communication.

Furthermore, grammatical borrowing (especially those concerning verbal morphology) is said to be hard to prove (Mithun 2012). The grammatical studies in this dissertation suggest that exploring presumed contact phenomena from a holistic approach can lead to better evidence for contact hypotheses. If presumed grammatical loans, like

158
the affixes *je-* and *-pe* occur more often with borrowed nouns and verbs, as was shown by the studies, then a stronger case for grammatical loan can be made.

### 7.2. Other findings

Some of the findings in the three studies reveal that loan verbs and their native counterparts can occur in the same constructions, that is, they can be adjacent to the same—borrowed or native-origin—items. This was true for the loans *konose* (‘know’), *vivi* (‘live’), *kree* (‘believe’), and *ajuda* (‘help’) which did not occur more frequently with presumed grammatical loans than their counterparts *kuaa*, *ko*, *rovia*, and *pytyvō*. It was argued that the reason these loans occur in the same constructions as their native-origin counterparts is because they take over some, but not all, the semantic meanings that the native-origin form originally had. This partial overlap between native-origin verbs and semantically-similar loans challenge the long held view that there are two mutually exclusive outcomes for loans: that either they will coexist with the native word or replace it (Haspelmath 2009).

Another important finding in one of the studies was speaker age as a predictor of the occurrence of the contact feature. In the studies on verb loans and the middle marker, preliminary observations of the data showed no influence of social factors on the feature under study, which is why social variables were not included in the analysis. However, in the study on differential object marking, social factors appeared to be relevant, and indeed, results showed that the younger the speaker the more likely they used the human direct object marker. In fact, age was a much better predictor than origin of the prefixed noun (loans favored the marker while native-origin nouns did not).
In the scenario of Guaraní-Spanish bilingualism, age can point to a contact effect because the younger the bilingual speaker the more Spanish dominant they are. However, as mentioned before, age was not a predictor of the use of je- as middle marker nor the use of the Spanish verb loans that were studied. This could be explained by the nature of the borrowing process. The object marker -pe was borrowed as part of the [DOM OBJhuman] construction in Spanish. However, nouns also appear without the object marker a, when they function, for example, as subjects. Following Goldberg (2006), it could be argued that the [DOM OBJhuman] is not a very fixed construction, as human direct objects can also appear without the case marker (in generic sentences, for example). Therefore, the occurrence of an object loan noun would not always trigger the occurrence of the human object marker. On the contrary, certain Spanish verbs almost always occur with certain prefixes or clitics, and thus, the stem and the prefix or clitic form very fixed constructions. Quejar-se (complain-MID), for example, occurs very rarely without the clitic se. This explains why it was borrowed as je-keha into Guaraní, and not just keha. As quejarse was already borrowed with the clitic, then the use of je- in verbs like jekeha does not get extended (as there is no room for extension). The suffix -pe, however, originated from a much less fixed Spanish construction; thus, the marker was not used in all constructions with human direct objects. Naturally, the use of -pe has room for growing and, as was shown, younger speakers are extending its use.

7.3. Limitations and future research

The studies in this dissertation showed that research on contact phenomena benefits from variationist methods. With such methods, patterns of use are drawn from
natural speech of diverse speakers, which are then quantified. The analysis and results are, therefore, representative of the speech community, and the linguistic phenomena that arise from everyday use can be described as they happen.

Exploring and quantifying the speech of different speakers provides information on how conventionalized are certain features. Conventionalization is central for many debates in contact linguistics, as it helps distinguish code-switching, on-line borrowing, and established loans. When a feature is used in similar ways across different speakers, it can be argued the feature has conventionalized, as was the case for the verb loans and the middle marker je-. When different speakers use the same feature differently, a change is taking place, as was the case with the object marker -pe. Analyzing variation (or lack of it) also shows that conventionalization is a continuum, as items can be less or more conventionalized (Schmid 2015).

However, there are two limitations of using variationist methods only. First, there is a limitation to how far back in time one can go. The second and third study explore grammaticalization processes, which can extend over many centuries. As long processes, these changes can be better visualized by comparing data from different points in time. This, of course, implies that written records be analyzed in addition to present-day spoken data.

The second limitation is concerned with data comparability. When aiming at exploring variation, interviewees usually use their casual style when speaking about their preferred topics. This freedom of topic choice can lead to a wide range of topics, and therefore, to a data set that is difficult to compare. For example, verbs played a key role in all three studies. However, speakers produced different verbs. Sometimes verbs that
showed interesting patterns were not produced by all speakers; sometimes the majority of tokens of one verb came from one speaker. Situations like these limited inter-speaker comparison.

Nonetheless, working with the corpus allowed us to identify interesting verbs and constructions which can be elicited in the future. Then, the next step for this research is to develop elicitation tasks based on the data that was extracted from the corpus. For example, in the object marker study, it was found that the DOM occurred more often when the object had a child referent. However, there were few verbs that occurred with adult objects and with child objects. Thus, to test the referent age factor, one has to control for subjects and verbs. Controlling for these and other factor can be achieved through elicitation tasks, as picture descriptions.

Another future development of this dissertation will be the exploration of the grammatical features described here in the Spanish of Guarani-Spanish bilingual speakers. If, for example, bilingual speakers used the object marker at a lower rate than monolingual Spanish speakers, this can be an additional piece of evidence for the contact hypothesis. These studies on Spanish will include an analysis of unscripted speech as well as an exploration of elicited speech. In sum, we believe that a mutual influence approach (studying the same feature in both languages in contact) will broaden our understanding of contact phenomena. I will also include bilingual speakers from more areas of Paraguay.
7.4. **Broader impact**

By conducting in-depth studies of loan verbs and grammatical borrowing, this dissertation contributed to the theory of language contact and to the literature on Guaraní. This project showed that loan verbs and sentence structures are sometimes borrowed together, and thus, pointed to the need for exploring contact phenomena in a holistic manner. Such an approach can advance our understanding of how the bilingual mind works.

In addition to these theoretical contributions, the research presented here intends to provide an empirical basis for the development of successful language policies in Paraguay, and to help revert the stigmatization of Guaraní speakers that use Spanish loans and structures. The descriptions of actual language practices of the Guarani-speaking communities are a necessary step to maintain the language, but, to this day, they have been absent from linguistic policies. It is hoped that by having shown that loans are the result of complex processes that require deep knowledge of how the donor and the recipient languages work, these studies will help debunk widespread myths on bilingualism and bilingual speakers.
Appendix

8.0. Gloss

1 First Person Active
1INACT First Person Inactive
1EXCL First Person Exclusive
1INCL First Person Inclusive
1IOBJ First Person Indirect Object Pronoun/Clitic
1OBJ First Person Direct Object Pronoun/Clitic
2 Second Person Active
3 Third Person Active
ABIL Abilitative
ANTIC Anticausative
AUG Augmentative
CAUS Causative
COMPL Completive
COP Copula
DEO Deontic
DES Desiderative
DIM Diminutive
DM Discourse Marker
DUR Durative
EMPH Emphatic
FEM Feminine
FUT Future
FRU Frustrative
IMP Impersonal
IMPR Imperative
IMPF Imperfect
INT Interrogative
INTEN Intensifier
IOBJ Indirect Object
MASC Masculine
MID Middle
NEG Negation
NMLZ Nominalizer
LOC Locative
OBJ Object
OBL Oblique
ONOMAT Onomatopoeia
PASS Passive
PL Plural
PRES Present
PST Past
REC.PST Recent Past
8.1. Speakers

Table 11. List of Guarani speakers, from youngest to oldest

<table>
<thead>
<tr>
<th>Name (Sex)</th>
<th>Age</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alicia (F)</td>
<td>18</td>
<td>San Juan</td>
</tr>
<tr>
<td>Aníbal (M)</td>
<td>18</td>
<td>Asunción</td>
</tr>
<tr>
<td>Diego (M)</td>
<td>26</td>
<td>San Juan</td>
</tr>
<tr>
<td>Cristino (M)</td>
<td>27</td>
<td>San Juan</td>
</tr>
<tr>
<td>Gonzalo (M)</td>
<td>29</td>
<td>San Juan</td>
</tr>
<tr>
<td>Laura (F)</td>
<td>30</td>
<td>San Juan</td>
</tr>
<tr>
<td>Milan (F)</td>
<td>31</td>
<td>San Juan</td>
</tr>
<tr>
<td>Melisa (F)</td>
<td>32</td>
<td>Asunción</td>
</tr>
<tr>
<td>Rodolfo (M)</td>
<td>32</td>
<td>Asunción</td>
</tr>
<tr>
<td>Raquel (F)</td>
<td>33</td>
<td>Asunción</td>
</tr>
<tr>
<td>Camilo (M)</td>
<td>33</td>
<td>Asunción</td>
</tr>
<tr>
<td>Marianela (F)</td>
<td>36</td>
<td>San Juan</td>
</tr>
<tr>
<td>Sara (F)</td>
<td>36</td>
<td>Asunción</td>
</tr>
<tr>
<td>Mirta (F)</td>
<td>37</td>
<td>Asunción</td>
</tr>
<tr>
<td>Jaime (M)</td>
<td>42</td>
<td>Asunción</td>
</tr>
<tr>
<td>Susana (F)</td>
<td>45</td>
<td>Asunción</td>
</tr>
<tr>
<td>Eusebio (M)</td>
<td>47</td>
<td>Asunción</td>
</tr>
<tr>
<td>Vidal (M)</td>
<td>47</td>
<td>San Juan</td>
</tr>
<tr>
<td>Silvia (F)</td>
<td>47</td>
<td>San Juan</td>
</tr>
<tr>
<td>Claudia (F)</td>
<td>48</td>
<td>San Juan</td>
</tr>
<tr>
<td>Marisol (F)</td>
<td>54</td>
<td>San Juan</td>
</tr>
<tr>
<td>Raúl (M)</td>
<td>54</td>
<td>Asunción</td>
</tr>
<tr>
<td>Ramiro (M)</td>
<td>58</td>
<td>San Juan</td>
</tr>
<tr>
<td>Antonio (M)</td>
<td>60</td>
<td>Asunción</td>
</tr>
<tr>
<td>Mariana (F)</td>
<td>60</td>
<td>Asunción</td>
</tr>
<tr>
<td>Donato (M)</td>
<td>62</td>
<td>Asunción</td>
</tr>
<tr>
<td>Malena (F)</td>
<td>63</td>
<td>San Juan</td>
</tr>
<tr>
<td>Natalio (M)</td>
<td>64</td>
<td>San Juan</td>
</tr>
<tr>
<td>Roberta (F)</td>
<td>65</td>
<td>San Juan</td>
</tr>
<tr>
<td>Leoncia (F)</td>
<td>68</td>
<td>Asunción</td>
</tr>
<tr>
<td>Rumilda (F)</td>
<td>68</td>
<td>Asunción</td>
</tr>
<tr>
<td>Pipo (M)</td>
<td>69</td>
<td>Asunción</td>
</tr>
<tr>
<td>Sebastián (M)</td>
<td>72</td>
<td>San Juan</td>
</tr>
<tr>
<td>Norberto (M)</td>
<td>73</td>
<td>San Juan</td>
</tr>
<tr>
<td>Nombre</td>
<td>Edad</td>
<td>Ciudad</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>----------</td>
</tr>
<tr>
<td>Reina (F)</td>
<td>73</td>
<td>San Juan</td>
</tr>
<tr>
<td>Nata (F)</td>
<td>74</td>
<td>Asunción</td>
</tr>
<tr>
<td>Julián (M)</td>
<td>75</td>
<td>Asunción</td>
</tr>
<tr>
<td>Ana (F)</td>
<td>77</td>
<td>San Juan</td>
</tr>
<tr>
<td>Carlota (F)</td>
<td>79</td>
<td>Asunción</td>
</tr>
<tr>
<td>José (M)</td>
<td>85</td>
<td>San Juan</td>
</tr>
</tbody>
</table>
References


Gimeno, María Cecilia. 2012. “Sistemas de Alineamiento En Guaraní de Corrientes (Familia Tupí Guarani, Argentina).” Universidad de Sonora.


Kallfell, Guido. 2016. ¿Cómo hablan los paraguayos con dos lenguas?: Gramática del Jopara. Asunción: CEADUC.


Michnowicz, Jim. 2015. “Maya–Spanish Contact in Yucatan, Mexico: Context and Sociolinguistic Implications.” In New Perspectives on Hispanic Contact :


Morinigo, Marcos Augusto. 1931. Hispanismos en el guarani. Colección de estudios indigenistas ; 1; Colección de estudios indigenistas ; 1. Buenos Aires,.


