

TYPOLOGICAL AND SOCIAL CONSTRAINTS
ON LANGUAGE CONTACT

AMERINDIAN LANGUAGES IN CONTACT WITH SPANISH

VOLUME I

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@let.uu.nl
<http://www.lotschool.nl>

Cover illustrations: The Building of the Babel Tower by Pieter Bruegel (1563);
Guaraní Indians by Ulrico Schmidl (1599, plates 6 and 11); Columbus's Second
Voyage, Anonymous (1584).

ISBN 978-90-78328-62-9
NUR 616

Copyright © 2008: Jorge Gómez Rendón. All rights reserved.

Typological and social constraints on language contact

Amerindian languages in contact with Spanish

VOLUME I

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel der Universiteit
op donderdag 2 oktober 2008, te 12.00 uur

door

Jorge Arsenio Gómez Rendón

Quito, Ecuador

Promotiecommissie

Promotor: Prof. Dr. P.C. Hengeveld

Copromotor: Dr. D. Bakker

Overige leden: Prof. Dr. P. Muysken

Prof. Dr. T. H. Stolz

Prof. Dr. W. Dietrich

Prof. Dr. M. A. Woidich

Dr. U. Ansaldo

Dr. E. Hekking

Faculteit der Geesteswetenschappen

Universiteit van Amsterdam

*Para mis padres,
Arsenio y Edith*

CONTENTS

Volume I

INTRODUCTION	1
PART I. THE THEORY	5
1 Taking a stand	7
1.1 On language and theories	8
1.2 The structure of this book	8
2 Views on language contact	11
2.1 What is language contact?	12
2.2 Communities and speakers in contact	14
2.3 Multilingualism and language contact	17
2.3.1 Diglossia and language contact	18
2.4 Social and historical aspects of language contact and change	19
2.5 Contact situations and outcomes	22
2.5.1 Language shift: second language acquisition and language loss	22
2.5.2 Language maintenance and mixing: borrowing and codeswitching	25
2.6 An explanatory model of contact-induced language change through borrowing	31
2.6.1 Some terminological distinctions	31
2.6.2 A functional explanation of contact-induced language change	32
2.7 Summary	44
3 Theories on linguistic borrowing	45
3.1. The theory of Functional Grammar	46
3.1.1. A functional view of language	46
3.1.2. A functional view of language contact	49
3.1.3. Standards of adequacy and language contact	50
3.1.4. Motivations of contact-induced language change	53
3.2. The Theory of Parts of Speech	55
3.3. Implications of the Parts-of-Speech Theory to Lexical Borrowing	58
3.4. Implicational Hierarchies and Scales of Borrowability	59
3.5. Scales of Borrowability: a critical overview	64
3.5.1. Lexical borrowability	65
3.5.2. Grammatical borrowability	71
3.5.3. Syntactic borrowability: calquing of word order patterns	74
3.6. Structural compatibility as a constraint on borrowing	77

3.6.1.	Morphological typology and structural compatibility	78
3.6.2.	Parts-of-speech typology and structural compatibility	80
3.7.	Summary	82
4.	The Research Program	83
4.1	A critical overview of studies on linguistic borrowing	83
4.1.1	Studies on linguistic borrowing in Latin America	85
4.2	A program of research on linguistic borrowing	86
4.3	Research questions and general hypotheses	93
4.3.1	Borrowing hypotheses from the Principle of Functional Explanation	94
4.3.2	Borrowing hypotheses from the Principle of System Compatibility	95
4.3.3	Borrowing hypotheses from the Scales of Borrowability	96
4.3.4	Borrowing hypotheses from the Parts-of-Speech Theory	97
4.3.5	Borrowing hypotheses from Language Typology	101
4.4	The Methodology	103
4.4.1	Data collection	103
4.4.2	Data processing	106
4.4.3	Problematic issues	113
4.5	Summary	117
PART II.	THE LANGUAGES	119
5.	Spanish	121
5.1.	Spanish in the Americas: <i>e pluribus unum</i>	121
5.1.1.	Hispanicization and diglossia	123
5.1.2.	Colonial expansion, dialectal leveling and Andalusian influence	124
5.1.3.	Dialectal areas of Spanish in the Americas	129
5.1.4.	Spanish in contact with Amerindian languages	130
5.2.	Spanish in Ecuador	134
5.2.1.	Dialects of Spanish in Ecuador	137
5.3.	Spanish in Paraguay	142
5.3.1.	Linguistic characterization of Spanish in Paraguay	144
5.4.	Spanish in Mexico	152
5.4.1.	Dialects of Spanish in Mexico	156
5.4.2.	Spanish in Querétaro	158
5.5.	Spanish: a typological characterization	161
6.	Ecuadorian Quechua (Quichua)	169
6.1.	The history of Quichua in Ecuador	175
6.2.	The dialects of Quichua in Ecuador	179
6.3.	Quichua: a typological characterization	184

6.4.	Borrowing hypotheses for Quichua	191
7.	Paraguayan Guaraní	195
7.1.	The history of Paraguayan Guaraní	200
7.2.	Language variation and language mixing in Paraguay	207
7.3.	Paraguayan Guaraní: a typological characterization	216
7.4.	Borrowing hypotheses for Paraguayan Guaraní	225
8.	Otomí	227
8.1.	The history of Otomí	235
8.2.	The dialects of Otomí	240
8.3.	Otomí: a typological characterization	245
8.4.	Borrowing hypotheses for Otomí	264
9.	Predictions in comparative perspective	267
9.1.	Predictions from the Principle of Functional Explanation	267
9.2.	Predictions from the Principle of System Compatibility	267
9.3.	Predictions from the Scales of Borrowability	268
9.4.	Predictions from the Theory of Parts of Speech	268

Volume II

PART III. THE ANALYSIS	269
10. Comparative analysis: lexical borrowing	271
10.1. Morpho-phonological adaptation of lexical borrowings	271
10.1.1. Quichua	272
10.1.2. Paraguayan Guaraní	277
10.1.3. Otomí	284
10.2. Linguistic borrowing in the corpora	290
10.2.1. Lexical borrowing in the corpora	293
10.3. The distribution of lexical borrowing in the corpora: parts of speech	294
10.3.1. Verbs	297
10.3.2. Nouns	298
10.3.3. Adjectives	301
10.3.4. Manner adverbs	303
10.3.5. Complex borrowings	305
10.4. The use of lexical borrowings in the corpora	306
10.4.1. The use of loan verbs	306
10.4.2. The use of loan nouns	309
10.4.3. The use of loan adjectives	317
10.4.4. The use of loan manner adverbs	331
10.4.5. The use of complex borrowings	334

10.5.	Dialectal variation in the distribution and use of lexical borrowings	336
10.6.	Bilingual performance in the distribution and use of lexical borrowings	341
10.7.	Summary	347
11.	Comparative analysis: grammatical borrowing	351
11.1.	Morpho-phonological adaptation of grammatical borrowings	351
11.1.1.	Morpho-phonological adaptation of prepositions and conjuncts in Quichua	352
11.1.2.	Morpho-phonological adaptation of conjuncts and determiners in Guaraní	353
11.1.3.	Morpho-phonological adaptation of prepositions and conjuncts in Otomí	354
11.2.	The amount of grammatical borrowings in the corpora	356
11.3.	Distribution and use of grammatical borrowings in the corpora	360
11.3.1.	Articles	360
11.3.2.	Conjuncts	364
11.3.3.	Prepositions	380
11.3.4.	Discourse markers	392
11.3.5.	Other parts of speech	400
11.4.	Dialectal variation in the distribution of grammatical borrowings	412
11.5.	Bilingual performance in the distribution of grammatical borrowings	415
11.6.	Summary	417
12.	Conclusions	421
12.1.	Spanish borrowing in cross-linguistic perspective: similarities and differences	421
12.2.	Social causes and linguistic factors in the modeling of borrowing behavior	425
12.3.	Language typology and contact-induced change	426
12.4.	Towards a research program on linguistic borrowing	429
13.	Bibliography	431
14.	Appendices	453
14.1.	Annotated samples	453
14.2.	Speakers	497
14.3.	Chi-square tables	501
	Summary in English	511
	Summary in Spanish (Resumen)	513
	Summary in Dutch (Sammenvating)	515

TABLES

3.1	Lexemes and syntactic slots in English	55
3.2	Parts-of-speech systems	56
3.3	The parts-of-speech hierarchy	57
3.4	Borrowing continuum based on Whitney 1881	72
3.5	Borrowing continuum based on Muysken 1981	73
3.6	Morphological Typology and Borrowability of form-meaning units	80
4.1	Levels of integration and spelling differences	107
4.2	Structure of data file	108
4.3	Table of frequencies for one informant (fragment)	109
4.4	Total for parts of speech in Imbabura Quichua (fragment)	112
5.1	Areas of Amerindian substratum in Spanish	133
5.2	Dialects of Spanish in Ecuador: distinctive linguistic features	140
5.3	Linguistic features of Colloquial Paraguayan Spanish	145
5.4	Phonetic and morphological features of Mexican Spanish dialects	157
5.5	Phonetic features of Spanish in Querétaro	159
6.1	Quichua vitality according to percentage of native speakers	171
6.2	Levels of monolingualism and bilingualism	172
6.3	Quichua monolingualism and illiteracy by age groups	173
6.4	Levels of bilingualism by gender	173
6.5	Uses of Quichua by socio-communicative per settings and province	175
7.1	Speakers above 5 years by area, sex and language (2002)	196
8.1	Total population and number of speakers per indigenous language	229
8.2	Total population and number of speakers in the Otopamean branch	229
8.3	Otomí Population in indigenous households by municipality in 2005	231
8.4	Classification of Otomí dialects per locality, state and linguistic features	243
10.1	Totals of borrowings and native tokens in the corpora	290
10.2	Totals of codeswitching per number of switches and tokens	291
10.3	Minimum and maximum percentages of borrowings for speakers	292
10.4	Borrowing percentages for dialects of each recipient language	292
10.5	Percentages of lexical and grammatical borrowing	293
10.6	Lexical and grammatical borrowing in terms of types and tokens	294
10.7a	Percentages of borrowings per lexical class	294
10.7b	Lexical borrowings vs. regular distribution of syntactic functions	295
10.8	Loan verbs from the total number of borrowings	297
10.9	Loan verbs from the total number of lexical borrowings	298
10.10	Loan nouns from the total number of borrowings	299
10.11	Loan nouns from the total number of lexical borrowings	299

10.12	Loan adjectives from the total number of borrowings	301
10.13	Loan adjectives from the total number of lexical borrowings	302
10.14	Loan manner adverbs from the total number of borrowings	304
10.15	Loan manner adverbs from the total number of lexical borrowings	304
10.16	Complex borrowings from the total number of borrowings	305
10.17	Distribution of loan verbs per syntactic slot	307
10.18a	Distribution of loan nouns per syntactic slot	309
10.18b	Noun borrowing vs. regular distribution of syntactic functions	308
10.19a	Distribution of loan adjectives per syntactic slot	318
10.19b	Adjective borrowing vs. regular distribution of syntactic functions	318
10.20	Distribution of loan manner adverbs per syntactic slot	331
10.21	Distribution of complex borrowings per syntactic function	334
10.22	General distribution of borrowings per dialect or sociolect	337
10.23	General distribution of codeswitching per dialect or sociolect	337
10.24	Distribution of parts of speech per dialect or sociolect	338
10.25	Distribution of functions of loanwords per parts of speech and dialect	340
10.26	General distribution of Spanish borrowings per level of bilingualism	342
10.27	General distribution of codeswitching per level of bilingualism	343
10.28	Distribution of loan words per parts of speech and level of bilingualism	344
10.29	Distribution of functions of parts of speech per level of bilingualism	345
10.30	Co-relation of bilingualism, borrowing, codeswitching and flexibility	347
11.1	Phonological accommodation of Spanish conjunctions in Quichua	352
11.2	Assimilated and non-assimilated prepositions and conjuncts in Otomí	355
11.3	Contribution of grammatical borrowing to overall borrowing	356
11.4	Grammatical borrowings (tokens) per category	357
11.5	Percentage of grammatical borrowings per categories	413
11.6	Percentages of grammatical borrowings per level of bilingualism	415
11.7	Percentages of complex borrowings per level of bilingualism	417

FIGURES

2.1	A model of language shift and endangerment	23
2.2	The Principle of Functional Explanation	33
2.3	The Principle of Functional Explanation	33
2.4	A causation model of contact-induced language change	35
2.5	Types of factors inducing language change	38
3.1	Standards of adequacy in a functional perspective	52
4.1	Example of functional adaptation	100
4.2	Example of functional specialization	100
4.3	Example of lexicalization	101
5.1	Meanings of <i>prestar</i> in Colloquial Paraguayan Spanish	150
5.2	Spanish-Guaraní continuum in Paraguay	151
6.1	Access to basic facilities in two Quichua-speaking highland provinces	183
7.1	Dimensions ordered by language choice	198
7.2	Language varieties within the branch of Guaraní languages	210
7.3	Combinatory possibilities of Spanish-Guaraní language mixing	215
11.1	Borrowing scale of grammatical items according to frequency	357
11.2	Muysken's continuum and frequency distribution in the corpora	358
11.3	Muysken's continuum and frequency distribution in each language	359
11.4	Language-specific distribution of grammatical borrowings	359

MAPS

5.1	The Languages of Ecuador	136
5.2	The Languages of Paraguay	143
5.3	The Languages of Mexico	153
6.1	Map of Quichua Dialects in Ecuador	180
7.1	Jesuitical missions in the seventeenth and eighteenth centuries	204
8.1	Languages spoken in the state of Querétaro, central Mexico	232
8.2	Present-day distribution of Otomí dialectal areas per state	242

ABBREVIATIONS

For the identification and parsing of the loanwords in the corpora, two subsets of labels were used: one for signaling the parts of speech in the source language (Spanish); the other for signaling the syntactic functions of the major parts of speech in the recipient language. In addition to these labels, several others were used for the morphemic glossing of examples. The following tables contain the full list of abbreviations.

ABBREVIATIONS USED IN THE ANALYSIS OF BORROWINGS

Part of Speech	Abbreviation	Syntactic Function	Abbreviation
Noun	N	Head of Referential Phrase	HR
Adjective	A	Modifier Referential Phrase	MR
Verb	V	Head of Predicate Phrase	HP
Adverb	D	Modifier Predicate Phrase	MP
Conjunct(ion)	C	Modifier of Modifier	MM
Subjunct(ion)	S	Coordinator	COORD
Determiner	T	Subordinator	SUB
Discourse marker	K		
Loan translation	L		
Numeral	M		
Preposition	P		
Manner Adverb	DM		
Time Adverb	DT		
Place Adverb	DP		
Noun Phrase	NP		
Verb Phrase	VP		
Adjective Phrase	AP		
Prepositional Phrase	PP		
Pronoun	R		
Interrogative Pronoun	RI		
Relative Pronoun	RR		
Complex	CX		
Interjection	I		

ABBREVIATIONS USED IN MORPHEMIC GLOSSING

Abbreviation	Gloss	Abbreviation	Gloss
1	First person	INF	infinitive
2	Second person	INFER	inference
3	Third person	INST	instrumental
ABL	ablative	INT	interrogative
ACC	accusative	INTS	intensifier
ACMP	accompaniment	INTERP	interpelative
ACT	actuality	LIM	limitative
ADIT	additive	LOC	locative
AFF	affirmative	MASC	masculine
ALL	allative	MIT	mitigator
BEN	benefactive	NEG	negative
CAUS	causative	NEUT	neuter
CMP	completive	NMLZ	nominalizer
COM	committative	OBJ	(in)direct object
COND	conditional	OBLG	obligative
CONJ	conjunction	PASSIVE	passive
COP	copulative	PERF	perfect
DAT	dative	PL	plural
DEF	definite	PSTPRF	past perfect
DEM	demonstrative	POSS	possessive
DEMCOP	copulative	PRED	predicative
DET	determiner	PRF	perfective
DIM	diminutive	PRS	present
DIST	distal	PRO	pronoun
D	dual	PROG	progressive
DUB	dubitative	PROL	prolative
DUR	durative	PROX	proximal
EMPH	emphatic	PST	past
EUPH	euphonic	PTCP	participle
EXCL	exclusive	PURP	purposive
EXT	existential	RECP	reciprocal
EVID	(direct) evidence	RECPST	recent past
FEM	femenine	REFL	reflexive
FOC	focus	REL	relative
FUT	future	REMPST	remote past
GEN	genitive	REP	reportative

GER	gerund
HAB	habitual
HON	honorific
IMPF	imperfective
IMPR	imperative
INCL	inclusive
INCH	inchoactive
INDEF	indefinite

S	singular
(SP)	Spanish
SBJ	subjunctive
SUB	subordinator
SUP	superlative
TEM	temporal
TMLS	timeless
TOP	topicalizer

ACKNOWLEDGEMENTS

Every book is a product of society, not only because it collects ideas from different authors but also because many people make their contribution to its final state. The present book is not an exception. The long and winding road that led to these pages has been full of blessings for me through the support and help of people in Europe and Latin America.

The University of Amsterdam and the Amsterdam Centre for Language and Communication (ACLC) provided me with all the material and financial means to carry out my investigation in the best possible way. I am especially grateful to my supervisor, Kees Hengeveld, for his wise orientation, unconditional support and warm friendship to me and my family during these years. To my co-supervisor, Dik Bakker, who patiently and enthusiastically guided me step by step in the stages of analysis and writing, and was always there for helping me attain the highest academic standards.

In the warm and cozy offices on the Herengracht I enjoyed the company of supportive and friendly people: Annerieke Boland, Marina Dyakonova, Eva van Lier and Niels Smit are some of them. I am grateful in particular to Rafael Fisher, who gave me his friendship and helped me out to find a flat for me and my family once we reunited in Amsterdam. My special thankfulness to Gerry Wanders, true friend and helpful colleague from the very day I stepped foot on the Herengracht to my last days in the Bungehuis. My family and I treasure her friendship as one of the best gifts from our stay in the Netherlands.

My first year of residence in Amsterdam would have been hard without the selfless support of Hella Olbertz, Roland Tweehuizen and Joost den Hahn. Hella contacted me in Ecuador and encouraged me to apply for the doctoral project in Amsterdam, helped me find the way in the milieu of Dutch culture during the early days of my stay in Europe, and introduced me to academic circles. Roland and Joost opened to me the doors of their charming flat on the Binnenbrouwersstraat.

Across the Atlantic I was fortunate to find people who gave me their support in plenty of ways. During my fieldwork in Paraguay I received the most valuable help from Lino Trinidad Sanabria, knowledgeable speaker of Ava ñe'ẽ and champion of the linguistic rights of Guaraní in Paraguay. In Mexico, José Antonio Flores Farfán generously shared his knowledge in the field of language contact and helped me continue with the writing of my dissertation. My six-month stay in Mexico City was possible thanks to a fellowship from Centro de Investigaciones y Estudios Superiores en Antropología Social-CIESAS. Once in Mexico, my work on the Otomí language was made possible by Ewald Hekking, who openhandedly shared with me his rich database and knowledge of the language and his cozy home in Queretaro. In this beautiful Mexican city I received also the help from several

friends: Aurelio, Jaap, Luz María and Lourdes deserve special mention. In my home country I received the valuable assistance of old and new friends. I am grateful to Iván Zapater, who accompanied me in my first transatlantic travel to Europe and showed me the ways of the Old World. My thankfulness to Marleen Haboud, for her unrestricted support since 1992 and her confidence in my academic accomplishments. She remains a great example to me for her socially committed dedication to the indigenous languages of Ecuador.

I owe my deepest gratefulness to my family for their unconditional and permanent support. To Arsenio and Edith, my beloved parents, from whom I learnt not only discipline and hard work but the most important things in life: love, honesty and solidarity. Words are never enough to express my gratitude to them for their wise counsel, selfless support and unending love. To Blanca and Felicidad, who were always full of best wishes and blessings for me in the achievement of my goals. They are not longer with me but I always carry their memory in my luggage. My thankfulness also to Liliana, Jorge Luis, Steve, Patricio and Xavier for their affectionate company and their words of encouragement.

And finally, to my dearly loved wife and son, Tania and Gabriel, for their long wait in Ecuador until our reunion overseas, for all the times I could not spend with them because of work, but above all for their understanding and unselfish company over these years of travels and absence. They are part and parcel of this book inasmuch as a man can speak his mind because others keep silent.

INTRODUCTION

Studies on language contact have been prolific in the last decades. The increasing interest of linguists in this particular field implies the recognition that languages do not develop independently from other languages, and that the outcomes of language contact result from adaptive answers of linguistic systems. From this point of view, contact linguistics offers the opportunity of studying the interaction of social motivations and linguistic factors in the process of language change: how nonlinguistic forces model human languages within the limits set by their structures. Nonetheless, one of the major shortcomings of most studies on language contact is their lack of a theory-driven approach and a solid empirical foundation, which reduces the cross-linguistic scope of their findings and the general reliability of their generalizations.

The present study deals with language contact from the perspective of linguistic borrowing. Its empiric foundation is an extensive corpus of spontaneous speech collected in the field. Its framework is the theory of parts of speech and the theory of contact typology. Because the main goal of this study is the identification of cross-linguistic regularities in borrowing, the recipient languages under scrutiny are different in their typological profile but similar in their contact with one donor language. In this way, differences in the outcomes of borrowing can be ascribed to differences in typology, just like similarities in the process of borrowing can be attributed to analogous contact situations. Accordingly, it is assumed that the comparison of borrowing tendencies in typologically different languages can shed light on how linguistic structure influences the outcomes of contact and the extent of such influence vis-à-vis nonlinguistic factors. The recipient languages selected for analysis are Quichua, Guaraní and Otomí while the donor language in contact with them is Spanish.

The first part of the book is theoretical in nature. It deals with the conceptual foundations for the analysis of linguistic borrowing. Crucial to such analysis is the development of a causation model of contact-induced language change, in which hierarchically ordered causes interact with each other at different levels. The model serves as a point of departure for the interpretation of linguistic and nonlinguistic factors in lexical and grammatical borrowing. Parts of speech, borrowability and morphological typology are discussed as linguistic factors modeling the outcomes of borrowing. All the theoretical elements are put together in a comprehensive research program which sustains the present investigation.

The second part describes the source language and the recipient languages in terms of their historical development, sociolinguistic status, dialectal variation and typology. The account of the historical development of each language provides a more accurate characterization of the intensity and duration of contact and the

expected degree of influence between the languages. The sociolinguistic description of the recipient languages in terms of their diglossic position and the societal levels of bilingualism in their respective speech communities enables a straightforward measurement of the pressure exerted by the source language on the recipient languages and the extent of borrowing. The classification of the languages in terms of parts of speech, morphological type, dialectal variation and other typological features sets the benchmark for the analysis of borrowing types. The historical, sociolinguistic and linguistic description of the languages unfolds in the framework of the causation model proposed in the first part and serves to make specific predictions about the borrowing behavior of each language.

The third part represents the analytic core of the book. It describes the findings from the analysis of corpora and compares these findings to the predictions made for each language in order to test the validity of the borrowing hypotheses. Lexical and grammatical borrowings receive individual treatment in terms of their contribution to overall borrowing, their morpho-phonological adaptation to the recipient language, and the uses to which borrowings are put in accordance with native or novel functional distinctions. The use of borrowings is tested for dialects and sociolects in order to determine the extent to which dialectal variation and bilingualism model borrowing behavior. The overall findings of lexical and grammatical borrowing are evaluated in the framework of the causation model and the contact-induced changes in the typological profile of the borrowing languages.

The main conclusions of this study point to the interplay of linguistic and nonlinguistic factors in the modeling of linguistic borrowing. The distribution of borrowings in any given language cannot be explained solely by either type of factors. The interplay of factors at different levels confirms the dynamic nature of the causation model proposed for the explanation of contact-induced changes. Also, the overall findings confirm that even if linguistic constraints can be overridden by nonlinguistic factors, the outcomes of borrowing are determined in principle by the structural possibilities of the participating languages. In sum, not everything goes in linguistic borrowing, because structural and other restrictions set the limits of language mixing. Typology seems to be a modeling factor even when structural limits are trespassed. This is due to the resistance of the basic typological parameters to change in both normal and contact situations. These parameters are largely preserved in the recipient languages of this study after hundreds of years of intense contact with the source language, even if incipient and moderate changes are attested in less crucial typological features.

The pressure exerted by the donor language on account of the hegemonic position of its speakers may induce major structural changes in the recipient language, but these changes are co-determined by the structural limits of its linguistic system, the level of societal and individual bilingualism, and the attitude of speakers towards language mixing. Cases of massive borrowing are therefore

those in which speakers refuse to abandon their language and adapt it to the discursive and communicative needs imposed by the dominant language. This is all the more evident in multicultural and multilingual contexts, in which the orientedness of language towards the accomplishment of communicative goals is at stake. In this perspective, the languages of this study are survivors of a long history of intense contact because they have been flexible enough to adapt to the new socio-communicative settings of the colonial society.

In addition, the present study demonstrates that scales of borrowing or hierarchies of borrowability are not cross-linguistically valid, and that typological, sociolinguistic and historical considerations are always necessary to make them more precise and refine their predictive capacity. For example, as demonstrated by one of the languages of this study, the often assumed predominance of lexical over grammatical borrowing can be reversed in a context of rapid language shift and increasing levels of bilingualism, provided grammatical borrowings accommodate to the structure of the recipient language. In all, any evaluation of scales or hierarchies of borrowing must be language-specific and consider both linguistic and nonlinguistic factors in accordance with a multi-level dynamic model of causation.

While the analysis addresses a number of issues about the relation between linguistic borrowing, language typology and bilingualism, it necessarily leaves several questions open. Some of them concern the relation between code switching and borrowing, the relation between phrasal borrowing and code switching, the influence of semantic restrictions or distributional rules on the use of loanwords, the influence of language loyalty on language mixing in situations of diglossia and intense contact, and the diachronic study of the borrowing process on the basis of historical records. These and other questions are part of an agenda for future research in the field of language contact.

PART I

THE THEORY

The first part of the book focuses on theoretical issues related to language contact, borrowing and typology.

Chapter 1 presents an introductory overview of the main goals of the book. Chapter 2 discusses various issues of contact linguistics and offers my personal views in relation to contact-induced language change, linguistic borrowing, and the influence of social and cultural factors on language change. I develop an explanatory model of contact-induced language change by identifying different types of causes and factors and their respective contributions to language change.

Chapter 3 is devoted to the discussion of theories on lexical and grammatical borrowing. The chapter pays special attention to the theory of Functional Grammar proposed by Dik (1997) and the theory of parts of speech (Hengeveld 1992). The chapter develops the implications of both theories for language contact in general and linguistic borrowing in particular. Further issues addressed in Chapter 3 concern implicational hierarchies, scales of borrowability, morphological typology and structural compatibility.

Chapter 4 presents the research program of the present investigation. The first section gives an overview of studies on linguistic borrowing in Latin America, the area in which fieldwork was conducted for this book. In the second section I motivate the selection of three Amerindian languages and the linguistic and nonlinguistic criteria for data collection. The third section offers a detailed discussion of the research questions and the borrowing hypotheses. The fourth section describes the methodology of research including the collection and processing of data, the description of the corpora, and the problems tackled in different stages of research. As an important innovation with respect to other studies on linguistic borrowing, the computational tools developed for corpus analysis and the encoding of grammatical categories are comprehensively described in the fourth section.

Chapter 1

Taking a stand

To state that language contact is as old as language itself may sound as an exaggeration to the ears of those who consider languages self-contained entities developing on their own, but it is less so if we think for a moment that it is not languages *per se* that are in contact, but speakers. In these terms, language contact is expected every time two or more groups of speakers meet, and thus language contact implies as much motion of people as transfer of languages. To what extent the borrowing of lexicon and grammar is rule-governed in language contact and reflects the internal organization of the languages involved, and to what extent social and cultural factors play a role in such process are the main questions I attempt to answer in this book.

The relatively recent idea that language contact is a window on linguistic structures has given a new impulse to contact studies over the last years. In this perspective, language contact mirrors the ways in which languages react as dynamic structures to their sociocultural environments. Imbued with the same spirit, I intend to give new insights into how languages react to other languages by accommodating their structures and their usage.

At the heart of this research is the debate about the typological constraints on language contact. Whereas some authors take the existence of these constraints for granted (e.g. Hill and Hill 1986), others downplay linguistic factors (e.g. Thomason and Kaufman 1988), and still others deny their contribution categorically (e.g. Thomason 2001). I do not pretend to settle the issue here. On the contrary, I will add fuel to the fire by showing that the typological profile of the languages in contact is relevant when it comes to explaining the linguistic outcomes of such contact, but that it is far from being the only factor involved. As nowadays no one can disregard the major role played by social factors in linguistic change, the real question is *how* typological and social factors interact. This study is a contribution to understand such interaction.

While the present research is framed in the overall debate of contact-induced language change, its results are limited to the specific cases analyzed here and should not be generalized across the board. Considering the variety of contact scenarios around the world, any statement made on the basis of the data and the analysis presented in this book should be mapped onto other languages and contact situations with extreme caution. It is hoped that similar studies be undertaken for other languages in order to enlarge the gamut of contact situations under examination.

1.1. On languages and theories

From the numberless aspects of language contact, the present study deals with borrowing, both lexical and grammatical. It is based on the investigation of extensive corpora of spontaneous speech collected for three recipient languages (Guaraní, Quichua and Otomí) which have been in contact with one donor language (Spanish) for the last four centuries with more or less intensity. The purpose is to identify what types of borrowing from the donor language occur in the recipient languages and how they are used. The choice of these languages is motivated by the fact that any systematic assessment of the output of contact is feasible only to the extent that the target languages are different from each other in their typological profile while the donor language is kept constant in each case. This procedure allows us to compare results and inquire into possible explanations that include typological and social factors. Further reasons for the choice of these languages are the large size of their speaking communities and the availability of good grammatical descriptions, all of which facilitates the collection and analysis of data.

Because the present study seeks to identify principles in the borrowing of lexical and grammatical elements and their use in the recipient languages, two different approaches have been adopted depending on the type of borrowing. For lexical borrowing, I take the concept of parts of speech as the tool for analysis. The theory of parts of speech developed by Hengeveld (1992) and Hengeveld *et al* (2004) offers a benchmark by virtue of its typological approach – required to understand the idiosyncrasies of the donor language and the recipient languages. This theory defines parts of speech on the basis of functional-syntactic criteria and classifies languages according to the use of lexical classes in syntactic slots. To the extent that it focuses on *major* word classes (i.e. verbs, nouns, adjectives, and adverbs) the theory is relevant for the analysis of lexical borrowing. For grammatical borrowing I take as a point of departure the hierarchies of borrowability proposed in several studies on language contact, typology and grammaticalization (cf. Muysken 1981b, 1999; Lehmann 1986; Croft 1990; Heine *et al* 1991; Bakker and Hekking 1999; Bakker *et al* 2008.). Hierarchies of borrowability show which word classes are borrowed more frequently than others. These hierarchies encompass lexical and grammatical borrowings and serve as comprehensive frames for testing hypotheses. The corpora of the recipient languages provide the empirical basis for such testing.

1.2. The structure of this book

Chapter 2 offers a critical review of a number of issues in the field of language contact, including theories on the interaction of linguistic and social factors and the types of contact outcomes. Special attention is paid to the discussion of similarities and differences between borrowing and codeswitching, since both categories

intersect in various ways. The first part of Chapter 3 sets the theoretical framework for the study of lexical borrowing in terms of parts of speech (Hengeveld 1992; 2004) and develops a number of hypotheses with respect to the borrowing of word classes. The second part of Chapter 3 presents an overview of hypotheses of grammatical borrowing and relates them to the case studies under scrutiny. Chapter 4 presents the research program in detail, including a description of the methods used in sampling, collecting, parsing and analyzing data, as well as other methodological issues relevant for the investigation.

A detailed study of the languages in contact is presented in chapters 5 through 8. Chapter 5 discusses the evolution of Spanish in Latin America and focuses on the contact areas of the Andes, Paraguay and central Mexico. Chapters 6 to 8 are devoted each to one recipient language. Therein I address the historical background and the linguistic factors that feature the contact situation of the recipient languages with Spanish. The discussion focuses on the ways native speakers adapt their languages to sociocultural pressures from the mainstream society and the various levels of bilingualism in the speech communities. Each chapter discusses the parts-of-speech system and other typological features of the language in question and develops specific borrowing hypotheses. The hypotheses are discussed in comparative perspective in Chapter 9.

Chapter 10 analyses the statistics of lexical borrowing according to the hypotheses from the theory of parts of speech. The distribution of parts of speech in the corpora and the use of lexical borrowings in the recipient languages are discussed thoroughly in that chapter. The analysis follows a comparative approach in order to identify frequencies and tendencies attributable to the typology of the participating languages, their dialectal variation and their levels of bilingualism. Chapter 11 analyses the statistics of grammatical borrowing and elaborates on bilingualism as a relevant factor in the borrowing and use of grammatical items. Chapter 12 presents the conclusions from the analysis of borrowing data in previous chapters and discusses the implications for language contact research. Annotated texts extracted from each corpus are provided in the appendices so that readers have enough material to compensate the fragmentary nature of the examples discussed in the analytic chapters.

Chapter 2

Views on Language Contact

This chapter reviews the relevant literature on language contact and presents my views on several contact-related topics. In the first section I discuss the main elements in any definition of language contact. The second section deals with language contact from societal and individual perspectives and how contact is approached differently from the society and the speaker. With both perspectives as points of departure, the third section discusses the relation between language contact and individual bilingualism on the one hand, and between language contact and societal diglossia on the other. The relationship between bilingualism and diglossia is central to the analysis developed in the next chapters since most speakers of Indian languages in Latin America are, to different degrees, bilingual in their native language and Spanish or Portuguese but continue to live in a diglossic state where the European language is socially dominant. In addition, I discuss how levels of bilingualism within the speech community determine the speakers' ability to incorporate items from other languages, the acceptance of incorporated items, and the attitudes towards language mixing. The fourth section deals with the social and historical factors of language contact. The discussion builds on the assumption that social factors as much as linguistic ones determine contact-induced language change. I demonstrate that the inclusion of social and historical factors in the analysis increases the predictive capacity of constraint models like those discussed in Chapter 3 and helps us outline their scope and limits. The fifth section approaches linguistic borrowing from two complementary points of view: processes and outcomes. I follow here a division adopted by authors such as Thomason (2001) and Winford (2005: 373-427) for the classification of contact-induced change. The use of both parameters to measure contact-induced change allows for a more dynamic view of linguistic borrowing, i.e. one that focuses on mechanisms *and* results. Different outcomes of language contact are linked to specific settings. Outcomes are grouped in three types, each with its own set of social and cultural factors: language mixing, language shift and language creation. The sixth section offers a critical discussion of the motivations and factors of language contact and change, in particular the interplay between linguistic and nonlinguistic (sociocultural) motivations and factors within a multi-causal and dynamic model. The chapter closes with a summary of the covered topics.

2.1. What is language contact?

Contact (adj./n.) (1). A term used in sociolinguistics to refer to a situation of geographical continuity or close social proximity (and thus of mutual influence) between languages or dialects. The result of contact situations can be seen linguistically, in the growth of loan words, patterns of phonological and grammatical change, mixed forms of language (such as creoles and pidgins), and a general increase in bilingualism of various kinds. In a restricted sense, languages are said to be 'in contact' if they are used alternately by the same persons, i.e. bilinguals. (David Crystal, *A Dictionary of Linguistics and Phonetics*, 2006: 102)

Considering the bulk of literature produced on language contact issues in the last decades, one cannot but be surprised that definitions of language contact are scarce. While some definitions are rather simplistic, others are more specific as regards the elements involved. The definition quoted above has, in my view, two advantages. On the one hand, it is explicit about the different meanings of the term 'contact'. On the other, it incorporates several elements of relevance such as geographical continuity, social proximity, alternating use, bilingualism and bilingual speakers. I do not intend here to provide my own definition of language contact but discuss the main elements any good definition should include by linking such elements to the specific contact situations analyzed in this book.

Any definition of language contact includes three basic elements, namely: two or more languages, the speakers of these languages, and a socio-cultural setting in which contact takes place. Of course, this is a simplification of facts since every contact situation is different, depending on a large number of variables going from the strictly linguistic to the social.

Contact involves two or more languages or dialects of one language. In the latter case we speak rather of dialect contact. In the Ecuadorian Highlands, for example, an intensive contact exists between speakers of urban and rural dialects of Spanish, and between the Mestizo speakers of these dialects and the Indian speakers of Spanish. Their contact led to a dialect continuum stretching from standard urban varieties of Spanish to interlanguage varieties of second-language Spanish spoken by Quichua native speakers in the cities, where they migrate seasonally for work. The same continuum is reported for dialects of Quichua,¹ with a standard variety used for instruction in classrooms on the one end, and highly Hispanicized varieties

¹ Henceforth Ecuadorian Quechua will be called simply 'Quichua'. Ecuadorian Quechua is classified as part of Quechua II (Torero 1964). Quechua II dialects share a number of traits with other varieties from Southern Peru, Bolivia and Argentina (Adelaar 2004: 185ss). The difference between Quichua and Quechua is explained in Chapter 6.

spoken by Indians in close contact with the Spanish-speaking society on the other end (Muysken 1985: 392).

Therefore, in any contact situation it is necessary to identify first whether the varieties in contact are languages from different families, from the same family, or dialects of one language. This is true even if mutual intelligibility is reduced enough to consider two dialects as different languages. This is the case of urban Guaraní dialects and ethnic dialects such as Mbya or Tavytera in Paraguay (cf. 5.3). In the present book I deal with languages from different families: Spanish, a Romance language of the Indo-European stock; Ecuadorian Quichua, a language of the Quechua family; Paraguayan Guaraní, a language of the Tupi-Guaraní family; and Otomí, a language of the Otomanguean family.

The second element in any definition of language contact is Janus-faced. On the one side are the speech communities; on the other, the individual speakers. A tendency prevails in language contact studies which focuses on languages (a systemic approach) and speech communities (a social approach). Individual speakers are generally set aside from the discussion, thereby obscuring the fact that speakers are the real agents of language contact. Considering both speech communities and individual speakers enables a more comprehensive interpretation of sociolinguistic factors such as speaker's perceptions and attitudes towards language contact and its outcomes (cf. 2.3). Moreover, a speaker approach opens a largely unexplored field in contact linguistics: the psycholinguistic processes at work when two or more languages or dialects are in contact.² I address the individual dimensions of bilingualism in the analysis of borrowing in Chapters 10 and 11.

The sociocultural setting is the third element of language contact. Sociocultural setting refers to a number of physical, social and cultural variables that make up the communicative situation of contact. The first of these variables is the geographical space of the speech community (their ethnic space) and the geographical space shared by speakers of both communities (their contact zone). The latter space may be embedded in the ethnic space or be created on occasion by the coming together of both communities. However, speakers of two languages need not share geographical space for language contact to occur (Thomason 2001: 2). English is disseminated in non-English speaking communities through the media. The speakers of these communities incorporate a number of English words and constructions in their language without being in contact with English speakers at all.³ A further element of the sociocultural setting is social space. This embodies a coherent set of practices (including verbal behavior) accepted in the speech community. In the context of the

² A pioneer study in this direction is Myers-Scotton (2006) in which contact-linguistic and neurolinguistic approaches are intertwined, with promising results for future research.

³ Likewise, speakers need not share time for contact to take place. The use of internet is illustrative in this respect. Separated by long distances, speakers communicate from their own spaces and local times through the cyberspace of the web.

present investigation linguistic data were collected from socially significant verbal practices (i.e. *not* elicited) inside specific geographical spaces (communities) in real time (face-to-face interactions).

Language contact may be defined from several perspectives, but any definition must incorporate the elements discussed above. Accordingly, when contact is mentioned in this study, it refers to the contact among individual (often bilingual) speakers from different speech communities who communicate with each other by using different linguistic strategies, one of which is language mixing (specifically, linguistic borrowing). The contact of people and languages develops within the social and cultural boundaries of the speech communities concerned.

2.2. Communities and speakers in contact

Agency in language contact may be analyzed from the perspective of the speech community or from the perspective of the speakers. Both approaches are not contradictory but complementary. Each sheds light on different processes of language contact. In this section I address first the notions of ‘speech community’ and ‘speaker’ to the extent that both have import to the processes and outcomes of language contact. Later I discuss the interface between the speech community and the speaker by bringing to light the relationship between social practices and individual linguistic behavior in language contact. Finally, I focus on the issue of language contact from the perspective of national societies and globalization, with particular reference to Latin America.

The term ‘speech community’ describes a group of human beings identified in terms of geographical and social spaces and the set of sociolinguistic practices which make them different from other groups (Crystal 2006: 427). This definition encompasses three elements (people, spaces and practices) which combine in different manners to characterize distinct speech communities. Space may be physical, geographical and social. Linguistic practices embody the linguistic behavior of speakers, including their language and the ways they use it for communication. Speech communities may be as different in size and character as nations, ethnic groups, immigrants, or groups of people sharing the same work or profession. These groups engage in linguistic contact with other groups of the same or different size: e.g. immigrant groups engage in contact with national societies just like ethnic groups take part in contact with other ethnic groups.

The socio-cultural setting analyzed in this book involves groups of Spanish-Amerindian bilinguals in contact with groups of Spanish monolinguals.⁴ This setting

⁴ Other settings, not addressed here, involve 1) Spanish monolingual groups in contact with Amerindian monolingual groups, and 2) groups of Spanish-Amerindian bilinguals in contact with each other. While the first setting is rather infrequent, except for a few cases of isolated

generally involves a minority group embedded in the larger speech community of the nation-state.

‘Speech community’ and ‘speaker’ are theoretical constructs often imbued with reductionism, hence the need to make both concepts specific in discussions of language contact. For one thing, the concept of ‘speech community’ should not lead us to overlook that speakers of flesh and blood are the ultimate agents of linguistic contact and change: individual speakers from different linguistic backgrounds exchange information by means of verbal signs when they engage in communication. Any individual speaker is characterized by sociolinguistic variables such as sex, age, ethnic background and education. The integration of these variables in the analysis makes linguistic variation emerge from seemingly uniform speech communities. Linguistic data for the present investigation were collected in socially and geographically identifiable speech communities (e.g. the Quichua speech community of Otavalo or the Otomí speech community of Santiago Mexquititlán). Each of these communities, however, includes a number of sociolects which deserve special consideration, especially because sociolectal variation in one speech community may surpass dialectal boundaries in certain cases.

Societal and individual aspects of contact are interconnected in complex ways. The interface between the speech community and the speaker is an ever-changing space of bidirectional influence where feedback from both sides is the rule. As noted above, a set of linguistic practices characterizes every speech community. These practices are the materialization of language usage, and individuals are raised in them as part of their socialization. In principle, individual linguistic behavior is determined by collective linguistic practices. Changes in these practices result in changes in individual linguistic behavior. In turn, provided certain conditions are met, changes in individual speech disseminate in the community and become collective linguistic practices. One condition for the spreading of individual changes in verbal behavior is the innovative role of the individual speaker in the speech community as determined by his/her political and economic position but also by his/her linguistic proficiency in higher and lower varieties in diglossic situations. Even if language contact does not require fluent bilingualism (Thomason 2001: 1), individual bilingualism in any degree is a trigger of language contact and change. It is the bilingual speaker who by innovating his/her speech with the inclusion of foreign lexical or grammatical elements triggers off a chain of similar speech acts leading to the incorporation of the same elements in the group’s linguistic pool.

communities which come into contact with colonizers in frontier zones (e.g. some groups of Wao speakers who have occasional contacts with timber merchants or oil workers in the Ecuadorian Amazon Lowlands), the second setting corresponds to contact between minority groups whose varieties are dialects of the same language (e.g. speakers of different Otomí dialects). Both settings are not studied in this book, however.

Summing up, contact-induced language change may be approached from the speech community and the speaker. Both standpoints are complementary for any satisfactory account of contact-induced language change. A speech-community approach views contact as a series of speech events with speakers of different languages, and language change as the outcome of those events.⁵ A speaker approach views contact as the coexistence of linguistic varieties in the speech of bilingual or multilingual speakers, and language change as the ways in which one linguistic variety influences the other in speech production. Hence the terminological distinction between ‘borrowing’ as used in historical and comparative linguistics (focusing on the speech community) and ‘transfer’ or ‘interference’ as used in second language learning (focusing on the individual speaker). A discussion of the term ‘borrowing’ and other related concepts is presented in section 2.5.2.

One final issue to be addressed here is language contact from the perspectives of national societies and globalization. I address this issue in the context of Latin America, the geopolitical space which concerns us here.

Since the emergence of the first nation-states in Europe in the sixteenth century, the sociopolitical space of the nation became a major locus of language contact in Europe and its colonies overseas.⁶ The building up of most European nations occurred at the expense of linguistic minorities which were and continue to be integrated in overall national projects – usually against their own will. The cases of Basque in Spain and Welsh in England are two examples. While the forced incorporation of linguistic minorities in the frame of the nation was not uncommon in Europe, it was the rule in the colonies.

A conservative estimate of the pre-Columbian population of the Americas gives some thirteen million people speaking over one thousand languages (Rosenblat 1954: 102). In South America, “the number of living languages is estimated to be around 500, but there is no doubt that they constitute a fraction of the languages spoken in South America at the arrival of the Europeans” (Tsunoda 2005: 21). All over the world the contact between European languages and native languages was accompanied by extensive language loss. As Tsunoda rightly notes, “colonization by European nations has exerted perhaps the most devastating damage in the way of language loss. The languages of the European powers spread to other parts of the world and exterminated, or at least diminished, a large number of aboriginal languages” (Tsunoda 2005: 4).

⁵ By speech events I understand here “a communicative exchange made meaningful by culturally-specific structures of participants, genres, codes and other elements” (Crystal 2006: 428).

⁶ Interestingly enough, this scenario repeats in other non-European nations. Japan and China are two cases in point. For an evaluation of language contact and endangerment in Japan and other parts of the world, see Tsunoda (2005: 17ff).

In spite of an emerging worldwide sociopolitical formation based on transnational flows of people and goods (globalization),⁷ the nation in Latin America continues to provide the geopolitical frame for language contact between the official languages of the nation-states and the native languages of the embedded ethnic groups. The national society remains the matrix of language contact in the three situations analyzed in this book. Differences are noticeable however. Thus, for example, bilingualism characterizes the Paraguayan society regardless of ethnic affiliation while Spanish-Amerindian bilingualism in Ecuador and Mexico occurs only among Indian ethnic groups. Similarly, Guaraní is official in Paraguay together with Spanish while the latter is the only official language in Mexico and Ecuador. Thus, the concepts of multilingualism, bilingualism and diglossia become relevant only against the backdrop of states and national societies. These concepts are discussed in the next section, as they are used throughout this book to illuminate social, cultural and linguistic aspects of contact in Latin America.

2.3. Multilingualism and language contact

Societies nowadays are multilingual without exception. This does not mean all of their members are multilingual too. In former European colonies, multilingualism is typical of non-European speakers while citizens of European descent are often monolingual. Latin America is a case in point. The majority of the population of Latin American countries is monolingual in Spanish or Portuguese. Notwithstanding this fact, it is estimated that some eight hundred Amerindian languages are spoken nowadays in the continent. The three countries in which fieldwork was conducted for this investigation have an important number of Amerindian languages, with Mexico ranking as the richest in multilingual diversity followed by Paraguay and Ecuador⁸. Differences in bilingualism are notable however. Approximately ninety

⁷ Globalization is the backdrop for several forms of language contact, for example, between English and Spanish in the United States or Peninsular Spanish and Latin American Spanish. The latter case involves Spanish-speaking Ecuadorian immigrants in Spain. These cases provide rich material for an investigation of language contact in contemporary transnational scenarios but go beyond the scope of this study.

⁸ There is a certain disagreement with respect to the real number of languages in each country, which probably results from internal sociopolitical conflicts. Thus, for example, *Ethnologue* (2005) lists a total of 298 languages for Mexico, while the (Mexican) *Instituto Nacional de Geografía y Estadística* speaks of only 30 languages. In Ecuador we find a similar situation with respect to the number of speakers, with Indian native speakers varying from one to three million depending on who provides the figures. For an overview of the linguistic diversity in Mexico, see Flores Farfán and Nava López, *La riqueza lingüística de México: un patrimonio seriamente amenazado* (2007). The Indian languages in Ecuador and the state of the art of their research are presented by Montaluisa and Álvarez, *Las lenguas indígenas en el Ecuador y el estado actual de sus investigaciones* (2004). For linguistic data from the last census in Paraguay, visit the website www.dgeec.gov.py.

percent of the national population of Mexico and Ecuador is monolingual while a similar percentage of the population in Paraguay is bilingual. The widespread use of Spanish and Portuguese in Latin America is closely associated with the sociopolitical role played by these languages in colonial and republican times. For the last five hundred years Spanish and Portuguese have been the dominant languages in Latin American and the linguistic basis for national projects. This means that European languages and Amerindian languages have coexisted in a typically diglossic condition for a long time.

2.3.1 Diglossia and language contact

Introduced first by the Arabist William Marçais in 1930 and disseminated by Charles Ferguson in his classic article (Ferguson 1959) the term ‘diglossia’ refers to the compartmentalized use of two languages or two dialects of one language in mutually exclusive settings. Typical of diglossia is that one of the languages (the high variety, H) occupies a politically dominant position with respect to the other (the lower variety, L) the difference lying on the degree of formality of each variety and its association with public or domestic environments. In this context, one language (H) is learned in schools and spoken in public settings while the other (L) is acquired at home and spoken exclusively in the family or the community. Usually, the speakers of one variety (H) are of higher socioeconomic status than the speakers of the other. Ferguson insisted on the opposition between diglossia and bilingualism as “the analogous situation where two distinct (related or unrelated) languages are used side by side throughout a speech community, each with a clearly defined role” (Ferguson 1964,; 429). However, recent studies have demonstrated that diglossia in fact coexists with societal bilingualism.⁹

Diglossia prevails all over Latin America to a greater or lesser degree, with Spanish or Portuguese as official languages and prestige varieties associated with culture and education. Even Paraguay, with its overwhelming number of bilinguals, is a diglossic society. In fact, the traditional characterization of Paraguay as a model bilingual society ever since Rubin (1968) has veiled the subordinate status of Guaraní and the actual compartmentalization of languages in this country.

Diglossia in Latin America is rooted in complex sociopolitical structures inherited from colonial times and reproduced with minor changes up to the present. In this diglossic context, Amerindian speakers learn Spanish or Portuguese in order to participate in the mainstream society and the market economy of their respective countries. The knowledge of Spanish or Portuguese helps people get jobs, buy and

⁹ For a terminological discussion and analysis of the literature on the topic with special attention to the Arab world, see Alan S. Kaye, “Diglossia: the state of the art”, *International Journal of the Sociology of Language* 152 (2001), pp. 117-119.

sell their products, or simply have access to public facilities and services provided by the government. No similar pressure exists for Spanish or Portuguese speakers in Latin America, which explains why multilingualism is a reality only for those whose native language is not Spanish.

Since the foundation of contemporary Latin American republics Amerindian peoples have experienced a permanent pressure for language shift. This pressure increased significantly in the twentieth century at the side of urbanization. The results are not uniform however. Some speech communities abandoned their native languages over the years while others maintained their linguistic heritage. The maintenance in this case was not gratuitous, and many languages experienced important changes in their structure. Language mixing is one of these changes. It includes the massive relexification of vocabulary as found in certain varieties of Nahuatl (cf. Hill and Hill 1986), the extensive use of code-switching strategies as typical of some lects of Paraguayan Guaraní (Gómez Rendón 2007b) or even the creation of mixed varieties (cf. Gómez Rendón 2008b). At the same time, those speech communities which have maintained their native languages show gradient levels of bilingualism among their speakers. One and the same speech community may have a wide variation of Spanish proficiency, from incipient bilingualism in older generations to fluent bilingualism in younger generations. This variation has resulted in a language continuum stretching from monolingual speakers of the native language (if any) to monolingual speakers of the national language (e.g. Muysken 1985: 392). On the other hand, attitudes towards language mixing vary from conservative stands that reject any form of borrowing and codeswitching (e.g. *Guaraníete* or pure Guaraní in Paraguay) to tolerant or even favorable positions with respect to language mixing (e.g. *Media Lengua*, a Spanish-Quechua mixed language in Ecuador). Attitudes and perceptions towards language mixing are often determined by social and historical factors rather than linguistic factors per se. The role of social and historical factors in a comprehensive understanding of language contact is the next topic.

2.4. Social and historical aspects of language contact and change

In communities of all sizes, from the tiniest villages to the biggest nations, language contact (which is itself a result of social history) has social consequences (Sarah Thomason, *Language Contact: An Introduction* 2001: 4)

This quotation highlights the importance played by nonlinguistic factors in the outcomes of contact as well as the nonlinguistic consequences of such outcomes. It also points out the need to include social and historical criteria in the analyses of

contact-induced language change. I address here how social history and nonlinguistic factors result in distinct linguistic outcomes.

As a means of communication, language is instantiated in society through individual verbal behavior. This instantiation is studied by sociolinguistics and makes the point of departure for contact linguistics. Language does not exist outside society and acoustic signals get their meaning only when used in communication. As a result, languages are subject to the specific conditions of their societal usage. Just like speech communities undergo transformations by the influence of external factors, languages experience changes.

An external factor influencing the drift of language change is the colonization of a human group by another group with a different language. In the last five hundred years language contact proliferated all over the world as a result of the expansion of Western civilization. The dissemination of the Spanish language is illustrative of this. Further external factors such as slavery and epidemics played a major role in the linguistic diversity of European colonies through the decimation of aboriginal speech communities. Clearly, the outcomes of contact can be properly understood only through social and historical motivations. Any explanation focusing on linguistic factors only falls rather short.

But linguistic factors do play a role in contact-induced change. Languages are not aggregates of sounds and words but sets of linguistic signs arranged in a system of rules and patterns. To this extent language change is expected to follow the paths of development determined by the linguistic systems involved. In other words, the linguistic system defines the scope of change in contact situations. Contact-induced change is not fundamentally different from internally motivated change; what makes it different are the factors intervening in each case. Internally motivated changes follow the paths traced by the system (e.g. the vowel shift in English or the consonant shift in Germanic languages) just like contact-induced changes occur within the boundaries of the system (e.g. the emergence of mixed varieties of Romani in different European countries).

There is no consensus among scholars concerning the ultimate reasons of language change. Some authors point out that linguistic forces are powerful enough to operate changes in language (Chomsky 1978). Others maintain that linguistic changes are mainly the result of some kind of language contact (Thomason 2001; Winford 2005). It is not my purpose here to take a stand in this debate but to illuminate the complex ways in which social factors interact with linguistic ones. My position with respect to the explanation of contact-induced change is explained in section 2.6. For the time being, suffice it to say that any assessment of the causes of contact-induced language change should make two crucial distinctions: one between linguistic and nonlinguistic causes; and other between primary and secondary causes. The explanation of contact-induced changes as described here is dynamic and relational in nature as it weaves different causes in one single mould.

To give an idea of how complex the intervention of factors may be in language contact and change, I describe a unique feature of contemporary Paraguayan Guaraní: the borrowing of Spanish articles.

Guaraní has been in contact with Spanish in the last four hundred years. As a result of this century-long contact, Paraguayan Guaraní shows numberless traces of Spanish both in the lexicon and the grammar (cf. Chapter 7 & 8). One of these traces is the presence of deictic forms *la* and *lo*. Most students of Paraguayan Guaraní classify these forms as articles (cf. Gregores and Suárez 1967; Trinidad Sanabria 1998; Krivoshein de Canese 2001). Given the close similarity between these forms and the Spanish articles plus the fact that pre-contact Guaraní lacked the category of articles, the contact explanation seems self-evident. After a closer look, however, contact does not explain everything and several questions remain without answer: how do we explain the presence of articles in a language which originally had none?; how do we explain that other languages without articles (e.g. Quechua) which have been in contact with Spanish as long as Guaraní have *not* borrowed articles at all?; and how are deictic forms *la* and *lo* used in Guaraní? Tentative answers to these questions have been discussed elsewhere (Gómez Rendón 2007b). In my analysis both forms originate in Spanish articles but they are not used exclusively as articles in Guaraní: other, more frequent uses are anaphoric, cataphoric and elliptical. Moreover, the functions of *la* and *lo* resemble those of native deictics. Some of these native deictics are not used any more in contemporary Guaraní while the rest co-occur with the Spanish deictics. This suggests that the use of *la* and *lo* may be motivated by sociolinguistic and stylistic factors. An inquiry into the origin of this unique case of article borrowing shows that lexical chunks (frozen noun phrases) inserted as code switches in bilingual discourse should have been the source for the insertion of Spanish articles. Clearly, a contact explanation of article borrowing in Guaraní is not self-sufficient. A deeper investigation is required to unveil all the intervening factors and the intricacies of the processes of insertion and re-functionalization of grammatical borrowings.

So far I have insisted upon the importance of integrating social and cultural factors in the explanation of contact-induced change but have not mentioned their nature and scope. Myers-Scotton (2002: 31f) provides a list of six factors contributing to bilingualism which are, in her view, the primary causes of language contact. These factors include military invasion and colonization, living in a border area or an ethnolinguistic enclave, migration for social and economic reasons, formal education, the spread of international languages and the emergence of ethnic awareness. While Myers-Scotton is right in stressing the interplay of factors, her statement that the cycle of language contact is ignited always by bilingualism is not entirely valid. Indeed, it is bilingualism that is ignited by language contact and different levels of bilingualism result in different linguistic outcomes.

2.5. Contact situations and outcomes

The outcomes of contact may be grouped in three general categories: language shift, language maintenance and language creation. Each setting has certain consequences for the speech community and the languages involved. Shift implies second language acquisition and results in the loss of a community's native language. Maintenance in contact settings involves second language acquisition without loss of the native language, and mixing of elements from both languages through borrowing and code-switching. Language creation results in the emergence of novel varieties such as bilingual mixed languages, pidgins and creoles. The discussion of shift and maintenance is relevant for the present study inasmuch as both processes are underway in the speech communities investigated. Language creation in the form of mixed languages has been reported for the Ecuadorian Andes (Muysken 1985; Gómez Rendón 2005, 2008b) and Paraguay (Gómez Rendón 2007b) but will not be addressed here.

In the following I pay special attention to borrowing as it is the central topic of the book. I discuss several definitions of borrowing and pin down differences from code-switching. Also, I discuss the relation between borrowing and bilingualism and the processes of linguistic adaptation of loanwords in the recipient language.

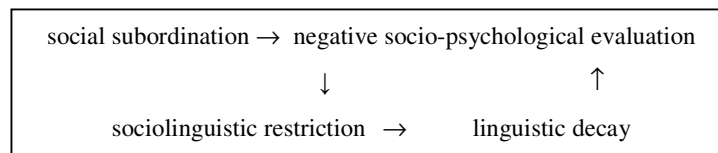
2.5.1 Language shift: second language acquisition and language death

The term 'language shift' describes the process in which one language – generally the native language – is replaced by another. Language shift may be described for individual speakers or speech communities. It may be gradual or sudden depending on a series of sociopolitical factors. At a societal level language shift is typically unidirectional as only one of the speech communities in contact abandons its native language for that of the other community. More often than not, the shifting community occupies a subordinate place as a result of colonization by a foreign group or domination by one sector of the same society. From this perspective shift is imposed on the subordinate community by the hegemonic group. There are cases in which the dominant group learns the language of the subaltern group (Latin-speaking groups in Greece are a good example) but these are rather exceptional. In either case language shift ends with the demise of one of the languages. In other words, language death is the end point of language shift.

But language contact *not always* results in language shift. The literature describes a large number of cases in which the subordinate group learned the language of the hegemonic group but did not abandon their own. A number of factors influence the decision of speakers to maintain or abandon their native language. These factors are also responsible for speeding up or slowing down the shift. Ethnolinguistic loyalty and positive attitudes towards one's language in

general promote maintenance. Negative evaluations of one's language usually trigger shift. Of course, negative evaluations are not gratuitous but the result of social subordination. Speakers whose social and economic status is lower as a result of their lack of employment and education usually think their language puts them in a disadvantageous position in relation to the speakers of the dominant language and view language shift as the best choice to gain social mobility. Language compartmentalization in diglossic settings usually leads to the loss of the subordinate language and triggers negative evaluations about this language's capacity to be an appropriate means of communication. To this extent social subordination is the beginning of a vicious circle of language decay that ends with the disappearance of the subordinate language. The circularity of the process has been sketched by authors like Dressler (1982: 325-325) to explain language endangerment, a process closely associated with language shift and death:

Figure 2.1 A model of language shift and endangerment



Language shift is not a necessary consequence of language contact but an expected result of social subordination. The position of the shifting speech community with respect to other speech communities is decisive. Thomason (2001: 23) identifies four positions in a contact situation which may influence shift or maintenance: indigenous superordinate; migrant superordinate; indigenous subordinate; and migrant subordinate. Each position is associated with either shift or maintenance: for example, an indigenous superordinate group will never shift but a migrant subordinate group will do so rapidly. In general terms, superordinate groups tend to maintain their language while subordinate groups usually shift to the language of the dominant group.

Focusing on Latin America, we find that language shift has been a steady process over the last five centuries. As a result of external and internal colonization in Latin America, ethnolinguistic groups occupy a subordinate position within their national societies. However, their linguistic reactions to subordination are not uniform: on the one side are speech communities that shift to national languages; on the other side are speech communities that maintain their native language in spite of having learned the official language. While language shift leads to societal monolingualism, language maintenance implies a diglossic distribution of languages across communicative spaces (cf. section 2.4).

In the speech communities of this study, language shift occurs differently depending on several factors. One factor inhibiting shift is the larger number of speakers of these languages in comparison with other minority languages. Another factor is the political position of the speech community in the mainstream society. A third factor is the ethnolinguistic loyalty and the awareness of the importance of language for the definition of ethnic identity in a multicultural society.

Quichua speakers make the largest non-Spanish speech community in the country and enjoy a strong political position.¹⁰ Both factors have certainly increased their ethnolinguistic awareness and slowed down language shift. A 1993 survey of the vitality of Highland Quichua (Buttner 1993) found that the native language was widely spoken at family and community levels across provinces while the great majority of Quichua speakers were bilingual to different degrees.¹¹ The same survey found that language shift is particularly visible in immigrants who move to the cities for work and learn Spanish in order to increase their socioeconomic mobility. Still, urban migration does not necessarily result in language shift. Urban Quichua speakers maintain their native language as a means of communication in domestic spaces. In other words, Quichua in the cities become an in-group language agglutinating speakers of the same sociolinguistic background and furthering group cohesion.

Otomí shows a higher degree of language shift than Quichua (Bakker and Hekking 1999: 6). The speed of this shift varies across dialects and areas. A major cause is the subordinate status of the Otomí speech community in relation to the Spanish-speaking society and other better positioned Indian groups (e.g. Nahuatl speakers). The lack of language revitalization and education programs adds to social subordination to setting the conditions for a rapid shift to Spanish.

The case of Paraguay is notoriously different from the other two and particularly interesting from a sociolinguistic point of view. Paraguay boasts a ninety percent of bilingualism among their national population. The native language (Guaraní) is spoken both in the cities and the countryside, even if there is a high degree of mixture with Spanish and the language occupies a subordinate position vis-à-vis Spanish. Still, language shift to this language is reduced to the minimum.

¹⁰ A reasonable estimate is 1.500.000 Quichua speakers in the highlands and the lowlands.

¹¹ Chimborazo and Imbabura are the provinces with the largest Quichua-speaking population in the country. Chimborazo is the largest with some 250.000 Quichua speakers, followed by Imbabura with some 70.000 speakers (source: www.abayala.org). These figures differ from those presented by *Ethnologue* (2005), for which speakers of Quichua in Chimborazo are 1.000.000 while Quichua speakers in Imbabura count 300.000. The differences lie on the fact that *Ethnologue* figures include the overall ethnic population, i.e. all those Indians who do not speak Quichua but consider themselves Quichua. Reasonable estimates for both provinces are around 350.000 speakers for Chimborazo and 200.000 for Imbabura. In both provinces, however, the number of bilinguals is different, with a higher degree of rudimentary bilingualism in Chimborazo.

Unlike the situation of Quichua in Ecuador and Otomí in Mexico, urban migration in Paraguay does not trigger shift but reinforces Guaraní-Spanish mixing through borrowing and code-switching (cf. Gómez Rendón forthcoming/b). As I show in the following, language mixture is typical of language maintenance in diglossic settings.

2.5.2 Language maintenance and mixing

The contact between one language in superordinate position and another in subordinate position does not end necessarily in language shift. Provided a number of conditions are met, subordinate groups can maintain their native languages even if the language of the hegemonic group continues to be the privileged means of communication in the larger society. How stable maintenance may be in diglossic settings remains unclear. If pressure on the subordinate group increases for some reason and the group's ethnolinguistic loyalty weakens as a result of migration or intermarriage, the conditions are set for a rapid shift to the language of the dominant group. So far there are no sociolinguistic techniques that predict this type of changes. What is clear from the literature is that languages do not remain the same after contact.

Contact-induced language change requires some knowledge of a second language at the level of the speaker and certain degree of bilingualism at the level of society for a rapid dissemination of innovative forms in the speech community. Speakers with higher or lower levels of bilingualism develop a number of communicative strategies,¹² a cover term of which is language mixing.

Language mixing refers to the mixture of lexical and/or grammatical elements of languages in contact.¹³ I prefer the term 'language mixing' to 'code mixing' for two reasons: 1) it is used also to refer to the mixture of registers or dialects; 2) 'code mixing' is used by some authors as equivalent to intrasentential codeswitching or switching of languages within the same sentence. The term 'language mixing' is therefore less ambiguous. It encompasses a great variety of linguistic phenomena associated with distinct sociolinguistic settings and distinguished on the basis of criteria such as phonetic accommodation, morphosyntactic integration, resemanticization, and frequency of use. Two of these phenomena are borrowing and code-switching.

¹² Assuming that bilingual speakers do not mix their languages is misleading. Quite the opposite, speakers with higher levels of bilingualism tend to mix their languages frequently with a variety of purposes (Thomason 2001: 53f). For an illuminating study of code mixing in the speech of Turkish immigrants in the Netherlands, see Backus (1996).

¹³ Another cover term equivalent to language mixing is offered by Muysken, who prefer to speak of language interaction "as a very general cover term for different, frequently highly innovative, results of language contact, both involving lexical items (as in code-mixing) and otherwise (e.g. phonological and syntactic interference)" (Muysken 2000: 1). Others like Holmes (1992: 34ff) make no clear distinction between code-mixing and codeswitching.

Borrowing and code-switching: a critical overview of definitions

The need to distinguish borrowing from code-switching was put forward first by Pfaff in a paper that summarized the state of the art on code-switching and borrowing at that time (1978: 295ff). The need to differentiate both phenomena has become more urgent since a great number of studies on borrowing and code-switching have appeared in the last decades.

Haugen pioneered a systematic study of borrowing in his article *The Analysis of Linguistic Borrowing* (1950). He defined borrowing as “the attempted reproduction in one language of patterns previously found in another” (Haugen 1950: 212). While he admitted that the term failed to describe the metaphor of mixture in proper terms, he believed it offered an advantage for its unambiguous use in linguistics as compared to others of ambivalent currency.¹⁴ The term however was never as explicit as Haugen thinks. Since its origin in nineteenth-century historical linguistics, *borrowing* was used as the dustbin for everything that could not be explained in terms of sound laws (Myers-Scotton 2002: 234). Over the years *borrowing* was associated with lexical elements and became synonymous of *loanword*.¹⁵ However, *borrowing* and *loanword* do not refer to the same things. Borrowing refers to the linguistic elements and the process of incorporation of these elements into the recipient language. Loanword refers to the linguistic elements proper. Still, some authors suggest a process behind such compounds as *loan blend*, *loan shift* and *loan translation* (Crystal 2006: 275). Henceforth I use *borrowing* or *loanword* to refer to linguistic units being borrowed and *borrowing process* to refer otherwise.

Two further comments about terminology are pertinent. One has to do with the use made by Weinreich (1956) of the term *interference* as synonymous with borrowing. In his use of the word, *interference* refers both to lexical and structural (grammatical) interference:

The ways in which one vocabulary can interfere with another are various. Given two languages, A and B morphemes may be transferred

¹⁴ “The metaphor implied is certainly absurd, since the borrowing takes place without the lender consent or even awareness, and the borrower is under no obligation to repay the loan. One might as well call it stealing, were it not that the owner is deprived of nothing and feels no urge to recover his goods” (Haugen 1950: 211). Some of these ideas led Johansson to propose the term ‘code-copying’ as a more felicitous term. For a discussion of his terminology and the implications for contact research see Johansson (1998).

¹⁵ Nowadays both terms are used interchangeably albeit the former is still the cover term. For example, Thomason speaks of lexical, morphological and syntactic borrowing (2001: 70-1). Some authors use grammatical borrowing and syntactic borrowing as synonyms (Campbell 1995) while others speak of the borrowing of word order patterns of one language into another (Heine 2005; Heine and Kuteva 2005).

from A into B, or B morphemes may be used in new designative functions on the model of A-morphemes with whose content they are identified; finally in the case of compound lexical elements, both processes may be combined [...] The parallelism with the formulation of grammatical interference is evident. Equivalence of designative function here corresponds to identity of grammatical function in the previous chapter. The separation of the grammatical and lexical aspects of interference presupposes, of course, that many morphemes do have a designative function distinct from their purely grammatical function. (Weinreich 1968: 47)

The second comment concerns a similar use of the term *interference* in Thomason (2001). This author considers interference a cover term for borrowing and shift-induced interference. The decisive criterion is the occurrence of imperfect language learning. When imperfect learning of a second language plays no role and no language shift takes place, the outcome of contact is the borrowing of linguistic features from another language. When imperfect learning does play a role and language shift is in progress, the outcome of contact is shift-induced interference produced by native speakers of one language in the language they are learning. In other words, borrowing is a mirror image of interference because the effects on language are similar but their direction is the opposite. Borrowing affects first the lexicon and then morphology, syntax and phonology, provided contact is intense enough. Interference begins with grammar and affects the lexicon only later, though not necessarily. Thomason admits possible exceptions to the direction of both processes. In particular she points out their simultaneous occurrence in certain contexts:

A possible exception to this generalization might occur if the shifting group is a superstrate, a socio-economically dominant group, rather than a substrate [...] since in most group shift situations it is not the dominant group that shifts, however, most cases of shift-induced interference support the basic prediction. In fact, it is fairly easy to find examples of mutual interference, borrowing by dominant-language speakers and shift-induced interference by subordinate-language speakers that directly illustrate the contrast between the two types of interference. (Thomason 2001: 75f)

Unlike the study of linguistic borrowing, that of code-switching is of relatively recent origin in linguistics. Still, it has received increasing attention by linguists and sociolinguists in the last decades for the social functions it performs and the insight it offers into the processing of language in the bilingual mind. Even if codeswitching is not the topic of this book, it is necessary to make a distinction between codeswitching and borrowing as two different mixing strategies in bilingual

discourse. From the discussion of several definitions I identify differences between codeswitching and borrowing on the basis of linguistic features. These are used as heuristic criteria for the analysis of language data in the frame of this study.

Gumperz (1981) defined codeswitching in broad terms as “the juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems” (1981: 59).¹⁶ More recently, Thomason has defined codeswitching as “the use of material from two (or more) languages by a single speaker in the same conversation” (2001: 132). Both definitions are equivalent in principle, but differences can be identified as well. First of all, Gumperz’ definition speaks of grammatical systems or subsystems while Thomason speaks of languages. While codeswitching occurs in languages, dialects and registers, a great number of sociolinguistic studies on codeswitching deal with occupational and domestic varieties rather than with languages in general.¹⁷ The second difference lies on the inclusion of more than two codes (be it languages, dialects or registers) in Thomason’s definition. The third difference is that codeswitching occurs within the scope of the conversation for Thomason but within the same speech for Gumperz. If we consider speech and conversation synonyms of speech event, both definitions are then roughly equivalent. What is crucial anyway is that codeswitching occurs within one exchange and not across turns.

Certain definitions of codeswitching make a distinction between intersentential switching (which occurs at sentence boundaries) and intrasentential switching (which occurs within the sentence). Others include (lexical) borrowing as one type of codeswitching (Muysken 2000: 32). In my view, borrowing and codeswitching are distinguished as separate phenomena on the basis of several linguistic criteria (cf. *infra*).

The relation of codeswitching to diglossia and bilingualism deserves some comment. Ideally, a diglossic situation in which the use of languages is compartmentalized impedes the emergence of codeswitching. To this extent diglossia and codeswitching exclude each other (Romaine 1989: 111). Recent studies show not only that both phenomena are not opposite, but also that bilingualism does not necessarily imply codeswitching. The results from our investigation provide additional evidence of this statement.

Borrowing and code-switching: differences and criteria for distinction

The discussion about the best procedure to differentiate borrowing from codeswitching is not settled. Still, several criteria have been put forward to establish

¹⁶ This definition has been adopted, among others, by Suzanne Romaine (1989: 111).

¹⁷ Sociolinguists use the term ‘style shifting’ as interchangeable with codeswitching (cf. Crystal 2006: 79).

such distinction (Poplack *et al* 1987; Romaine 1989; Poplack and Meecham 1998; Thomason 2001).¹⁸ The following is an overview of the relevant literature.¹⁹

One way to make a distinction between codeswitching and borrowing is by establishing the bilingual or monolingual condition of the speaker. If language mixing occurs in monolingual speech, the process at work is borrowing.²⁰ If mixing phenomena occur in bilingual speech, the process involved is codeswitching. The reason is simple: for codeswitching to occur, the speaker must know both linguistic systems; for borrowing to occur, only one system is required.

Another way to distinguish codeswitching from borrowing is the frequency of foreign elements. Foreign elements that disseminate in the speech community and recur in individual speech become established borrowings as opposed to code switches that are more idiosyncratic to the extent they serve different discursive, social and psychological purposes.²¹ There exists, however, a special type of borrowings identified in the literature as nonce borrowings (cf. Poplack and Meechan 1995) or single occurring elements (Myers-Scotton 2002: 153ff). Nonce borrowings are characterized as occurring only once in discourse and being of limited distribution in the speech community. They are single (content) words perfectly integrated to the morphology and syntax of the receiving language despite their non-recurrence in individual speech. Unlike established borrowings which are fully accommodated to the phonological system of the recipient language, the phonological integration of nonce borrowings is incomplete. The question is therefore how to distinguish nonce borrowings from code-switches.²² For some authors the answer is the structural integration of the foreign elements (Poplack *et al* 1987). Often referred to as *nativization*, the integration of foreign elements is the third criterion to distinguish borrowing from code-switching: only singly occurring

¹⁸ Poplack and Meechan (1998), on the other hand, maintain that singly occurring forms are nonce borrowings and not code switches. Far from being definitive, their proposal leaves a number of questions without answer, such as how to distinguish between established borrowings and nonce borrowings.

¹⁹ One related topic not addressed here for reasons of space but relevant to the effects of contact-induced language change is the idea that codeswitching results in borrowing through the crystallization of complex lexical items or chunks. For a discussion of the possible causal relation between both phenomena and their relation to contact-induced language change, see Backus (2005) and Field (2005). Field is particularly clear about the non-causal relation not only between codeswitching and borrowing but also between both phenomena and contact-induced language change (p. 341s).

²⁰ Bilingual speakers too may produce monolingual discourse if the use of languages is compartmentalized as typical of diglossic situations. Cf. *supra*.

²¹ Types of codeswitching are, among others, topic switching, metaphorical switching, switching for affective functions or simply switching for flagging group identity. For a discussion of codeswitching types, see Romaine (1989: 112ff) and Holmes (2001: 34ff).

²² In a different perspective, nonce borrowings may result simply from the smallness of typical corpora. Thus, either they are infrequent borrowings – an early attestation of a new loanword – or indeed an instance of a rare code switch (Dik Bakker, p.c.).

elements adapted to the phonological, morphological and syntactic patterns of the recipient language are borrowings; those which fail to adapt are code switches.

However useful these criteria are for a distinction between (nonce) borrowings and code-switches, both of them are far from providing definitive answers. The reason is twofold. First, the frequency of occurrence of foreign elements at individual and societal levels – which is an index of their nativization – is difficult to measure with accuracy. Second, phonological nativization may be a function of the speaker's bilingualism and newer loanwords may not be nativized anymore. This view is sustained by Thomason (2001: 135) and corroborated by my data of Quichua and Guaraní, where recent Spanish loanwords are widespread in social discourse and adapted to the morphology and syntax of the recipient language even if they remain phonologically unintegrated.

Further criteria for distinguishing borrowing from codeswitching are Sankoff and Poplack's (1981) free morpheme constraint and equivalence constraint. The first constraint establishes that "a switch may not occur between a bound morpheme and a lexical form unless the lexical form has been phonologically integrated into the language of the morpheme" (Romaine 1989: 115). The equivalence constraint "predicts that code-switches will tend to occur at points where the juxtaposition of elements from the two languages does not violate a syntactic rule of either language" (*op. cit.* 116). Both constraints have been shown to fail in the case of hybrid forms (cf. Eliasson 1990; Moffat and Milroy 1992) and my data corroborate the non-applicability of these constraints in several cases.

In the cross-linguistic analysis of borrowing pursued in this study the following criteria were followed in order to identify foreign elements in discourse:

- a) Morphological and syntactic integration in the recipient language, including participation in inflectional and derivational processes and native word order patterns and constructions.
- b) Phonological adaptation to the recipient language, including raising and lowering of vowels, observance of stress patterns and syllable structure, and other phonotactic criteria.
- c) Resemanticization of foreign elements in the recipient language.
- d) Frequency of occurrence of foreign elements in one speaker and across speakers.
- e) Frequency of occurrence of word classes in one speaker and across speakers.

Despite the overall applicability of these criteria, there are frozen idiomatic expressions which fail to make a clear case for borrowing or code-switching. These and other problems in the analysis of the data are addressed in Chapter 4.

2.6. An explanatory model of contact-induced language change through borrowing

In this section I resume the discussion about the role of linguistic and nonlinguistic causes in the explanation of contact-induced language change. After some terminological distinctions, I outline a multi-causative model for the explanation of borrowing and discuss the types of causes involved (motivations, factors and conditions) and their interplay in the shaping of linguistic outcomes.

2.6.1 Some terminological distinctions

In functional explanations there is an indiscriminate use of terminology referring to the causes of language change. Reasons, motives, motivations, factors, constraints and triggers are some of the most used terms. They are used interchangeably more often than not and without any previous discussion. The impression I have from reviewing most of the literature on the topic is that authors usually take for granted what these terms refer to.²³ My position is that any investigation into the causes of contact-induced language change must define the use of these terms in the frame of an explanatory model.

The model presented in this section is functional in nature as it explains language change in general and borrowing in particular as a result of a series of causes motivated by contact among languages. The model is based on certain terminological distinctions that seek to identify the causes of contact-induced language change in more precise terms. These distinctions are based on the following definitions:

- **Cause:** a cover term for any nonlinguistic *circumstance* or any linguistic *element* which produces or prevents changes in language. Causes are classified in primary, secondary and tertiary. Primary causes correspond to the ultimate explanations of language change, i.e. those circumstances which first unchain a series of events leading to change. Secondary and tertiary causes add to primary ones to advance or prevent changes in language.
- **Motivation:** a term referring to primary causes. Motivations are nonlinguistic and include social, geographical and communicative. Motivations influence secondary and tertiary causes and induce language change even if these causes are absent.

²³ Most books on contact linguistics do not include a glossary of terms. And if they do, they do not provide full explanations of such terms. Thomason's *Contact Linguistics* (2001), for example, includes an extensive glossary of terms but none of those mentioned above even though they are used throughout the book.

- **Factor:** a term referring to secondary causes. Factors are linguistic (the system of language) and sociolinguistic (speech community, language loyalty, linguistic self-perception, etc.). Factors are those circumstances or elements which *inhibit* or *promote* language change. Factors do not act on their own but interplay with motivations (primary causes) to produce language change. Their influence may be either increased by promoting conditions (tertiary causes) or reduced by inhibiting conditions (tertiary causes). Inhibiting conditions are sometimes referred to as constraints in the literature while promoting conditions are often equivalent to triggers. Here I make a fundamental distinction between factors (secondary causes) on the one hand, and triggers and constraints (tertiary causes) on the other.
- **Conditions:** a term referring to tertiary causes. Conditions are linguistic (speech events, word frequency in the recipient language) and sociolinguistic (speaker variables such as age, gender or education). Conditions are those circumstances or elements which *speed up* or *slow down* language change. They are classified as positive or negative accordingly. Conditions do not act on their own but interact with factors in such a way that the influence of factors is increased or reduced.

The hierarchy of causes goes from primary (motivations) to secondary (factors) to tertiary (conditions). There is a general primacy of nonlinguistic causes over linguistic ones at all levels. This predominance is based on the notion that nonlinguistic circumstances are the ultimate causes of contact-induced language change. Nonlinguistic causes are also distinguished from linguistic ones in that the speaker is aware of the influence of nonlinguistic causes on his/her linguistic behavior (e.g. identity, loyalty, prestige) while linguistic causes usually act beyond the speaker's consciousness (e.g. markedness, inflection, paradigmaticity, etc.). On the other hand, the interplay of causes is not excluded provided the primacy of nonlinguistic causes is observed. This interplay is not always symmetrical. Motivations may influence factors but not the opposite, but factors and conditions may influence each other. Motivations, factors and conditions of linguistic borrowing are discussed in the following section in the context of an explanatory model of contact-induced language change.

2.6.2. A functional explanation of contact-induced language change

Functional explanations of language change are based on the notion that languages are *not* autonomous entities evolving on their own but the result of socio-communicative needs. Not leaving aside the inherently systematic organization of language as a coherent set of elements governed by rules and patterns, functionalism in linguistics privileges a holistic view that comprehends not only rules and patterns

but also concrete uses and communicative needs as determined by social praxis. In this perspective language is viewed as an individual behavior anchored in social practices. Consequently, changes in language are interpreted as adaptations of the linguistic system to the changing circumstances of society, which determine the communicative needs of individual speakers and speech communities. These adaptations are by no means random but obey the constraints of the linguistic system. In other words, changes in the linguistic system are externally motivated but internally ruled. This premise sustains most functional views of language change and embodies the paradigm of the present investigation, the goal of which is to provide support to it through the typological and sociolinguistic analysis of empirical data.

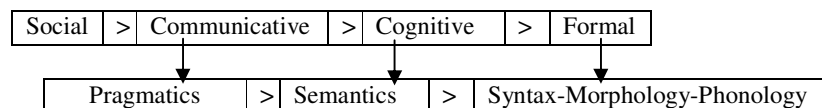
The model presented here is framed in the Principle of Functional explanation as elaborated by Dik (1986). This author studied the different elements that enter into a functionalist explanation of language change and grouped such elements in different categories ordered from the nonlinguistic to the linguistic. These hierarchies form the basis of the Principle of Functional Explanation.

Figure 2.2 The Principle of Functional Explanation (adapted from Dik 1986)

Nonlinguistic:	Social	>	Areal	>	Discourse	>
Linguistic	Pragmatic	>	Semantic	>	Formal	>
Formal	Syntactic	>	Morphological	>	Phonological	.

The Principle of Functional Explanation consists of one hierarchy containing three different subhierarchies of nonlinguistic, linguistic and formal causes. In this model nonlinguistic causes have primacy over linguistic ones. Within linguistic causes, the functional, pragmatic and semantic factors are placed higher in the hierarchy than formal factors of syntactic, morphological and phonological character. Bakker and Hekking (1999) offer a contact-induced change interpretation of the Principle which extends the model discussed here. For these authors “the higher factors give the motivation for languages to adopt and incorporate external elements [while] the lower factors provide the constraints on processes of language change while at the same time motivating still lower factors in a cascade-like way” (Bakker and Hekking 1999: 4). In Bakker *et al* (2008) the Principle is collapsed in one hierarchy in somewhat different terms. The linguistic part of this hierarchy corresponds to the levels of grammar as shown below:

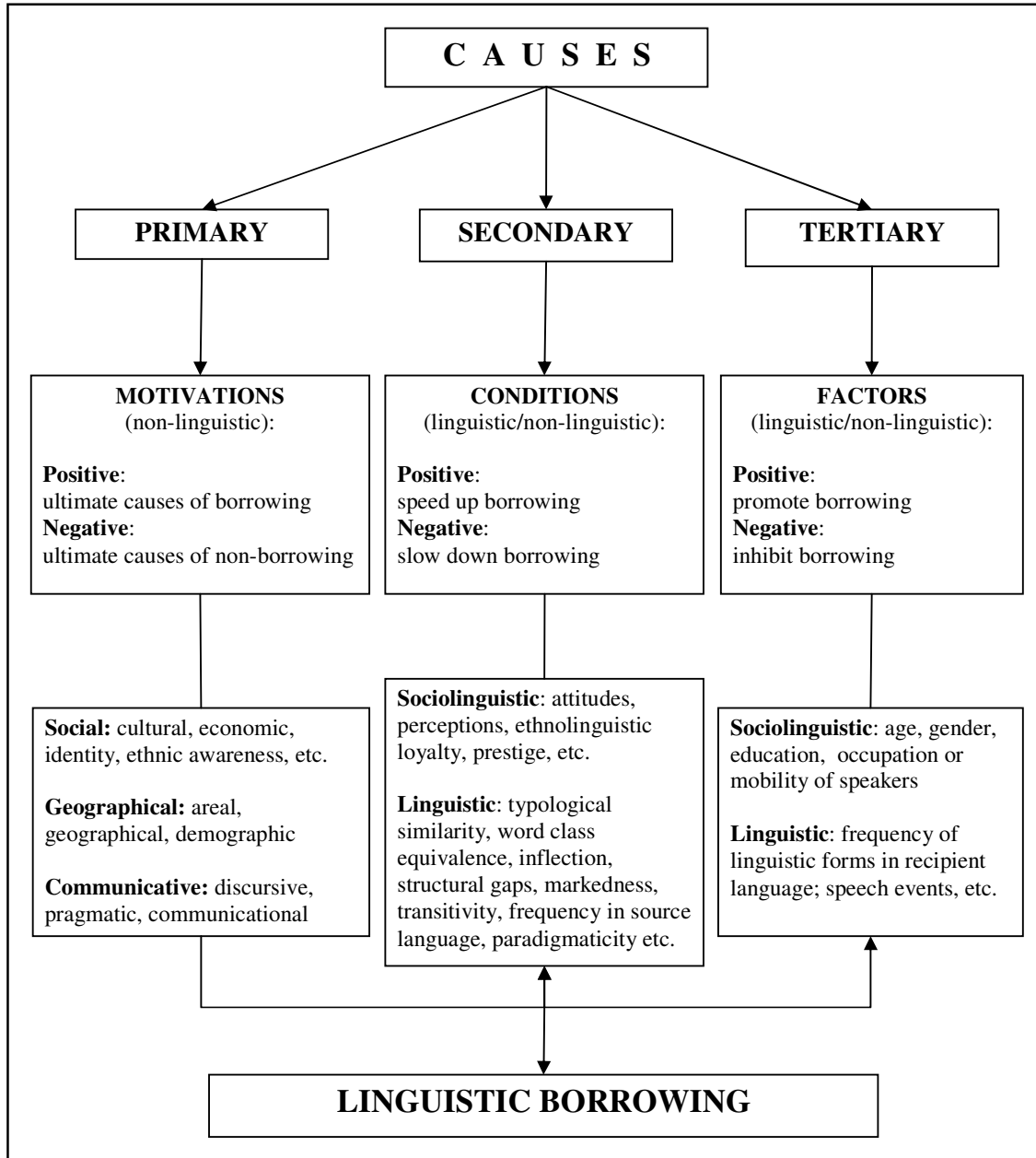
Figure 2.3. The Principle of Functional Explanation



Both hierarchies distinguish between nonlinguistic and linguistic causes but still are different. Bakker *et al* do not include areal and discursive factors in the nonlinguistic subhierarchy. As far as formal factors are concerned, both hierarchies include three levels corresponding to the subsystems of grammar. Also, both hierarchies place social causes at the top and formal (linguistic) factors at the bottom, with communicative and discursive causes in the middle. The same arrangement of causes is preserved in the model outlined in Figure 2.4. The main characteristics of this model are summarized as follows:

- a) The model makes two crucial distinctions: one between primary, secondary and tertiary causes; and another between motivations, factors and conditions. Each causal element occupies a place in the model which corresponds to a place in the hierarchies proposed by Dik (1986) and Bakker *et al* (2008).
- b) At the higher level, motivations are classified into social, physical or discursive. At lower level, factors and conditions are grouped in linguistic and sociolinguistic.
- c) The nonlinguistic-linguistic distinction traverses all the levels of causation and separates motivations from factors and conditions. At the same time, the linguistic-sociolinguistic distinction establishes a further division within factors and conditions.
- d) The model is dynamic to the extent that feedback is permitted at different levels of causation. In general, motivations and factors work together, and so do factors and triggers. However, motivations may either intensify or cancel the contribution of linguistic and sociolinguistic factors and conditions. Also, the effect of factors may be intensified or weakened by conditions, just like these may be intensified or eventually canceled by factors.
- e) At the lowest level, conditions do not effect changes directly but act through factors. Similarly, factors induce language change through motivations. This means that motivations determine the eventual effects of factors and triggers and the final shape of contact-induced change.
- f) Even though motivations, factors and conditions are ordered in a hierarchy, the model enables the interplay of causal elements provided the hierarchy is observed. This interplay reflects the dynamics and the multi-causality of the model, with different elements contributing to effect changes in language but each at its own level.

FIGURE 2.4. CAUSATION MODEL OF CONTACT-INDUCED CHANGE



In the following I illustrate each type of cause by means of examples focused on linguistic borrowing in order to show the interplay of motivations, factors and conditions.

Motivations of language change

Motivations are by definition nonlinguistic. They may be of three kinds, namely, social, geographical and communicative. Social motivations are those which have to do with changes in the structure of human groups. An example of a social motivation for language change is the reorganization of a colonized society by the colonizers. This reorganization may concern, among other things, the use of language in education. Inca rulers in the Andes used to bring young children of the local elites to their schools in Cuzco where they were raised in the Inca language and became bilingual in their parents' language and Quechua (Manheim 1991). Social reorganization in colonial settings may also affect the use of language in public spheres. After the conquest of Tawantinsuyu, Spanish replaced Quechua as the official language of the former Inca Empire and became obligatory in official transactions, even for those who did not speak the language – who were the great majority of the Indian population in the first decades of colonization – thereby introducing an important pressure for the learning of Spanish. Other social motivations include migration and social mobility. In these cases individual speakers usually adopt bilingualism as a strategy for accommodating to the state of affairs, with the result of their languages influencing each other in different ways. Forced migration was common during the Inca rule in the Andes. It consisted in the uprooting of insurgent populations from their original places to other parts of the empire with the purpose of suppressing rebellions or helping the Inca take control of the new colonies. This practice explains the existence of a few Aymara words in several dialects of Ecuadorian Quichua but also the occurrence of Cañari words in Bolivia in spite of the thousands of miles that separate the respective speech communities.¹ This practice had important demographical consequences resulting in the transformation of the linguistic landscape of the northern Andes.

No less important for language change are geographical motivations. These include areal, demographical and geographical proper. Areal motivations are related to the distribution of peoples and languages in a geographical space. They determine contact between peoples who speak dialects of the same language or languages from different families depending on their distribution over a specific territory. The distribution of languages of the Guaraní family along the eastern Atlantic coast of

¹ Cañari was one of the nine languages spoken in Ecuador before Inca invasion. Although no grammar or dictionaries of these languages exist, their presence is well recorded by history, toponymy and anthroponymy. The most extensive study so far on the aboriginal languages of Pre-Inca Ecuador is due to Jacinto Jijón y Caamaño (1940).

South America called the attention of the first Portuguese settlers, who became aware of the similarities across these languages and used one of them (Tupi) for the colonization of the Atlantic coast and faraway places in the heart of the Amazon basin (Holm 1989). Over the years Tupi became the *lingua franca* of large areas in Brazil and influenced non-Guaraní languages. In the case of borrowing areal motivations explain the occurrence of allochthonous substrata in languages with a long history of contact with neighboring peoples. The existence of loanwords from Tsafiki (Barbacoan) in Imbabura Quichua is explained by the areal distribution of Barbacoan languages in Northern Ecuador in the past (Gómez Rendón and Adelaar, forthcoming).

Demographical motivations induce language change to the extent that the size of speech communities determines the rate and diffusion of contact-induced changes. Demographical motivations also influence group cohesion and affect ethnolinguistic loyalty and awareness. Demographic motivations are crucial in cases of language death as a result of a rapid demographical collapse caused by extinction or genocide. Reports on catastrophic events influencing language change are not uncommon. Dixon (1991: 241) mentions the extinction of the Tamboran language as a result of a volcanic eruption. Similarly, glottocide in Africa has been reported within the Khoisan language family and is responsible for the high rates of language shift and death until today in the area (Tsunoda 2005: 43). In the Americas, extinction and glottocide as motivations for language contact and change resulted from epidemic outbreaks in the first years of European colonization or slavery in rubber estates during the last half of the nineteenth century (Trujillo 1998: 460).

Geographical motivations proper have to do with the milieu of speech communities, the use of natural resources and the patterns of settlement. Geographical motivations determine linguistic processes such as dialectalization, language death or language contact with other speech communities. The speech community of Sia Pedee (Chocoan) in Ecuador is illustrative of this. Colombian Sia Pedee speakers migrated to Ecuador motivated by the exhaustion of resources in their original milieu by non-Indian colonization (Prodepine 1999). Because of their separation from the larger Sia Pedee community in Colombia and their everyday contact with speakers of Spanish, the Ecuadorian variety of Sia Pedee shows high degrees of Spanish borrowing and codeswitching accompanied by rampant levels of Hispanicization in the younger generations. According to the last sociolinguistic survey, the process of language shift in the Sia Pedee community will be completed in two generations with the eventual demise of the native language (Gómez Rendón 2006c).

Finally, communicative motivations as primary causes of contact-induced change encompass a rich gamut of causes which have to do with the transmission and receipt of information among speakers of one or different linguistic systems (languages, dialects, sociolects, registers). Communicative motivations include

discursive such as those determined by the organization of messages through language; pragmatic motivations such as those associated with the use of the linguistic system in specific social contexts; and communicational motivations such as those determined by the need to transmit messages in order to perform tasks in social contexts. An example of how communicative motivations induce language change in contact situations is illustrated by the first European conquerors in the Americas. They usually seized young male Indians in order to teach them Spanish and train them as translators. These *lenguas* (Sp. tongues) played a decisive role in the conquest as mediators between the Spanish monolingual conquerors and the Amerindian monolingual population. Later on, the *lenguas* became linguistic leaders in their native communities and agents of language changes induced by contact. Another situation in which communication motivated language change is the trade of African slaves. Slave trade usually began with the uprooting of entire speech communities and their moving overseas. During their transportation and their subsequent settlement speakers from different linguistic backgrounds found themselves forced to communicate with each other for practical reasons. The result was the emergence of a number of pidgins and creoles used by slaves for in-group communication.

Nonlinguistic motivations are the ultimate causes of language contact and deserve special attention in any model of contact-induced language change. For the sake of analysis I have separated social from areal, geographical and demographic motivations, but all of them work together in scenarios of contact and should be considered as acting concurrently.

Factors of language change

Factors are one type of secondary causes which effect language change through motivations and are influenced by the latter. Factors may be linguistic and sociolinguistic and promote or inhibit contact-induced changes. The different types of factors are sketched in Figure 2.5.

Figure 2.5 Types of factors inducing language change

	LINGUISTIC	SOCIOLINGUISTIC
INHIBITING	INHIBITING LINGUISTIC FACTORS	INHIBITING SOCIOLINGUISTIC FACTORS
PROMOTING	PROMOTING LINGUISTIC FACTORS	PROMOTING SOCIOLINGUISTIC FACTORS

The attitude of speakers towards their language is one of the most important sociolinguistic factors influencing borrowing. Attitudes generally include sociolinguistic self-perception, ethnolinguistic loyalty, and linguistic awareness. The importance of sociolinguistic factors is well exemplified in the case of the Quichua-speaking community of Imbabura. The positive attitude of Imbabura Quichua speakers towards their native language and their ethnolinguistic loyalty are crucial factors in the maintenance of Quichua as compared to other Quichua communities with higher rates of language shift. Attitudes toward language mixing can be influential as well. Some speakers of Jopara (a heavily Hispanicized variety of Guaraní) disdain their speech for being ‘corrupted’ and cultivate purism through the invention of neologisms. In this case a negative attitude towards language mixing inhibits contact-induced change. On the other hand, language mixing is considered positive in certain multilingual settings to the extent that it facilitates intercultural communication.

The prestige associated with foreign elements is a further sociolinguistic factor promoting linguistic borrowing. Zimmerman notes that lexical borrowing from Spanish in Otomí was induced by the prestige associated with linguistic forms of the European language in native discourse (Zimmerman 1999: 299-305). Of course, prestige is relative to the position of one of the languages in contact and results in diglossic use. This is obviously the case of Spanish and Otomí in Mexico. However, my view of prestige differs from Zimmermann’s in one important respect. From his analysis it becomes clear that prestige is a primary cause of linguistic borrowing whereas my model has prestige only as a promoting factor less influential than other social and historical factors.

Linguistic factors inhibit or promote language change in contact situations. One factor can operate in both directions depending on the presence or absence of the linguistic feature in question. As noted above, the literature on language contact often treats inhibiting linguistic factors as constraints. Thus, typological distance between the source language and the recipient language is an inhibiting factor in the case of borrowing: the lack of a grammatical category in the recipient language may inhibit the borrowing of items from this category. Accordingly, a postpositional language shall not borrow prepositions. Also, a language without grammatical gender shall not borrow gender markers as distinctive, productive units of meaning, even if it borrows masculine or feminine nouns or adjectives. Similar constraints are operative when phonological distinctions are absent in the recipient language which are otherwise present in lexical elements from the source language. In this case borrowings undergo a process of phonetic accommodation to the phonology of the recipient language. These examples should not be read, however, as if inhibiting factors prevent languages from borrowing lexical or grammatical elements which may be alien to their linguistic systems. There are a great number of cases in which foreign elements are borrowed in spite of their non-compatibility with the linguistic

system of the recipient language. The borrowing of Spanish prepositions in Otomí, a language without this category, is an example. The borrowing of Spanish articles in Guaraní is another. More powerful (social) causes are at work in these particular cases.

Inflection is often mentioned in language contact studies as a linguistic factor inhibiting borrowing. The argument is that borrowing elements from inflectional languages is particularly difficult because form and meaning are not univocally equivalent, i.e. one bound form corresponds to several grammatical meanings (e.g. aspect, person and number). In contrast, agglutinative languages do show equivalence between form and meaning so that one morpheme usually corresponds to one meaning. Therefore, it is assumed that agglutinative languages shall borrow bound morphemes from inflectional languages only seldom. Of course, it is not only a question of morphological typology but also of the relative social position of one language with respect to the other. Nonlinguistic motivations may induce changes even if opposite to the morphological profile of the recipient language. Guaraní and Quechua have borrowed many verbs from Spanish but not bound forms of verbal inflection. These forms have been borrowed, however, in cases of long-term contact including Southern varieties of Quechua (cf. Campbell 1987; Campbell 1993; Carranza-Romero 1998; Thomason 2001). In general, inhibiting linguistic factors should be understood as forces which *resist* but not cancel borrowing, the final outcome depending on a number of other motivations and factors.

Linguistic factors that promote contact-induced language change include, among others, typological similarity, structural gaps, markedness, word class equivalence and frequency in the source language.

It is often assumed that typologically similar languages offer better structural conditions for borrowing. Typological similarity is no doubt operative in borrowing, but in-depth studies are required to establish the specific contribution of typology. Similarly, structural gaps favor borrowing to the extent they provide blank spaces to be filled by elements from another language. While linguists usually explain a number of contact phenomena by means of structural gaps, their explanatory value is controversial to say the least. The notion of ‘gap’ is relative and may lead to misinterpretations, as it suggests that some languages are more ‘complete’ and developed than others without a certain category or linguistic element.²

² Because the lack of certain linguistic features in one language is determined only with reference to the features of another language, the notion of gap implies structural *dissimilarities* between two different linguistic systems rather than structural insufficiencies in one of them. It is relevant to speak of a non-structural communicative or socio-cultural gap only when structural dissimilarities between two languages in contact produce communicative failures or the unsuccessful transmission of messages between speakers of these languages. Also, because any function in language implies the use of linguistic material to perform a communicative task it is relevant to speak of functional gaps only when structural

Markedness is another linguistic factor often mentioned as promoting borrowing. In principle, foreign items are more borrowable when they are not marked for a given linguistic feature. Take the example of marked nouns versus unmarked nouns in Spanish. According to the markedness argument, singular nouns are more borrowable than plural nouns because their word structure is unmarked and therefore more transparent to speakers of the recipient language. A further promoting factor of linguistic nature is paradigmaticity. The argument holds that the openness of a lexical or morphological class in the recipient language facilitates the borrowing of elements belonging to such class. In this perspective, open classes (e.g. nouns, verbs) are more borrowable than closed classes (e.g. pronouns, articles).³ This is related in turn to word class equivalence. Traditional linguistic theory considered word classes as cross-linguistic categories. However, several authors have demonstrated that lexical classes are language-specific (Schachter 1985; Hengeveld 1992; Hengeveld *et al* 2004). Hengeveld (1992) for instance shows that lexical classes are unevenly distributed in a sample of sixty languages. Hengeveld's theory of parts of speech is discussed in detail in section 3.2.

The last promoting factor of linguistic nature mentioned here is frequency in the source language. Because frequently used forms are more 'visible' to the borrower and more relevant from a communicative point of view, it is not unwise to assume that their borrowability is greater than that of less frequent forms which are less instrumental in communication. Obviously, any validation of this hypothesis requires a corpus-based study of frequencies in the source language.

Muysken and van Hout (1994) have evaluated most of the aforementioned factors thorough the statistical analysis of Spanish borrowings in Bolivian Quechua. Their conclusions are the following:

dissimilarities result in a failure by one language to convey the semantic and pragmatic meanings expressed by the linguistic structure of the other language in contact. This failure results in borrowing alien linguistic material to transmit the intended meanings. This view goes beyond a univocal semantic relation between form and meaning because two different forms may have the same referent but each is associated with a different set of cultural and pragmatic values. This is indeed the origin of couplets, composed of native and borrowed lexemes which signify one and the same object but from different perspectives. In the case of Otomí in contact with Spanish, Bakker and Hekking show that "more often than not, the [Spanish] loan turns out to be semantically more specific than the original element, at least in the reading in which it is borrowed" (Bakker and Hekking 1999: 3). While borrowing can be considered enrichment in this case, it is neither structurally nor functionally "necessary" for the borrowing language. Furthermore, the use of the loanword is felicitous only to the extent that it is associated with the cultural value given by the Spanish-speaking society. It becomes clear, therefore, that the notion of 'gap' is potentially misleading and should be used only with extreme caution.

³ Still, I have reported pronoun borrowing in Imbabura Media Lengua (Gómez Rendón 2005; 2008b) and article borrowing in Guaraní Jopara (Gómez Rendón 2007b).

“The B values show that paradigmaticity is the strongest structural factor in our model. The second strongest structural factor is inflection in the donor language. Frequency has a (somewhat weaker effect), whereas peripherality has a clear effect, but opposite to what we predicted. [...] The most difficult categories to borrow consist of functional elements that are nominal in nature and form tightly organized subsystems. [...] We can conclude that the constraints model, operating on the basis of a comparison between a donor and a recipient language corpus, seems to be a promising way of studying the process of lexical borrowing. The results may be interpreted in such a way as to set up a new hierarchy of borrowability” (Muysken and van Hout 1994: 60-61).

Muysken and van Hout warn us about generalizing these conclusions to language pairs other than Spanish and Bolivian Quechua. Still, their analysis sheds light on the effective incidence of linguistic factors on the borrowing process. While these factors have been often addressed in previous studies, none of them has pondered their contribution in quantitative terms on the basis of a corpus. My purpose here is to advance a quantitative analysis of linguistic factors on the basis of large corpora of spontaneous speech collected for three typologically different languages.

Conditions that speed up or slow down language change

The last category of causation corresponds to conditions. Conditions speed up or slow down contact-induced language change in specific situations. They occupy the lowest position in the hierarchy and their influence is mediated by motivations and factors. Because conditions do not operate on their own, motivations and factors intervene every time a condition is at work. To this extent conditions differ from factors by a) their dependent action, and b) the degree of influence they exert on the borrowing process.

From the point of view of the individual speaker, sociolinguistic conditions such as age, gender, education and spatial mobility influence the degree borrowing. Generally speaking, older generations are more conservative in their linguistic usage than younger ones, and women usually preserve more archaic forms in their speech than men. The majority of Quichua monolinguals are older women who have never left their home communities. Traditionally, gender-based views of language change in the Andes consider women as depositories of the linguistic heritage of their communities and men as innovators and shifters. But gender roles are rapidly changing in the Andes and today it is not uncommon to find bilingual young women who migrate to the cities on a regular basis for trade or waged-labor and become agents of language change in their communities. In fact, the role of women was

decisive in the emergence and dissemination of Media Lengua in Imbabura (Gómez Rendón 2005: 46).

The role of education is decisive for contact-induced language change. In the Andes, educated Quechua speakers are usually more conservative than non-educated speakers. To be sure, purism is characteristic of literate speakers. The majority of members from institutions meant to preserve the purity of the Quechua language in the Andes, such as the *Academia Mayor de la Lengua Quechua del Cosco* in Peru or the newborn *Academia de la Lengua Quichua* in Ecuador, are Spanish-Quichua bilinguals with university education.⁴ On the other hand, the position of literate speakers in Paraguay is ambiguous with respect to the use of conservative or innovative forms. Interestingly enough, purism may produce the opposite of the desired effect because illiterate speakers unaware of non-mixed choices prefer to switch to the dominant language instead of mixing (Floyd 2005). Fortunately, the effects of purism practiced in academic circles are limited to these circles for the most part. Purism is more influencing if fostered by the speech community itself and motivated by social circumstances such as ethnic awareness or ethnolinguistic loyalty. In this case purism is not simply a condition but a factor inhibiting language contact and change.

A good candidate for a linguistic condition is word frequency in the recipient language. It is different from word frequency in the source language (cf. *supra*) for the position it occupies in the hierarchy: frequency in the recipient language is a condition of contact-induced change while frequency in the source language is a factor. The reason for such a distinction is that word frequency in the recipient language is further determined by factors such as markedness or paradigmaticity (Muysken and van Hout 1994: 54).

The frequency of an element in the recipient language may influence borrowing in two ways: first, if a native form is very frequent in the recipient language, it may be resistant to be replaced by a borrowing; second, if a native form is very frequent in the recipient language but has two or more meanings, a borrowed form may take over the less common meaning. The first prediction has been demonstrated substantially by Muysken and van Hout (1994: 53). The second prediction is harder to test. To the best of my knowledge there is no statistical study of a bilingual corpus which analyses the semantic specialization of borrowings in the recipient language. The occurrence of doublets from two different languages might be considered indirect evidence. Traditional Nahuatl used doublets as a stylistic strategy (Silver and Miller 1997: 108). Contemporary varieties of this language keep using doublets, but in this case one member of the doublet comes from Spanish and its meaning

⁴ Academic purism may result in neologisms which violate the language's own rules of word formation. For an overview of purism and language contact as part of linguistic ideologies in the Andes, see Howard (2007: 345-348).

differs slightly from that of the native element. Similar findings are reported for Otomí, one of the languages of this study (cf. Bakker and Hekking 1999).

2.6 Summary

From the assumption that any definition of language contact must integrate linguistic and nonlinguistic elements, I approach contact phenomena from a perspective that links sociocultural settings to linguistic outcomes. The interaction of societal and individual aspects is reflected on the ways in which individual bilingualism and societal diglossia shape language usage and the communicative strategies of speakers in multilingual situations. These strategies induce either language shift or language maintenance but in either case respond to specific social and cultural motivations. Besides, each strategy is associated with different linguistic mechanisms. Shift involves primarily second-language acquisition while maintenance in contact situations involves language mixing in the form of code-switching and borrowing. Both mixing mechanisms consist in the simultaneous use of elements from two (or more) languages in the same speech event. Various criteria separate code-switching from borrowing, the most important one being the adaptation of foreign elements to the phonological and morphosyntactic structure of the recipient language.

An important part of this chapter was devoted to discuss an explanatory model of contact-induced language change. In order to determine the specific weight of causes and their individual contributions, I made two fundamental distinctions: one between primary, secondary and tertiary causes; and another between linguistic, sociolinguistic and nonlinguistic causes. Primary causes were identified with nonlinguistic motivations while secondary and tertiary causes were classified in factors and conditions of linguistic and sociolinguistic nature. Factors inhibit or promote contact-induced change. Conditions speed up or slow down changes. The model was characterized as dynamic and multi-causal, with primary, secondary and tertiary causes influencing each other and nonlinguistic, sociolinguistic and linguistic causes concurring to shape the outcomes of contact.

Chapter 3

Theories on Linguistic Borrowing

This chapter addresses the theoretical issues relevant to the analysis and interpretation of the data. The chapter is divided in three sections. The first section makes the reader familiar with the approach of Functional Grammar (henceforth FG), which provides a broader context for the theory of parts of speech developed in the second section. In a functional perspective the structure of human languages is best understood in relation to their communicative function. Following this premise, I address the theoretical and methodological principles of a functional approach from the standpoint of FG in section 3.1.1. How language contact is viewed from a functional perspective is the topic of section 3.1.2. The relevance of language contact for a functional theory of language is demonstrated in section 3.1.3 with respect to the standards of observational, descriptive and explanatory adequacy and the search for patterns of language contact. The empirical data of this study offer a solid testing ground not only for theories of contact but also for general theories of language. The data come from contact varieties, the social and typological nature of which differs from that of languages commonly analyzed in the literature. Section 3.1.4 deals with the motivations of language change and their relation to language contact within a functional approach that puts socio-communicative needs in the foreground.

The second section of this chapter discusses in detail the theory of parts of speech developed by Hengeveld (1992) and Hengeveld *et al* (2004) and the points in which this theory differs from other related proposals. A preliminary outline of the implications of the parts-of-speech theory for lexical borrowing follows in the third section. These implications are the point of departure for the predictions made in the following chapters about the three languages of the corpora. The fourth section explores the relationship between implicational hierarchies on the one hand, and scales of borrowability on the other. The fifth section offers a critical overview of language-contact hypotheses from the literature because they provide the backdrop against which borrowing data are analyzed in the chapters that follow. Of particular interest for the analysis are the scales of borrowability proposed by several authors during the last fifty years (Haugen 1950; Moravcsik 1978; Campbell 1989; Field 2002; Bakker *et al* 2008). The relevance of these hierarchies lies on their encompassing of lexical and grammatical borrowing, which allows for a unified treatment of linguistic borrowing. The ultimate purpose of this section is to incorporate the hypotheses from the parts-of-speech theory and the scales of borrowability within a coherent testing framework that includes not only structural but also functional factors.

3.1. The theory of Functional Grammar

Functional Grammar (Dik 1997) is one of several functional theories of language developed in the second half of the twentieth century in response to the increasing influence of formalist theories, in particular Generative Grammar. Different from formalist approaches by its stress on communication and usage, FG has a number of things in common with these frameworks, not least its effort to map the systematic organization of language and the basic assumption of underlying structures on which utterances are mapped. On these grounds functional theories like FG have been called 'structural-functional grammars' (Butler 2002: 1). In what follows I sketch the main principles of a functional view of language and pinpoint the specificities of FG in so far as they diverge from those of other functional theories.

3.1.1. A functional view of language

Functional views of language characterize language primarily as an instrument of social communication. Two corollaries of this assumption are that (i) language use is the point of departure for any theory of language; and (ii) language is part of a general social capacity of human beings to deal with other human beings and become part of human society. Accordingly, social context, communication and usage model functional views of language. Dik (1997) articulates this approach through the concepts of communicative competence and natural language use:

In the functional paradigm a language is in the first place conceptualized as an instrument of social interaction among human beings, used with the intention of establishing communicative relationships. Within this paradigm one attempts to reveal the instrumentality of language with respect to what people do and achieve with it in social interaction. A natural language, in other words, is seen as an integrated part of the communicative competence of the Natural Language User (Dik 1997: 3)

A functional stance calls to define what 'function' is in the first place. In his review of the multiple meanings of 'function', Nuyts (1992: 60) has shown that the primary communicative role of language goes hand in hand with other functions of informative, intentional, socializing and contextualizing nature. Therefore, when FG and other similar theories attribute themselves a 'functional' character, they imply two things: first, they understand language in its primary function of instrument of social behavior; second, they analyze the system of language *always* as a function of the uses it performs in society.

Viewing language in its instrumentality for human interaction not only brings the study of language back to social praxis; it also calls for an object of study that

integrates structure and use. In this perspective the Chomskyan division between internal language and external language is unnecessary and creates an illusory opposition. The knowledge of the language system and the use thereof in communication are indeed complementary. Dik explains this integrative view of functionalism in the following terms:

A theory of language should not be content to display the rules and principles underlying the construction of linguistic expressions for their own sake, but should try, wherever it is possible at all, to explain these rules and principles in terms of their functionality with respect to the ways in which these expressions are used. (Dik 1997: 4)

This leads to a major issue of discontent between formalists and functionalists in the study of language: the autonomy of linguistic knowledge and the consequent autonomy of grammar. As a matter of fact any conception of language as a set of rules for its own sake entails inevitably the independence of grammar (or syntax for that matter) from the setting in which speech occurs. On the contrary, a functional view of language as firmly rooted in society implies that linguistic knowledge is *not* independent and can be adequately explained only to the extent that it is linked to its instantiation in speech. According to Croft (1995: 491ff), the idea of autonomy can be unfolded in two related issues that are at the heart of the divide between formalism and functionalism: the concepts of self-containedness and arbitrariness. Arbitrariness implies that the rules governing the structure of linguistic expressions are not determined by the rules governing the use of those expressions. Self-containedness maintains that rules of structure are organized in a closed system that cannot be affected from outside, i.e. from external factors involved in language performance.

In a functional perspective, the system of language rules is determined by the use to which linguistic expressions are put, and to this extent it is modeled from outside. As a result, the causes of language change are less internal than external to language. Any theory of grammar that boasts a truly *functional* approach should therefore prioritize pragmatics, semantics and discourse as the interface between language usage and language structure and submit syntactic, morphological and phonological levels of language organization to the scope of the former. This was the main goal of FG from its inception and continues to be the motivation behind its latest developments.¹

¹ The new model of FG is called Functional Discourse Grammar and has been under development in the last lustrum. See Hengeveld (2004) and Mackenzie (2004) for the latest overviews of this model.

Complementary of the previous discussion are the features of the functional paradigm outlined by Dik (1997). This author sketches the main characteristics of the functional paradigm “by answering a number of questions concerning the nature and functioning of natural languages” (Dik 1997:4). Some of these questions have been already addressed and will not be mentioned here except for those concerning the cognitive dimension, language acquisition, and language universals.

As Butler (2004: 37ff) shows in his comparative study of functional linguistic theories, FG is deeply concerned with a cognitive dimension. This concern has two implications: first, the seat of the system of rules governing structure *and* use is the mind of the Natural Language User (NLU); second, this system of rules makes up the “communicative competence” that enables speakers to use language for communicative purposes in a felicitous way (Hymes 1972). Dik insists that “competence” does not refer only to rules of structure but also to rules of use. While the position of the cognitive dimension is outstanding in the outline of FG, most of the paths leading to a cognitive understanding of language organization and function remain unexplored. One of these paths concerns the processing of the lexicon in the bilingual mind. Lexical processing will be explored here in relation to the influence of bilingual performance on linguistic borrowing.

Insufficiently explored in FG is also the acquisition of language by the child, though recent attempts in this direction have been made on the basis of linguistic typology (cf. Boland 2006). The conception of language acquisition in a FG perspective is basically constructionist (Butler 2004: 40). It focuses on the idea that language acquisition “develops in communicative interaction between the maturing child and its environment” (Dik 1997: 7). While the role of the environment is decisive, Dik insists that FG does not exclude genetic factors in language acquisition but downplays their role in the process. In this aspect he clearly parts company with formalist views.

Another major issue addressed by Dik is the explanation of language universals from a functional point of view. Dik makes it clear that the existence and nature of language universals is satisfactorily explained only if commonalities across languages stem from (i) the biological and cognitive blueprint shared by language users, and (ii) the common purpose of communication in social interaction. In Dik’s words, “one should like to be able to understand the pervasive common properties of languages in terms of the external factors which determine their nature” (1997: 7). This view is coherent with the idea that the ultimate causes of language change are found in cognitive and social factors.

In sum, the functional paradigm allows for an appropriate and comprehensive account of language contact phenomena in so far it defines language change within the limits of a system (with regularities and patterns based on structure) while characterizing socio-communicative factors as the primary causes of change.

3.1.2 A functional view of language contact

In this section I develop my approach to language contact from a functional perspective on the basis of three assumptions. The first assumption of this study is that *the communicative motivation that leads speakers to take part in verbal interaction within a speech community is also operative when speakers of two or more languages are involved in social behavior, regardless of the relative position of the languages with respect to each other*. In this view, the ultimate reason for contact between two or more language communities is communicative in the broadest sense of the word. In addition, the study of language contact and its varied output has far-reaching import for a theory of human communication and for models of verbal interaction like the one sketched for FG (Dik 1997: 8ff).

The second assumption is relevant for understanding language contact as a discourse-driven phenomenon: as a system of rules for structure and use, language is *not* autonomous, self-contained or monolithic. Language is considered an open system (Berthalanffy 1968) defined in these terms: *an organized array of elements (lexical, morphological, syntactic, etc) in dynamic interaction with each other and with the environment (other languages and speakers) for the purpose of human communication*. Defining features of language are: (i) non-additivity, the whole of language is more than the aggregation of parts; (ii) goal-orientedness, language is always oriented to the accomplishment of a communicative goal; (iii) equifinality, any state is determined not only by the initial conditions but above all by the nature of the process, so that identical states may grow out of different conditions. This characterization of language becomes evident in contact phenomena insofar as no adequate explanation of them can be based (i) on purely linguistic analysis, (ii) without considering the foremost communicative goal of language, or (iii) considering the typological outline of the languages in contact apart from the role of cognition and bilingualism.

The third assumption is that the agents of language contact are *speakers* and not languages per se. Every linguistic choice is the product of a psycholinguistic process which should not be oversimplified. When looking for typological constraints on borrowability, it is therefore assumed that the ultimate decision is up to the speaker. This statement implies that structural rules governing the borrowing process may be superseded by psychological and sociological factors. This leads to consider bilingualism another factor in the definition of borrowability. Speakers with different degrees of bilingualism show different amounts of borrowing. The borrowing of linguistic forms is not always a matter of consensus, and most forms enter the language through their being borrowed by few speakers or even just one. Monolingual speakers often learn non-native lexicon by imitating bilinguals for reasons of linguistic fashion. In such cases the role played by bilinguals in the borrowing process is crucial, since they become speech models for their

monolingual peers. In general, monolinguals and bilinguals access (borrowed) lexicon and grammar in different ways. Accordingly, it should be possible to test the differential access and use of borrowings by bilinguals and monolinguals. The fact that the use of foreign elements depends on the speaker's level of bilingualism explains why the origin of lexemes is perceived differently according to their degree of assimilation into lexicon, and why two individuals may not agree on the origin of the same lexeme. Of course, this situation does not exclude that reasonably bilingual speakers with intuitions about lexical classes in language A borrow lexemes from language B provided they fit the parts-of-speech system of A (cf. section 3.3).

3.1.3 Standards of adequacy and language contact

To the extent that a theory of language provides the parameters for the description of human languages, it must follow a number of standards of adequacy that make the baseline for any satisfactory account of language phenomena. Standards of adequacy as identified first by Chomsky (1965) include observational, descriptive and explanatory. There is general consensus about these standards nowadays, although not everybody agrees on their relative importance. Different approaches prioritize different standards of adequacy. Thus, observational adequacy is often downplayed by formalist theories, which consider explanatory adequacy as the ultimate goal of linguistic theory. For empirically oriented theories, however, observational adequacy is a benchmark insofar as the first step for descriptively and explanatorily adequate analyses is a comprehensive account of data. Different approaches entail also different views of what is descriptively adequate. Chomsky maintains that a theory of language is "descriptively adequate to the extent that it correctly describes the intrinsic competence of the idealized native speaker [and] makes a descriptively adequate grammar available for each natural language" (1965: 24). This is only partially true for FG, because this theory does not assume such a thing as an idealized native speaker and views competence exclusively as a communicative capacity.

In a functional perspective the three standards are related to each other, although not in the way assumed by Chomsky, i.e. with explanatory adequacy as a self-sufficient goal. For one thing, no linguistic theory that downplays observational and descriptive adequacies can boast sound foundations. Both adequacies imply the need of a model to rely on linguistic facts and to be empirically based. This shows that the relation between the standards is hierarchical in the sense that explanatory adequacy can be accomplished only if descriptive adequacy is previously attained, and the latter can be attained only if observational adequacy has been met before.

In addition to the aforementioned standards Dik proposed three standards according to which any linguistic theory is (i) *pragmatically* adequate when it observes the rules and principles of verbal interaction, (ii) *psychologically* adequate

when it is attuned to the psychological models of production and comprehension of linguistic expressions, and (iii) *typologically* adequate when it is capable of providing a grammatical description of any human language (Dik 1997: 13ff). These standards are related to each other insofar as typological adequacy is a pre-requisite for psychologically and pragmatically adequate theories.

While FG studies make use of data in diachronic perspective, none of them mentions a *diachronic* requirement. To include diachronic factors as part of any theory of language will account for the wide array of phenomena concerning language change and language contact (Bakker 1998: 1). In a similar way, acquisitional and areal criteria are required to account for phenomena of areal distribution and language acquisition. Acquisitional factors influence the patterns of language development and linguistic variation (cf. Boland 2006). Areal factors play an important role in the development of linguistic areas, in which the distribution of language features is determined by the sharing of geographical and sociocultural spaces by speakers of typologically different languages that become increasingly similar to each other as a result of contact.

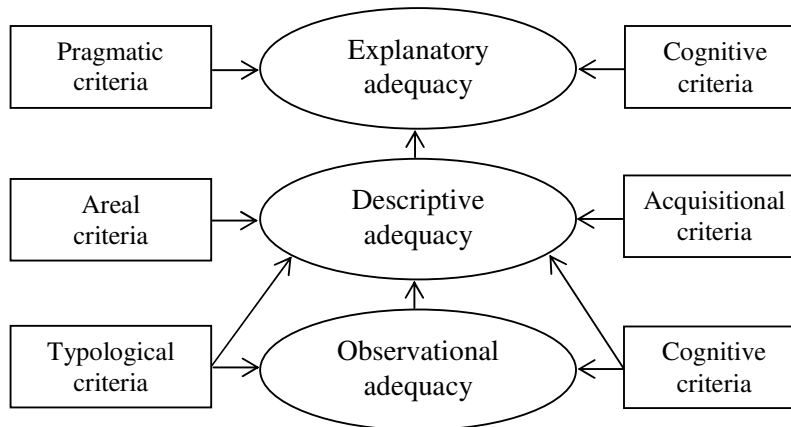
The pragmatic, psychological, typological, diachronic, areal and acquisitional specifications underlying the three standards of adequacy are criteria for the application of the general standards. Figure 1 sketches how I view these standards and their relationship. Thus, explanatory adequacy is the ultimate goal provided it relies on descriptive and observational adequacies and takes the six relevant criteria into consideration.²

The data analyzed in the following chapters provide a solid ground for testing the theory from different angles and attaining the necessary standards of adequacy. The analysis of lexical and grammatical borrowing will clarify the relation between lexicon and grammar and the influence of lexical structure on lexical accessibility (psychological adequacy). Linguistic borrowing will also shed light on the socio-communicative motivations that encourage natural-language users to incorporate foreign elements into their languages, and the pragmatic motivations of linguistic choice in bilingual speech (pragmatic adequacy). Furthermore, the study of borrowing will enhance the scope of analysis of several theories of language to the extent that borrowing includes contact phenomena often characterized as marginal, performative elements (typological adequacy). This is visible not only in borderline cases where massive borrowing results in restructured varieties which cannot be ascribed to either of the two parent languages (see the concepts of non-additivity and equifinality in section 3.1.2 above) but also in less dramatic scenarios of language

² Another useful principle that may be considered a rule of thumb in FG is that any comprehensive explanation of language facts “should strive for the lowest level of abstractness which is still compatible with the goal of typological adequacy” (Dik 1997: 16), and therefore it must be neither too concrete nor too abstract. This principle relates the typological requirement to the main standards of adequacy.

variation such as codeswitching in bilinguals or dialect formation in speech communities. Finally, the study of borrowing will trace the paths of language change and help the theory account for the development of the languages analyzed here (diachronic adequacy).

Figure 3.1. Standards of adequacy in a functional perspective



An adequate account of language contact phenomena is of interest for the development of any linguistic theory. Moravcsik (1978) develops the rationale for the inclusion of language contact phenomena in a theory of language by taking as a point of departure the notion of ‘actual linguistic utterance’ and the constraints on language interaction:

“[G]iven that the basic assumption that linguistic theory is to explain all the logically non-necessary facts about how human beings communicate in terms of orally articulated sounds, it follows that linguistic theory has to be able to characterize the concept of “actual human linguistic utterance” within the class of logically possible sets of human linguistic utterances such as those constituting a language, a dialect, and some particular style, and to impose constraints on various subsets of human languages correlated with the temporal and interactional relations of their speakers” (Moravcsik 1978:98)

The integration of contact phenomena and bilingualism into FG may help the model fulfill the standards of adequacy in a more comprehensive way. However, since any functional approach to language cannot be explanatorily adequate unless it also

sheds light on the motivations for language change, I explore now the ways in which these motivations interact with each other in a functional perspective.

3.1.4 Motivations of contact-induced language change

Inasmuch as linguistic expressions are ultimately determined by the uses to which they are put in verbal interaction, it is natural that the first cause of contact-induced language change is external to language itself. In similar terms, the fact that human language is an instrument of communication and exists only in relation to the accomplishment of this functionality implies that language change is a mechanism of adaptation to the communicative needs of language users. By ‘communicative needs’ I refer not only to processing needs such that linguistic expressions must be structured in a way that maximizes their effective parsing and facilitates the information flow. I also refer to social and cultural conditions, which become particularly relevant in multilingual situations like the ones prevailing in the speech communities of the languages analyzed here.

Language change follows naturally from verbal interaction. However, language change is not a random process, as language itself is *not* a set of unconnected elements but an organized array (see Section 3.1.2). Language is an open system with structure. It differs from a closed system in its interaction with the environment and the disturbing effects it suffers as a result of such interaction. It is precisely to these disturbances that natural language users react by adapting their language *along the parameters* set by linguistic structures. The present study seeks to identify these parameters for the case of languages that borrow elements from other languages. Linguistic borrowing illustrates the adaptive strategy of assimilation of foreign elements to the morphosyntactic matrix of the receiving language.

Nevertheless, any solutions to communicative pressures are necessarily provisional, as the environment constantly disturbs the system and language users look for the best adaptive alternatives within the limits of the linguistic system to preserve its stability. While this process is largely subconscious because speakers normally do not monitor their speech at the level of linguistic structure, there is also deliberate manipulation of language material as a mechanism of contact-induced language change (Thomason 2001, 149). Deliberate decision may be at work not only in garden-variety lexical borrowing but also in cases of relexification resulting in the emergence of mixed languages (Muysken 1997; Gómez Rendón 2005, 2008b). Counter to common assumptions deliberate change also influences grammatical borrowing provided levels of bilingualism are high (cf. Golovko 2003).

There is an ongoing tension between the requirements from verbal interaction (e.g. multilingualism) and the requirements from structure (e.g. effective formulation), and between external and internal factors. Communicative needs meet at the crossroads of internal and external motivations, and these concur in exerting

pressure on the language system. The outcome is that “language at any particular time is the result of competing motivations” (Butler 2004: 14; cf. Dubois 1985: 343ff). In language contact terms it is better to speak of internal and external factors simultaneously influencing the makeup of the languages in contact. For Dik the only possible, though provisional solution to competing motivations is a compromise:

There is thus continuous competition between different functional prerequisites; the actual synchronic design of a language is a compromise solution, a precarious balance in efficacy with respect to different functional prerequisites. (Dik 1986: 21f)

This “precarious balance” results from the compromise between the homeostatic tendency of language structure to remain unchanged and the transformative force of adaptation of the system to socio-communicative needs. Borrowing and imperfect second-language learning are two cases in point. On the one hand, there is no unnecessary borrowing, in the sense that “borrowing takes place with the borrowing individuals having some purpose in mind” (Moravcsik 1978: 102), and borrowings are assimilated to the matrix of the recipient language in order to minimize the chances of structural disruption (typological shift). On the other hand, the imperfect learning of a second language (L2) usually leads to the transfer of native-language (L1) features to the target language. In this case the structure of L1 resists to the changes effected by the structure of L2. Still, the tendency of language structure to remain unchanged not always holds on. Provided that socio-communicative needs have disrupting effects on the system (i.e. relexification or massive borrowing) and these cannot be prevented, there may be some type of restructuring. An illustration of this is the so-called matrix language turnover (Myers-Scotton 1998) in which the morphosyntax of the source language replaces that of the borrowing language and gives rise to mixed varieties such as Mednyj Aleut or Ma’a. While these are dramatic scenarios, the effects of contact can vary from less to more disrupting depending on the contact situation.

It is often claimed by linguists from the formalist school that the argument of competing motivations is non-falsifiable. According to such claim, the resort to functional factors in the explanation of language facts is plagued with vacuity because one of an endless number of functional motivations may always be invoked (Newmeyer 1998: 150). Furthermore, functional explanations would often produce circular arguments involving change and adaptation where there is no possible way to know which one caused the other (Butler 2004: 22). While the risks behind a deliberate use of functional factors in the explanation of language change are many, I consider it the duty of everyone who resorts to functional arguments to chart all the external factors that are considered operative in language change in a hierarchical arrangement that shows their degree of influence on such change and their interplay

either as co-determining instances or members of cause-effect chains. Complementarily, sociolinguistic factors should be related to cognitive factors in meaningful ways - in the case of borrowing through linking sociocultural factors to individual bilingualism, i.e. determining how the cognitive makeup of bilinguals is modeled by the social uses of language. An accurate weighing of external factors and their integration within a coherent frame is the only manner of not ‘explaining away’ language change from a functional perspective. The mapping of external factors for the specific sociolinguistic situations of the languages scrutinized in the following chapters is given in Chapter 2 and will be substantiated in the discussion of sociolinguistic factors in Chapter 4.

3.2 The theory of Parts of Speech

Part of the theoretical framework underlying this research is the theory of parts of speech as elaborated by Hengeveld (1992) and Hengeveld *et al* (2004). This theory provides a basis for the analysis of lexical borrowing. There are three basic differences from other theories of parts of speech. One is that Hengeveld’s theory is concerned only with what is usually termed “major parts of speech”, including nouns, verbs, adjectives and manner adverbs. Another difference is that Hengeveld defines parts of speech primarily on *syntactic* grounds. The basic syntactic unit is the phrase, which can be either referential (noun phrase) or predication (verb phrase). Two main slots are identified within each phrase, one for heads and one for modifiers. The possible combinations for English are shown below.

Table 3.1. Lexemes and syntactic slots in English

	Head	Modifier
Referential Phrase	Noun	Adjective
Predicate Phrase	Verb	(Manner) Adverb

Note that the predicate phrase modifier is the manner adverb rather than the adverb in broad terms. The reason is that only manner adverbs modify the head of predicate phrases while other adverbs modify the sentence as a whole (Hengeveld *et al* 2004: 6). The third difference in Hengeveld’s approach is that parts of speech are defined according to their non-predicative uses. Accordingly, “verbs are characterized by the fact that they have no non-predicative uses, i.e. they can be used predicatively *only*. Non-verbal lexemes, on the other hand, may have additional predicative uses, but their defining use is a non-predicative one” (Hengeveld *et al* 2004: 6; my emphasis). After testing the occurrence of lexical classes in syntactic slots in a sample of typologically and geographically distant languages, it was found that some languages have one lexical class for one syntactic slot (differentiated systems);

others have more than one lexical class for two or more slots (flexible systems); and still others lack lexical items to fill syntactic slots (rigid systems). English is a typical example of a differentiated system, with a separate lexical class for each syntactic slot, as opposed to languages like Samoan in which any lexeme can be used in any syntactic position without any derivation mechanism involved. Parts-of-speech systems from the most flexible to the most rigid are charted in Table 2.

Table 3.2. Parts of speech systems

<i>Part of Speech System</i>			<i>Head Pred. Phrase</i>	<i>Head Ref. Phrase</i>	<i>Modifier Ref. Phrase</i>	<i>Modifier Pred. Phrase</i>
Flexible	1	Samoan, Tagalog	Contentive			
	2	Quechua, Guaraní	Verb	Non-verb		
	3	Ket, Miao, Tidore	Verb	Noun	Modifier	
Differentiated	4	Basque, Japanese	Verb	Noun	Adjective	Manner Adv.
Rigid	5	Kisi, Wambon	Verb	Noun	Adjective	-
	6	Krong, Navaho	Verb	Noun	-	-
	7	non-attested	Verb	-	-	-

The table does not include intermediate systems, i.e. languages that do not fit into one basic type but share features of two types, provided these are adjacent in the classification. For flexible languages, this situation occurs when derived lexemes cannot be used in all the syntactic slots in which their base lexemes are used. This is the case of Lango (Western Nilotic), where a large class of adverbs (type 4) co-occurs with an open class of modifiers used in predicate and referential phrases (type 3). In the case of rigid languages, an intermediate system is attested when “the last class of lexemes on the hierarchy that is relevant for that language [is] a small closed class of items” (Hengeveld et al 2004: 25). An example of an intermediate rigid language is Tamil (Southern Dravidian). This language has no lexical modifiers of predicate phrases (type 6) and only a small class of adjectives (type 5). Due to restrictions on directionality only contiguous systems conflate to form intermediate types. In relation to system of parts of speech, typologically possible languages are therefore limited to seven discrete types (1-7) plus six intermediate types (1/2, 2/3, etc.). While combinatorially possible, other intermediate types (e.g. 1/7, 2/5, etc.) are not typologically viable. The classification of the languages of this study according

to their parts of speech in chapter 5 through 8 will make the above clear. According to Hengeveld *et al* (2004) the only parts-of-speech system that has not been attested is type 7, one in which the noun-verb distinction is absent, with verbs as the only lexical class. Iroquoian languages, particularly Tuscarora, have been considered examples of this type (Sasse 1988; Broschart 1991) but recently it has been demonstrated on a solid basis that these languages *do* make a distinction between nouns and verbs (Mithun 2001: 397ff). The existence of flexible languages in their most extreme form (type 1) has been questioned as well. A recent debate around the existence of lexical classes in Mundari (Austro-Asiatic) shows that different views on the issue of flexibility stem from different interpretations of semantic and morphological phenomena (i.e. polysemy vs. vagueness, zero-conversion vs. systematic flexibility) but also from whether flexibility is seen as a gradual phenomenon not necessarily involving the whole lexicon of a language.³

In general, parts of speech show a left-to-right hierarchy and systems are ordered according to this hierarchy. This means, for instance, that a language with a specific lexical class for predicate phrase modifiers *always* has individual lexical classes for the syntactic slots located to the left of this slot (system 4). No languages are attested that have a specific lexical class for predicate phrase modifiers and none for referential phrase modifiers. Hengeveld (1992) shows that the combination of syntactic positions with lexical classes is not random: it can be captured in an implicational hierarchy.

Table 3.3. The parts-of-speech hierarchy

Head of	Head of	Modifier of	Modifier of
Predicate	> Referential	> Referential	> Predicate
Phrase	Phrase	Phrases	Phrase

One feature of the hierarchy in need of explanation is directionality. The hierarchy's left-right order not only implies a *sequence* but also specifies a *direction* (expressed by >). In other words, the hierarchy shows not only an order of elements (x, y, z) but also the specific direction of this order (x>y>z). In practical terms this means that directionality determines the path of the processes of lexicalization and grammaticalization. It predicts, for example, that if a rigid language becomes more differentiated by replacing morphosyntactic strategies with a new lexical class, the latter will follow the last lexical class attested in that language. To be specific, if a type-6 language without adjectives and adverbs like Hixkaryana (Carib) created a new lexical class in its system of parts of speech, it would be the adjective and not the adverb, since the latter is not the lexical class immediately following the last attested lexical class in the language (i.e. nouns). Directionality is also relevant in

³ Cf. *Linguistic Typology*, Vol 9, Issue 3, December 2005.

explaining the case of bilingual speakers of typologically different languages who borrow items from one language to the other. In these cases directionality may be a constraint to the flexible use of borrowings. It might explain why, for instance, speakers of Quechua (type 2) do not use adverbs of Spanish (type 4) in adjective or noun positions even though this is permitted by the existence of a non-specialized lexical class of non-verbs.

On the one hand, positive evaluations of the model focus on its potential to account for language variation in a straightforward way “as the outcome of a process of successive syntagmatic and paradigmatic expansion” (Anward 2000, 8). On the other hand, most of the critiques⁴ deal with a) its restriction to only four major parts of speech; b) its ignoring of conventional lexical semantic differences and small syntactic categories; and c) the methodology behind the classification. Ongoing contributions from such fields as language contact and language acquisition aim at testing the model on an empirical basis. This will be a decisive step to determine its validity for capturing homogeneities and heterogeneities in the systems of parts of speech of languages around the world.

3.3 Implications of the Parts-of-Speech Theory to Lexical Borrowing

When two languages come into contact, linguistic material is exchanged between the source and the recipient. In the present study Spanish is the source language while the recipients are Guaraní, Quechua and Otomí (cf. Chapter 5). The general hypothesis to be tested in this study is that *the parts-of-speech systems of the languages involved in the borrowing process are relevant to determining the type of borrowed lexical classes and the functions to which they are put*. More specifically, the parts-of-speech system of the recipient language co-determines the borrowing of lexical classes from the source language and their functional adaptation in the recipient language. The implications from Hengeveld’s theory to lexical borrowing include a set of subhypotheses to be tested on data from either of two standpoints: the perspective of the source language, with emphasis on the identification of lexical classes and their frequency; and the perspective of the recipient language, with emphasis on the use and function of borrowed lexemes. Individual subhypotheses concerning the possible language contact situations are reviewed in Chapter 4; the implications for each of the languages of the sample are developed in Chapters 6 through 8.

⁴ See Petra Vogel and Bernard Comrie (eds.) *Approaches to the Typology of Word Classes*, for a critical evaluation of the model from different viewpoints.

3.4. Implicational hierarchies and scales of borrowability

The use of implicational hierarchies to capture parametric variation across languages became a common practice in typology ever since Greenberg (1966). In a similar fashion, efforts to capture parameters that model ‘borrowing’ and ‘transfer’ between languages have resulted in a number of scales of borrowability since the late nineteenth-century (cf. Whitney 1881). A critical discussion of scales of borrowability is presented in section 3.5. Scales of borrowability not only imply patterns in the preferences of borrowing and indicate *quantitative* tendencies; they also indicate specific paths of change and outline the *qualitative* nature of the borrowing process. It has been suggested that these interpretations are not contradictory and may be applied separately (cf. Van Hout and Muysken 1994: 41). The purpose of this section is to elucidate the links between implicational hierarchies and scales of borrowability. This task is relevant inasmuch as this study seeks to outline testable scales of borrowing from the hierarchy of parts of speech discussed in the last section.

Implicational hierarchies and scales of borrowability have many things in common, but many differences too. Both are the offspring of propositional logic and depend on formal deductive mechanisms (cf. Croft 1990: 49). Both seek to describe ranges of possibility as much as predict impossibilities in languages. Both indicate some type of asymmetry or unidirectionality. And both originate in the broader concept of markedness. But what is the relative status of implicational hierarchies and scales of borrowability in the context of a linguistic theory that aims to fulfill the standards of descriptive and explanatory adequacy? Van Hout and Muysken (1994: 41) pose several questions about the relationship between both concepts but leave them unanswered because their interest is quantitative rather than typological or historical. Because the present study is framed both in typology (section 3.2.) and language change (section 3.1.3.) it is necessary to explore the status and the relationship of these constructs.

Formally, typological hierarchies and scales of borrowability represent chains of implications arranged in consecutive order so that one element presupposes the others located before on the chain. Also, hierarchies and scales are different from each other in the following aspects: a) implicational hierarchies have originated from the observation of a relatively large number of languages while scales of borrowability have been proposed mostly on a language-pair basis, often supported by impressionistic rather than statistic evaluations; b) hierarchies boast a wider scope and applicability than scales because they conflate several implicational universals; c) hierarchies typically refer to different grammatical categories and processes (e.g. accessibility, definiteness, etc.) whereas scales of borrowability refer mainly to parts of speech and no proposals have been made so far to account for other grammatical parameters; finally, d) hierarchies have been given a dynamic

interpretation in order to account for language change from the perspective of diachronic typology (Greenberg 1978) while scales have been associated with synchronic stages of the languages in contact, which is due perhaps to the lack of historical linguistic records that enable the linguist to trace tendencies over time.⁵

To illustrate the point about the empirical basis of scales of borrowability, a few authors may be taken as examples. Often quoted as the first student of language contact who proposed a scale of borrowability, Whitney (1881) gives an impressionistic evaluation of borrowings from (non-specified) languages into English in support of his scale, saying “it has been comparatively easy to add adjectives and verbs to nouns because of the direct convertibility of our nouns into adjectives (a *gold* watch, a *leather* medal, etc) and of our nouns and adjectives into verbs (to *tree* a raccoon, to *grass* a plot of ground, to *brown* a complexion, to *lower* a price) without any change of form” (Whitney 1881: 20, his emphasis).⁶ However scanty and biased his evidence seems today, it is not radically different from the evidence presented seventy years later by Haugen in support of his own claims. Haugen presents three sets of statistical data for English borrowings in Swedish and Norwegian but provides no information about the way his corpora were gathered or the criteria considered in the analysis of loanwords. And yet, the data seem to be enough for him to state confidently that “all linguistic features can be borrowed, but they are distributed along a scale of adoptability [Haugen’s term for borrowability] which somehow is correlated to the structural organization” (Haugen 1950: 224). Thirty years later Singh (1981: 113f) provides a similar scale of borrowability for English loanwords in Hindi, without any satisfactory substantiation or systematic analysis. Few years later Thomason and Kaufmann (1988) put forward a scale of borrowability in five stages and their proposal became soon widely accepted as a general reference for the study of languages in contact. The same scale appeared recently in Thomason (2001: 70), with a few minor changes. Although these authors link borrowing levels to specific contact scenarios in a coherent way, the data they provide in support of their proposals are more anecdotic than meticulous, and their approach is less empirically founded than assumed. I further comment on Thomason and Kaufmann’s scale in the next section, in particular on Thomason’s statement that typological parameters do not govern contact-induced language change.

The foregoing discussion implies that scales of borrowability could hardly apply to languages other than those used for their formulation. The fact that too often English (or some other Indo-European language) is the language considered as

⁵ A notable exception is Karttunen (1978) for the case of Nahuatl in contact with Spanish. Indeed, the abundant Nahuatl record from the first years of the Spanish conquest to the present makes this contact situation rather exceptional and without parallel in other areas where a colonial language has coexisted for several centuries with a native language.

⁶ By “add adjectives and verbs to nouns”, Whitney means those cases in English in which zero conversion occurs such that no derivational mechanisms are at work.

source or recipient also limits the range of applicability of scales in a significant way.

The limited scope of scales of borrowability as proposed in the literature takes us to the next issue: the implicational nature of hierarchies in comparison to the seemingly contingent nature of scales as evinced by the various scales proposed. Many scales grow out of case studies of specific language pairs for which individual corpora have been analyzed. However, the formulation has often followed an inductive method without a theory-driven approach that enables a fine-grained analysis of data. Few of those who propose a scale of borrowability on empirical grounds make all the steps clear in their analyses. Others take it for granted that parts of speech are synonymous with lexical classes as defined by traditional grammarians for European languages. Only a theory-driven approach to the study of borrowing combined with empirical data shall provide students of contact with predictive devices and explanations about why scales are the way they are. We should not forget Moravcsk's warning in this respect:

“Even though constraints on borrowing, as we have just seen, can serve to explain observations about similarities and differences within and among languages, such constraints themselves are also in need of explanation” (Moravcsk 1978: 118)

This means that scales of borrowability and implicational hierarchies are themselves further explananda. That is, they represent only part of the explanation. Notice that for any explanation of language contact phenomena to be satisfactory, it must incorporate language-internal and language-external factors in a coherent theory of borrowing (Chapter 2).

From the above it becomes clear that a major difference between implicational hierarchies and scales of borrowability is the limited scope of scale not in terms of languages but also of grammatical categories. Most scales of borrowability include parts of speech only. Exceptional are those scales (called ‘hierarchies’ by some of their proponents) which cover lexical and grammatical borrowing alike (Cf. van Hout and Muysken 1994; Field 2002). On the other hand, the number of scales proposed for lexical borrowing exceeds by far the number of scales of grammatical borrowing. The preference for the former type stems from the traditional view that the grammatical apparatus of a language can hardly be affected by contact with other languages, no matter how long and deep this may be. Arguably, the limited scope of scales of borrowability is determined by the nature of borrowing itself, i.e. the nature of the linguistic material exchanged in most contact situations. This interpretation, however, bars the way to what is perhaps the most interesting aspect of the study of contact outcomes: the use made of loanwords in the recipient language. This side of borrowing has been hardly explored in contact linguistics in spite of its potential

contribution to a theory of language (e.g. how languages adapt their structures to changing environments) and language change (e.g. to what extent lexical and grammatical borrowing lead to typological shift or restructuring). As I show in Chapter 4, any theory of borrowing must include not only a model of borrowability (the conditions and the limits of what is borrowable or not) but also a model of borrowing usage (the ways in which languages adapt alien material).

Not less important is the difference between hierarchies and scales as far as diachrony is concerned. A Greenbergian dynamicization of typology (cf. Greenberg 1978) uses the tools for describing synchronic variation (universals and hierarchies) in the description of language stages and clears the way for a wider diachronic perspective. Despite language contact studies are closely related to historical linguistics, studies on borrowing have been largely focused on synchronic description and assumed that borrowing implies language change without any specification of this change. The cause of such blindness to diachronic description is implicit in the notion of 'borrowing'. Traditionally, borrowing has been understood as a product rather than a process. There is a warning against this bias already in Haugen (1950: 213f). Still, his long list of outcomes of borrowing shows his preference for the description of products over processes. Whether this preference is due to terminology or the lack of historical linguistic records available is unclear. What is fairly clear is that borrowing can be described both synchronically (products) and diachronically (processes) and that both sides are complementary to a large extent. In fact, language variation mirrors language change in that the parameters governing today's languages are equivalent to the limits of language change over time. Boland claims in this respect that "implicational hierarchies or markedness scales established for describing adult language variation are thus hypothesized to be reliable predictors of universals in language acquisition" (Boland 2006: 16). As it seems, the diachronic study of borrowing will contribute to unveil the relation between patterns in language variation and patterns in language change, and explain the nature of language universals. This study can profit from other fields of linguistics like stratigraphic analysis and grammaticalization theory. Anyway, the question remains whether scales of borrowability allow for a diachronic reading or not. In other words, it still is unclear whether there is a stepwise process of borrowing, according to which one part of speech is borrowed before others or, more radically, some parts of speech cannot be borrowed unless others have been before.

For typological hierarchies the assumption is that a diachronic interpretation is not only possible (cf. Greenberg 1978) but also feasible (cf. Heine 1991). The hypotheses underlying the diachronic interpretation of hierarchies hold that a) the nature of languages remains the same across time and the languages spoken in the

past are similar in nature to the languages spoken in the present⁷; b) patterns of variation across languages in the present mirror patterns of language change in the past, in a somewhat modified version of the idea that ontology recapitulates phylogeny. To the extent that a scale of borrowability may be deduced from a typological hierarchy, there is nothing that prevents the scale from being interpreted along diachronic lines, especially if borrowing is considered a gradual process of incorporation of alien material. This position was adopted, among others, by Moravcsik (1978), who assumes an equivalence between hierarchies and scales, on the one hand, and synchronic and diachronic interpretations, on the other⁸.

Some studies in language contact report instances of abrupt change and restructuring in which a language changes substantially within the time span of one or two generations as a result of massive borrowing and without consideration of lexical classes (cf. Muysken 1985; Gómez Rendón 2005). Despite these cases seem to run counter to a stepwise interpretation, they are not essentially different from other scenarios and represent one of the ends of the scale. Thus, a provisional statement would be that all things being equal, lexical borrowing proceeds by steps and may eventually lead to grammatical restructuring (Karttunen 1976; Campbell 1987; Fauchois 1988). Of course, things are not always equal and a large number of non-linguistic factors may intervene to determine the course of the borrowing process. Unfortunately the lack of systematic studies in the field of diachronic contact linguistics prevents us from making any decisive statement.

The four differences discussed above between typological hierarchies and scales of borrowability may be translated as deficiencies in the following terms: a) scales of borrowability have a limited applicability, derived as they are from language-pair studies; b) scales of borrowability are of limited scope in that their formulation, though empirically based, is not theoretically driven; c) scales of borrowability proposed so far have been applied mostly to the lexicon and only exceptionally to grammar, without any consideration of the use of borrowings in the recipient language; and d) scales of borrowability have been interpreted synchronically although their potential for a diachronic analysis is great.

In view of these deficiencies the present study seeks to model borrowability through: 1) a *comparative analysis* of borrowings across language pairs; 2) a *theory-driven approach* to borrowing in the framework of the parts-of-speech theory; 3) a comprehensive account of *lexical and grammatical borrowing* and their usage in the recipient language; and 4) a theoretical framework for language change through borrowing. The inclusion of non-linguistic factors such as the duration and type of

⁷ The hypothesis of uniformitarianism, which according to Croft derives from biology and geology (1990: 204, 274)

⁸ See, for instance, how she interprets her fifth statement. “No inflectional affixes can belong to the set of properties borrowed from a language *unless* at least one derivational affix also belongs to the set” (Moravcsik 1978: 112; my emphasis).

contact and the levels of individual and collective bilingualism complements the theory and allow for a multi-sided evaluation of data. As part of this research program the following section explores critically several scales of borrowability as a backdrop for the presentation of my model in Chapter 4.

3.5. Scales of borrowability: a critical overview

A number of proposals have been put forward in the literature on language contact to account for the occurrence of borrowings across languages. The study of linguistic borrowing and scales of borrowability grew out of the discussion about the existence and the status of mixed and Creole languages in the second half of the nineteenth century (Cf. Whitney 1881; Schuchardt 1882). Some hierarchies or scales of borrowability⁹ are established on the basis of a quantitative analysis of language-pair corpora; others are deduced from theoretical frameworks and claimed to be applicable cross-linguistically.¹⁰ The former hierarchies lack a comprehensive framework for the analysis and interpretation of data; the latter hierarchies need an empirical foundation that corroborates their claims in a relevant way.

As a matter of fact, most scales of borrowability include major parts of speech and function words. Few studies on borrowing analyze other grammatical categories such as word order (Campbell 1995: 136ff) or utterance modifiers (Matras 1998: 281ff). In our perspective, scales of borrowability should include not only content words (major parts of speech) but also function words and grammatical elements such as derivational and inflectional affixes. In principle it is possible to make a distinction between lexical borrowing (content words) on the one hand, and grammatical borrowing (agglutinative and fusional affixes) on the other, with function words occupying a place of transition between lexicon and grammar. In this study function words are considered part of grammatical borrowing and analyzed independently from the four word classes identified in the parts-of-speech theory outlined in section 3.2. Still, it is clear that any distinction between lexical and

⁹ Alternative terms are ‘hierarchies of adoptability’ or ‘hierarchies of receptivity’. For Haugen, “all linguistic features can be borrowed, but they are distributed along a SCALE OF ADOPTABILITY which somehow is correlated to the structural organization” (Haugen 1950: 224). Receptivity, in turn, is defined as the “capacity of absorbing words of foreign origin” (Vočadlo 1938: 170).

¹⁰ It is possible to distinguish between hierarchies of borrowability and scales of borrowability on the basis of their theoretical or empirical origin. Hierarchies of borrowability would be hypothetical models of borrowing with a number of falsifiable predictions. Scales of borrowability would describe a specific distribution of elements (parts of speech) in a language pair which may be predicted and tested in *similar* pairs. Both are in principle falsifiable but only hierarchies might be applicable to a large number of typologically different languages. It may be possible also to link hierarchies to universals of language, as done Moravcsik (1978). I have decided to use the term ‘scale of borrowability’ in this book in order to avoid terminological confusion.

grammatical borrowing is only schematic because borrowing implies a continuum stretching from content words (easiest to borrow) to affixes (hardest to borrow).

The non-existence of clearly defined boundaries between lexical and grammatical borrowing raises the question about the nature of borrowing. Borrowing is more a continuum of forms than a process with individual stages as represented by hierarchies. The concept of 'cline' as developed in grammaticalization theory is therefore more appropriate to describe this continuum. In fact, there are clear correspondences between clines of lexicalization (Hopper and Traugott 1993: 7) and scales of lexical borrowing, especially from a diachronic perspective. Also, hierarchies of grammatical borrowing can be inferred from clines of grammaticality. The commonality of features between grammaticalization and borrowing suggests that the latter should be conceived as a continuum and their analysis calls for a unified theory of lexicon and grammar.

In the same way that grammaticalization theory helps to define borrowability in more accurate terms, linguistic typology contributes to a better understanding of the limits of borrowing. In this case, the typology of the languages involved in the borrowing process helps to define constraints on borrowable elements. Thus, the morphological typology of the recipient language is described according to criteria of synthesis and fusion in order to predict, on the basis of such criteria, what elements a language may borrow. Likewise, source and recipient languages may be classified according to their typology of parts of speech so as to predict the type of loanwords transferable from one to the other. The typological approach to the borrowing of parts of speech diverges from traditional analyses in that these assume a univocal correspondence between the parts of speech of the source language and those of the recipient language. This assumption is misleading when typologically different languages and functional adaptation of borrowings are considered. The student of borrowing who works only on the parts-of-speech system of the source language may not find any trouble in establishing quantitative scales but will certainly fail to explain the use of borrowed elements in the recipient language. A typological approach to borrowability on the basis of parts of speech provides a comprehensive framework for the analysis of data from the perspective of both source and recipient languages.

3.5.1. Lexical borrowability

The literature on language contact describes lexical borrowing as the most widespread type of linguistic transfer. Every human language may be said to have borrowed one or more words from other language(s) at some point of its history. Several reasons have been adduced for the prominence of lexical borrowing in contact situations. First, lexical borrowing accomplishes the extension of the denotational capacity of the recipient language insofar as "the classes of words most

closely involved with the culture of a language are the *content words*” (van Hout and Muysken 1994: 42; their emphasis). Second, the perceptual saliency of content words on the basis of their phonetic shape makes lexical borrowing more prominent. Third, the semantic transparency of content items makes lexical borrowing more frequent than grammatical borrowing (Field 2002: 36).

Lexical borrowing is defined as the transfer of *content* words as opposed to the transfer of *function* words and morphemes (grammatical borrowing). There is a consensus among scholars that nouns, verbs and adjectives are content words, although their distribution is not the same across languages. The classification of adverbs as content words is disputed however. If adverbs are defined as *verb* modifiers, then their class is smaller than the class of adverbs defined as broader modifiers. Adverbs defined as verb modifiers include only manner adverbs because other subclasses have a wider scope than the verb. Additionally, manner adverbs in some languages form a relatively open class different from the closed set of time and place adverbs. Only manner adverbs form open classes¹¹ as opposed to other types of adverbs which are closer to function words. The classification becomes more problematic from a cross-linguistic perspective, because certain languages lack adverbs as a separate lexical class and use other lexical classes (verbs, nouns, adjectives) or non-lexical strategies instead. This explains why some scales of borrowability consider adverbs lexical borrowings while others put them on the grammatical side.

Scales of lexical borrowability

Regardless of their theoretical or empirical foundation, all scales of borrowability agree that nouns are by far the largest class of content items that languages borrow in contact situations. Explanations for the primacy of nouns include their perceptual saliency and semantic transparency and the fact that borrowed nouns expand the language’s referential capacity. From their study of English loanwords in Canadian French, Poplack *et al* (1988: 64) conclude that one factor influencing the large presence of borrowed nouns in their corpus is their low level of structural integration in the discourse of the recipient language and their quality of being the word class that carries most of the lexical content. The openness of the noun class as compared to other parts of speech is indeed a factor but it must be assessed in relation to other lexical classes and subclasses. There are languages in which nouns are grouped in

¹¹ As shown above (3.2.), the parts-of-speech theory that makes the theoretical framework for this study restricts adverbs to the subclass of manner adverbs. The reason is that only manner adverbs modify heads of predicate phrases (verbs) while other adverbs modify larger constituents such as clauses or sentences (Hengeveld 1992: 71f). To this extent, the borrowing hypotheses derived from this theory include only manner adverbs. For other adverbs, a number of predictions can be made on the basis of traditional hierarchies of borrowability.

clearly restricted subclasses while in other adjectives and manner adverbs conflate with nouns in one large class of non-verbs. Considering that Poplack was studying typologically similar languages, her ‘openness’ assumption may be misleading if applied to other language pairs. In addition, it remains to know to what extent the distribution of borrowed nouns is determined by the distribution of native nouns in discourse. The question can be answered only on a language-specific basis.¹² Poplack shows that univocal correspondences cannot be claimed in the distribution of parts of speech between borrowings and native items:

If borrowing into the various grammatical categories mirrored monolingual tendencies, we would expect to find comparable proportions of native and borrowed forms in each part of speech. However, the predilection for borrowing nouns exceeds by more than a factor of five the frequency of this category in French [...] Thus we may confirm that nouns have a particular propensity to be borrowed, over and above their frequency of occurrence in the host language” (Poplack *et al* 1988: 63f).

This predilection supports Moravcsik’s view (1978: 111) that noun borrowing is a universal of language contact and languages can borrow further lexical material only if nouns are borrowed first. Therefore, her position not only assumes precedence in time but suggests also “the possibly related phenomenon of a language always having a larger number of borrowed nouns than the number of borrowed items in another lexeme class” (p. 111). In addition to the examples quoted by Moravcsik in support of her claim, there are others like English loans in Hindi (Singh 1981), Spanish nouns in Otomí (Hekking and Muysken 1995) and Quechua (Muysken 1981; Gómez Rendón 2006a), and English nouns in Prince Edward Island French (King 2000).

A further factor that may influence the distribution of borrowings in word classes is the type of contact between the intervening languages and their relative position in society. An important number of borrowing situations in the literature on contact involve language pairs composed either of two European languages, or one European language as the source language and one non-European language as the recipient. The question is whether the outcomes of these situations can be generalized to more “exotic” scenarios. The contact between Spanish and Quechua and Spanish and Otomí illustrate diglossic situations where speakers of one language – generally the recipient – are subject to sociopolitical domination by speakers of another language – typically the source of borrowing. One could argue that

¹² A small sample of native discourse in each of the languages of this research was analyzed for the distribution of parts of speech in order to serve as a point of reference in the evaluation of borrowing preferences (Cf. section 4.3).

situations of political dominance force speakers of the dominated group to adopt the language of the mainstream society alongside their native language (bilingualism) or simply replace the latter with the dominant language (shift). Thus, it is possible that the outcomes of contact are not the same in situations where the speech communities enjoy a sociopolitical balance. The contact situations of English and French in Canada and Prince Edward Island are illustrative of this. While the intrusion of English in the local French-speaking culture is evident, English speakers and French speakers share a common Western heritage. Interestingly, the distribution of noun borrowing is closely similar to that of contact situations involving a European language and a non-European language (cf. Bakker and Hekking 1999, for Spanish and Otomí). Can we conclude from this that the overwhelming frequency of nouns in borrowing is a universal of language contact?

To the best of my knowledge, the only reported cases in which verbs are borrowed more frequently than nouns are the Brazilian languages Tariana (Aikhenvald 2002: 224) and Hup (Epps, forthcoming). In the case of Hup the linguistic purism dominant in the borrowing speech community restricts the entrance of Tukano nouns but not the borrowing of Portuguese nouns. The reason for such preference would be that nouns are more salient than verbs because these occur in complex forms while nouns can be easily “smuggled in” (Epps, forthcoming).

Arguably, noun borrowing is less frequent in situations involving two culturally similar groups with a long history of contact because there are few objects unknown to either group. On the other hand, for two culturally different groups that scarcely had contact in the past the need to adopt items referring to new physical objects surpasses other considerations. This suggests that explanations of the distribution of noun borrowing should include diachronic and cultural factors. The extreme case of Hup verb borrowing shows that strong predictions fail if there are factors of language ideology (perceptions and attitudes) influencing the mechanisms of contact. Still, the idea that social and cultural factors determine the scope of borrowing in each contact situation does not exclude the existence of linguistic constraints on the outcomes of contact.

As regards loan verbs, their position in the scales of borrowability is not fixed. Some hierarchies consider verbs as the second largest lexical class (cf. Haugen 1950; Thomason and Kaufmann 1988). Others put them either after adjectives (Whitney 1881; Muysken 1981; Singh 1981) or consider both as coterminous (Field 2002). Moravcsik represents the most extreme position because she considers that “a lexical item whose meaning is verbal can never be included in the set of borrowed properties” (Moravcsik 1978: 111). The empirical evidence available goes counter this statement: not only are verbs borrowed in many contact situations, but their number is also relatively high. A less strict interpretation of Moravcsik suggests a different scenario: verbs are borrowable items but they are always subject to native

mechanisms of derivation. This suggests that loan verbs might be used as non-verbs. The analysis of parts of speech presented in section 3.2 points in a similar direction.

While the evidence confirms the borrowing of verbs across typologically different languages, it is still notable that verbs are borrowed with less frequency than nouns. Several explanations have been put forward to explain this. Most have to do with the fact that verbs, unlike nouns, are not purely content items but carry structural information, which would make them more difficult to borrow than nouns, since their borrowing would require a knowledge of the source language beyond the lexicon. The degree of such knowledge depends on the syntactic and morphological constraints of the source and the recipient languages: for example, in order to borrow verbs from a fusional language like Spanish, in which verb roots are mixed with (derivational and) inflectional morphology, speakers of an agglutinative language like Quechua must know the structure of the Spanish verb. An example of Media Lengua illustrates this point:

- 1) *muy* *pokito* *disayuno-ta* *da -li-k* *ka-rka*
 very few.DIM breakfast-ACC give-DAT-DUR to.be-PST.3S
 ‘he used to give us a miserable breakfast’

Imbabura Media Lengua is a mixed language composed of Quechua grammar and Spanish lexicon which is spoken in the Ecuadorian Andes. It is the result of the intense contact of Quechua speakers with the Spanish-speaking society. Media Lengua speakers are proficient in Ecuadorian Andean Spanish and a local variety of Quechua. Example (1) contains, among other things, the Spanish verb root *da-* ‘give’, which has been borrowed along with the cliticized form of the indirect object pronoun for third person *le-*. Both elements do not form a frozen unit. The root and the clitic are assigned individual functions and meanings: *da-* replaces Quechua *ku-* ‘give’; *li-* indicates dative case. For Media Lengua speakers to identify the Spanish verb root and its cliticized pronoun correctly, a nearly native command of the language is required. This is indeed the case. The above suggests that the structural properties of the source language and the level of bilingualism of borrowers are important factors shaping the outcomes of contact.

Adjectives are next on the list of lexical categories for their borrowability. Several studies have shown that adjectives are not a monolithic, undifferentiated category (Dixon 1982; Schachter 1985, Bhat 1994). In some languages they are classified in the same category of nouns while in others they behave like verbs. These facts have challenged the universality of the adjective category from a typological point of view. Conservative views (Croft 1991; Bhat 1994) consider adjectives as prototypical modifiers. Others (Baker 2003) define adjectives less in terms of their prototypical nature than in opposition to nouns and verbs. Whatever

the case may be, adjectives are a problematic category in terms of their cross-linguistic variation and the related implications for a theory of parts of speech.

Students of language contact do not agree either on the position of adjectives along a scale of borrowability. They agree on placing adjectives immediately next to verbs but not on their relative position. Many claim that adjectives are more borrowable than verbs and put them before them on hierarchies (cf. Whitney 1881; Muysken 1981; Singh 1981; Field 2002). Others (Haugen 1950) invert the order and state that adjectives are only a peripheral category. The relative position of verbs and adjectives depends more on the part-of-speech systems of the languages involved and less on the inherent borrowability of either class (Romaine 1995: 65). Languages without a clear-cut morphosyntactic distinction between adjectives and nouns borrow these lexical classes from languages which do make such a distinction and use both in exactly the same distribution. Since no case studies provides a classification of the parts-of-speech systems of the languages participating in the borrowing process, no typological criteria are available to evaluate the relative position of these lexical classes on the hierarchies proposed.

The class of adverbs proves not less problematic. The reasons have to do again with the gamut of lexical and morphological variants involved under the label “adverb”. As shown above, adverbs have subclasses with different morphological and syntactic behaviors which make their grouping in one single class a matter of convention rather than categoriality. In distributional terms, only manner adverbs are verb modifiers proper. Other subclasses modify adjectives or even other adverbs, and still others modify clauses and sentences. In morphological terms, adverbs are similar to adjectives in several ways. In some cases adverbs are produced from adjectives by adding a derivational morpheme (e.g. English *-ly*, Spanish *-mente*); in other cases no derivation is required and the same form may be used adjectivally or adverbially. The variation within the adverb category makes it clear why any attempt to make valid generalizations on the borrowing of adverbs is doomed to fail. Of the aforementioned authors, only Whitney 1881 and Haugen 1950 show the position of adverbs on the hierarchy explicitly (immediately after the verb). Others do not mention adverbs at all or assume they are included under adjectives (Muysken 1981).

For adjectives and adverbs, considerations of the typological profile of the donor and the recipient languages are required to evaluate their contribution to borrowing. Furthermore, a typological consideration of adjectives and adverbs challenges by itself the universality of scales of borrowability and restricts their application to the limits imposed by the typology of the languages in question. In other words, any hierarchy should be applied only to the specific donor-recipient pair considered in a particular contact situation and not across the board as in the universals proposed by Moravcsik (1978). Of course, this does not necessarily mean that hierarchies are useless predictors. From a set of implications derived from the

theory of parts of speech I show in Chapter 4 that hierarchies of borrowability refine their predicting capacity through the inclusion of typological criteria concerning the morphological type and the system of parts of speech of the language pairs considered. The issue of typological compatibility will come up clear in the following discussion about the borrowability of grammatical elements.

3.5.2. Grammatical borrowability

Less numerous but also less rigorous proposals have been made about grammatical borrowing. In this context, Campbell (1993) admitted that “grammatical borrowing has been both neglected and abused in studies of syntactic change” (1993: 91). Positions on this issue range from the statement that grammatical items can be borrowed almost without restriction (e.g. Wackernagel 1926-8: 8; Thomason 2001: 63) to the idea that grammatical borrowing is not possible at all (Sapir 1921: 203). Intermediate positions are represented by Weinreich (1953: 25), according to whom grammatical borrowing is possible only to the extent that the donor and the recipient languages are structurally compatible. Grammatical borrowing refers not only to the transfer of function words and bound morphemes but also to syntactic borrowing.¹³ In this section I review several issues concerning grammatical borrowability as a backdrop for the subsequent discussion of structural incompatibility.

Scales of grammatical borrowability

Scales of borrowability cover a continuum stretching from lexicon to grammar. Some authors (e.g. Muysken 1981) include grammatical borrowing on the right end of this continuum. In this perspective grammatical borrowing is an extreme case of borrowing associated with contact situations more intense than those leading to lexical borrowing.¹⁴

The study of grammatical borrowing has awakened the interest of many students of language since the late eighteenth century (Gyarmathi 1799, quoted in Campbell 1993: 91). However, few have undertaken a systematic research into the mechanisms involved in the process of borrowing. Proposals concerning grammatical borrowing count many. Some authors put forward general tendencies while others promote the latter to the status of universals without much consideration of extralinguistic factors and on the basis of a limited number of

¹³ It is worth noting that some authors (particularly Harris and Campbell 1995; but also Heine and Kuteva 2005 to some extent) equate syntactic borrowing with grammatical borrowing.

¹⁴ This may be taken as a rule of thumb in contact linguistics but there is one exception, i.e. the borrowing of word order patterns, a phenomenon supposedly found in all cases of languages in contact (Heine 2005).

contact situations. I discuss hereunder two scales proposed in the literature of language contact.

The first scale of grammatical borrowability was proposed by Whitney (1881) as part of a broader scale of linguistic borrowing. This author considers grammatical borrowing an extension of lexical borrowing along a continuum:

Table 3.4. Borrowing continuum based on Whitney 1881

Lexical borrowing	Grammatical borrowing				
	Function words			Affixes	
...	Prepositions	Conjunctions	Pronouns	Derivational	Inflectional
>	>	>	>	>	

The proposal makes two basic distinctions: one between lexical and grammatical borrowing, and another between function words and affixes. A further distinction separates derivational from inflectional affixes on the basis of the nowadays common idea that inflectional morphology is less borrowable than derivational morphology (e.g. Weinreich 1953; Moravcsik 1978; Field 2002). This is the view held by Moravcsik, who states in her fifth hypothesis that “no inflectional affixes can belong to the set of properties borrowed from a language unless at least one derivational affix also belongs to the set” (Moravcsik 1978: 112). According to Campbell (1995: 135) this hypothesis is absolutely false, as there are several cases in which inflectional morphemes have been borrowed without derivational ones being previously borrowed.¹⁵

The second scale I want to discuss here is the one put forward by Muysken (1981: 130) and Muysken and van Hout (1994). This scale is embedded in a continuum of borrowability stretching from lexicon to grammar.

Table 3.5. Borrowing continuum based on Muysken 1981

Lexical borrowing	Grammatical borrowing						
	Preposition	Coordinating Conjunction	Quantifier	Determiner	Pronoun	Clitic Pronoun	Subordinating conjunction
...							
>	>	>	>	>	>	>	

¹⁵ Campbell cites, among others, the example of Bolivian Quechua, in which the Spanish plural inflectional morpheme *-s* has been borrowed but “apparently without any borrowed Spanish derivational affixes” (1995: 135). While this may be true for Bolivian Quechua, it is not the case for Ecuadorian Quichua, which has borrowed, apart from the Spanish plural, at least two derivational morphemes, the agentive *-dur*, from Spanish *-dor*, and the diminutives *-itul/-ita*, from Spanish *-ito/-ita*. The borrowing of both affixes is clearly motivated by the borrowing of unanalyzed lexical chunks. For an analysis of grammatical borrowing in Imbabura Quechua, a Northern dialect of Ecuadorian Quechua, see Gómez Rendón and Adelaar (forthcoming).

Unlike Whitney's, this scale include only function words but *not* affixes on the assumption that function words are more prone to borrowing than affixes. This assumption can not be considered a universal constraint because there are well-attested cases in which affixes (viz. bound morphology) have been borrowed without accompanying function words (viz. free morphology). by Heath (1978) reports a case of the widespread diffusion for the aboriginal languages of Arnhem Land. Heath's "diffusible" categories include case affixes, derivational verbal affixes, verbalizers and the like, while "non-diffusible" categories are independent pronouns, bound pronominals (pronominal clitics) and demonstrative stems and adverbs, which are precisely those categories Muysken sets higher on the scale of borrowability. Haugen (1956) entertains a similar idea, according to which "function words, which only occur as part of utterances, are seldom borrowed" (1956: 67). In general, no precedence of function words (free morphemes) over affixes (bound morphology) may be claimed with universal value.

Another issue related to the borrowing of function words is their function in the borrowing language. One of the universals of language contact proposed by Moravcsik maintains that the borrowing of a function word implies the borrowing of "its linear order with respect to its head" (Moravcsik 1978: 112). This means that function words are always borrowed along with the corresponding syntactic pattern and function. In other words, no prepositions are borrowed which function as postpositions in the recipient language. There are two objections to the terms in which this hypothesis is formulated. On the one hand, this claim is counterintuitive in the sense that it excludes the borrowing of function words between languages with different syntactic patterns (Campbell 1995: 136). On the other hand, while counterexamples to the hypothesis are hard to find, lack of evidence is not sufficient proof. Still, a potential counterexample is the borrowing of the Spanish feminine article *la* in Paraguayan Guaraní, a language originally without articles. Not only the borrowing of *la* violates the requirement of structural compatibility; the use of the article as anaphoric and cataphoric pronoun in Guaraní breaks the word order patterns of Spanish article in Spanish.¹⁶

To complement his scale, Muysken lists a number of general ruling principles such as: 1) content words are easier to borrow than function words; 2) words that belong to structured paradigms are more difficult to borrow than words that do not belong to a structured paradigm; 3) case-assigning words are more difficult to borrow than words *not* assigning case; and 4) morphologically complex words are

¹⁶ For an analysis of the Spanish article in Paraguayan Guaraní, see Gómez Rendón (2007b). An alternative explanation of its use might be the homophony of the feminine article with the pronominal clitic.

more difficult to borrow than simple words. Principles (1) and (4) have been already mentioned in this chapter; principles (2) and (3) may be considered their extensions.

The idea of paradigms as structural constraints on borrowing is recurrent in the literature of language contact up to date. Case assignment has been used recently as a borrowing parameter by Myers-Scotton in her four-morpheme model. Comparing Muysken and Myers-Scotton with respect to the effects of case assignment, however, sheds a major discrepancy between them. Muysken considers case assignment as preventing verbs and prepositions from being borrowed easily while Myers-Scotton identifies content words as morphemes whose case-assigning condition precisely furthers borrowing. Notice also that Muysken makes a distinction between coordinating and subordinating conjunctions. In principle coordinating conjunctions are easier to borrow than subordinating ones. This hypothesis, however, does not hold for Otomí (Bakker and Hekking 1999) which borrows as many coordinating as subordinating conjunctions from Spanish. Still, the scale of borrowing and the general principles sketched by Muysken provide useful parameters for identifying the types and frequencies of grammatical borrowing.

The last issue I want to address here concerns the precedence of lexical borrowing with respect to grammatical borrowing as implicit in all scales of borrowability. Moravcsik (1978) promoted the precedence of lexical over grammatical borrowing to the status of a universal of language contact, and it has been admitted explicitly or implicitly ever since. The founding of this principle is the empirical fact that lexical items usually are more prominent in discourse than function words or affixes and therefore more available for borrowing. Nevertheless, “there would seem to be no inherent connection between prior lexical borrowings and grammatical loans, and hence no theoretically significant implications, even if this claim should prove true” (Campbell 1995: 134). Furthermore, the attestation of lexical precedence does not exclude scenarios where grammatical categories are borrowed in the absence of lexical items. The well-known case of Finno-Ugric influence on Russian in the absence of lexical loans adds to other recently documented cases, including the borrowing of syntactic patterns without lexical elements known as calquing.

3.5.3. Syntactic borrowing: calquing of word order patterns

Syntactic borrowing is often used as a synonym of grammatical borrowing (e.g. Campbell, 1989; 1995). Here we restrict this term to the calquing by one language of the word order patterns from another language while considering ‘grammatical borrowing’ a cover term for any kind of non-lexical borrowing. It has been suggested that syntax is resistant to change through contact, but the facts show that syntactic borrowing is ubiquitous and results from other borrowing processes at the level of lexicon and grammar (cf. Heine and Kuteva 2005). In this perspective

syntactic borrowing is considered “the epiphenomenal product of processes whereby meaningful structures are reinterpreted as some other structures” (Heine 2005: 60). Syntactic borrowing will be dealt with here to the extent that it occurs in our corpora as a result of the borrowing of lexical items and function words.

The borrowing of word order patterns is relevant to a theory of language change insofar as deviant word orders in languages cannot be explained by internal changes only. They usually originate in syntactic calques from neighboring languages and have two forms. On the one hand, previously marked word orders become unmarked as a result of the calquing of similar syntactic patterns from contact languages: in other words, a markedness shift occurs as a result of the frequency of the once marked pattern in the recipient language (Campbell 1995: 136ff). On the other hand, as a result of the transfer of the meanings encoded in lexical items borrowed from a contact language, “some structure is reinterpreted as some other structure, with the result that a seemingly new word order arises” (Heine 2005: 65). Both types of syntactic change may cause a typological shift in the borrowing language, but only the second type implies lexical borrowing. An example of reinterpretation of structures is found in Paraguayan Guaraní. The example below involves the reinterpretation of the native finite verb *ojapo* ‘it makes’ within a time adverbial construction:

- 2) *o-japo* *mokõi* *ary* *o-mano-ma-va'ekue* *che-ru*
 3-do two year 3-die-already- NMLZ.PST 1S-father
 ‘It is two years that my father died’
 Sp. ‘hace dos años murió ya mi papá’

Interestingly, the speaker of (2) is an educated bilingual man who feels proud of speaking what he considers “pure” Guaraní (*Guaraníete*). He was not aware of the extent to which Spanish had influenced the way he builds phrases and sentences in Guaraní. From the gloss it is clear that (2) is a perfect copy of the Spanish sentence, even though no lexical item from this language is involved. First, the Guaraní postposition *guive* ‘from, since’ has been replaced by a Spanish-modeled construction based on the finite verb *hace* ‘it makes’. Second, the adverbial phrase has been fronted for emphasis – an uncommon mechanism in classical Guaraní. Third, the subject has been placed in sentence-final position, an exceptional strategy in traditional Guaraní in spite of its relatively free word order. According to Heine (2005) cases like the transformation of *ojapo* from the status of a finite verb to that of an adposition on the model of Spanish are instances of contact-induced syntactic change through “a process of grammaticalization as it can be observed in situations that do not involve language contact” (Heine 2005: 71).

Of the authors who have proposed scales of borrowability, only Thomason (2001: 70) identifies syntactic borrowing explicitly as contact-induced change in

word order. On her four-stage borrowing scale, syntactic borrowing is coterminous with intensity of contact. Syntactic changes do not occur in the first stage (casual contact), but they appear increasingly in the following stages. The scale goes from an increased usage of previously rare word orders in the second stage to their fixation as unmarked word orders in the third stage, and the occurrence of “sweeping changes” in relativization, coordination, subordination, comparison and quantification in the fourth stage (Thomason 2000: 71).

Well-documented cases of contact-induced syntactic change are mentioned in Weinreich (1956) and Moravcsik (1978). The first author identifies syntactic borrowing as ‘interference in word order’ in the context of widespread bilingualism (Weinreich 1956: 38). Although syntactic borrowing not necessarily implies bilingualism, it is the natural result of having two linguistic systems in contact. Moravcsik, in turn, speaks of “the borrowing of syntactic constituent-ordering rules” for a handful of language families including Ethiopian Semitic, Cushitic, Assamese, Indo-European, Tibeto-Burmese, Dravidian and Bantu languages. Syntactic borrowing is part of the sixth universal of language contact proposed by Moravcsik, according to which “a lexical item that is of the ‘grammatical type’ (which type includes at least conjunctions and adpositions) cannot be included in the set of properties borrowed from a language unless *the rule that determines its linear order with respect to its head is also included*” (Moravcsik 1978: 112, my emphasis). The copying of word order patterns as a result of the borrowing of function words has been previously attested for one of the languages of this research (Quichua). There is a close link – i.e. a grammatical relationship – between the borrowing of Spanish subordinating conjunctions such as *porque* ‘because’ or *si* ‘if’ and the abandoning of native Quichua SOV word order for a Spanish-like SVO pattern (Gómez Rendón 2007a). Similar developments have been identified in cases of massive lexical borrowing as in Media Lengua (Muysken 1985), where the dropping of the Quechua accusative marker *-ta* on Spanish-origin items is related to an increasing frequency of SVO word orders. The difference in both cases lies on the speakers’ level of bilingualism: most Ecuadorian Quechua speakers are only partially bilingual; Media Lengua speakers are full-fledged bilinguals. Evidence of a second link between syntactic borrowing (interference) and bilingualism is found in Paraguayan Guaraní, where the copying of Spanish word order patterns with or without Spanish loanwords is common in bilingual discourse.

The study of syntactic borrowing is underdeveloped in relation to other types of borrowing, for which reason conclusive statements cannot be made as long as a comprehensive collection of data and new analytic approaches are not available. In this book I do not develop an analysis of syntactic borrowing. Still, I wanted to highlight the relevance of it for a comprehensive evaluation of the data in the following chapters.

3.6. Structural compatibility as a constraint on borrowing

Structural compatibility is one of the most used and abused concepts in contact linguistics. Proposals of structural compatibility are based on the notion that only structurally compatible languages can borrow from each other, which means that the typology of languages constrains their ability to borrow lexical and grammatical elements.

In his review of the topic, Campbell (1995: 123f) quotes several authors, from Meillet (1914) and Weinreich (1953) to Bickerton (1981) and Aitchinson (1981), who maintain in one form or another that borrowing (or interference for that matter) is possible only between structurally similar languages. Campbell reviews a large number of cases in which typologically different languages have been in contact, with the resulting exchange of grammatical material from one another. In concluding Campbell states that

Such examples as those presented here show that the structural-compatibility requirement in any absolute sense is incorrect. It is as a general tendency or preference that we may expect the claim to hold, but how is it to be framed? To be very useful in a theory of change, it would require an explicit notion of what “shared syntactic [grammatical] similarity” is and how one determines it. Essentially at stake here is how social factors can overcome structural resistance to borrowing” (Campbell 1995: 125).

The statement is crucial for the position granted to social factors without the exclusion of structural conditions. As shown in the previous chapter, social factors not only downplay other factors when it comes to borrowing but also trigger language change in a more general sense. At the same time, structural (typological) factors remain a backdrop before which changes are displayed and signal potential ways of development for language change. While the non-universal validity of the criterion of structural similarity is well documented in the literature, the failure to characterize this criterion in more specific terms led to its invalidation as a powerful predictor.

The ambiguity in the treatment of structural compatibility is best exemplified by Weinreich (1953: 64-5) in his assessment of structural constraints on interference. After a thorough discussion supported with empirical evidence, Weinreich summarizes his findings in a table of structural factors that stimulate or hinder interference at the phonological, morphosyntactic and lexical levels. The result is a collection of general criteria that have lost all of their predictive capacity: stimuli for interference are broadly characterized as “any points of difference between two systems” (1968: 64); inhibitors, on the other hand, are system stability and intelligibility without further specification. While these constraints are general

enough to be valid for all types of interference, not less general are the factors corresponding to different levels of language organization. Consider for example the type of interference that Weinreich calls “abandonment of obligatory categories” in grammar. He does not mention inhibiting factors but only one stimulating factor, i.e. the co-existence of “very different grammatical systems”. From recent studies of creolization we know that phenomena such as grammatical simplification and restructuring are the result of a very long chain of socio-historical events that unchain a complex of linguistic processes.¹⁷ In cases of lexical interference such as the “the specialized retention of an ‘indigenous’ word after borrowing of an equivalent” (Weinreich 1968: 64), the avoidance of semantic confusion and the elimination of “superfluous terms” are considered stimulating and resisting factors, respectively. Still, some studies of couplets (borrowed item vs. native item) have shown that motivations and factor influencing their formation and use go beyond semantic vagueness and language economy (e.g. Montes de Oca 2004: 70-84).

Structural similarity may take different shapes depending on the structure used as measuring stick. In morphology, for instance, structural criteria are agglutination or polysynthesis but also bound and free morphemes. In the lexicon, lexical categories, parts of speech and semantic categories are structural parameters. Therefore, it is necessary to specify the kind of structure we have in mind when speaking of structural compatibility.

To make the notion of compatibility a predictive device for this study, I restrict myself here to the criteria of a) morpheme type and b) parts of speech, and articulate recent proposals in this field with the notion of structural compatibility. The concept of parts of speech concerns lexical borrowing. In turn, the concept of morpheme type as a structural criterion for cross-linguistic compatibility bears relevance for lexical and grammatical borrowing. Both concepts are discussed below on basis of two different studies (Field 2002; Hengeveld *et al* 2004).

3.6.1. Morphological typology and structural compatibility

The Humboldtian classification of languages according to morpheme types has been often criticized for its failure to capture the real complexity of languages. Yet, it remains a useful parameter to attempt a preliminary classification of languages provided several other criteria are taken into consideration. The classification of languages into isolating, agglutinative and fusional languages proves especially valuable in the field of contact linguistics because it marks the boundaries of change and the outcomes of contact. Field (2002) has recently introduced morpheme types

¹⁷ The case of Ecuadorian Quechua is illustrative in this respect. The language lost in the last four centuries several obligatory categories that are preserved nowadays only in central Peruvian varieties. For an insightful study of this process of gradual restructuring, see Muysken (forthcoming).

as criteria for the identification of typological constraints on contact outcomes. I summarize hereunder the two principles of his proposal and bring them under the light of the main topic of this section: the relation between borrowability and structural compatibility.

According to Field (2002: 27f) the classification of languages based on morpheme types¹⁸ takes as points of departure: 1) an index of synthesis which shows the lesser or greater correlation between morpheme and word (i.e. how many morphemes build a word); and 2) an index of fusion which shows the amount of lexical or grammatical information contained in one morpheme (i.e. how semantic information is mapped on morphological material). When both indexes are considered, languages are of three types, namely: *isolating-analytic*, i.e. those languages which exhibit a univocal correlation between morpheme and word (one morpheme per word) as well as one semantic unit per morpheme; *agglutinating-synthetic*, i.e. those languages which exhibit a many-to-one correlation between morpheme and word (two or more morphemes per word) but still assign one semantic unit per morpheme; and *fusional-synthetic*, i.e. those languages which not only exhibit a many-to-one correlation between morpheme and word but also assign several semantic units per morpheme. Each language type has its own form-meaning units: isolating-analytic languages have independent words; agglutinating-synthetic languages have independent words, roots and agglutinating affixes; and fusional-synthetic languages have all of the above plus fusional affixes. The type of form-meaning units that may be borrowed from one language depends on the inventory of units of the recipient language. According to Field (2002: 42) this can be captured in two complementary principles:

The Principle of System Compatibility (PSC)

Any form or form-meaning set is borrowable from a donor language if it conforms to the morphological possibilities of the recipient language with regard to morphological structure.

The Principle of System Incompatibility (PSI)

No form or form-meaning set is borrowable from a donor language if it does not conform to the morphological possibilities of the recipient language with regard to morpheme types.

These principles allow us to chart all the possible form-meaning units that are borrowable from one language to another depending on the morphological typology of the recipient language. The following table adapted from Field (2002: 42) summarizes all compatible and incompatible units of a donor language with respect to the recipient language.

¹⁸ This classification is inspired originally in Comrie (1989).

Table 3.6. Morphological Typology and Borrowability of form-meaning units

Typology of recipient language	Compatible forms of donor language	Incompatible forms of donor language
Fusional-synthetic	Independent words, roots, agglutinating affixes, fusional affixes	Zero (all forms of donor language are compatible)
Agglutinating-synthetic	Independent words, roots, agglutinating affixes	Fusional affixes of donor language only
Isolating-analytic	Independent words, roots (analyzed as discrete words in an isolating recipient language)	Any affix of donor language (including agglutinating and fusional forms)

Table 6 shows that borrowable form-meaning units range from free and bound morphemes, when the borrowing language is fusional-synthetic, to free morphemes and one set of affixes, when the borrowing language is agglutinating-synthetic, and free morphemes, when the borrowing language is isolating-analytic. The case of roots is somewhat ambiguous as they are bound morphemes but analyzed as independent words when the recipient language is isolating.

Summing up, the morphological profile of the borrowing language constrains the type of form-meaning units that may be incorporated from a source language. In principle no restrictions apply for the borrowing of lexical or grammatical items provided they match the morpheme type of the recipient language. The principles of system compatibility and system incompatibility set the limits of borrowing and antecede scales of borrowability. In other words, these principles provide the general rules for borrowing while the scales make specific predictions about lexical and grammatical categories within the space delimited by the principles. The constraints on borrowing based on morpheme types and lexical-grammatical categories may be further refined if the systems of parts of speech of the borrowing languages are included. The next section discusses the contribution of the theory of parts of speech to the refinement of the notion of structural compatibility and the resulting constraints on borrowing.

3.6.2. Parts-of-speech typology and structural compatibility

In section 3.2 I presented the theory of parts of speech developed by Hengeveld (1992) and Hengeveld *et al* (2004) and showed how languages are classified according to their parts-of-speech systems on the basis of two criteria: a) the undifferentiated use of lexical classes; b) the use of alternative strategies to replace absent lexical categories. Accordingly, languages may be classified in three basic

types: flexible languages which use one lexical class in different syntactic slots; differentiated languages which use one lexical class in one syntactic slot; and rigid languages which use morphosyntactic strategies instead of lexical items to fill one or more syntactic slots. Intermediate systems are expected when derived lexemes cannot be used in all the syntactic slots in which their base lexemes are used, in the case of flexible languages; or when the last class of lexemes on the hierarchy is a small closed class of items, in the case of rigid languages.

The general implications of Hengeveld's theory of parts of speech for borrowing were discussed in section 3.3 and are summarized as follows:

The parts-of-speech systems of the languages involved in the borrowing process are relevant to determining the type of borrowed lexical classes and the functions to which they are put in the recipient language. More specifically, the parts-of-speech system of the recipient language co-determines the borrowing of lexical classes from the source language and their functional adaptation in the recipient system.

Like the principle of system (in)compatibility, the implication of parts-of-speech theory for borrowing is based on the typology of the recipient and the donor languages. While the system of parts of speech of the donor language determines the lexical classes *available* for borrowing, the system of parts of speech of the recipient language determines what lexical classes are *borrowable*.

The matching of the principle of system (in)compatibility with the implications of the theory of parts of speech may help to predict what types of lexical borrowing are permitted across typologically different languages in the following terms:

- 1) Lexical flexibility in the parts-of-speech system of the recipient language increases borrowability to its maximum when this language is fusional-synthetic.
- 2) Lexical rigidity in the parts-of-speech system of the recipient language decreases borrowability to its minimum when this language is isolating-analytic.
- 3) Lexical flexibility or rigidity in the parts-of-speech system of the recipient language increases or decreases borrowability when this language is agglutinative-synthetic.

That is, lexical flexibility increases the borrowability of lexical items determined by the morphological type of the recipient language while lexical rigidity reduces the borrowability of these items. Flexibility and rigidity act therefore as factors promoting or inhibiting borrowability of lexical items from the source language in accordance with the morphology of the recipient language.

3.7. Summary

Linguistic constraints on lexical and grammatical borrowing can be ordered according to their applicability from the more general to the more specific: the principles of system compatibility and incompatibility, which determine the borrowable types of lexical and grammatical units on the basis of their conformity to the morphological profile of the recipient language; the scales of borrowability, which make predictions about the borrowing of word classes in terms of precedence in time and frequency; and the theory of parts of speech, which determine what content words are borrowable depending on the parts-of-speech systems of the recipient language.

These complementary sets of constraints are systematized in principles, hierarchies and hypotheses and are all interconnected. Hypotheses on lexical borrowing are applicable only in the framework of borrowing continua established by hierarchies. Hierarchies of borrowability are applicable only in the framework of the morphological profile of the recipient language. The following chapters show how these constraints interact and how I incorporate them in the analysis of the data.

Chapter 4

The Research Program

This chapter outlines the research program underlying the present investigation and discusses a number of questions related to the methods used in the collection and analysis of data. The chapter is organized in three sections. The first section reviews studies on linguistic borrowing and the methodologies used for the investigation of this linguistic phenomenon. I outline a program of research on linguistic borrowing oriented to solving methodological shortcomings on the basis of an in-depth analysis of three contact situations in different areas of Latin America which involve one donor language (Spanish) and three typologically different languages (Paraguayan Guaraní, Ecuadorian Quichua and Mexican Otomí). The choice of languages is substantiated on a theoretical and methodological basis. The third section sets the main questions guiding this research and how I intend to provide answers to them. Afterwards I discuss the general hypotheses to be tested on the corpora collected for the aforementioned languages. The hypotheses are developed from the premise that the typology of the languages in contact co-determines the degree and the form of lexical and grammatical borrowing and their functions in the recipient language (cf. section 3.3). The hypotheses will be further developed for each of the investigated languages in the following chapters. The last section addresses a number of methodological issues concerning the process of data collection, the setup of the corpora and their characteristics, the representation of data and the criteria used in the statistical analysis. The discussion of some research problems and their solutions rounds off the chapter.

4.1 A critical overview of studies on linguistic borrowing

Linguistic borrowing was first studied in the nineteenth century as part of comparative and historical linguistics. Early studies viewed linguistic borrowing as a random phenomenon influenced by countless non-linguistic motivations which, albeit interesting, lacked relevance for linguistic theory. A notable exception is the classical study on “language mixture” by Whitney (1881) in which he sketched a number of regularities of linguistic borrowing. In the early twentieth century Meillet (1921), Vočadlo (1938) and others advanced the research on linguistic borrowing in more precise terms. However, the most important step toward a systematic study of borrowing was taken by Haugen (1950). Haugen not only addressed current issues in the field but provided a thorough classification of borrowings which somehow survives up to date. The second breakthrough in the study of linguistic borrowing is the work of Thomason and Kaufmann (1988). According to Myers-Scotton (2002:

236), the innovative aspects of their proposal are: a) the distinction between interference and borrowing; b) the linking of borrowing types to intensity of contact through a borrowing scale; and c) the inclusion of lexical and structural borrowing in one scale. Their contribution was certainly decisive in establishing linguistic borrowing as a central issue in contact linguistics.

Most of the aforementioned studies do not grow out of corpus-based investigations. Instead, they are collections of findings from different case studies. Many of these studies are not comparable on account of the differences in social, cultural and historical aspects of the contact settings,¹ but also because the methods followed in the collection and analysis of data are not standardized. Case studies on linguistic borrowing concern individual pairs of donor and recipient languages in different contact situations all over the world. They are based on corpora gathered from oral or written sources. A serious shortcoming of case studies is their preference for analyzing linguistic borrowings from dominant (usually European) languages in native languages. This bias is explained by the fact that lexical and grammatical borrowing in colonial and neocolonial settings has a greater impact on native languages, but also by the focus on imperfect learning and acquisition in the study of non-standard varieties of dominant languages, such as Spanish, Portuguese or English in the Americas. A further shortcoming of case studies is that sociolinguistic information concerning bilingualism and other speaker factors are not considered in the analysis. Moreover, an important number of case studies make no functional distinction between code-switching and borrowing. As I have shown, the distinction between both phenomena is important in diglossic situations because it implies different levels of bilingualism and distinct communicative answers to the pressure exerted by contact. Finally, the great majority of case studies on linguistic borrowing have not been conducted in a theoretical framework that provides the researcher with tools for analysis and hypotheses about expected outcomes in specific situations.

The findings of an increasing number of case studies are used to support distinct views on language contact (cf. Romaine 1989; Thomason 2001; Myers-Scotton 2002) and different scales of borrowability (cf. Haugen 1953; Muysken 1981; Singh 1981; Bakker and Hekking 1999; Field 2003), their results being often interpreted in an ad hoc manner. It has been repeatedly claimed, for example, that borrowings are grouped in recognizable word classes and certain word classes are preferred over others. Findings that confirm these claims are numerous in the literature, but much less numerous are studies which explain these claims in the framework of linguistic theory. The fact that authors resort to ad hoc explanations has resulted in a larger number of borrowing scales without sufficient theoretical foundations. The

¹ A notable exception for the treatment of data and the consideration of several intervening factors is Stolz and Stolz (1996).

universals of language contact proposed by Moravcsik (1978) are an exception, not only because of the typological principles underlying her proposal but also because of their arrangement in a coherent set. Still, it was not difficult to demonstrate that these universals are disconfirmed by evidence from many contact situations around the world. Other explanations are less consistent though potentially promising: e.g. frequency and transparency as factors influencing the preferential borrowing of open classes over closed classes. In addition, the study of the influence of typology on contact-induced language change has a long tradition in linguistics (Meillet 1921; Vočadlo 1938; Weinreich 1968). All in all, explanations of linguistic borrowing are not theory-driven and therefore lack a systematic treatment.

4.1.1 Studies on linguistic borrowing in Latin America

Studies on linguistic borrowing in Latin America have focused on the influence of Spanish (and Portuguese) on Amerindian languages.² Some of these studies are based on isolated examples extracted from grammars and dictionaries, while others result from a corpus-based investigation. In both cases findings are not discussed in a coherent framework.

One of the earliest reports on linguistic borrowing in Amerindian languages is Boas (1931), who presents an inventory of Spanish elements in Modern Nahuatl (Uto-Aztecan). More recent studies on Nahuatl are Hill and Hill (1986) and Flores Farfán (1999), both of which show the impact of Spanish-origin borrowings on the lexicon and the grammar of the native language up to the emergence of mixed sociolects. Studies on lexical borrowing for other languages of the Uto-Aztecan family from the perspective of lexical acculturation are Silver and Miller (1997) for Mountain Pima and Comanche, and Campbell (1987) for Spanish influence on Pipil. From an ethnolinguistic point of view Brown (1994) provides an interesting survey of Spanish and English lexical acculturation in Native American languages.

For the influence of Spanish on Otomí (Otomanguean) there are several reports by Hekking and Bakker (1998, 1999, and 2007) plus a comprehensive study by Hekking (1995) on language shift and restructuring in the Otomí dialect of Santiago Mexquititlán. These studies are corpus-based and typology-oriented and stress the relevance of typological factors for borrowing. Further studies on native Mexican languages are Brody (1976) on Spanish-origin particles borrowed as discourse

² The influence of Amerindian languages on the local and regional varieties of Spanish is comparatively less studied. Worthy of mention are Haboud (1998) on Ecuadorian Andean Spanish influenced by Quichua; Krivoshein de Canese and Corvalán (1987) and Dietrich (1995) on Paraguayan Spanish influenced by Guaraní; and Flores Farfán (1998; 2000) on Mexican Spanish influenced by Nahuatl and learned as a second language by native speakers of native Mexican languages. Still, none of these studies are corpus-based and only few are theory-driven. The result is that most of their findings remain at a purely descriptive level.

markers in various Mayan languages; and Knowles-Berry (1987) on linguistic decay in Chontal (Mayan) as a result of Spanish borrowing. Both authors do not base their findings on corpora but on isolated data collected through elicitation techniques.

Studies on linguistic borrowing in the Andes are comparatively numerous. Notable are Muysken and van Hout (1994), a model-based statistical investigation of a corpus from Bolivian Quechua; Carranza-Romero (1998) on the borrowing of Spanish derivational and inflectional morphemes in Peruvian Quechua; Muysken (1979; 1981; 1985; 1997) on Spanish borrowing and relexification in Ecuadorian Quichua (Media Lengua). A more specific study of linguistic borrowing is Hulleman and van Gein (1998), who analyze the borrowing of Spanish-origin diminutive constructions in Cochabamba Quechua (Bolivia). Hekking and Muysken (1995) compare the distributions of word classes in Spanish borrowings in Bolivian Quechua and Mexican Otomí. Similarly, Stolz and Stolz (1996) report on the borrowing of Spanish function words on different continents.

The contact between Spanish and Guaraní in Paraguay has remained relatively understudied despite its relevance for the relation of bilingualism and borrowing. Until recently all the studies of the influence of Spanish on Guaraní were due to one single author. Morínigo (1936) is an extensive compilation of Spanish words in Guaraní accompanied with examples, glosses and ethnographical explanations. Morínigo (1959) and Morínigo (1982) address the phenomenon of Spanish grammatical borrowing in Guaraní from two slightly different perspectives. A recent corpus-based study of lexical and grammatical borrowing in Paraguayan Guaraní from a typological perspective is Gómez-Rendón (2007b). While many sociolinguistic studies insist on the unique condition of Paraguay as the only bilingual country in Latin America, systematic analyses of the impact of Spanish on Guaraní are scarce and limited to isolated cases.

4.2 A program of research on linguistic borrowing

Considering the theoretical and methodological shortcomings of most studies in linguistic borrowing it was therefore necessary to set up a research program that solves these deficiencies by

- a) identifying a set of constants and variables in order to have control over the data and arrive at valid language-specific and cross-linguistic conclusions;
- b) working from a corpus-based perspective in order to base the analysis on realistic data; and
- c) interpreting the data in the framework of a linguistic theory that predicts the behavior of borrowing in different typological and sociocultural settings.

The research on linguistic borrowing which is reported in this book follows these guidelines. In the following I explain how the guidelines translate into a set of parameters.

Borrowing versus shift-induced interference

This investigation focuses exclusively on borrowing and the ways Spanish influences Amerindian languages. The study of the influence of these languages on Spanish through imperfect learning is not considered here. The focus on borrowing implies that all the speakers interviewed are native speakers of one Amerindian language while their second language, either learned or acquired, is a variety of Spanish spoken in their area of origin or residence. Accordingly, the native language is the dominant language of the speech community to which the speakers belong, while Spanish is used mostly in transactions with the mainstream society, usually though not always, outside the borders of the speech community. The Spanish proficiency of the speakers interviewed varied depending on such variables as gender, age and formal schooling.

Borrowing versus code-switching

Spanish-origin elements present in the native language are either borrowings or code-switches.³ Each type has its own linguistic features and may be distinguished from the other according to several parameters discussed in section 2.5.2.1. Since my purpose is to identify constraints on the borrowing of Spanish elements in the native language and their accommodation to the recipient language, only the first type of elements (borrowings) is considered for analysis. Code-switches were identified and labeled according to their length (single-word or complex) but they were not included in the analysis. In order to assess the overall contributions of code-switching and borrowing in the corpus of each language, texts were analyzed for number, length and type of code-switches. This helped us measure the differential contributions of foreign elements in the samples and identify relations between them. Likewise, it helped us establish a distinction between single-word borrowings and single-word code switches, on the one hand, and between complex borrowings and complex code switches, on the other.

Considering the controversial nature of single-word code switches and their relation to borrowing (section 2.5.2.2) it was necessary in several cases to conduct a fine-grained analysis of the phonological and morphosyntactic integration of the foreign elements in the native discourse before assigning them to either category.

³ Arguably, nonce borrowings represent an intermediate stage between established borrowings and code switches, thereby forming a continuum between codeswitching and borrowing.

This procedure also helped us make a clearer distinction between complex borrowings (e.g. frozen constructions lexicalized as single units in the recipient language) and complex code switches (e.g. chunks of foreign material inserted in the recipient language). I do not pretend in this way to settle the issue of the distinction of both types of mixing phenomena. My purpose was purely instrumental to the extent that the application of several criteria of structural accommodation could help us draw a dividing line between borrowings and code switches for analysis.

Lexical and grammatical borrowing

Different from a number of case studies on linguistic borrowing that focus on the lexicon, this study includes both lexical and grammatical elements. This inclusion seeks to integrate the findings in one single model of linguistic borrowing. This model aims at explaining linguistic borrowing as the outcome of contact between two typologically different languages on the basis of four sets of constraints determined by a) the principle of functional explanation ; b) the principles of system compatibility and incompatibility; c) the scales of borrowability; and d) the parts-of-speech systems of the languages in contact. Each constraint produces individual hypotheses that predict the higher or lower probability for a linguistic element to be borrowed from one language provided certain conditions are met. The treatment of lexical and grammatical borrowing within a single model is based on the premise that each type of borrowing is determined by time and intensity of contact and other sociocultural motivations. It implies a continuum stretching from lexical to grammatical borrowing as time and intensity of contact increase. While the view of borrowing as a continuum is based on the concept of 'scale of borrowability', the matching of borrowing types with degrees of contact is inspired by the borrowing scale proposed first by Thomason and Kaufmann (1988) and refined later by Thomason (2001).

Typologically different languages as recipients of borrowing

Most studies on linguistic borrowing are investigations of individual cases in different parts of the world. Only a few analyze borrowing in cross-linguistic perspective. This perspective is required when the purpose is to find structural constraints on the borrowing process as in the present investigation. In order to find cross-linguistic constraints on borrowing it is therefore necessary to analyze more than one language but also typologically different languages. This procedure helps us deal with the immense variety of languages of the world. While it is realistically unfeasible to collect and analyze all the languages representing the world's typological variety, it is clearly viable to begin with a sample of languages whose typological profiles are different from each other.

The typological criteria for the selection of the languages of this study include language family, morphology, lexical classes, types of affixation, types of adpositions, and word order. The languages selected were Guaraní, Quichua and Otomí. These languages meet the conditions of typological variation along the aforementioned parameters. Each language belongs to a different family, though all of them are spoken in the Americas. Guaraní is a Tupi language of the Tupi-Guaraní family, spoken by five million people in Paraguay and the Argentinean province of Corrientes. Quichua is a language of the Quechua family, spoken by one million people in the Andean Highlands of Ecuador. Otomí belongs to the Otomanguean branch of the Otopamean language family and is spoken by three hundred thousand people in different states of Central Mexico. As regards their morphological typology, Guaraní is originally a polysynthetic language while Quichua is typically agglutinative and Otomí more analytic than the other two at the level of the sentence but synthetic at the level of the phrase (Bakker *et al* 2008). Also, the three languages differ from each other in their systems of parts of speech: Guaraní and Quichua are flexible languages, but the former shows a larger number of word classes which are used predicatively; Otomí, on the other hand, is a rigid language without adjectives. The type of affixation in these languages makes them different too: Guaraní has both prefixes and suffixes (Gregores and Suárez 1967); Quichua has only prefixes (Cole 1982); and Otomí uses both plus numerous clitics (Hekking 1995). Guaraní and Quichua, on the contrary, are typically postpositional languages; Otomí diverges from them in that it uses prepositional constructions to link elements within the noun phrase. Basic word order is another point of divergence among these languages. Guaraní shows a relatively free word order, with a preference for SVO order (Gregores y Suarez 1967). Otomí has VOS and VS as basic orders, although the frequency of SVO constructions is increasing in usage today (Lastra de Suárez 1994; Hekking 1995). Finally, Quichua is a typical verb-final language, even though the occurrence of SVO constructions has increased over the last years as a result of contact with Spanish. A detailed characterization of the typology of these languages is presented in Chapters 6 through 8.

In addition to linguistic reasons, the investigation of these languages offers good fieldwork conditions because a) they are vital in their number of speakers; b) they are spoken in vast areas of their respective countries; and c) they are relatively well described, with a number of grammars and dictionaries, some of which date from the first years of the Spanish colonization.

One language as the source of borrowing

A fundamental methodological premise of the present investigation is that the recipient languages must be typologically different in order to produce cross-linguistically valid conclusions, and the source language must be kept constant for

all the recipients and typologically different from them. These conditions are met satisfactorily by Spanish.

As a result of the worldwide expansion of the Spanish Empire in the fifteenth and sixteenth centuries, the Spanish language is present in the five continents today. Except for the territory of today's Brazil, which became part of the Portuguese Empire in the early fifteenth century, Spanish was the official language in Central and South America during the three centuries of Spanish colonization. In addition, Spanish was also spoken in several areas in the southern United States until the end of the nineteenth century. The linguistic heritage of Latin American republics reflects the dominance of the Spanish language, and the countries in which the investigation was conducted are no exception. Spanish remains the official language and the largest in terms of speakers in Ecuador, Paraguay and Mexico. Spanish has a long history of contact with hundreds of native languages all over Latin America. At the same time, Spanish has remained typologically distinct from native languages in spite of their substratum and adstratum influence.⁴ Spanish remains a fusional-synthetic language with prepositions and flexible word order, and distinct lexical classes for individual syntactic functions. In all, the sociolinguistic and linguistic conditions of Spanish in the Americas allow for the investigation of linguistic borrowing from one language into typologically different languages. In this way the input to borrowing is kept constant and the foundations are laid for cross-linguistic generalizations.

Spanish America as a sociocultural region

In my model of contact-induced language change the ultimate motivations for borrowing are essentially nonlinguistic. Therefore, any cross-linguistic analysis of linguistic borrowing requires that sociocultural motivations be similar enough to allow for comparison. Keeping the sociocultural variable constant is unfeasible in any realistic study given the enormous variety of national, regional and local societies and cultures found in a vast region like the Americas. Nevertheless, by taking Spanish America as the geographical space for the contact between the source language (Spanish) and the native languages, an important degree of social and cultural unity is warranted. This unity is substantiated by a series of historical events and the resulting sociolinguistic facts.

⁴ Arguably, the Spanish varieties spoken in Ecuador, Mexico and Paraguay are not the same. In fact, it is possible to find a number of lexical and morphosyntactic differences. However, these dialects remain mutually intelligible and typologically similar to each other and to Peninsular Spanish. While this is not the case of Spanish varieties spoken by non-native speakers – like many forms of Indian Spanish described in the literature (e.g. Flores-Farfán 2000) – these were not considered for the present investigation.

The Spanish rule in the Americas lasted over three centuries and left sociocultural imprints in the continent. The identification of Spanish America as a cultural region is based on the cultural heritage shared by all Spanish-speaking countries in the continent. This heritage expresses in a number of facts, from legal and administrative apparatuses to architecture and religion. Notice that a focus on similarities does not neglect differences, which are many and very important. Differences are firmly rooted in the heritage of numberless Indian cultures, many of which survive to the present and became the basis for the foundation of nations such as the Guaraní in Paraguay, the Inca in Peru or the Aztec in Mexico.

As regards the language, the great majority of countries in Spanish America are diglossic societies. In this context Spanish is the language of prestige while the native languages are usually excluded from public spheres. This is the case of Quichua, Otomí and even Guaraní in their respective countries. The official status of Guaraní in Paraguay does not make it different from other native languages in sociolinguistic terms.

In sum, the sociocultural heritage of all Spanish-speaking countries in Latin America and the condition of dominance of Hispanic culture over native cultures allow a controlled comparison between the contact situations analyzed in this book.

A corpus of spontaneous speech from a representative group

Not being based on the investigation of a corpus, most studies of linguistic borrowing take as material for their analysis a collection of isolated examples from the languages in question or a sample of written texts extracted from other sources. The approach of this study is the opposite. I have analyzed individual corpora for the three languages of the sample. These corpora were collected in situ according to a number of criteria to be explained in section 4.4.1. No elicitation was used in the process and speech events were recorded in socially and culturally relevant settings. In doing so I sought to reduce speech monitoring and de-contextualization of verbal exchanges to the minimum. Speech monitoring is an important factor influencing the number and type of borrowings in situations where the source language is used by speakers of the recipient language for their interaction with people from outside of their speech communities, especially if the languages are in a diglossic situation. Accordingly, the corpus of each language is comprised of spontaneous speech in face-to-face interactions.⁵

A further criterion for the setup of the corpora was the inclusion of a representative group of speakers from the speech communities under study. By including *lectal* variation into the sample I could chart the speech of different

⁵ In few cases, however, second-hand material from other sources was to be used in order to cover other registers or sociolects to which the researcher could not have access.

subgroups: men and women, older and younger generations, and literate and illiterate speakers. The rationale was twofold: the requirement of representativity of the sample; the integration of the time variable in contact-induced language change. The latter criterion requires some comment here. Because changes in language are in most cases gradual and take place within the time span of several generations, it is necessary for any study of language contact to plot changes diachronically as well. Ideally, only a longitudinal study over a time span of several decades would meet this requirement. Because such a study is out of the question here, the next option was to simulate time in the sample through charting the speech of individuals from different age groups. This procedure enabled us to find, for example, that older generations, usually more conservative than younger ones in their linguistic usage, prefer borrowing over code-switching. The findings were supported by historical information from earlier sources, when available, which confirmed the gradual entrance of foreign elements in the form of code switches as the intensity of contact and bilingualism increased.

The collection and analysis of corpora meeting the criteria of sociolinguistic and diachronic representativity are time-consuming tasks. In our case, the samples required between up to fifty nine speakers depending on the language and the average text length normally surpassed 1500 tokens. Accordingly, the resulting corpora differed in size, from 60.000 to 110.000 tokens.

Contact-induced language change in diachronic perspective

Linguistic borrowing as a contact-induced language change is a process and calls for a diachronic view. The process is visible in the way foreign elements are incorporated into the recipient language: from their occurrence in the idiolects of bilingual innovators to their subsequent spreading among other speakers and finally to the speech community as a whole. The process is also reflected in the gradual accommodation of foreign elements when used over a longer period of time: from their non-assimilation at the phonological level to their full integration into the phonological system of the recipient language. In the absence of similar corpora for previous stages of the language, the process can be mapped, to a certain extent, either by recording the speech of individual speakers from different age groups, as argued above, or by studying earlier sources in the form of grammars and dictionaries. For the languages of the sample there exist no pre-contact corpora that serve as a yardstick for comparison. Instead, we have a series of linguistic descriptions prepared since the early years of the Spanish conquest by members of the clergy for evangelization purposes. The availability of grammatical descriptions is not the same for the three languages, however. Fray Pedro Cáceres and Alonso Urbano wrote the first grammatical description and the first dictionary of Otomí in 1580 in 1605, respectively. Fray Ruiz de Montoya published a grammar and

dictionary of Guaraní only in 1640. The sources appeared even later for Ecuadorian Quichua: the first grammatical sketch was published only in the mid-eighteenth century (Anonymous 1760) while the first dictionary came up a few years later (Velasco 1787). All these sources were used as a reference for earlier stages of the language and served to keep track of early borrowings in the languages.

A theoretical framework providing analytic tools and testable hypotheses

As indicated at the beginning of the chapter, most studies of linguistic borrowing do not base their analysis of data on a specific theoretical framework. Their interpretation of data is obscured by the use and abuse of *ad hoc* linguistic explanations without previous hypotheses about the expected number, type and use of borrowings. The present study seeks to fill this gap by working within the theoretical framework of linguistic typology and sociolinguistics and avail of their respective analytic tools and hypotheses. The concepts and implications of the theoretical framework were amply discussed in Chapter 3. In the following section I develop several hypotheses by taking as a point of departure the premise that the typology of the languages in contact co-determines the degree and the form of lexical and grammatical borrowing.

4.3 Research questions and general hypotheses

On the basis of the model of contact-induced language change developed in section 2.6 I assume that nonlinguistic motivations are the primary cause of linguistic borrowing in any contact situation and the outcomes of contact thus motivated are modeled by linguistic and nonlinguistic factors and conditions. Theoretically, any contact-induced change is possible provided that a number of nonlinguistic circumstances are met. Still, research on language contact shows that not all possible changes are attested and that the outcomes of contact are regular and systematic to a great extent. Therefore, the central question to be answered is how regularities in contact-induced language change are influenced by structural factors derived from the typological features of the languages in contact. For this purpose I investigate the number, type and functional adaptation of Spanish lexical and grammatical borrowings were investigated in three typologically different Amerindian languages (Guaraní, Quichua and Otomí). The research questions may be detailed as follows:

- Do linguistic factors play a role in the borrowing process of Spanish elements into Guaraní, Quichua and Otomí, and if so, to what extent and under what conditions?
- More specifically, do the typological profiles of these languages play a role in the borrowing process, and if so, to what extent and under what conditions?

- And even more specifically, do the lexical and grammatical categories of these languages play a role in the borrowing process, and if so, to what extent and under what conditions?

In other words, the main goal is to identify the linguistic factors that promote or inhibit borrowing of certain lexical and grammatical categories and the linguistic conditions that speed up or slow down borrowing. The influence of linguistic factors and conditions will be confronted with the influence of nonlinguistic motivations in each contact setting. Generally speaking, the interaction between nonlinguistic motivations and linguistic factors and conditions is expressed in the following terms:

Native speakers of language R (recipient) who also speak language S (source) with different degrees of proficiency are driven by nonlinguistic circumstances to incorporate a lexical or grammatical feature of their second language (S) into their native language (R). This feature either is available in language R or not. In the first case borrowed features either replace an already existing feature in language R or make it more specific. In the second case, the paradigm of features in R is either extended or adapted by the entrance of features from S. The chance for any feature from S to be borrowed by speakers of R is co-determined by nonlinguistic and linguistic factors. Linguistic factors are the typological similarity between R and S, the equivalence between word classes in R and S, the frequency of borrowed features in S, the translatability of features from S, the paradigmaticity of word classes in R, etcetera. Similarly, chances for any feature from S to be borrowed more rapidly by R are increased by the frequency of the native feature being replaced with a feature from S.

The influence of nonlinguistic motivations and linguistic factors and conditions on the outcomes of contact is expressed in the form of hierarchies. For linguistic borrowing these hierarchies represent arrangements of lexical and grammatical elements ordered according to the higher or lower probability of borrowing from one language to another. An extensive discussion of hierarchies or scales of borrowability was presented in Chapter 3. In the following I present a number of hypothesis based on: a) the Principle of Functional Explanation; b) the Principle of System Compatibility; c) the scales of borrowability; and d) the Hierarchy of Parts of Speech. Each hypothesis is based on a hierarchy of linguistic factors or elements and will be tested on the corpora of the languages.

4.3.1 Borrowing hypotheses from the Principle of Functional Explanation

From the Principle of Functional Explanation (cf. 2.6.2) the principal hierarchy of factors governing the borrowing process is H.1: pragmatic factors are the most

decisive in any contact situation, followed by semantic and formal ones. In this perspective, pragmatic and semantic factors are promoters of borrowing while formal factors act mainly as constraints.

H.1

Pragmatic factors	>	Semantic factors	>	Syntactic-Morphological-Phonological factors
-------------------	---	------------------	---	--

This hierarchy of factors is translated in more specific terms by positioning pragmatic markers on the top of the scale in relation to other linguistic elements. This is expressed in subhierarchy 1.1 below. Discourse markers include basically topic and focus markers.

H.1.1

Discourse markers		>	Other linguistic elements
Topic marker	Focus marker		...

4.3.2 Borrowing hypotheses from the Principle of System Compatibility

A second hierarchy predicts the probability that a foreign element may be borrowed easier than others. The hierarchy is based on the classification of morphemes in free and bound. The prediction states that free forms are more prone to borrowing than bound forms in a contact situation. The term ‘morpheme’ is a cover term including not only grammatical forms such as inflectional or derivational affixes but also free lexical morphemes such as nouns or adjectives. Free morphemes and bound morphemes are roughly equivalent to lexical and grammatical classes, respectively. Exceptions are function words, which are free grammatical morphemes. Some clitics also share characteristics of free and bound morphemes and therefore occupy an intermediate position in the hierarchy.

H.2

Free morpheme	>	Clitic	>	Bound morpheme
---------------	---	--------	---	----------------

Because languages have different morphological profiles, this hierarchy is insufficient to account for all cases of linguistic borrowing. It is therefore necessary to include the morphological type of the languages in contact on the basis of the Principle of System Compatibility. According to this principle, “any form or form-meaning set is borrowable from a donor language if it conforms to the morphological possibilities of the recipient language with regard to morphological structure” (Field 2002: 42). In other words, if the recipient language is fusional-

synthetic, it may borrow virtually any foreign element, including free forms such as independent words and bound forms such as roots, agglutinating and fusional affixes. If the language is isolating-analytic, it may borrow only free forms while most bound forms (roots and affixes) are not borrowable in principle. In these terms subhierarchy H.2.1 below makes the predictions of H.2 more specific by establishing which languages have fewer difficulties in borrowing elements from another language on the basis of their morphological type:

H.2.1

Fusional-synthetic	>	Agglutinating-synthetic	>	Isolating-analytic
--------------------	---	-------------------------	---	--------------------

4.3.3 Borrowing Hypotheses from the Scales of Borrowability

The split between lexicon and grammar is the basis of a third hierarchy. This is at the same level of H.1, and both are considered to interact with each other. The hierarchy orders lexical and grammatical elements according to their degree of borrowability. From the borrowing scales discussed in Chapter 3, lexical elements are represented as more borrowable than grammatical elements and occupy the first place:

H.3

Lexical elements	>	Grammatical elements
------------------	---	----------------------

An extension of this hierarchy concerns class type. Classes are grouped in open, half-open and closed depending on the number of elements and the possibility that other (foreign) elements are incorporated. The incorporation of borrowings to closed classes is more difficult than their incorporation to half-open and open classes. In general terms open classes and closed classes correspond to lexical and grammatical elements, respectively. Half-open classes are a matter of content because they are halfway between lexicon and grammar. The open-closed constraint on classes is known as paradigmaticity (cf. 2.6.2.2). This is summarized in the following subhierarchy:

H.3.1

Open class	>	Half-open class	>	Closed class
------------	---	-----------------	---	--------------

A further subhypothesis extends the predictive value of H.3 by specifying the degree of borrowability of lexical and grammatical classes. Subhypothesis H.3.2 derives from the scales of borrowability discussed in section 3.5. To visualize the relation

between the central hypothesis and its subhypotheses they have been conflated into one single hierarchy with subhierarchies in descending order:

H.3.2

Lexical elements		>	Grammatical elements	
Open class	>	Half-open class	>	Closed class
Noun > Verb > Adjective > Adverb		>	...Adpositions...	> ...Auxiliary >...> Article

The basic prediction from H.3.2 is that elements of open lexical classes are more borrowable than elements of closed grammatical classes. The subhierarchy has two interpretations depending on the language: a) from the perspective of the source language it implies that elements of this language which correspond to major parts of speech are in general easier to borrow than elements which correspond to other parts of speech; b) from the perspective of the recipient language it implies that the parts of speech of this language determine the borrowability of a foreign element – which may or may not correspond to an equivalent class in the source language – depending on the openness or closedness of the target class. Both interpretations derive from the hypothesis that the systems of parts of speech of the languages involved in the borrowing process are relevant to determine the type of borrowings. I consider the second interpretation more relevant to the analysis pursued here because it determines not only the possibility for a foreign element to enter in a certain class but also how this element is used in the recipient language. It is clear that the typological profile of both languages in contact is relevant, but that of the recipient language is decisive.

4.3.4 Borrowing hypotheses from the Parts-of-Speech Theory

The theory of parts of speech developed by Hengeveld (1992) and Hengeveld *et al* (2004) defines parts of speech primarily on syntactic grounds and considers the phrase as the basic syntactic unit. Phrases can be referential (noun phrase) or predicational (verb phrase). Each phrase is composed of two slots, one for heads and one for modifiers. Along these parameters the theory establishes the existence of three types of languages: flexible languages, with one lexical class for two or more syntactic functions; differentiated languages, with one lexical class for one syntactic function; and rigid languages, with no lexical classes for one or more syntactic slots and morphosyntactic strategies used instead.

This classification of languages according to parts of speech has an important consequence for lexical borrowing: the types of lexical items that may be borrowed in a given contact situation depend, among other things, on the flexibility or rigidity of the languages in contact. Accordingly, the general hypothesis is formulated as follows:

- H.4** *Contiguous borrowing hypothesis:* the typological distance between the source language and the target language is bridged in the order given by the parts-of-speech hierarchy.

This hypothesis is further specified in three subhypotheses making specific predictions about the order and frequency of lexical classes in the borrowing process:

- H.4.1.** The more to the left a lexical class, the greater the number of lexemes borrowed: heads are borrowed more often than modifiers and modifiers of referential phrases more often than modifiers of predicate phrases.

From their leftmost position in the hierarchy, verbs make the largest class of expected borrowings in any contact situation, followed by nouns, adjectives and manner adverbs. Notice there is a crucial difference between this prediction and the one based on hypothesis H.3.2 where nouns occupy the first place in the hierarchy. The great majority of scales of borrowability have nouns as the most borrowable lexical class of all. The privileged position of verbs in Hengeveld's hierarchy is explained by the predication-oriented approach of Functional Grammar (Dik 1998) and the fact that the class of verbs is present in all parts-of-speech systems except for contentive type-1 languages (cf. Table 3.2). Only the analysis of data will demonstrate which prediction holds true.

- H.4.2.** Languages that borrow lexemes from one lexical class of the source language also borrow lexemes from previous lexical classes of the hierarchy. Therefore, a language that borrows modifiers of referential phrases borrows heads of referential and predicate phrases as well but not necessarily modifiers of predicate phrases.
- H.4.3.** Flexible and rigid languages borrow a larger number of lexemes from the lexical class immediately after the last differentiated class attested in their own system. Languages which distinguish verbs and non-verbs (Type 2) will borrow nouns in larger numbers because these are the lexical class that follows the last differentiated class of their system (verbs).

Hypothesis H.4 establishes the general conditions for lexical borrowing on the basis of a typological parameter, namely, the system of parts of speech. It does not exclude borrowings from lexical classes outside the parts-of-speech system of the recipient language, provided they belong to contiguous classes. It constrains lexical borrowing in terms of contiguity of classes while allowing different uses of borrowed lexemes in the recipient language. The predictions from hypothesis H.4

and its subhypotheses can be tested only across typologically different languages since the parameters for comparison are not absolute but relative with respect to the recipient language. The borrowing data from three typologically different languages makes such comparison feasible in this study.

Additional hypotheses can be tested for specific language pairs. Because the three contact situations of this study include Spanish, the following hypotheses concern only language pairs in which the source is a differentiated language (like Spanish) and the targets are flexible (Guaraní, Quichua) or rigid (Otomí).⁶ Hypothesis H.4 establishes the general parameters for lexical borrowing while the following hypotheses define the parameters of incorporation of borrowed items into the parts-of-speech systems of the recipient languages. Therefore, they concern use and function rather than order and frequency.

H.5. *Functional adaptation hypothesis:* the syntactic distribution of borrowed lexemes in the recipient language follows the distribution of lexical classes in this language. For instance, a flexible language of type 3 that borrows adjectives may use them as modifiers of referential and predicate phrases because it has a general class of modifiers performing both syntactic functions. This prediction is relevant not only for languages with flexible and rigid systems but also for languages with differentiated systems. In any case, the typological configuration of the recipient language does not undergo changes.

⁶ A slightly different interpretation of Hengeveld's hierarchy is provided by Bakker *et al* (2008) in the following terms: "We expect a target language T of type 1-4, which has all four syntactic positions available, to borrow all four types N, V, A and MAdv from a source language S without much constraint. When T is more flexible than S, there are two possibilities: functional adaptation or functional specialization. According to the first, more liberal hypothesis, borrowed elements will be treated as if they belonged to the lexicon of T: e.g., in a type-2 language, borrowed adjectives from a type 3-5 language may be used as heads of referential phrases apart from being used as modifiers. According to the second, less liberal hypothesis, borrowed elements will figure only in their original function. If T is less flexible than S then we only expect specialization among the borrowed elements in the relevant area, e.g. in a type-3 language some [Verb, Non-verb] elements borrowed from a type-2 source language may be used exclusively as heads and others exclusively as modifiers of referential phrases. On the rigid side of the scale, i.e. T languages of types 5-7, we expect to find low numbers of elements from an S language which have an 'unknown' part of speech, and specialization for elements which are borrowed into one of their original classes. E.g. a type-6 language will in principle not borrow a [A, MAdv] element from a type-4 language, and it will borrow [Verb, Non-verb] elements from a type-2 language only in the function of heads of predicate and referential phrases" (Bakker *et al* 2008).

Figure 4.1 Example of functional adaptation

DIFF-4: SOURCE	VERB	NOUN	ADJECTIVE	ADVERB
FLEX-3: TARGET	VERB	NOUN	MODIFIER	
SYNTACTIC SLOT	HEAD PREDICATE PHRASE	HEAD REFERENTIAL PHRASE	MODIFIER REFERENTIAL PHRASE	MODIFIER PREDICATE PHRASE
FLEX-3: TARGET	VERB	NOUN	MODIFIER	

- H.6.** *Functional specialization hypothesis:* borrowed lexemes fill only the syntactic slots they occupied in the source language, i.e. the borrowings from one lexical class specialize in their original syntactic slots. Accordingly, borrowed adjectives and adverbs are used only in their respective positions of modifiers of referential and predicate phrases but not interchangeably as if they formed one lexical class. This hypothesis applies only to flexible recipient languages – which become gradually differentiated in the process and undergo a typological change by which new lexical classes emerge.

Figure 4.2 Example of functional specialization

DIFF-4: SOURCE	VERB	NOUN	ADJECTIVE	ADVERB
FLEX 2: TARGET	VERB	NO-VERB		
SYNTACTIC SLOT	HEAD PREDICATE PHRASE	HEAD REFERENTIAL PHRASE	MODIFIER REFERENTIAL PHRASE	MODIFIER PREDICATE PHRASE
FLEX 2: TARGET		NOUN	ADJECTIVE	ADVERB

- H.7.** *Lexicalization hypothesis:* borrowed lexemes fill empty slots in the parts-of-speech system of the target language. If a language does not have a lexical class to fill a syntactic slot, borrowings of a lexical class come to fill that slot. This hypothesis is applicable only to rigid recipient

languages – which become gradually differentiated in the process and undergo a typological change by which former morphosyntactic strategies are replaced by new lexical items.

Figure 4.3 Example of lexicalization

DIFF 4: SOURCE	VERB	NOUN	ADJECTIVE	ADVERB
RIG 6: TARGET	VERB	NOUN	---	---
SYNTACTIC SLOT	HEAD PREDICATE PHRASE	HEAD REFERENTIAL PHRASE	MODIFIER REFERENTIAL PHRASE	MODIFIER PREDICATE PHRASE
RIG 6: TARGET	VERB	NOUN	ADJECTIVE	ADVERB

Hypotheses 5-7 have to do with the use of borrowed lexemes in the recipient language. They imply and exclude each other in several ways. Hypotheses 6-7 are applicable only if hypothesis 5 is not fulfilled, because the former imply a change (lexical differentiation) in the parts-of-speech system of the recipient language while the latter implies no modification in this system. The foregoing hypotheses must be specified for each of the recipient languages of this study according to their typological characteristics. Likewise, it is necessary to specify the typology of the source language. This is done in Chapter 5, where I discuss the historical, sociolinguistic and typological aspects of Spanish.

4.3.5 Borrowing hypotheses from language typology

There is a number of hypotheses from the linguistic typology which concerns contact-induced language change and linguistic borrowing. While many typological parameters have been discussed in the literature on contact, only some of them have proved universal to be applicable to any language and, by extension, to any languages in contact. The fact that typological parameters are the hardest to change under normal circumstances leads to the following general hypothesis and subhypotheses formulated elsewhere (Bakker *et al* 2008):

- H.8 The longer a typological parameter takes to change without a strong external pressure, the longer it takes to change with such a pressure.

This hypothesis predicts that if a typological parameter takes a longer period of time to change in non-contact situations, the same parameter will take longer to change in contact situations. The distribution of lexical classes in syntactic slots is one of such parameters. This means that the typological shift predicted by hypotheses H.6 and H.7 is unlikely to occur over a short period of time. Further subhypotheses derived from H.8 concern constituent order, word order patterns and analyticity.

- H.8.1 Borrowed elements agree with the morphosyntax of the recipient language and are easier to borrow when their basic syntactic position in terms of Head-Modifier relations in the recipient language is the same as in the source language. Thus, adpositions are borrowed in their original syntactic position, if this is available in the syntactic matrix of the recipient language. Therefore, a postpositional language will not borrow prepositions from a prepositional language.
- H.8.2 The frequency of the constituent order patterns in the recipient language may change in the direction of orders attested in the source language. This does not mean that new orders are introduced but those already existing gain in importance with respect to others. For example, a VSO language in contact with a SVO language may change its basic order to SVO (an unmarked alternative order) but not to SOV (a marked alternative order).⁷

The contact-induced syntactic change predicted by H.8.2 takes place when previously marked word orders in a language become unmarked as a result of the calquing of similar syntactic patterns from another language. Therefore a markedness shift results from the high frequency of once marked patterns in the recipient language (Campbell 1995: 136ff).

- H.8.3 Languages tend to borrow elements which express an already existing function in more analytic terms. Thus, if the recipient language marks a possessive relation by inflectional affixes, it may borrow an adposition that expresses possession in the source language.

This subhypothesis predicts a gradual shift from synthesis to analyticity as a result of contact. However, this shift will take longer to occur according to H.8. Evidence for a shift to analyticity is found in two of the languages of the sample. Synthetic and analytic strategies may coexist in one and the same language in the form of double marking, and their coexistence indicates that the language is halfway in the shift to

⁷ Siewierska (1998:493) shows on the basis of a sample of 171 languages that 63% of the VSO languages have SVO as an alternative order and only 13% have SOV as an alternative.

analyticity. Although there are cases of change in the opposite direction, i.e. from analyticity to synthesis,⁸ the borrowing of grammatical forms is more unlikely to happen according to H.2 and H.3.

While the hypotheses of linguistic typology are tested in the concluding chapter, the other hypotheses are tested in Chapters 10 and 11 on the statistics produced from the analysis of borrowing in the corpus of each language.

4.4 The Methodology

The first part of this section describes the process of data collection and the setup of the corpora. The selection of informants and the fieldwork conditions are discussed for each language. The second part describe all the steps in the processing of data and the conventions used for the identification of borrowings and functions in the recipient language. The last part discusses several problems encountered in the fieldwork and the analysis of data as well as the ways in which these problems were dealt with.

4.4.1 Data collection

The present investigation of constraints on linguistic borrowing is based on the analysis of corpora of spontaneous speech collected for Guaraní, Quichua and Otomí. The setup of individual corpora that meet the requirements of representativity and naturalistic speech was crucial to obtaining high-quality material for the analysis.

The data for Quichua and Guaraní were collected by the author during several fieldwork visits to Ecuador and Paraguay between 2004 and 2006. The data for Otomí are part of a larger corpus collected over the last decade by Ewald Hekking. For the three languages the same guidelines were followed in data collection with a view to obtaining comparable corpora with similar format.

The collection of data favored spontaneous speech over monitored registers such as those produced through elicitation or writing. Spontaneous speech was recorded in social contexts different from those of sociological surveys and ethnographic interviews. Recording events usually involved several participants in domestic settings, with one leading speaker and other taking the floor to make comments. However, only the contributions of the main speaker were considered for analysis. The reason for recording speech in these settings was to encourage the spontaneous participation of speakers and increase their involvement with their narratives. The duration of the recording sessions varied in each case from a quarter

⁸ One of the most striking examples is Sri Lanka Malay, which has evolved from analytic to synthetic as a result of contact with Tamil. Cf. Adelaar 2003.

of an hour to one or two hours in some cases. Before each recording session the author registered all contextual and biographical information relevant for the interpretation of data. This was done on a field notebook according to accepted ethnographic standards. The sociolinguistic information collected during the researcher's stay in the communities was registered in the same way. These data served to contextualize the speech events recorded and the process of data collection as a whole. Because a questionnaire was not applied, speakers set their own narrative agendas and covered a wide range of topics in the same session. Some issues turned up repeatedly, however, depending on the language and the social context. For Imbabura Quichua speakers, for example, the topic of interethnic relations between Indians and Mestizos and their work for the local *haciendas* were recurrent topics. For Guaraní speakers the favorite topics were the local usage of Guaraní and its importance for the Paraguayan identity. The variety of topics enabled the author to identify Spanish borrowings according to semantic fields.

The fieldwork in Ecuador was carried out during two visits of several months in 2004 and 2005. The geographical space covered during both stays corresponds to the Quichua-speaking areas of Imbabura in the Northern Highlands and Bolívar in the Central Highlands. These were selected for linguistic and sociolinguistic reasons. On the one hand, both varieties, though mutually intelligible, show a number of differences at the levels of lexicon and grammar, most probably due to different Pre-Inca substrata (Caranqui in Imbabura; Puruhá in Bolívar). On the other hand, both Imbabura and Bolívar show large numbers of Quichua speakers as compared to other districts of the Highlands. However, the vitality of Quichua seems stronger in Imbabura and their speakers show more positive attitudes towards their language than speakers of Bolívar Quichua. Moreover, Imbabura Quichua speakers show their higher levels of bilingualism due to their literacy and more active participation in the national society. Still, both dialects are representative of Highland Ecuadorian Quichua in linguistic and sociolinguistic terms. The total number of communities investigated in Imbabura and Bolívar is twelve.⁹

The fieldwork for Guaraní was carried out both in rural and urban areas, because the language is spoken by the great majority of the population all over the country. Accordingly, one part of the corpus includes language data collected in the cities of Asunción (capital), Encarnación (Itapúa District) on the Argentinean border, and Pedro Juan Caballero (Amambay District) and Ciudad del Este (Alto Paraná District), both on the Brazilian border. The second part of the corpus was collected in small towns and villages in the districts of Cordillera, Alto Paraná, Misiones, Caaguazú and Paraguarí.¹⁰ The selection was motivated by the fact that

⁹ A list with speakers per province, community and other sociolinguistic information is included in the Appendices.

¹⁰ See list of speakers of Paraguayan Guaraní in the Appendices.

differences between conservative and innovative dialects often correspond to rural and urban milieus. In general, urban speakers show higher levels of bilingualism and literacy and their sociolects are more innovative (preference of mixing strategies such as borrowing and codeswitching). On the other hand, rural speakers show lower levels of literacy and bilingualism and the influence of Spanish on their speech is also low. The distribution of linguistic borrowing and codeswitching according to urban and rural areas is discussed in Chapters 10 and 11.

The data for Otomí were collected by Ewald Hekking in several periods of fieldwork between 1993 (cf. Hekking 1995) and 2004. The Otomí corpus includes data of two dialects spoken in the state of Querétaro on the central plateau of Mexico. The dialects belong to the branch of north-western Otomí and are spoken in the town of Santiago Mexquititlán and the village of San Miguel de Tolimán, at a distance of two hours by car from the district capital, Querétaro. Santiago Mexquititlán is located in the southern part of the state of Querétaro, with a population of approximately 15.000 inhabitants, most of whom are Otomí speakers. San Miguel de Tolimán is located in the northern part of the state, with a population of around 700 inhabitants. Both dialects are mutually intelligible but there is not regular contact between their speech communities. In both cases, however, the native language is losing ground to Spanish.

Considering the fragmentation within the Otomí branch of Otopamean languages and their ongoing process of dialectalization, it was necessary to restrict the corpus to one Otomí-speaking region. The dialects of Santiago and Tolimán are representative of the Otomí varieties spoken on the central plateau (e.g. Mezquital Otomí). The size of the Otomí-speaking population in Santiago Mexquititlán and Tolimán is an important factor of dialect differentiation and influence on the vitality of the native language vis-à-vis Spanish.

For Otomí, additional corpora were available from the work by Hekking on other aspects of the language. His collection of picture-elicited stories in particular served to outline sociolectal differences in an important number of speakers. In addition, Hekking has collected a large corpus of local Spanish spoken by Otomí native speakers. While not used in the analysis, this corpus helped to identify the distinct levels of bilingualism in the communities under study. Hekking and Bakker (2005) have published the results of a study of shift-induced interference from Otomí in the local Spanish of Santiago Mexquititlán following the same format of the present investigation.

The fieldwork activities described above resulted in individual corpora for the three languages. The data were collected from a representative sample of speakers

following several sociolinguistic criteria.¹¹ First, the corpora include texts produced by men and women, even though it was not always possible to balance the number of contributions by gender in every community. Similarly, the data for each language correspond to speakers whose ages range between 20 and 75 years at the time of recording, roughly distributed in three age groups: one group of younger speakers (20-40); another group of medium-age speakers (40-60); and one group of older speakers (>60). The age variable helped to identify differences in borrowing strategies and the accommodation of borrowings to the recipient language.

The variables of gender and age are complemented by those of literacy and bilingualism. In general, speakers were grouped into those with a higher degree of bilingualism and those with a lower degree. An accurate classification of speakers according to their proficiency in Spanish was not feasible however, because of the lack of Spanish samples for every speaker. The introductory data accompanying each speaker file and the sociolinguistic information from the fieldwork notes filled the gap to some extent. In association with bilingualism, literacy proved to be a useful criterion to establish the influence of Spanish on the native language. Speakers were classified as either literate (those who read and write in Spanish, in the native language or both) or illiterate. More often than not, differences across speakers were observed at the level of schooling (e.g. some completed the elementary school while others did not). Further information collected for as many speakers as possible was their community of origin, if different from the current place of residence.

Information for all of these variables is available for the great majority of speakers of the samples. Additional information about work, migration, time of residence and age of second language acquisition/learning is available only for two languages (Otomí and Quechua) but not for every speaker. The fragmentary nature of these data prevented us from using them in the analysis.

4.4.2 Data processing

The steps followed in the processing of data include the parsing of borrowings and code switches, the assignment of borrowings to lexical classes in the source language, and the assignment of borrowings to syntactic functions in the recipient language. These steps are described individually in the following.

The speech recorded in individual sessions was transcribed with the help of native speakers and digitalized through a word processor. The resulting transcripts were individual texts in computer-readable form for each speaker. The transcription

¹¹ The full list of the speakers who contributed to each corpus is included in the Appendices along with relevant information about their dialect, community of origin, age, education and level of bilingualism.

was phonological and paid attention only to distinctive features. A system of standardized spelling was used in each case. The selection of the spelling system was not unproblematic however, because of their lack of phonological accuracy and the existence of two or more competing systems. Only noticeable dialectal differences were coded in transcription. Because Spanish borrowings are differently pronounced by speakers according to their levels of bilingualism and the degree of phonetic accommodation to the recipient language, differences in pronunciation were registered with a phonetic orthography¹². This procedure helped to associate phonetic differences with higher or lower levels of bilingualism and to distinguish borrowings from code switches in many cases. The following are two Spanish loans pronounced with different degrees of integration by speakers of Quichua and Otomí:

Table 4.1 Levels of integration and spelling differences

Sp. <i>cosecha</i> ‘harvest’	[kuziça]	(Quichua)	high integration
		[kuzeça]	↓
		[kozeça]	
		[koseça]	low integration
Sp. <i>vecino</i> ‘neighbor’	[bisinu]	(Otomí)	high integration
		[bisino]	↓
		[besinu]	
		[besino]	low integration

Because one and the same loanword may show different levels of phonetic accommodation depending on the speaker, differences were coded only if clearly divergent from the standard Spanish pronunciation. Accordingly, integrated loanwords such as [kuziça] and [bisinu] were coded as {kuzicha} and {bisinu}¹³ while others like [kozeça] or [besinu] were coded with the standard spelling due to their phonetic similarity to native Spanish pronunciation. A large number of non-integrated loanwords were either nonce borrowings or single-item code switches.

Each transcription was accompanied with relevant information about the speaker, the place and date of recording, the language, the person who recorded the session, and, in some cases, the topic or topics of the session. The standard structure of the resulting files for each speaker is as follows:

¹² Notice that the phonetic realization of a word may be different across Spanish dialects and is definitely distinct in some cases from the pronunciation in Peninsular Spanish.

¹³ Because it is a phonetic transcription, the grapheme {z} corresponds to the voiced sibilant sound.

Table 4.2 Structure of data file

IDENT OF INFORMANT:
 GENDER:
 AGE:
 LOCATION:
 EDUCATION:
 < further social parameters >
 TARGET LANGUAGE:
 SOURCE LANGUAGE:
 RECORDED BY:
 TOPIC:
 \$
 < text >

In the next step Spanish borrowings were identified between slashes, and code switches between square brackets. With this procedure the morphological material of the recipient language was separated from the borrowings, and lexical borrowings were isolated from native bound phonemes. The following example (1a) is extracted from a text produced by a speaker of Paraguayan Guaraní. Spanish borrowings and code switches are indicated in slashes and square brackets, respectively. The gloss and free translation is given in (1b).

- (1a) /maestro/kuéra nombo'eséi pe Guaraní ha sy ha tuvakuéra ai oimo'ã /que/ pe imemby péicha oñepyrũ Guaraníme ha'e ij/atrasado/taha, ha'e noiporã mo'ãi [en el nivel de los otros]

- (1b)
- | | | | |
|----------------------|-----------------------|-----------|----------------|
| <i>maestro-kuéra</i> | <i>n-o-mbo'e-sé-i</i> | <i>pe</i> | <i>Guaraní</i> |
| teacher-PL | NEG-3-teach-VOL-NEG | DEM | Guaraní |
-
- | | | | | | | |
|-----------|-----------|-----------|-------------------|----------------|------------|-----------|
| <i>ha</i> | <i>sy</i> | <i>ha</i> | <i>tuva-kuéra</i> | <i>oi-mo'ã</i> | <i>que</i> | <i>pe</i> |
| and | mother | and | father-PL | 3.think | that | DEM |
-
- | | | | | |
|----------------|---------------|-----------------|-------------------|-------------|
| <i>i-memby</i> | <i>péicha</i> | <i>o-ñepyrũ</i> | <i>Guaraní-me</i> | <i>ha'e</i> |
| 3-child | thus | 3-begin | Guaraní-LOC | 3S |
-
- | | | | |
|--------------------------|-------------|------------------|---------------|
| <i>ij-atrasado-ta-ha</i> | <i>ha'e</i> | <i>no-i-porã</i> | <i>mo'ã-i</i> |
| 3-stay.behind-INCH-REL | 3S | NEG-3-good | COND-NEG |
- [en el nivel de los otros]
 [at the level of the others]

‘Teachers do not want to teach Guaraní, and parents think that if their children begin to learn Guaraní, they will stay behind and will not be at the same level of other children’

Code switches were identified in the transcriptions in order to separate them from borrowings. Code-switches are excluded from the analysis by default. However, their total number and average size in each corpus were determined so as to identify their contribution to discourse. Code switches such as the bracketed stretch in (1a) are distinguished from complex lexical borrowings in which two or more constituents make up a frozen expression.

Parsed texts were entered into a computer program developed especially for this purpose by Dik Bakker. The program counts all the borrowed tokens (separate word forms) and groups them into types (different forms).¹⁴ The resulting tables contain all the native and foreign words in a text as well as their absolute and relative frequencies. Table 4.3 is part of a larger table produced for a Quichua speaker (AC) who produced a text with a length of 2740 tokens and 32 code switches. There are 642 borrowings (23,43%) grouped in 406 different types (25.39%).

Table 4.3 Table of frequencies for one informant (fragment)

Frequencies for informant: AC

TARGET (Quechua)		SOURCE (Spanish)	
-ca s_Q	35	-hora- q_q	1
-ca s_S	15	-horas- q_q	1
-ca q_S	2	-lado- q_q	2
-ca q_Q	1	a S_S	2
-camacuna s_S	1	administrador- Q_q	1
-caman s_Q	1	agua S_S	2
-chishcanchicho s_Q	1	agua Q_Q	2
-chu s_Q	1	agua- Q_q	2
-cpi s_S	1	aguanta- Q_q	1
...			
Types	1193 (74.61%)	406 (25.39%)	1599
Tokens	2098 (76.57%)	642 (23.43%)	2740
TTR	1.759	1.581	1.714

¹⁴ The notion of type as it is used here refers to the form without morphological or semantic considerations of complexity or homonymy. Therefore, no lemmatization has taken place. This has a great influence on the interpretation of this measure, especially when comparing languages of different morphological types.

Q_Q Typ	693	79
Q_Q Tok	1175	115
Q_q Typ	228	225
Q_q Tok	388	406
S_Q Typ	89	25
S_Q Tok	107	28
S_q Typ	33	74
S_q Tok	38	89
...		

Similar tables were produced for each speaker and one general table for the whole corpus of each language. Table 4.3 presents native forms on the left column and borrowed forms on the right. The numbers on the right hand side of each column correspond to the number of times one form occurs in the text (tokens). While one single form may have several correspondences according to the morphological environment, these were not considered as distinct types. The analysis of the morphological environment helped to identify: a) the word class of the borrowed form; b) the influence of the native morphology on the distribution of borrowed forms in individual word classes; and c) the morphological integration of borrowed forms into the recipient language. In notation, upper case letters (Q, G, O, and S) stand for free forms and lower case letters (q, g, o, and s) for bound forms; both are separated by a low hyphen representing the borrowing or the native form in question. The order of native and borrowed forms is alphabetical. A summary of frequencies is given at the end of the list, including the numbers and percentages of types and tokens for borrowed and native forms as well as the type-token ratio.¹⁵ Tables also include the number and size of code switches and the percentages of these forms in the overall texts. The frequencies of the morphological environments of loanwords and native forms are provided as well. The table of frequencies for the overall corpus of a language corpus includes all the above-mentioned data plus the number of speakers who have used a certain type.

The program was designed in such a way that Spanish borrowings are added to a dictionary. Any new form in the input that matches a borrowing type in the dictionary is automatically parsed in slashes and allocated to its corresponding

¹⁵ The program provides other statistics such as TTR2, the type-token ratio where the number of types is divided by the square root of the number of tokens. This figure is somewhat more stable and less dependent on the actual length of the text (cf. van Hout and Muysken 1994).

type¹⁶. In this way the dictionary can register the number of tokens of one type across the texts but also the speaker frequency of types.

In addition to automatic parsing and type assignment, the dictionary can solve problems originated in the use of alternative spellings for borrowed forms in the source language, when these spellings were used to mark idiolectal variations. Alternative forms are specified in the dictionary as equivalent to one existing type, as shown in (2) for Spanish *cosecha* ‘harvest’. In this way the dictionary can take three orthographically deviant forms as belonging to the same type and prevents the doubling of lemmas.

- (2) *cosecha* ‘harvest’
kuzicha > cosecha
kuzecha > cosecha
kozecha > cosecha

Because the goal was not only to identify the number of Spanish borrowings in the samples but also to group them in lexical classes and identify the functions they perform in the recipient language, the final step in the processing of data was the enrichment of texts with information about the syntactic position of the borrowed tokens in the recipient language.

The codification of parts of speech and syntactic functions was done by adding two codes on the right hand side of each borrowed form: one for the part of speech of the borrowed form in Spanish, and other for the syntactic slot the borrowed form occupies in the recipient language¹⁷. After this procedure, the stretch illustrated in (1a) looks as (3).

- (3) */maestroNHR/kuéra nombo'eséi pe Guaraní ha sy ha tuvakuéra ai oimo'ã /queC/ pe imemby péicha oñepyrũ Guaraníme ha'e ij/atrasadoAHP/taha, ha'e noiporã mo'ãi [en el nivel de los otros]*

In (3) Spanish *maestro* ‘teacher’ is coded as a noun (N) that functions as the head of a referential phrase (HR). The Spanish conjunction *que* is coded as (C) without further specification because of its functional equivalence in Guaraní. Finally, Spanish *atrasado* ‘stayed behind’ is coded as an adjective (A) that functions as the head of a predicate phrase (HP). While the loan noun (N) is used in its original

¹⁶ The device should be applied with caution because lexical items may be ambiguous and no device can make decisions with full certainty. In other words, the results must be checked on hindsight in each case. The pattern-match procedure scores around 90-95% security for Otomí but considerably lower for Quichua. This may be due to the amount of overlap between the two phonological systems involved, to the extent they are reflected in the spelling systems.

¹⁷ A list with all the codes used for labeling loanwords is given in the Appendices.

syntactic position (head of a referential phrase), the loan adjective (A) is used in a position not typical of Spanish (head of a predicate phrase). The codes are contained in the dictionary and assigned automatically by default unless otherwise specified in the text. With this information, the dictionary entries for the borrowings identified in (3) are describes as follows:

- (4) atrasado A (MR, HR, HP)
- (5) maestro N (HR, MR)
- (6) que C (C)

The entries contain a) the standard form plus other alternative forms, if such is the case; b) the part of speech to which the borrowing corresponds in Spanish (Adjective, Noun, Conjunction, etcetera); and c) the syntactic slot(s) filled by the borrowing in the recipient language (for the case of *atrasado*, these positions are three but only one for *que*). The program produces totals per part of speech and syntactic function for the individual speaker and the whole corpus of the language. Table 4.4 shows a fragment from the survey of parts of speech for Spanish borrowings in the corpus of Imbabura Quichua. Only the lexical class of nouns is given for illustration.

Table 4.4. Total for parts of speech in Imbabura Quichua (fragment)

PoS	SynFnc	TYP	TOK	NSP
N	HR	515	990	9
N	HR, MR	36	174	6
N	MR	25	27	5
N	MP	8	11	6
N	HR, MP	4	45	4
N	HP, HR	2	5	2
N	TOTAL	590	1252	

The table has five columns: the first for the part of speech, the second for the syntactic function, the third for types, the fourth for tokens and the fifth for speakers. The statistics in Table 4.4 are to be read as follows:

A total number of 1252 Spanish nouns were found for all Imbabura Quechua speakers (9). These nouns correspond to 590 different types. Of this number 515 nouns were used *only* in their prototypical function of heads of referential phrases. Further functions modifier of referential phrase (36 types), modifier of predicate phrase (4 types) and head of predicate phrase (2 types). Single functions include modifier of referential phrase (25

types) and modifier of predicate phrase (8 types). If syntactic functions are related to number of speakers, all speakers (9) used Spanish nouns in their prototypical syntactic position (head of referential phrase) while only five used them as modifiers of referential phrase and two used them as head of referential and predicate phrases simultaneously.

The foregoing description of the processing of data showed the choices I made and the methodological and theoretical criteria underlying these choices. As a whole, data processing was a collective effort which lasted over two years with the active participation of several research partners apart from the author of this study.

4.4.3 Problematic issues

The collection and processing of data in case studies of language contact often a number of difficulties. The overall theoretical scheme of this investigation, the criteria for the selection of languages and the setup of their respective corpora were instrumental in solving many of these problems.

Difficulties in data collection

The following questions challenged the collection of data from the beginning of this investigation:

- 1) How to account for variation in a representative sample?
- 2) How to obtain samples of naturalistic language in a relatively short period of time?
- 3) How to obtain samples of spontaneous speech in settings where several individuals participate at the time and a number of distracting factors internal to the communicative setting are at work?

The main problem to be tackled was the selection of one or more varieties out of a great number of dialects and sociolects spoken in different speech communities. For Paraguayan Guaraní dialectal variation was not much of a problem thanks to the large number of Paraguayan citizens who have Guaraní as their mother tongue and the long history of efforts towards standardization. Consequently, for Guaraní I speak rather of sociolects (urban and rural). Things are not as straightforward for Quichua and Otomí. In general both languages have undergone – and continue to undergo – a striking process of dialectalization whereby differentiated varieties emerge in the time span of few generations. For Highland Quichua linguists have identified at least eight different dialects (cf. Buttner 1993; Ethnologue 2005). For Otomí this number increases dramatically, with over forty dialects scattered around the plateau and the highlands. Many of these varieties have hardly been described in

linguistic terms and generally lack grammars and dictionaries. It is true that efforts have been made recently in both countries to chart dialectal variation and to standardize the spelling systems, but most have failed by a resistance from the speakers themselves. No standardized spelling is available up to date for all Otomí dialects and the spelling unification for dialects of Highland Quichua is far from being accepted.

The dialects for this investigation are: Santiago Mexquititlán and San Andres de Tolimán for Otomí; and Imbabura and Bolivar for Quichua. The dialectal distance is different for each language. The Otomí dialects are fairly close to each other as shown by isoglosses (Lastra 2007: 46ff) while the Quichua dialects are farther apart. (cf. Chapter 6). On the other hand, the Guaraní varieties selected for this investigation are the rural and urban sociolects.

The selection of varieties in the case of Quichua and Otomí is appropriate in my opinion because of two reasons: they show a minimal degree of divergence from other dialects of the same language; and their sociolinguistic conditions are representative of speech communities in the respective areas. Other dialects do not meet these requirements and their selection would have strongly biased the results of this research. Let us take Ixtenco Otomí and Pastaza Quichua as examples. Ixtenco Otomí is geographically separated from the bulk of Otomí dialects of the central plateau and shows the highest degree of divergence from other dialects, among other things, because of a pervasive Nahuatl influence. In the same way, Pastaza Quichua is geographically isolated from the highland dialects and is influenced by a century-long contact with Amazonian languages such as Záparo and Achuar, the speakers of which use Quichua as a lingua franca for interethnic communication (cf. Trujillo 1998; Gómez Rendón 2006d).

Equally important in the collection of data was the question of how to record naturalistic language in a relatively short period of time. In the context of the present investigation I made two fieldwork stays of several months in Ecuador and Paraguay. At the time of starting fieldwork I had over six years of experience in various Highland Quichua communities, some of which participated in the study. With this background I did not have to learn the language to conduct the recording sessions myself and benefited from my previous knowledge of the field, which made longer stays unnecessary. In all the communities, however, I took some weeks to become familiar with the speakers and their families before I could record sessions in informal settings. The help I received from local assistants was invaluable in this case. They served as brokers between me and the families and contributed to reduce the disturbance caused by the presence of a non-native Quichua speaker in a setting usually reserved for communication among native speakers. While I did not cancel in this way the influence of my presence on the spontaneity of the verbal exchanges, I consider it was the closest approach to an ideal setting I could attempt.

Fieldwork in Paraguay was more time consuming. My stay in Paraguay from December 2004 to March 2005 was the first I made to this country. So I took me at least one month to get to know people and places and at least another two months to be familiar enough with the language for basic verbal communication with Guaraní speakers. Under these circumstances, the assistance from local people was necessary in the first months to get sessions recorded in accordance with the standards of the project. Local assistants, particularly in small towns and villages, were decisive for a successful collection of the data. Their full names and their communities appear in the Appendices. In each community I trained assistants in recording techniques but did not participate in the sessions. In those few communities where my stay was long enough and my level of the language enabled me to communicate with basic fluidity, I conducted sessions myself.

The fieldwork for Mexico was conducted by Ewald Hekking following the same standards. His proficiency in the language, his many-year fieldwork experience in Otomí communities, and his permanent residence in the country were helpful in the collection of naturalistic samples. In addition, Hekking has written extensively on Otomí since 1984 and participated in the process of revitalization of Otomí in close collaboration with the speech communities. In over twenty years of work on Otomí, Hekking collected a large corpus of the language, a small part of which was selected for this investigation along the lines explained in section 4.2.

A further difficulty in the collection of naturalistic language data was the recording of individual speech in collective settings. This problem was particularly relevant for Quichua. Most sessions recorded in Quichua communities involved the target speaker and his/her close relatives. While this may be considered a disturbance of the recording setting, it served to create an informal environment, foster the participation of speakers and encourage them not to monitor their speech – which usually occurs when non-native speakers are present. Local assistants participated as speakers or bystanders in most recording sessions. For the transcription of data it was therefore necessary to separate the main speaker's discourse from that of other participants. The texts analyzed contain only the contribution of the target speaker. Other contributions were included only if there were two leading speakers whose contributions cannot be separated. I am aware that this procedure simplifies the data by extracting parts of discourse from their context. However, for the sake of the present investigation, the main goal of which is to identify the influence of Spanish on the native language of the speakers and *not* analyze speech events, the procedure is fully justified. An integral transcription of the recordings could be undertaken however, in order to produce a comprehensive corpus appropriate for other types of analysis.

Difficulties in data processing

Notational issues in particular became problematic for the analysis of the data. Some of these issues were mentioned already in previous sections. I address here two of them: 1) the spelling system used for phonological transcription; and 2) the computer-based analysis versus the manual analysis of data.

Because this study does not analyze language data at the level of phonetics, a phonological transcription was considered sufficient and time-saving given the size of the corpora. The spelling system used for transcription was the standardized versions of the respective dialects. When these were not available, the choice was to use the most widespread spelling. As mentioned above, standard spellings were not always available. For Ecuadorian Quichua there is a standard spelling since 1980. The main characteristic of this spelling is the replacement of the graphemes <k> for <c> and <w> for <u>, or <g> depending on the environment. Recently the old graphemes <k> and <w> have been reintroduced, but their use is not generalized yet and many books and textbooks continue to use the old graphemes. Considering the dialectal variation in Highland Quichua, the use of either grapheme is not an arbitrary choice. For example, post velar /q/ occurs in Bolívar along with velar /k/ while only the velar realization is found in Imbabura. Neither spelling makes a difference but represents both realizations with <k>. Similarly, the grapheme <l> may be ambiguous as it corresponds to the flap [ɾ] in Bolívar and the lateral [l] in Imbabura. For the sake of uniformity throughout the corpus, I decided to use the 1980 alphabet for all the transcriptions but introduced graphemic distinctions when differences were relevant for the analysis, i.e. when they concerned Spanish loanwords accommodated to the phonological system of Quichua.

The spelling issue in Guaraní is much less problematic, albeit not entirely exempt from controversies. Since the mid nineties a standardized spelling is used for writing books and textbooks in Guaraní and taught in schools all over the country. Still, there are people, especially in the countryside, who keep using the older system. The differences between the current spelling and the old one have to do with the representation of nasals and affricates. Nasalization is represented by the dieresis <¨> in the old system while the current spelling uses the tilde <~>. While the new spelling is used by most Guaraní speakers nowadays, it is not exempt from a number of problems, some of which encourage hot debates in academic circles in Paraguay. The use of the so-called *pusó* /' / is but one example. However, it is clear that disagreements have not barred the diffusion of the new spelling system, and this situation made the transcription process more straightforward.

The spelling issue in Otomí is still more problematic. The dialectalization of Otomí varieties is the main obstacle to the accomplishment of a standardized spelling. In fact, Otomí has been written differently by speakers, linguists and

teachers (Lastra 2007: 19). In the context of this study, the solution was to use the most widespread spelling. This is the spelling system developed by Hekking for Querétaro Otomí, which is being used in the states of Hidalgo and Mexico as well. The dialects of Santiago Mexquititlán and San Miguel de Tolimán use Hekking's spelling. Besides, they are not phonetically different from each other, which facilitated the transcription process considerably. More problematic was the writing of foreign words due to the absence of a standardized form and the different degrees of phonological accommodation of loanwords according to the level of bilingualism of the speaker and the age of the loanwords. As explained above, differences in pronunciation were noted only if they were clearly deviant from the Spanish standard form. Likewise, different notations of one word were solved by introducing them in the corresponding entry of the digital dictionary. In this way the program could recognize different phonetic realizations of the same lexeme and allot them to the same entry.

A further problematic issue was the analysis of the data by hand. Although the computer program designed by Bakker was used to manage the corpora in an easier and more systematic way, an important part of the coding analysis was done by hand. During the first stages when the dictionary was under construction, borrowings were identified manually along with their lexical classes and syntactic functions in the recipient language. While these tasks became less time consuming as the dictionary was gradually enriched by new input, manual analysis was still necessary in order to identify the syntactic functions of loanwords as these cannot be assigned by default but only in the broader context of the phrase and the sentence. Prototypical functions usually predominate over non-prototypical ones, but this is not always the case. Table 4.4 shows this clearly: loan nouns in Quichua occupy three different syntactic positions (HR, HP, and MP) and these conflate with each other in three different combinations. In the end, the time spent in manual analysis, including the assignment of lexical classes and syntactic positions and the checking for consistency took about three hundred hours for the corpus of each language.

4.5 Summary

This section described the research program and the methodology of investigation of linguistic borrowing. I developed a research program on the basis of nine criteria that served as guidelines in the different stages of this study. A large part of this chapter focused on the development of hypotheses from the theoretical framework discussed in Chapter 3. The last section dealt with the methodology used in the collection and processing of data and the problematic issues faced in the process.

PART II

THE LANGUAGES

The second part of the book focuses on the sociolinguistic, historical and typological aspects of the languages of this investigation. I pay attention therefore not only to the recipient languages (Guaraní, Otomí and Quichua) but also to the source language (Spanish).

Chapter 5 presents a general overview of Spanish in the Americas and individual descriptions of the regional varieties in contact with Guaraní, Otomí and Quichua. The first section deals with: the variation and unity of Spanish, its sociolinguistic status, its origin in the expansion of the Spanish Empire, the process of dialectal leveling in the first century of colonization, the classification of dialectal areas, and the contact between Spanish and Amerindian languages. The next three sections focus on Ecuadorian Spanish, Paraguayan Spanish and Mexican Spanish, respectively. For each variety I discuss the number of speakers, identify dialectal and sociolectal differences in the areas of research, describe contact varieties, and list the main features at phonetic, morphosyntactic and lexical levels. The fifth section deals with the typological classification of Spanish according to several features relevant for the analysis of linguistic borrowing, in particular the classification of parts of speech in the terms of Hengeveld (1992) and Hengeveld *et al* (2004).

Chapters 6, 7 and 8 concentrate on Quichua, Paraguayan Guaraní, and Otomí, respectively. For the sake of comparison the structure of the three chapters is uniform and includes the following information. The first section deals with the geographical distribution of the language, the number of speakers, the vitality of the language and other sociolinguistic information. The second section addresses the history of the language. The third section describes the dialectal distribution, with a focus on the varieties selected for this study. The final section discusses the typological classification of the language according to the parameters considered for the classification of Spanish. It includes the language-specific predictions of borrowing according to the typological classification of each language.

Chapter 9 discusses the borrowing hypothesis in comparative perspective on the basis of the general hypothesis of Chapter 4 and the language-specific predictions of the previous chapters.

Chapter 5

Spanish

Spanish is spoken today by 332 million people in Latin America, Europe and Africa. This number does not include 23 million speakers in 22 countries where the language is not official¹. In geographical distribution Spanish is the fourth language after English, French and Russian.² Varieties of Spanish are counted by dozens, perhaps even hundreds according to some authors (cf. Resnick 1975)³ but intelligibility among them remains to a great extent.

In extension and number of speakers the Americas are the largest continent of those in which Spanish is spoken as a first language.⁴ Peninsular Spanish was brought to the continent by European colonizers and developed there peculiar characteristics as a result of its internal evolution and its century-long contact with indigenous languages.

5.1. Spanish in the Americas: *e pluribus unum*

Following Lapesa (1992: 269) and others (Alba 1992; Moreno de Alba 2004) I use here the expression “Spanish *in* the Americas” instead of “American Spanish”.⁵ The variation of Spanish in the Americas prevents us from qualifying it as a monolithic, indivisible entity. Notwithstanding this, all American dialects remain mutually intelligible and show in essence the same typological characteristics. This situation is underlined by Lapesa when referring to the issue of variation versus unity in Spanish: “at all levels of the language we do not find complete unity, but at all levels

¹ Jaime Otero, *La demografía de la lengua española* (1999). Online publication of *Centro Virtual Cervantes*, available at http://cvc.cervantes.es/obref/anuario/anuario_99/otero/.

² According to Otero (1999) Spanish is spoken over 11,990,000 km² equivalent to 7.2% of the world area.

³ For the Americas, Rona gives a number of 23 different dialectal varieties of Spanish (Rona 1964: 215-226). Resnick (1975) puts forward a classification along phonological parameters, according to which the number of dialects of Spanish in the continent would be as many as 276, as noticed by Canfield (1978: 170).

⁴ This distinction is relevant in my view if we consider that the number of people learning Spanish as a second language is estimated around twenty million plus another million who speak Spanish as a lingua franca in interethnic communication. Cf. Otero 1999.

⁵ “American Spanish” is the standard term in English as opposed to “Peninsular Spanish”. In Spanish, the distinction proposed here is accomplished by the use of prepositions *de* and *en*. The term *español de América* assumes an indivisible entity that was carried to the Americas and remains essentially undifferentiated. The term *español en América*, on the contrary, implies the particularities of the language as spoken in the continent and does not conceive Spanish as an indivisible unity across national boundaries.

we do find a common ground which continues to be much stronger than any particularities” (Lapesa 1966: 307; my translation).

Authors have characterized “Spanish” as a complex diasystem composed of a number of Spanish languages along diachronic, diatopic and diastratic parameters (Rona 1969; Alba 1992; Montes de Alba 2004). For Bartoš (1971: 14ff) Spanish varieties in the Americas should be considered national languages enclosed in the boundaries of nation-states. According to this view, it is just a matter of qualifying the term “Spanish” with the patronymics of each country in order to obtain such aggregates as Mexican Spanish, Paraguayan Spanish and the like. This classification of Spanish dialects is a common practice among specialists and non-specialists. Still, evidence demonstrates that national boundaries do not necessarily match linguistic ones. Border varieties illustrate this situation clearly. The Spanish of Chiapas (southern Mexico) and the Spanish of northern Guatemalan qualify as one single dialect on account of phonetic and lexical commonalities based on Mayan influence. The Spanish of Pasto (southern Colombia) and the Spanish of Tulcan (northern Ecuador) represent one single dialect in similar terms. Both cases show that cover terms such as ‘Mexican Spanish’ or ‘Ecuadorian Spanish’ are useful for general purposes but requires further specification for a more accurate description, especially of those countries that are less homogeneous in linguistic terms. Such specification implies, among other things, dividing line between highland and lowland varieties along phonetic, morphosyntactic and lexical parameters. In fact the highland-lowland distinction in Hispanic America is the results of different colonization patterns.

This chapter does not endorse a diasystemic view of Spanish⁶ and considers that cross-dialectal unity prevails all over Hispanic America.⁷ Accordingly, the term “Spanish” is used in a broader sense to refer to all dialects and sociolects spoken in the areas of study. Terms such as “Spanish in México” or “Spanish in Ecuador” are used instead of their counterparts “Mexican Spanish” or “Ecuadorian Spanish” so as to leave the door open to dialectal considerations when these help to explain the findings of this investigation.

Hispanicization through colonial expansion is described in sections 5.1.1 and 5.1.2. Spanish dialectal variation in the Americas and Amerindian influence are discussed in sections 5.1.3 and 5.1.4. Further sections focus on the Spanish dialects of the areas under study, including Highland Ecuador (section 5.2), Paraguay

⁶ Notice that a diasystemic view is prevalent among students of Quechua as well. However, I consider this view to be suitable for the description of linguistic variation in this case, because of the marked fragmentation and divergence of Quechua varieties.

⁷ For Anderson (1991) the linguistic commonality in the former Spanish colonies laid the foundations for the Independence movement through the dissemination of revolutionary ideas in print form (pamphlets, diaries, books, etc) which would have been impossible, according to Anderson, if dialectal variation had been too large.

(section 5.3) and Highland Mexico (section 5.4). The typological features of Spanish which are crucial for the analysis of loanwords are described in section 5.5.

5.1.1. Hispanicization and diglossia

At the end of the twentieth century Spanish native speakers in the Americas were 294 million people, unevenly distributed in twenty different countries (Otero 1999). The great majority of these countries were traditional Spanish-speaking areas because of their former circumscription in the Spanish Empire for over three hundred years. The United States of America became an important Spanish-speaking area since the second half of the twentieth century as a result of migration from Hispanic American countries.⁸ Spanish in Latin America is official in administration and education in nineteen countries and co-official with another language (Guaraní) in one country (Paraguay)⁹. While most of these countries have a larger numbers of Amerindian speakers, Spanish is dominant in all of them and co-exists with indigenous languages in a diglossic distribution (cf. section 2.3.1).

The vitality of Spanish in the Americas is strengthened by its official status and the ongoing Hispanicization of native peoples through formal schooling, media and urban migration. The Hispanicization of native peoples began from the early years of colonization and went hand in hand with evangelization. The process speeded up after the wars of Independence and the emergence of the nation states in the early nineteenth century. As part of their goals of national unity, the new republics sought to homogenize their citizens by reducing ethnic and linguistic differences to the minimum (Anderson 1991; Radcliff and Westwood 1999). Language policies were implemented by all administrations, regardless of their conservative or liberal affiliation, in order that non-Spanish speakers learned Spanish as a means of communication with the mainstream society while abandoning their native

⁸ The Spanish-speaking population in the United States was around 29 million by 1997, which represented 11 % of the total population (Morales 1999). This number is increasing rapidly as a result of demographic growth and ongoing migration flows. Official estimates give 13.8% of Spanish-speaking population in the United States for 2010 (Silva Corvalán 2000). These numbers exceed by far the size of the Spanish-speaking population in a dozen of Hispanic American countries. Although the vitality of Spanish in the United States is strong enough for it to be the second language in the country, it is clearly losing ground to English. In general, the sociolinguistic status of Spanish with respect to the English in the United States remains diglossic, even in the states with large numbers of Spanish speakers (cf. Silva Corvalán 2000).

⁹ It is important to stress the national character of bilingualism in Paraguay, because in several countries Spanish is official along with other Amerindian languages *only* in their respective areas of influence. For example, the 1998 Ecuadorian Constitution establishes that Spanish is the only official language of the country while it is co-official with indigenous languages in their respective areas. Although co-official status encourages the use of native languages in education and other local affairs, it is rather restrictive and does not affect the diglossic situation of the indigenous languages in the Americas.

languages or restricting their use to domestic spaces. Linguistic standardization became a primary goal for policy-makers in the nineteenth century and remains a major concern in today's language policies in Latin America (Gonzalez Stefan 2000). This is clearly exemplified by the design of most bilingual education programs in Latin America. As evidenced by the increasing Hispanicization of native populations, these programs become, at best, instruments to facilitate the learning of Spanish in early stages of schooling. At present, Indian movements all over the continent are claiming the linguistic rights of native peoples. Still, the dominant position of Spanish vis-à-vis Amerindian languages remains largely intact. Ironically, the only difference from the old establishment is that Spanish becomes the mother tongue of an increasingly large number of indigenous people whose ethnolinguistic identity is dissolved on the way.

5.1.2. Colonial expansion, dialectal leveling and Andalusian influence

Spanish entered the Americas for the first time in 1492 with Columbus and his crew. But the presence of Spanish after the first arrival was to be ephemeral because Taíno Indians murdered all soldiers Columbus left at Hispaniola. It is only after Columbus' next voyages (1493 through 1497) and the first large migration of peninsular immigrants to the West Indies (ca. 1500) that a considerable number of Spanish colonizers came to settle in the Americas on a permanent basis. From the West Indies the colonization of the continent proceeded to the West (Mexico) and the Southwest (Central America) almost simultaneously. Arias Dávila founded Panama City in 1519 and two years later the Aztec Empire was defeated. The foundation of Cartagena de Indias took place in 1533. Quito, Lima, Bogotá and other major cities along the Andes were founded immediately afterwards. Venezuela was colonized in 1547, nine years after the foundation of Asunción in present Paraguay. The foundation of large urban center in Uruguay, Argentina and Chile took place only in the second half of the sixteenth century because the colonization of the southernmost territories was a long and expensive enterprise. In all, the colonization of the continent took over two centuries, from the first voyage of Columbus to the conquest of the last Araucanian stronghold in southern Chile. Of course, this does not mean that every corner of the continent was eventually settled. In fact, several areas in the Amazon basin remained unexplored today.

Because the colonization of the Americas was not uniform in time and space, the Spanish language that arrived at the continent was not uniform either. This fact explains the dialectal variation existing nowadays. According to Moreno de Alba,

“Es innegable, por tanto, que el español llevado a tierras americanas por los conquistadores y colonizadores no fue exactamente el mismo para las Antillas (fines del siglo XV) que

para el cono sur (fines del XVI y todo el XVII) En más de un siglo, la lengua cambia. Puede pensarse incluso que algunas peculiaridades lingüísticas de las diversas regiones hispanoamericanas tengan su explicación, entre otros factores, en la fecha del inicio de su colonización” [It is indisputable that the Spanish language taken to the Americas by conquerors and missionaries was not the same for the Antilles (in the end of the fourteenth century) as for the southern areas (in the end of the sixteenth century and throughout the seventeenth century). Over a century the language changed. We may even think that some linguistic particularities of the Hispanic American areas may be explained, among other factors, by the time their respective colonization began] (Moreno de Alba 2004: 13; my translation).

In the same year Columbus arrived at the Antilles, the Catholic Kingdom of Castilla expelled the Arabs from their last stronghold in Granada and Antonio de Nebrija published the first Spanish grammar. From 1492 onwards a series of changes in the language took place uninterruptedly, with important consequences for the final configuration of the varieties spoken first in the Peninsula and later in the Americas. According to Alonso (1962: 85-102), these changes were initiated as early as the fourteenth century but took shape only in the sixteenth century, that is, during the first century of Spanish colonization. The changes that molded Spanish to its present shape were mainly phonological. One of them was the merge of the voiced and voiceless palatal fricatives /j/ and /ç/ in the voiceless palatal fricative, by virtue of which [mujer] ‘woman’ became [muçer].¹⁰ Another change was the merge of the bilabial fricative /v/ and the bilabial stop /b/ in one voiced bilabial phoneme. Changes of morphological nature took place along with phonological ones: e.g. the replacement of verbal inflectional forms of second person plural (-ades, -edes, -ides) with shorter forms (-áis, -éis -ís). Simultaneously, the expansion of Castilian Spanish in areas of the Peninsula where other languages such as Basque and Arabic were spoken, as a result of the political predominance of the Castilian Kingdom, encouraged the entry of numberless lexical borrowings. All these changes made the linguistic landscape of the Spanish Peninsula by the middle of the sixteenth century one of effervescent transformation.

If the structural changes in sixteenth-century Peninsular Spanish passed to the Spanish colonies overseas, where did Spanish in the Americas get those features which make it different from peninsular varieties, and especially from Castilian Spanish? Why did Spanish in the Americas not follow the same path of Castilian

¹⁰ Later in the seventeenth century a voiceless velar fricative /x/ replaced the voiceless alveolar fricative , so that [muçer] finally became [muxer] as in present-day Spanish. I am indebted to Wolf Dietrich for calling my attention to the consecutive order of these changes.

Spanish and produce one homogeneous language instead of a great number of dialects? An answer to this question is possible only if sociohistorical causes and linguistic factors are considered side by side.

Assuming that waves of colonists from the Peninsula to the Americas were uninterrupted for over three hundred years, there is no reason for Spanish in the Americas to have become different from Peninsular Spanish. But this differentiation indeed occurred as a result of nonlinguistic and linguistic causes. To begin with, Spanish colonists came from different parts of the Peninsula and spoke different dialects of Spanish. Spanish historians have identified eight dialects spoken in the Peninsula by the turn of the sixteenth century, but we cannot be absolutely sure of their number (Catalan 1989). Some dialects were more widespread and politically dominant than others (e.g. Castilian Spanish). In certain cases Spanish dialects were spoken along with other languages such as Basque. In sum, no unified Peninsular Spanish existed at the time of the American colonization, just like no Peninsular Spanish exists today. The Spanish brought to the Americas was heterogeneous not only from a diachronic perspective but, most crucially, from a diatopic perspective, that is, from the point of view of the different dialects brought to the Americas by Peninsular colonists. For several scholars, notably Alonso (1967), the early convergence of dialects in the Americas, particularly in the Antilles, laid the foundation for Spanish in the Americas. This was accomplished through a process of leveling by which the dialects represented in the American speech community gradually lost their differences. The basis of Spanish in the Americas would be therefore a sort of average of the early peninsular dialects. The concept of dialect leveling is close to the notion of koinecization. De Granda characterizes the first stage of Spanish in the Americas as one of koinecization, in which “through a series of linguistic accommodations the initial heterogeneity of languages converges at a final stage that may be called the Spanish of America” (De Granda 1994: 26; my translation). According to Danesi (1977), koinecization explains certain characteristics of American Spanish, such as *seseo* (the alveolar simplification of the phonological alveolar-dental distinction as occurs in Castilian and other Spanish dialects):

“Leaving aside sociological factors, there is a linguistic explanation for the phenomenon of *seseo* in America. What appears to have occurred, in our opinion, is that *seseante* speech was adopted by colonists from all parts of the Peninsula as a result of phonological systems in contact; that is, the opposition /θ – s/ soon came to have a low functional load in areas of mixed speech and thus became less resistant to merger: /θ – s/ = low yield → /s/” (Danesi 1977: 1992-3).

While the explanation is convincing, Danesi assumes wrongly that the contact of different phonological systems is a purely linguistic factor. Quite the opposite, the contact stems from a sociological motivation: the coexistence of speakers from different speech communities in the same geographical and social space (cf. section 2.2). The question is less *why* the merger yielded /s/ and not /θ/ rather than *how* it took place. The explanation is nonlinguistic and lies on the demographical disparity in the leveling process.

Demographic motivations played a decisive in the configuration of Spanish in the Americas. It is a well-known fact that speakers from different parts of Spain were not evenly represented in the first waves of colonization. From the analysis of historical documents linguists have established that the contribution of Andalusian dialects was by far the most important, especially in the first quarter of the sixteenth century. From the analysis of a large corpus of demographic data corresponding to twenty percent of the total population that migrated to the Americas during the so-called Antillean period (1492-1519) Boyd-Bowman concludes that:

“En la época primitiva o antillana, el grupo más numeroso en cada año, y en todas las expediciones, fueron con mucho los andaluces, de los cuales más de 78% procedían de las dos provincias de Sevilla (1259-58%) y Huelva (439-20%). [...] de cada tres colonizadores, por lo menos uno era andaluz; de cada cinco, uno era oriundo de la provincia de Sevilla; de cada seis, uno se llamaba vecino o natural de la ciudad del mismo nombre” [During the Antillean stage the largest group in the expeditions was by far that of Andalusians. Of this group over 78% came from the two provinces of Seville (1259-58%) and Huelva (439-20%) [...] of every three colonists, at least one was Andalusian; of every five, one had been born in the province of Seville; of every six, one was inhabitant of Seville or had been born there.” (Boyd-Bowman 1964: ix; my translation).

On the basis of these figures scholars have proposed an Andalusian origin to explain several characteristics of American Spanish varieties, including the aforementioned *seseo*. The Andalusian hypothesis became widely accepted over the years, but disagreement persists now about the time span and the scope of the Andalusian influence. The major presence of Andalusian speakers in the Americas is documented for the first years of colonization but not for the entire sixteenth century. Moreover, there is no systematic study of the demographic composition for periods later than 1520, which prevents us from making any conclusive statements. It is certain that Andalusian dialects influence Spanish in the Americas through the alveolar simplification of /s/ and /θ/, but several other phonological and phonetic characteristics of Andalusian dialects (e.g. the aspiration of /s/ in coda position, the merger of /l/ and /r/, the weakening of voiced fricatives in intervocalic

environments, etc.) are not widespread across the continental but restricted mainly to coastal regions including the Antilles, the Mexican coast, Panama, and the littoral regions of Colombia, Venezuela, Ecuador, Peru and Chile. Other regions such as the Mexican Plateau, the Andean Highlands from Venezuela through Chile, and Paraguay, do not show those features. Toscano (1953) classifies Ecuadorian Spanish spoken in two clearly identifiable dialects: the coastal variety with a number of Andalusian traits and the highland variety with few or none of them. Candau (1987: 634), for example, did not find Andalusian features in the Spanish of the southern highlands of Ecuador.

The accessibility of lowland cities was an influencing factor in this case. Linguistic historians consider that speech innovations brought from Spain through Andalusia and Canarias found rich soil in coastal cities as a result of their permanent contact with the metropolis. On the other hand, the language spoken in plateaus and highlands preserved old Castilian features, many of which are archaic from the point of view of modern Spanish. This polarity is attested in several countries: Veracruz (port city) and Mexico City (highland capital); Cartagena (port city) and Bogotá (highland capital); Guayaquil (port city) and Quito (highland capital); Lima (lowland capital) and Cuzco (highland city). Interestingly, most administrative centers of the Spanish Crown were not located on coastal areas but in the interior, except for Lima, capital of the Viceroyalty of Peru, and Buenos Aires, seat of the Captainship of Rio de la Plata. The obvious question is, therefore, why non-coastal centers including the capital of the Viceroyalty of New Spain and several capitals of *Audiencia* were not influenced by Andalusian dialects considering their position as centers of administration and culture. De Granda (1991) maintains that Andalusian features spread across the Empire with different intensity during the sixteenth century; some became deeply rooted in specific areas while others disappeared. The reasons are both sociopolitical and demographic. On the one hand, for Andalusian features to prevail, an important input of Andalusian speakers was required on a permanent basis, a condition that could be met only in coastal cities. On the other hand, it is likely that the dominant dialect in most administrative centers was not Andalusian but Castilian, since most officials of the Crown came from this area, in particular from Toledo and Madrid.

A major sociolinguistic motivation to distinguish between lowland and highland dialects was the Pre-Columbian influence on Spanish language and culture, notably the Aztec in Central America and the Inca in the Andes, both of which had their areas of influence in the highlands. According to Rosenblat the strong articulation of consonants in highland varieties of Spanish go against the internal development of the language and should be explained by an external force, which, in his view, cannot be other than the influence of indigenous languages such as Nahuatl in Central America and Quechua in the Andes (Rosenblat 1967: 150). Moreno de Alba

(2004) summarizes the process of substratum influence proposed by Rosenblat in the following terms:

“Esta influencia tuvo que darse en ciertas condiciones: el lento y complejo proceso de hispanización, la acción del mestizaje, iniciado desde la primera hora y prolongado hasta hoy, el bilingüismo de amplios sectores indios. Habría que distinguir dos momentos: en el primero se cumplen los cambios que estaban en marcha en el español y se está todavía dentro de una básica y amplia unidad del español americano. El segundo momento, iniciado tímidamente al principio, alcanza su fuerza en las generaciones siguientes, debido sobre todo al bilingüismo y a la penetración de voces indígenas. El fonetismo de las tierras altas prueba sin duda que una influencia extraña puede contrarrestar las tendencias propias de la lengua” [This influence occurred in certain conditions: the long and complex process of Hispanization; the miscegenation process that began from the first contact up to date; the bilingualism of large Indian populations. It is useful to distinguish two stages: in the first stage the changes already in progress in the Spanish language took their final shape, still within the basic unity of the American Spanish; the second stage, which developed only partially at the beginning, gained momentum in the following generations due to the bilingualism and the penetration of Indian loanwords. The phonetic characteristics of the highlands are proof that a foreign influence can counteract the language’s own tendencies.] (Moreno de Alba 2004: 104f; my translation).

The hypothesis of the Pre-Columbian influence has been challenged more than once (e.g. Lope Blanch 1969, 1972). From my point of view the evidence analyzed by Rosenblat is conclusive, let alone the large number of recent studies on the influence of Amerindian languages in regional varieties of Spanish (cf. Haboud 1998 for Ecuadorian Highland Spanish; Dietrich 1995 for Paraguayan Spanish). The question of the influence of Indian languages on Spanish will be addressed again after sketching the dialectal areas of Spanish in the Americas in the following section.

5.1.3. Dialectal areas of Spanish in the Americas

Several classifications have been advanced since the first dialectal map of American Spanish proposed by Henríquez Ureña in 1921. This author put forward a division of dialects in five areas: 1) the bilingual zones of the southern and southwestern United States, Mexico and Central America; 2) the Hispanic Antilles, the coast and the plains of Venezuela and northern Colombia; 3) the Andes of Venezuela, the interior and the Pacific coast of Colombia, Ecuador, Peru, northern and central Bolivia, and

northern Chile; 4) Central and southern Chile; and 5) Argentina, Uruguay, Paraguay and southeastern Bolivia. His criteria for classification were geographical (territorial contiguity), cultural (a shared heritage) and historical (indigenous substrata). Unlike later proposals (Rona 1964; Resnick 1976) this classification does not take phonological and phonetic criteria into consideration. In fact, most critiques of Henríquez' classification focus on his failure to notice linguistic data. For Rona (1964) dialects are linguistic facts in themselves, and therefore any dialectal classification must be based on linguistic facts *only*. For Rona, nonlinguistic criteria can be proposed only to the extent they support a classification established on linguistic grounds. In these terms, Rona undertakes a different classification by drawing isoglosses along four linguistic criteria which, in his opinion, are the only ones a linguist can be certain of, namely: the phonetic realization of /y/ as [ʒ] (*žeísmo*); the phonological neutralization between /y/ and /ɲ/ (*yeísmo*); the use of pronouns *vos* (*voseo*) and *tú* (*tuteo*); and the verbal paradigm of both pronominal forms. The result is a number of 23 dialectal areas. The dialectal map resulting from the isogloss method looks strikingly different from the map drawn by Henríquez Ureña. Rona groups together the Antilles, the Atlantic coast of Venezuela and Colombia, the eastern half of Panama, and Mexico, excluding the southeastern states of Chiapas, Tabasco, Yucatan and Quintana Roo. Second, Rona does not consider Ecuador as one single area but distinguishes the Andean Highlands from the rest of the country. Third, he considers Paraguay a self-standing area distinct from Rio de la Plata (Buenos Aires and surroundings).

With minor changes, Rona's classification has been widely accepted in linguistic and nonlinguistic circles. Other classifications (cf. Resnick 1975) consider as many as 16 linguistic features and produce a much larger number of dialectal areas. No classification is definitive, though, as every one depends on the number and type of linguistic features considered, and whether linguistic factors are viewed as independent or interlinked with nonlinguistic ones. Apart from the four features mentioned by Rona (*žeísmo*, *yeísmo*,¹¹ pronominal *voseo* and *tuteo*, and the associated verbal paradigms) any reliable classification must include another type of historical linguistic data, which Rona sets aside but is certainly a valuable yardstick for the measurement of dialectal divergence in the Americas: the influence of the Amerindian substratum. This is the topic of the next section.

5.1.4. Spanish in contact with Amerindian languages

For any Spanish speaker in the Americas one of the most striking – and often embarrassing things – when traveling to another country is the discovery that

¹¹ In fact, *žeísmo* and *yeísmo* are not independent but phonetic phenomena in complementary distribution.

meanings of words change simply by crossing the border, or that many words that sound Spanish have no meanings to him or her. Anecdotic as it may sound, this fact is evidence of the lexical complexity of Spanish varieties in the continent, one that goes well beyond the phonetic features we described in the previous section. Only a minor part of this complexity is due to the occurrence of archaic Spanish forms in American dialects.¹² The largest part can be explained only by a century-long influence of Amerindian languages on Spanish (cf. Haboud 1998; Olbertz 2005; Palacios Alcaine 2005b). Native languages contributed to the configuration of Spanish in the Americas mainly with their lexicon, although their influence on several aspects of grammar is not unimportant in a number of dialects. Several of the Amerindian languages that contributed to the lexical pool of Spanish in the Americas disappeared a few decades after the first invasions (e.g. Taíno). Others died in the long process of Hispanicization of native peoples (e.g. Chibcha). Others survive with great vitality up to the present (e.g. Nahuatl, Quechua, and Guaraní). The different fates of Amerindian languages have determined the type and degree of their influence in each region. In these terms, the analysis of language death or maintenance of Amerindian languages in the Americas can help us explain synchronic differences across American dialects of Spanish, whether they are derived from substratum or from adstratum influences.

Let us begin with those languages that experienced a premature death. The first Amerindian language the Spaniards found in the West Indies was Taíno, an Arawak language spoken at Hispaniola (today's Dominican Republic and Haiti). Several Taíno Indians were taught Spanish and became interpreters for the Spaniards in their occupation of the Antilles and the Caribbean coasts of Venezuela. The use of Taíno was viable because the language was spoken in several islands and showed a close resemblance to other languages of the Arawakan family. None of the languages once spoken in the Antilles has survived. Many Caribbean Indians died in the years following 1492 by epidemic diseases. The rest perished as a result of slavery and genocide. There is no grammatical description or dictionary of Taíno but the words the Spaniards borrowed from this language in the early times of colonization. Some of these words describe endemic flora and fauna (e.g. tobacco, maize, etc.) while others refer to objects (e.g. hammock) or social institutions (e.g. cacique). Given the short time span of contact with Spanish and the small number of bilingual Taíno Indians who survived, the influence of this language on Spanish remained purely lexical. Because the Antilles were the first area of Spanish occupation and none of the local Amerindian languages survived, the contact with these languages was

¹² Moreno de Alba (2004: 262) rightly warns us about the ambiguity of the notion 'archaism', which is always applied with reference to Peninsular usage. There is nothing archaic about these forms from the perspective of the speakers themselves, of course. This author notes that, even if so-called archaic forms abound almost in every American dialect of Spanish, there are only a few of them used at a continental level (2004: 267).

comparatively short and the influence it exerted on Spanish did not go farther. According to the borrowing scale of Thomason (2001: 70), Taíno-Spanish contact can be classified as type 1 (casual contact).

A longer contact with Amerindian languages induced deeper changes in regional varieties of Spanish. Early in the colonization of the Northern Andes, Spanish entered in contact with several Pre-Columbian languages in addition to Quichua. Quichua coexisted as a *lingua franca* with these languages since the Inca occupation of present Ecuador until the second half of the sixteenth century (Gómez Rendón and Adelaar, forthcoming). These languages also coexisted with Spanish in the first century of colonization and left noticeable traces in the lexicon and the phonology.¹³ Other languages coexisted with Spanish for an even longer period of time and thus exerted a more decisive influence. Some of these were Muisca in the Colombian Andes (Adelaar and Muysken 2004: 81f) and Tupi in the province of Rio de la Plata and part of the Amazon lowlands.¹⁴ Spanish speakers living in these areas spoke Muisca and Tupi in order to communicate with native peoples, as can be deduced from a number of borrowings from these languages into local Spanish. These contacts can be classified somewhere between type 2 and type 3 in Thomason's scale in attention to the type of changes induced in Spanish. However, they belong to type 2 (slightly more intense contact) if intensity is the yardstick. In addition, there are differences in the direction of influence. Contact-induced changes in Ecuadorian Highland Spanish took place by the agency of non-native speakers while the changes induced by contact with Chibcha and Tupi were mostly due to Spanish native speakers. In terms of van Coetsem (1988: 3) the influence of Pre-Columbian languages on Ecuadorian Highland Spanish makes a case of donor-language agentivity, in which speakers of the source language bring about the changes. On the other hand, the influence of Chibcha and Tupi on Spanish is a case of recipient-language agentivity, in which speakers of the recipient language implement the changes. Distinguishing both types of agentivity in contact-induced language change allows us to identify the direction of the influence and the different processes at work (Winford 2005: 373ff). Furthermore, the distinction is helpful when it comes to explaining substratum and adstratum influences from Amerindian languages coexisting to date with Spanish.

The last group of languages considered here are those which not only survived colonization but remain strong for their number of speakers. I do not include under this category a large number of Amerindian languages that are still living but have slight or no influence on local varieties of Spanish. Three languages have

¹³ In a similar way these languages left their imprints on local varieties of Quichua as a result of contact (e.g. aspiration of plosives in onsets).

¹⁴ Tupi was perhaps the *lingua franca* with the widest distribution, as it was spoken not only in Spanish territories but also in the Portuguese Empire, notably along the Atlantic coast of Brazil and the interior.

significantly contributed to the shape of regional varieties of Spanish in terms of distribution, number of speakers and duration of contact: Nahuatl in Central Mexico; Quechua in the Andes; and Guaraní in Paraguay and Northern Argentina. Today these languages have speech communities including millions of speakers. A great number of them are also bilingual in Spanish, with different degrees of proficiency. Many idiosyncrasies of the Spanish spoken in the areas of influence of these languages are explained by the bilingualism of its speakers. Monolingual Spanish speakers have played a minor role, if any, in the transfer of lexical and structural features of these languages. These are cases of source-language agentivity, and the changes induced in Spanish by Amerindian contact correspond roughly to types 2 and 3 in Thomason's scale. However, an additional distinction is required. The influence of Amerindian languages on Spanish is one of adstratum in those areas in which the two languages coexist today but one of substratum in a number of areas where the languages coexisted in the past. It is therefore possible to group Spanish dialects in five areas according to Amerindian influence. This is shown in the following table.

Table 5.1 Areas of Amerindian substratum in Spanish

Substratum areas	Languages (language family)
1. Mexico and Central America	Nahuatl (Uto-Aztecan), Mayan
2. Dominican Republic, Cuba, Puerto Rico, Northern Venezuela and Northern Colombia	Taino, Carib (Arawakan)
3. Andes of Colombia, Ecuador, Peru, Bolivia, Northern Chile	Quechua, Aymara (Andean)
4. Central and Southern Chile	Mapuche
5. Paraguay and Northern Argentina	Guaraní (Tupi-Guaraní)

The Spanish varieties in contact with the Amerindian languages of this study belong to the first area (Otomí), the third area (Quichua), and the fifth area (Guaraní). Substratum and adstratum influences are discussed in the following sections. In this respect it is worth noting that substratum phenomena in South America are few in comparison to adstratum phenomena. While substratum phenomena imply extinct vernaculars, adstratum phenomena implies a long-term contact with living indigenous languages.¹⁵

¹⁵ I am indebted to Wolf Dietrich for calling my attention to this fact as well as to the implications of both types of phenomena for the contact situations analyzed in this book.

5.2. Spanish in Ecuador

As the official language in Ecuador, Spanish is spoken all over the country. Of a national population of 13,363,593 people (CEPAL 2005), approximately 90 percent is Spanish monolingual (12,000,000) while the rest is bilingual in Spanish and one of the nine indigenous languages of Ecuador, with different degrees of proficiency. Spanish speakers are unevenly distributed in the country, with the Coast and the Highlands containing the bulk of the Spanish-speaking population (95%), and the three major cities (Quito, Guayaquil and Cuenca) taking in 50% of the total number.

The sociolinguistic situation in Ecuador is clearly diglossic. Spanish is the dominant language and the only one used for administration purposes. The situation has not changed since colonial times. The 1998 Constitution grants Indian peoples the right to use their own languages as co-official in their respective territories, but learning and speaking Spanish is simply a matter of fact for them. The process of Hispanicization begins early in the life of non-Spanish speakers. Children in native communities often grow up listening to their parents speaking Spanish. They listen to Spanish in the media and in the public spheres. Furthermore, the process of Hispanicization is reinforced in the schools. Indeed, the process continues despite bilingual education programs implemented since the late eighties (Yáñez Cossio 1995) simply because such programs were designed as a bridge to the acquisition of Spanish (cf. *supra*). It is not surprising, therefore, that most speakers of indigenous languages in present Ecuador speak Spanish with more or less proficiency depending on such factors as age, gender and time of schooling. Correspondingly, bilingual speakers have become functionally monolingual as they do not use their native languages anymore nor speak them to their children (Buttner 1993). In all, bilingualism is growing in rural indigenous communities while Spanish monolingualism is the rule for an increasing number of indigenous immigrants in the cities (Haboud 1998).

In the context of such a steady process of Hispanicization, it is obvious that interferences in the Spanish of non-native speakers become part of local Spanish once the process of language shift is completed. The outcome is therefore double: interferences in the Spanish interlanguage of non-native speakers (adstratum); and interferences crystallized in the local varieties of Spanish of native speakers with different ethnolinguistic background (substratum). Adelaar and Muysken (2004) summarize this situation in the following terms:

“If we try to imagine how this influence was exerted, the most plausible scenario is one of second-language learning by Quechua speakers in a sociolinguistically complex environment. The particular stratification of variable elements within the Spanish target-language speech community affects the process of acquisition of these elements.

This stratification is crucial because it may reflect, in part, stages of interrupted or incomplete L2 acquisition at an earlier point in time. As time goes on, the products of intermediate and advanced interlanguage grammars (A and B)¹⁶ are incorporated into the native speech community (C and D), but most often as vernacular, non-standard forms. In a synchronic perspective, then, native speakers of the target vernacular end up producing outputs that seem like interlanguage outputs. The particular interlanguage features which come to be adopted as non-standard features in the Spanish target speech community serve as models, at a later stage, for new learners” (Adelaar and Muysken 2004: 592).

While shift-induced interference is more visible today as a result of the rampant levels of Hispanicization of native populations, it was present from early colonial times. This is reflected on the substratum of several highland dialects, where the bulk of the non-Spanish population (Quichua) lived and continues to live. Still, it is necessary to underline the fact that the agents of these changes are originally speakers of other languages, *not* Spanish native speakers. The reason for this is the ethnic bias of bilingualism in Ecuador: native speakers of Spanish *never* learn an indigenous language and therefore cannot transfer features from these languages to Spanish. Except for a number of Spanish colonists in the first years of colonization, who learned Quichua to communicate with native people, one-sided bilingualism persists to date.

The presence of Spanish in today’s Ecuadorian Highlands dates back to the early 1530s when Sebastian de Benalcázar founded the first cities on his march for the conquest of Quito, the last Inca stronghold in the Northern Andes. After the defeat of the last Inca generals, the process of colonization proceeded rapidly. By 1600 the principal cities of the Ecuadorian Highlands had been founded. The evangelization and subsequent Hispanicization of native peoples began immediately after the last Indian uprisings were suppressed by the mid 1500s. The Highlands concentrated most of the Indian population in the Real Audiencia de Quito during colonial times. Today, the demographic distribution of the Indian population is more or less the same. Quichua speakers make up the largest ethnic group, with an approximate number of 1,500.000 speakers in nine of the ten highland provinces. These circumstances provided an ideal setting for contact between Spanish and Quichua, the outcomes of which are reflected in both languages.

¹⁶ Letters A-E stand for different types of Spanish speakers with influence from Quechua: A – Quechua speakers learning Spanish; B – stable Quechua-Spanish bilinguals; C – Spanish monolinguals living in bilingual communities; D – Spanish monolinguals living in bilingual regions; and E – Spanish monolinguals living in bilingual countries (Adelaar and Muysken 2004: 590).

Map 5.1. The Languages of Ecuador

Source: *Ethnologue* 2005

The process of colonization of other regions in Ecuador was different. Gonzalo Pizarro and Francisco de Orellana explored the Amazonian lowlands early in the 1540s, with the eventual discovery of the Amazon River in 1542. The first Spanish settlements in the Amazonian lowlands were founded shortly afterwards. First the Dominicans and later the Jesuits and the Salesians took up the evangelization of native peoples scattered in the vast regions of the jungle. The presence of Spanish was early in the Amazon basin but it consisted of few Spanish settlements, some of which had to be founded once and again after being devastated by Indian raids. Consequently, the number of Spaniards in the area was considerably smaller and the Indians continued to live scattered all over the jungle. In other words, demographical and geographical factors prevented a widespread contact of languages. Only in recent years the Amazon lowlands have experienced an important growth of Mestizo settlers, with the corresponding increase in language contact and the raise of bilingualism and Hispanicization.

The Pacific lowlands were colonized gradually too. Except for the city port of Guayaquil and its surroundings, vast extensions of land remained largely unexplored until the early nineteenth century (Ayala Mora 1993). The pattern of scattered settlement among the native peoples from the Pacific lowlands barred the colonization of the area: Spaniards could not profit from the local Indian workforce, as they certainly did in the Highlands, where the bulk of the native population was concentrated¹⁷. In this context the Pacific lowlands did not experience the same process of language contact as the Andes. While contact with indigenous languages was sporadic, contact with Spain and other coastal cities through the port of Guayaquil was permanent. Therefore, the linguistic influence on Lowland Spanish and Highland Spanish were different.

5.2.1. Dialects of Spanish in Ecuador

Of the historical developments just described, two distinctive dialects emerged in the eighteenth century: Highland Spanish and Lowland Spanish.¹⁸ Ecuadorian Highland Spanish (*español andino ecuatoriano*) has been the object of several studies in the last years (Niño Murcia 1995; Haboud 1998; Olbertz 2002; Palacios Alcaine 2005), most of which focus on its non-standard characteristics resulting from intense contact with Quichua. Ecuadorian Highland Spanish is the source of borrowings in the Quichua of Imbabura and Bolívar. Ecuadorian Littoral Spanish (*español costeño* or *español litoral ecuatoriano*) is part of Equatorial Littoral Spanish (*español ecuatorial ribereño*), a group of dialects spoken along the Pacific coast of Colombia and Ecuador and the northern coast of Peru (Zamora and Guitart 1982). Ecuadorian Littoral Spanish has received considerably less attention from students of contact, even though the influence of an important population of African descent in the Pacific Lowlands provides material for the study of African substratum. In the following I compare the major linguistic features of both dialects and explain their possible origins in language contact phenomena. The features of *seseo* and *yeísmo* (cf. section 5.1.1) are excluded from the discussion as they are shared by both dialects and neither can be ascribed to contact.

Since Toscano (1953) it is usual to characterize Ecuadorian Highland Spanish for its strong articulation of consonants (Sp. *fuerte consonantismo*) as opposed to the

¹⁷ The current demographic makeup of the Pacific lowlands is the result of migration of Mestizos and Indians from the Highlands during the twentieth century.

¹⁸ Spanish spoken in the Amazonian lowlands by monolinguals is not different from Highland Spanish. There are important differences, though, in the Spanish produced by bilinguals whose first language is Amerindian (e.g. Shuar, Cofan, etc.). In fact, one may find different varieties of second-language Spanish depending on the linguistic background of the speakers.

weak articulation typical of Ecuadorian Littoral Spanish.¹⁹ Translating this impressionistic assessment to phonetic terms, it is possible to identify three features of Ecuadorian Highland Spanish: 1) the raising of medial vowels /e/ and /o/ to [ɪ] and [ʊ] (Sp. *mesa* ‘table’ /mesa/ → [mɪsa]; Sp. *carro* ‘car’ [kaɾu]); 2) the relaxation and eventual elision of unstressed vowels (Sp. *pues* ‘thus’ /pues/ → [ps]); 3) the fricativization of the trill /rr/ (Sp. *carro* ‘car’ /karro/ → [kaɾo]) and the lateral alveopalatal /ʎ/ (Sp. *calle* ‘street’ /kaʎe/ → [kaʒe]). Because none of these features occurs in Ecuadorian Littoral Spanish, authors assume that their origin is substratum and adstratum influences from Quichua. The explanation is valid for the raising of medial vowels, to the extent that Quichua does not have /e/ or /o/. It is less satisfactory for the phenomena of vowel elision and fricativization. As pointed out by Adelaar and Muysken (2004: 591f), the fact that similar fricativization phenomena are found in many Quichua varieties do not confirm their origin in this language, in particular because more conservative Ecuadorian varieties do not show this feature. Moreover, vowel elision and fricativization occur in other Spanish dialects far from the Andes (e.g. the Mexican central plateau).²⁰ It is more reasonable to assume that fricativization in the Ecuadorian Highlands is a *Sprachbund* phenomenon: both languages have influenced each other to such an extent that they begin to share structural features, one of which is fricativization. Such interpretation is suggested by Adelaar and Muysken (2004: 592). Taken together, the elision of unstressed vowels and the fricativization of trills led to the occurrence of consonant clusters not occurring in other Spanish dialects.

One of the most visible influences of Quichua substratum on Ecuadorian Highland Spanish at a morphological level is the widespread use of diminutives, even in lexical items that generally do not take them, such as adverbs and, most importantly, pronouns. A similar use is not found in the lowlands. In the same way, the simplification of the clitic pronouns *la* and *lo* (Sp. *leísmo*) occurs across the Highlands (cf. Zamora and Guitart 1982: 224) but only for certain areas of the

¹⁹ The vowel salience in Ecuadorian Littoral Spanish is the product of two phonetic processes not found elsewhere in Ecuador: the aspiration and eventual elision of /s/ in coda position and the elision of /d/ in intervocalic position. While the first process results in CV-type syllables, the second produces diphthongs from the fusion of two syllables. Altogether, these processes cause vowels to stand out in a string of sounds. Sánchez Méndez (1998: 80ff) analyzed historical documents of the seventeenth century in order to find traces of the aspiration of /s/. His findings are surprising: the aspiration of /s/ was common not only in the Coast but in vast areas of the Highlands, including Quito. Later, this phonetic feature disappeared from the highland dialect but remained in the littoral dialect. This confirms to a certain extent de Granda’s proposal that all the characteristic features of today’s American dialects disseminated all over the continent, but not all of them were preserved uniformly depending on a series of factors internal to the language (De Granda (1991: 38).

²⁰ Interestingly enough, these and other dialects where elision of unstressed vowels and fricativization occur, belong to highland varieties. This confirms in part the highland-lowland distinction.

Ecuadorian Pacific Coast. The Coast either prefers the use of *lo* (Sp. *loísmo*) or makes a distinctive use of pronominal clitics as in Peninsular varieties. Other features are shared by Lowland and Highland dialects alike: the use of *voseo*, albeit particularly frequent in the northern Highlands; and the alternation of *tú* and *vos* (Quillis 1992: 603). The lack of number and gender agreement between articles and nouns is found also in both regions. In the Highlands this lack is typical of lower sociolects of Quichua –Spanish bilinguals. In Lowland Spanish this feature is also typical of lower sociolects, but its origins cannot be due to shift-induced interference, because speakers are Spanish monolingual. In the latter case the lack of number and gender agreement might be explained by two substrata: 1) a Quichua substratum based on the intensive migration from the Highlands to the Lowlands in the second half of the twentieth century; 2) a substratum influence from African languages.²¹ African substratum has been called for to explain the occurrence of a syntactic feature typical of certain sociolects of Ecuadorian Littoral Spanish: double negation. Whatever the case may be, the lack of in-depth studies on the topic prevents us from making conclusive statements in this respect. On the other hand, syntactic developments typical of Ecuadorian Highland Spanish have received more attention by linguists in the last years. In this context, two of the most characteristic syntactic features of Ecuadorian Andean Spanish (gerund constructions with perfective meaning and the use of tenses with evidential value) have been explained, satisfactorily in my view, by the influence of Quichua morphosyntax (cf. Haboud 1998; Olbertz 2002).

Another point of differentiation between highland and littoral dialects of Spanish in Ecuador is the lexicon. Quichua lexical borrowings occur even in the higher sociolects of Highland Spanish, but their presence is very limited in the Pacific Lowlands. Besides, it is hard to find Quichua borrowings of cross-dialectal use. The Quichua word *wawa* ‘child’ used by monolingual and bilingual speakers all over the Highlands occurs is virtually inexistent in Ecuadorian Littoral Spanish. In general, Ecuadorian Littoral Spanish lacks words of indigenous origin because Spanish in the Pacific Lowlands did not coexist with any native language as it did in the Highlands.

Table 5.2 below summarizes the major linguistic features from both dialects of Spanish. There is a clear-cut distinction between the two dialects at all linguistic levels. Spanish in the Highlands is modeled by substratum and adstratum influences from Quichua. In the Pacific Lowlands, substratum influence of African languages is a potential influencing factor, although conclusions are only speculative for the lack of studies in the field. The Andalusian influence on the phonetics and phonology of Ecuadorian Littoral Spanish is much more relevant, especially on the aspiration of

²¹ Lipsky (1987) found lack of number and gender agreement in the Black Spanish of the Chota Valley.

voiceless sibilant /s/ in coda position and the elision of /d/ in intervocalic environments as is typical of other coastal varieties of Spanish in the Americas.

Table 5.2 Dialects of Spanish in Ecuador: distinctive linguistic features

Level	Ecuadorian Highland Spanish	Ecuadorian Littoral Spanish
Phonetics/ Phonology	<ul style="list-style-type: none"> • Elision of vowels in unstressed syllables • Fricativization of the trill /rr/, realized often as a voiced sibilant /ʀ/ or a voiced sibilant in lower sociolects. • Fricativization of the lateral palato-alveolar /ʎ/, realized as the voiced alveolar /ʒ/ (<i>lleísmo</i>) 	<ul style="list-style-type: none"> • Aspiration or elision of /s/ in coda position • Aspiration of fricative velar /x/ as [h] • Elision of /d/ in intervocalic position. • Occasional alternation of /l/ and /r/ without elision
Morphology	<ul style="list-style-type: none"> • <i>Leísmo</i> across communities and strata • Simplification of gender and number distinctions in clitics • Lack of number and gender agreement (typical of non-native Spanish speakers) • <i>Voseo</i>, widespread in the Highlands and coexisting with <i>tuteo</i> in higher sociolects • Frequent use of diminutives on nouns, adjectives and adverbs 	<ul style="list-style-type: none"> • Limited use of <i>voseo</i> (prevalent in indicative constructions) • Lack of number and gender agreement, especially in lower sociolects • Reduced <i>leísmo</i>; pronominal clitics preserve number and gender distinctions in some areas while <i>loísmo</i> is prevalent in others
Syntax	<ul style="list-style-type: none"> • Use of gerund constructions with perfective meaning • Use of tenses with evidential value (reportativity and sudden discovery) • Use of SOV order (only in the Spanish of Quichua bilinguals) 	<ul style="list-style-type: none"> • Double negation in some sociolects (probably of African origin)
Lexicon	<ul style="list-style-type: none"> • Borrowings from Quichua and other Pre-Columbian languages 	<ul style="list-style-type: none"> • No native borrowings

As regards the Spanish varieties of Imbabura and Bolívar, where samples for the Quichua corpus were collected, differences are minor because both are highland subdialects. The demographic composition and the levels of bilingualism of Quichua speakers are more relevant for these varieties. According to the 2001 census,²² Imbabura and Bolívar have a population of 344,044 and 169,370, respectively. This difference in size is partly explained by higher rates of growth in Imbabura as compared to Bolívar (2.01% vs. 0.34% in the period from 1990 to 2001). The pattern of settlement in both provinces also determines of the influence of Quichua on Spanish. Half of the population of Imbabura lives in the cities but only one third of the population of Bolívar lives in urban centers. Because the Quichua population is concentrated in the countryside, a less intense contact between Spanish and Quichua is expected in Bolívar. Evidence for this is the fact that roughly half of the population of Otavalo (the second largest city in Imbabura) is Quichua-speaking while only a minor percentage of the population speaks Quichua in Guaranda (capital of Bolívar). In addition, Imbabura Quichua speakers are more bilingual than Bolívar Quichua speakers and thus influence local Spanish more deeply, because the access of indigenous speakers to the local variety of the dominant language determines their degree of influence on this.²³ This access is determined not only by demographical factors but also by education. In these terms, the higher accessibility of native Spanish to Quichua speakers in Imbabura (especially in the cities) cancels the contact effects of a larger indigenous population and explains why Imbabura Spanish is essentially the same as Bolívar Spanish.

Even with different settlement patterns, the presence of an important number of Quichua speakers in both provinces leads to expect noticeable linguistic effects as a result of the crystallization of non-native features and their transfer to local Spanish. If the number of bilinguals is large enough to disseminate these features in the target-language speech community, these might be traced also in the speech of Spanish monolinguals. While several data from our fieldwork in Imbabura corroborate this assumption, the lack of a systematic corpus of local Spanish in these provinces prevents us from making conclusive statements. Still, Imbabura Spanish and Bolívar Spanish are closely similar, and thus any difference in the outcomes of Spanish borrowing between Imbabura Quichua and Bolívar Quichua cannot be ascribed to differences in the input.

²² Source: Instituto Nacional de Estadística y Censos, www.inec.gov.ec. May 2007.

²³ The stratification of features influences the outcomes of second language acquisition. It is determined by the accessibility of the native variety of the target language. Accessibility, in turn, is determined by nonlinguistic factors such as geography (e.g. distance from urban centers), demography (e.g. smaller number of native speakers) or sociocultural factors (e.g. segregationism). An additional factor is the stability of bilingualism. In Ecuador and other Andean countries bilingualism is only a bridge to Hispanicization. The data from the sociolinguistic survey conducted between 1992 and 1993 in Ecuador point to this direction (cf. Buttner 1993).

5.3. Spanish in Paraguay

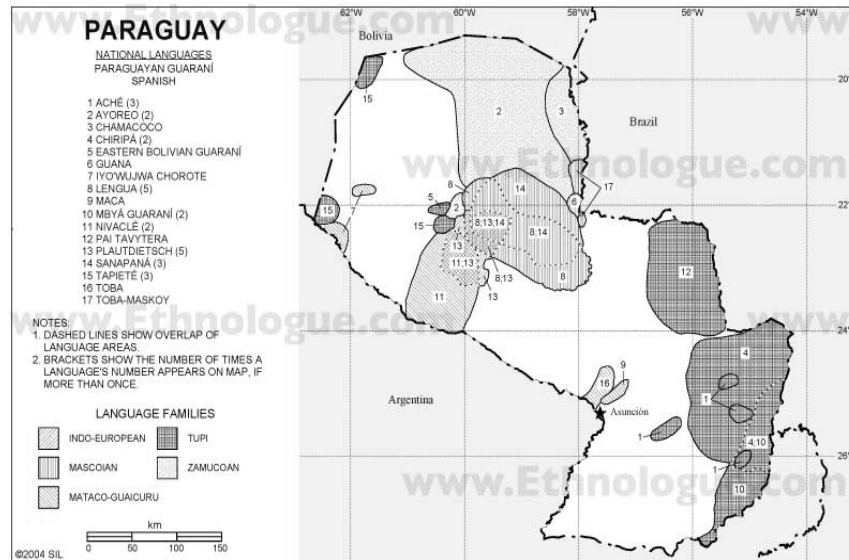
Spanish in Paraguay is co-official with Guaraní. Paraguay is the only country in Latin America where Spanish is co-official with another language, notably an Amerindian language. According to the 2002 census, Spanish monolingual speakers make a small sector of the population (6%) while Spanish-Guaraní bilinguals are the largest (59%), followed by Guaraní monolinguals (27%) and speakers of other languages (8%). In all, Guaraní is spoken by 86% of the population including monolinguals and bilinguals, while Spanish is spoken only by 65%. Generally speaking, Spanish is prevalent in the cities while Guaraní prevails in the countryside. Considering the high rates of urban-rural mobility and the shared knowledge of cultural traits in both areas, this partition is definitely artificial. The population in Paraguay is rather evenly distributed in urban and rural areas despite that urbanization rates have been lower than in other countries.²⁴ These facts might explain, to a certain extent, the slow process of Hispanicization in Paraguay, which in countries like Ecuador or Mexico is concomitant with high rates of urbanization.

Apart from Spanish, other European languages spoken in Paraguay include Portuguese, German (Plattdeutsch), Italian and Ukrainian. Portuguese has become stronger in Paraguay in the last decades as a result of an intensive contact with the Brazilian society. Portuguese is spoken today along with Spanish and Guaraní in border cities like Pedro Juan Caballero and Ciudad del Este. Paraguay also has seventeen indigenous languages from four different families (Tupi, Mascoian, Mataco-Guaicuru and Zamucoan). Most of their speakers are bilingual in Spanish but their native languages are not as robust as Guaraní.

The majority of Spanish monolinguals lives in the cities, especially in Asuncion, and belongs to middle and upper classes. Bilingual speakers are distributed in the cities and the countryside. Bilingualism varies across speakers and areas, from incipient in poor rural zones to coordinate in educated middle classes of the capital. In general, higher levels of bilingualism, not found in other Spanish-speaking countries, are present in Paraguay. Still, Paraguay is essentially different from other Latin American countries in one respect: Spanish continues to be the higher variety vis-à-vis Guaraní, in spite of the co-official status of the latter. This means that bilingualism in Paraguay is essentially diglossic (Krivoshein de Canese 1999: 2).

²⁴ According to CEPAL (2005) this situation will change dramatically in the coming years: urban resident will make 64% of the total population by 2015, which corresponds to an increase of 25% as compared to 1990.

Map 5.2. The Languages of Paraguay



Source: Ethnologue 2005

Guaraní is associated with national identity, solidarity and intimacy, while Spanish is associated with social mobility and job opportunities. The law orders that administrative and legal proceedings be bilingual, but in practice they are conducted in Spanish and translated to Guaraní only if necessary. Spanish is dominant too in the audiovisual media and the press, despite that an increasing number of publications in Guaraní appear every year in the form of popular literature or school texts. The position of Spanish in education is pretty much the same, even though Paraguay boasts a long tradition of bilingual education. In practice, Spanish shares the classroom with Guaraní in primary education while clearly prevailing over this language in secondary and tertiary education.

The history of Spanish in Paraguay is different from other countries. The main reasons are demographic: different colonization patterns plus a marked disparity in the number of Indians and Spanish settlers throughout colonial times. The development of Spanish in Paraguay is not a self-contained process but is closely related to the history of Guaraní. A full account of the historical and social processes involved in the configuration of the linguistic landscape of present Paraguay is presented in Chapter 7. For the time being, suffice it to say that nothing like a widespread Hispanicization took place in colonial times in Paraguay, because the cities remained the only strongholds of the European language until the late nineteenth century.

Spanish in Paraguay is highly uniform across geographical areas. In contrast, differences are important across social strata. This means that language variation in Paraguay is largely diastatic.²⁵ In this aspect Paraguay differs from Ecuador and Mexico, where variation is both diatopic and diastatic. On the other hand, the influence of native languages other than Guaraní on Spanish is minimal. Accordingly, differences in the sociolects of Paraguayan Spanish should be attributed to contact with Guaraní and the bilingualism of large sectors of the population.

5.3.1. Linguistic characterization of Spanish in Paraguay

Spanish in Paraguay is no doubt one of the most interesting dialects in terms of lexical and structural idiosyncrasies. Differences from other varieties may be so great in certain cases that intelligibility is compromised. The motivations and factors modeling the emergence of this unique variety boil down to contact with an indigenous language.²⁶ The Guaraní influence on Spanish has become a serious problem for educational policy makers in Paraguay²⁷ while it provides a fertile ground to test sociolinguistic theories and study the linguistic outcomes of bilingualism for linguists and other students of language.²⁸

Paraguayan standard Spanish is similar to other national standards in Latin America, with the difference that it is used only in formal situations involving administration, education and mass communication. That not every Spanish speaker in Paraguay is proficient in the standard demonstrates the width of the gap between

²⁵ The different forms of Spanish interlanguage spoken by indigenous speakers and first-generation immigrants are not included.

²⁶ In fact the influence between Spanish and Guaraní is reciprocal, so that both languages converge in quite a few aspects. In this context Melià (1998) proposes the existence of a third language in Paraguay, which is neither Spanish nor Guaraní. The idea has been challenged by several authors in Paraguay, for whom it is just another case of language contact in Hispanic America.

²⁷ From my analysis of the interviews in the *Atlas Lingüístico Guaraní-Románico* (2002) I conclude that the position of most Paraguayans towards the introduction of Spanish in Guaraní is tolerant to some degree while their attitude towards the introduction of Guaraní in Spanish is negative without exception. A number of authors (cf. Krivoshein de Canese 2000; Trinidad Sanabria 2005) consider this type of mixture and the resulting differentiation of Paraguayan Spanish from other national varieties a major obstacle for social and cultural development.

²⁸ Since the first in-depth study of bilingualism in Paraguay (Rubin 1968) multitude of papers and books have been published on this and other related topics. Various studies on linguistic and sociolinguistic aspects of Spanish in Paraguay appeared in the two volumes of *Sociedad y Lengua: Bilingüismo en el Paraguay* edited by Graziella Corvalán and Germán de Granda (1982). More recent studies are Dietrich (1995; 1996), de Granda (2000; 2004); Palacios Alcaine (2001; 2003) and Shaw (2004).

the colloquial language and the standard, which in Paraguay is larger than in other countries.

Table 5.3 lists the most salient features of colloquial Spanish in Paraguay. Following Krivoshein de Canese and Corvalán (1987:15) I call this variety ‘Colloquial Paraguayan Spanish’ (CPS) to distinguish it from the national standard (Paraguayan Spanish, PS). A large part of the data presented here comes from the contrastive study made by both authors between Spanish and Guaraní (*op. cit.*). The table includes the main features of CPS at phonetic, phonological, morphological, syntactic and lexical levels. It should be underlined that these features are *not* the outcome of imperfect language learning. Of these features, those of relevance for language contact are further elaborated in order to explain the phonetic shape of Spanish loanwords in the Guaraní corpus.

Table 5.3 Linguistic features of Colloquial Paraguayan Spanish

Level	Feature	Example
Phonetics and Phonology	• Aspiration (and eventual elision) of /s/ in coda position	<espinazo> ‘spine’ [espináso] → [ehpináso] → [epináso]
	• Aspiration [h] in positions where Old Spanish had the voiceless labiodental [f]	<huir> ‘to flee’ [fuir] → [hoyo]
	• Replacement of the voiced bilabial stop [b] with the fricative labiodental [v]	<burro> ‘donkey’ [bufo] → [vufo]
	• Pre-nasalization of the voiced bilabial stop [b] as [mb] in onsets	<bromista> ‘jester’ [bromista] → [mbromista]
	• Aspiration of the voiceless fricative labiodental [f] as [h]	<función> ‘function’ [funsión] → [hunsión]
	• Vowel /u/ realized as tensed central [i] in diphthong /ue/	<puerta> ‘door’ [puerta] → [pierta]
	• Elision of /d/, /s/, /n/, /l/, /r/ in word final position	<pared> ‘wall’ [pared] → [paré]
	• Elision of /n/ accompanied by nasalization of the vowel segment	<camión> ‘truck’ [kamión] → [kamiõ]
	• Insertion of glottal stop [ʔ] in intervocalic position	<caí> ‘I fell (off)’ [kaí] → [kaʔí]

	<ul style="list-style-type: none"> • Replacement (or eventual elision) of stop /t/ with /h/ in consonant clusters 	<Atlántico> ‘Atlantic’ [atlántiko]→[aɫántiko]→[alántiko]
	<ul style="list-style-type: none"> • Nasalization of velar /g/ in onsets as /ŋ/ 	<Miguel> [migel] →[miŋel]
Morphology	<ul style="list-style-type: none"> • <i>Voseo</i> 	<i>Vos</i> sos buena conmigo ‘You are good to me’
	<ul style="list-style-type: none"> • Use of Guaraní <i>nde</i> (2S) as a vocative 	¡Nde, qué cosa rara! ‘Hey, that’s weird!’
	<ul style="list-style-type: none"> • No gender distinction in pronominal clitics (<i>leísmo</i>) 	<i>Le</i> quiero a mi hijo ‘I love my son’
	<ul style="list-style-type: none"> • Doubling of pronominal objects, sometimes without agreement 	<i>Le</i> encontré <i>a ello</i> en casa ‘I found her at home’
	<ul style="list-style-type: none"> • Use of demonstratives before possessive adjectives 	<i>Aquel otro</i> tu hermano esta afuera ‘One of your brothers is outside’
	<ul style="list-style-type: none"> • Double marking of possession: possessive adjective + complement 	<i>Se</i> perdió <i>de mi</i> mi canasto ‘I lost my basket’
	<ul style="list-style-type: none"> • Replacement of definite articles with demonstrative adjectives 	<i>Este</i> padre de tu amigo vino hoy ‘Your friend’s father came today’
	<ul style="list-style-type: none"> • Lack of number and gender agreement between articles, adjectives and nouns. 	<i>Lo</i> ladrillo bien cocinada The well-cooked bricks
	<ul style="list-style-type: none"> • Use of Guaraní suffix <i>-kue</i> instead of Spanish prefix <i>ex-</i> 	La mi novia <i>kue</i> ‘My ex girlfriend’
	<ul style="list-style-type: none"> • Use of Guaraní quotative and reportative particles <i>ko</i> and <i>niko</i> 	Si es así <i>nikó</i> ya podé venir no más ‘If it is as you say, then just come’
	<ul style="list-style-type: none"> • Use of mitigating particles of Guaraní in imperatives: e.g. <i>na</i>, <i>mi</i>, <i>ke</i>. 	Quedatena un poco más conmigo ‘Please, stay a bit longer with me’
	<ul style="list-style-type: none"> • Use of Spanish <i>todo</i> ‘all’ to mark perfectivity. 	¡Tu hijo creció <i>todo</i> ya! Your son has grown!

	<ul style="list-style-type: none"> • Use of Guaraní verbal particles <i>kuri</i> and <i>ra'e</i> for recent and distant past 	Comí <i>kuri</i> con ellos 'I've just eaten with them' Cuando llegaste, yo salí <i>ra'e</i> 'When you arrived, I had long left'
Syntax	<ul style="list-style-type: none"> • Verbal forms of <i>voseo</i> 	<i>Vos sabés</i> que te quiero 'You know I love you'
	<ul style="list-style-type: none"> • Lack of prepositions in nominal complements 	Mandó hacer una <i>casa dos pisos</i> 'He had a two-storey house built'
	<ul style="list-style-type: none"> • Elision of the head noun in phrases whose complement indicates origin 	<i>De</i> Tobatí no son miedosos 'People from Tobatí are not afraid'
	<ul style="list-style-type: none"> • Use of prepositions for direct objects 	Me piso <i>por</i> el pie 'He treaded my foot'
	<ul style="list-style-type: none"> • Use of preposition <i>en</i> for motion verbs and stative verbs alike 	Me fui <i>en</i> la iglesia 'I go to the church'
	<ul style="list-style-type: none"> • Lack of possessive forms for relatives 	Ese es el hombre <i>que su</i> vaca se perdió. 'That is the man whose cow got lost'
	<ul style="list-style-type: none"> • Use of preposition <i>mediante</i> instead of <i>para</i> in causative constructions. 	<i>Mediante</i> que llovió creció el maíz 'Because it rained, the maize grew'
Lexicon	<ul style="list-style-type: none"> • Lexical items borrowed from Guaraní without phonological or semantic change 	<i>mita'i</i> 'child' <i>tajachí</i> 'policemen'
	<ul style="list-style-type: none"> • Lexical items borrowed from Guaraní with phonological change only 	<i>acané</i> 'fool' < <i>akāne</i> 'stinky head'
	<ul style="list-style-type: none"> • Lexical items borrowed from Guaraní with phonological and semantic change 	<i>ra'i</i> 'friend' < <i>ra'y</i> 'son'
	<ul style="list-style-type: none"> • Spanish lexical items with meanings calqued on Guaraní semantics 	<i>prestar</i> 'lend' = 'borrow, lend, use'
	<ul style="list-style-type: none"> • Hybrid words with morphemes from Guaraní and Spanish 	<i>platami</i> 'tip' < money+DIM(G) <i>yaguarear</i> 'betray' < dog+INF(Sp)

Not all features listed are exclusive of CPS. The aspiration of /s/ in coda position and the use of *voseo* with its corresponding verbal forms occur in several dialects of Rio de la Plata and other lowland and highland areas of the continent (cf. 5.1.1). In a similar way, *leísmo*, though less widespread, occurs also in Highland Ecuador (Zamora and Guitart 1982: 167). What makes these features characteristic of CPS is their widespread dialectal and geographical distribution. Thus, while the aspiration of /s/ occurs only in the lowland dialect of Ecuadorian Spanish it occurs in all sociolects of CPS. Similarly, *voseo* coexists with *tuteo* in Highland Ecuador, but its use is exclusive of other forms of second person singular in CPS. In sum, the aforementioned features characterize CPS as a national dialect. This is possible because there is a high degree of dialectal uniformity.

Some features listed in Table 5.3 are explained as internal developments of the language whereas others are products of internal and external factors. Examples of internal development are the retentions of older Spanish forms. The initial aspiration in a word like *huir* ‘to flee’ is documented for sixteenth and seventeenth century Spanish (Alarcos Llorach 1981: 257). Because this phenomenon is not reported for Guaraní, contact with this language cannot be the explanation in this case. In a similar way, the aspiration of the voiceless fricative labiodental [f] → [h] is explained as retention from older Spanish. However, the absence of [f] in Guaraní is a condition reinforcing aspiration in this case. Language-internal and language-external factors converge also in the elision of /d/, /s/, /n/, /l/ and /r/ in word-final position. This elision is attested in other varieties of Spanish (e.g. Antillean) but none of these sounds occur in coda position in Guaraní. In contrast, the bilabial-labiodental distinction /b/ - /v/ cannot be explained as archaic because such a distinction disappeared from peninsular Spanish before the time of the American conquest (Moreno de Alba 2004: 18). The adstratum influence from Guaraní is a decisive factor in this case because both sound are phonemic in this language. The remaining phonetic-phonological features of CPS are equally explained by contact with Guaraní. The pre-nasalization of the voiced bilabial stop is explained by the allophonic occurrence of /b/ as [mb] in Guaraní. The nasalization of /k/ as /ŋ/ is also allophonic in this language. The insertion of the glottal stop in intervocalic position is determined by the Guaraní rule prohibiting the diphthongization of vowels.²⁹ The nasalization of vowels after elided nasal segments resembles the elision of nasals in word-final position in other dialects, with the difference that the preceding vowel is

²⁹ Gregores and Suárez explain this phenomenon as follows: “The non-diphthongal transition is a special characteristic of Guaraní, particularly noticeable to Spanish speakers, because the diphthong is the more frequent transition between higher unstressed vowels and lower ones in Spanish. For instance, the Spanish word *piola* is always [pióla] in Spanish; when it occurs, as a loanword in Guaraní, it is [pi ó la]: there are two phonetic syllables and very little difference in prominence between the vowels (due to stressed [o]), and the last about twice as long as one vowel [...]” (Gregores and Suárez 1967: 54).

not nasalized as in CPS. The same process is involved in the phonological accommodation of Spanish loanwords ending with /n/. In all, evidence points to changes induced by contact with Guaraní.

In morphology, all the features of CPS – except for *voseo* and *leísmo* – are induced by contact with Guaraní. They can be classified in two types: those which calque Guaraní morphology and those which use Guaraní morphology. The first type includes: the lack of number and gender agreement in the noun phrase, induced by the non-marking of these categories in the indigenous language; the double marking of possession, modeled on similar construction in Guaraní; and the use of the Spanish *todo* ‘all’ to indicate perfectivity. The second type is illustrated by the use of demonstratives before possessive adjectives and the replacement of definite articles with demonstrative adjectives. Among the calqued features commonly interpreted as grammatical borrowings are the use of Guaraní particles *ko* and *niko* for quotative and reportative clauses, Guaraní mitigating particles for imperatives, past-tense particle *kue*, pronominal *nde* for second person singular vocatives, and particles *kuri* and *ra’e* for recent and distant past. The case of demonstratives is particularly interesting, because Guaraní borrows Spanish articles *la* and *lo* as demonstratives and uses them with more or less the same distribution as their native counterparts (Gómez Rendón 2007b). For this particular case – but also for several others – one may speak of convergence between colloquial Paraguayan Spanish and Guaraní through the reciprocal borrowing of elements from overlapping grammatical categories, with the semantic value they have in one of the languages, most likely the dominant language in the mind of bilingual speakers (i.e. Guaraní).

With the exception of verbal *voseo*, CPS syntax shows evidence of changes induced by contact with Guaraní. The mechanism in all cases is the same: the calquing of syntactic structures from Guaraní through elision of constituents or re-functionalization of native material. No influence of Guaraní on CPS word order is reported, probably because the indigenous language shows a relatively flexible word order and its most frequent pattern (SVO) overlaps with that of Spanish. This is the opposite to the situation in Ecuadorian Spanish, where a tendency towards verb-final order is induced by Quichua SOV pattern. The syntactic features listed in Table 5.3 are typical of CPS. However, similar constructions have been reported for Spanish interlanguages produced by native speakers of other Indian languages with typological characteristics similar to Guaraní. Such is the case of prepositions used in direct object complements or the simplification of the distinction *a-en* (Flores Farfán 2004). The difference lies on the fact that these (and other) features are *not* transitory outcomes of imperfect learning but have crystallized in the colloquial speech of the bilingual Paraguayan community.

The lexicon of GPS is remarkably influenced by Guaraní. The lexemes in Table 5.3 by no means exhaust all Guaraní lexical borrowing in CPS. In addition, there is a large number of native Spanish items whose meaning is calqued on that of

equivalent items in Guaraní. The case of *prestar* is one of the most interesting. The following examples are taken from Krivoshein de Canese and Corvalán (1987: 78f):

- 1) CPS: Estoy *prestando* este cuchillo ‘I am using this knife’
 PS: Estoy *usando* este cuchillo
 PG: *Aipuru* aína ko kyse
- 2) CPS: *Presté* este cuchillo de Pedro ‘I borrowed this knife from Pedro’
 PS: *Tomé prestado* este cuchillo de Pedro
 PG: *Aipuru* ko kyse Perúgui
- 3) CPS: *Préstame* un poco tu cuchillo ‘Please lend me your knife’
 PS: *Préstame* tu cuchillo por favor
 PG: *Eipurúkami* cheéve nde kyse

Only the third construction in CPS has an equivalent semantic meaning in PS while the other two are expressed in PS through a different verb (*usar*) and a verbal periphrasis (*tomar prestado*). From the Guaraní gloss it becomes clear that CPS is calquing Guaraní semantics. The result is one single word used with three different meanings, whereas standard (Paraguayan) Spanish uses three different words for each meaning. Figure 5.1 illustrates this ambiguity.

Figure 5.1 Meanings of *prestar* in Colloquial Paraguayan Spanish

CPS		PG		PS
prestar	→	puru	→	usar (to use)
			→	recibir en préstamo (to borrow)
			→	dar en préstamo (to lend)

These examples suffice to demonstrate that Guaraní influence on CPS is one of far-reaching consequences. Different from the outcomes of other situations of Spanish-Amerindian contact (e.g. Quichua or Nahuatl) the outcomes in this case are at all levels of linguistic structure. Considering formal schooling and socioeconomic status, it is true that sociolectal differences are many. Nevertheless, most features of Table 5.3 recur across CPS social strata. Guaraní influence is more extensive in varieties with a high degree of lexical and structural borrowing. These varieties make a case for language intertwining (Bakker 1994). They are colloquially known in Paraguay as *castení* (a hybrid of *castellano* and *Guaraní*). The following fragment from *Ramona Quebranto* (Ayala 1989), a novel written in *castení* gives an idea of the type of language mixing involved. Guaraní borrowings appear in italics.

- 4) Cuando baja agua limpiamo ¡rovya! Cualquiera no mira hata que pasa necesidá. ¡Dónde pa en otro parte alguno va repará por nosotros? La ecuelita de mi hijo oñeinundá, pero ndaipóri la problema, porque veterano kuera guerra Chaco preta su galpón a maetra, y santa pacua. [...] ¡Che memby kuéra trabaja má fácil aquí! Petei ovendé chicle calle Palmape; upe otro, canillita; ha otro, lutrabota centrope. ¡Naumbréna! Che aexplicá bien, pero no entendéi voi, porque su cabeza oiko en otra parte, y no e Chacariteña. (Margot Ayala 1989: 89ff)³⁰.

Notice that the whole novel is written in phonetic spelling for the purpose of capturing colloquial speech as accurately as possible. Several of the aforementioned phonetic features of CPS appear in the text, among others, the aspiration of /s/ in coda position in *hata* ‘until’ (< *hasta*). The text contains a few Guaraní loanwords (*rovya* ‘glad’, *peteĩ* ‘one’) and one function word (*upe* ‘that’). Grammatical borrowings include, among others, the plural marker *kuera*, the postposition *-pe*, the singular first-person pronoun *che* and the negation *-i*. Finally, there are two code switches to Guaraní, one verb phrase (*ndaipóri*, ‘there is not’) and one noun phrase (*che memby kuera*, ‘my sons’). Spanish native speakers cannot understand the above passage in its full meaning unless they speak Guaraní as well. The question is therefore how to classify this variety: ‘Guaraníticized’ Spanish or hispanicized Guaraní? In order to answer this question we must know the matrix language of the mixture. This has been undertaken elsewhere (Gómez Rendón, forthcoming/a).

The occurrence of these varieties along with more conservative ones suggest the existence of a dialect continuum between Spanish and Guaraní in similar terms to those proposed by Muysken for Highland Ecuador (Muysken 1985). This continuum has standard Paraguayan Spanish and standard Paraguayan Guaraní on its ends, with intermediate varieties according to their level of Hispanicization or Guaraníticization. This is shown in the following figure.

Figure 5.2. Spanish-Guaraní continuum in Paraguay

GUARANÍTICIZATION				HISPANICIZATION		
→	→	→		←	←	←
PS	CPS	CASTENI	~	JOPARA	CPG	PG

³⁰ “When waters recede, things become clear. We don’t care until we need. Who is going to take care of us? My son’s school was flooded, but that is not a problem because the Chaco veterans lend their building to the teacher and that is it! My sons work easier here. One of them sells chewing gum on Palma Street. The other sells newspapers. And the third works as a shoeshine boy in the downtown. It is something I don’t like much. I explained it to him once and again but he doesn’t get it, he is daydreaming somewhere else, he is not one from Chacareña slum anymore.” Free translation.

The arrows indicate the direction of the mixing process. At one point the mixture becomes so enmeshed as a result of borrowing and codeswitching that we cannot tell which language provides the morphosyntactic matrix. Intermediate varieties are essentially instable mixed lects that might crystallize as a distinct third language (Melià 1975). While our analysis confirms in part the emergence of a third language, it is by no means conclusive and further research on Spanish-Guaraní mixing is required. Hispanicized Guaraní varieties, which make the second part of the continuum, are analyzed in Chapter 7.

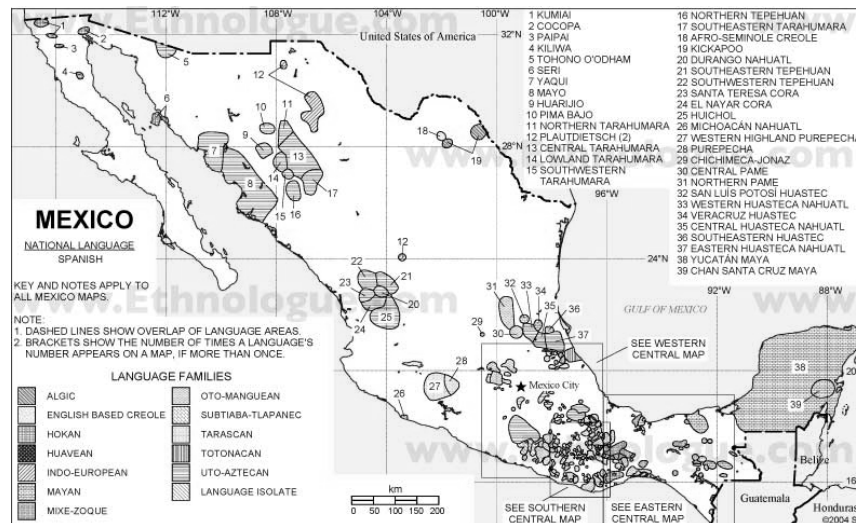
5.4. Spanish in Mexico

Spanish is the only official language in Mexico. The vast majority of the country's population speaks Spanish as their first or second language. By 1997 Spanish speakers in Mexico amounted to 94,275,000, which correspond to 98.5 % of the total population.³¹ The ongoing process of Hispanicization leads to assume that this percentage was even higher by 2005, for which date Mexico's population was 106,147,000 according to CEPAL estimations. Mexico is therefore the country with the largest Spanish-speaking population in the world, the bulk of which lives in the central and northern plateaus and the Caribbean and Pacific coasts. The central and northern plateaus are home to the three biggest cities in Mexico concentrating one third of the country's population: Mexico City in the central plateau, with a metropolitan area of 20,000,000 million speakers (20% of Mexico's population); Guadalajara, also in the central plateau, with a population of 4,300,000 inhabitants; and Monterrey, in the northern plateau, with an estimated population of 4,200,000. The unbalance between the cities and the countryside in demographic terms is more noticeable in Mexico than in the other countries as a result of the rampant levels of uninterrupted urbanization since the end of the nineteenth century. Like in Ecuador, the urbanization process in Mexico led to the racial miscegenation and Hispanicization of rural immigrants in the cities. This explains why even native speakers of Amerindian languages speak Spanish nowadays. The bulk of rural immigration to the cities is composed of Indian people who speak Amerindian languages. Spanish in Mexico coexists with about sixty languages from ten different families, including Uto-Aztecan, Otomangue, Mayan, Tarascan, Totonaco-Tepehua, Mixe-Zoque, Tequistlateco-Jicaque, Huave, Yuma-Seri and Algonquian. In 1997 Amerindian speakers above five years of age amounted to 6,044,547, out of a national population of 97,483,412 inhabitants. In all, Amerindian speakers represented 6.2% of the country's population. Of them, roughly 82% were bilingual

³¹ From Jaime Otero, *Demografía de la lengua española*, in the 1999 *Anuario* of Centro Virtual Cervantes. This figure does not include Mexican-origin immigrants in the United States, most of which maintain Spanish as their language in domestic and community settings.

in Spanish (4,924.412) with different degrees of proficiency, and 18% (1,002.236) were monolingual. According to Ortiz Álvarez (2005: 65), only one percent of the total Mexican population did not speak Spanish by 2000.³²

Map 5.3. The Languages of Mexico



Source: *Ethnologue* 2005

Ortiz Álvarez (2005: 74) notes that bilingualism among Amerindian speakers increased steadily during the twentieth century, with a yearly average growth of 1.6%. From 1930 to 2000 the indigenous bilingual population increased from 1,065.924 to 4,924.412 (i.e. 362% in 70 years)³³. These figures confirm an unchecked process of Hispanicization.

Bilingualism and monolingualism are different depending on gender and age. Most Amerindian monolinguals are found among elders and the number of monolingual women is generally higher than the number of monolingual men. This is reflected also in lower rates of literacy among women as compared to men. Differences in bilingualism are also important. From case studies such as Hekking (1995) and Hekking and Bakker (2005) one concludes that an important number of bilinguals are subordinate: their command of Spanish is limited to oral communication in informal settings while their reading and writing skills in the

³² Compare Quichua monolingualism in Highland Ecuador, estimated about 8,7% in 1993 (Buttner 1993: 69). Quechua monolingualism is much higher in Peru and Bolivia.

³³ The southern states of Oaxaca, Chiapas and Veracruz have the largest concentration of bilinguals in the country. In addition these states show the largest number of monolinguals.

language are minimal. Most bilinguals of Indian descent speak a variety of Spanish with interferences of their respective languages and their knowledge of standard Mexican Spanish is poor. In general the Spanish of Amerindian speakers is known as 'Indian Spanish'. Indian Spanish varieties are generally stigmatized and become an obstacle for the social mobility of their speakers (Flores Farfan 2000). Still, linguistic features characteristic of Indian Spanish have entered regional varieties of monolingual Spanish in predominantly indigenous areas. (e.g. Comiteco Spanish in Chiapas).

Sociolinguistically, Mexico does not differ from other Hispanic American countries, if perhaps for the number of Amerindian languages in contact with Spanish. Like in any other corner of Hispanic America, the knowledge of Spanish in Mexico provides an easier access to public services and clears the way for the effective participation in the market economy. In sum, Spanish is the socially and politically dominant variety. Spanish pervades education, administration and the media. The rapid integration of non-Spanish speakers to the national society through early Hispanicization is a factor common to all educational policies implemented in Mexico since the Independence, especially during the *Porfiriato* (1886-1911) and the post-revolutionary period.³⁴

Dating back to the 1960s, indigenous bilingual education is older in Mexico.³⁵ There are bilingual programs at local and district levels (e.g. Otomí-Spanish bilingual education in the state of Hidalgo) but most of them work on their own, in the absence of an encompassing national policy. With a few exceptions, these programs have been mostly *transitional*, because they view indigenous languages as instruments to help pupils acquire literacy skills and basic knowledge while adapting to the Spanish-speaking society. Only the last years have witnessed an emerging awareness among policy maker about the need of bilingual schooling to match the goals of national education and language maintenance³⁶.

³⁴ José Vasconcelos (1882-1959) is the undisputed epitome of this integrationist ideology in Mexico. Similar views in other Hispanic American countries prevailed from the second half of the nineteenth century (e.g. Sarmiento's ideas in Argentina). Stating the importance of biological, cultural and linguistic miscegenation (*mestizaje*) for the Latin American republics, these ideologies set in motion state apparatuses for the integration of non-Hispanic ethnic groups. In most cases the result was the Hispanicization of ethnolinguistic minorities. In others the outcome was their physical extermination.

³⁵ In most Andean countries indigenous bilingual education began in the late seventies. Legislation on bilingual education was passed in 1978 in Colombia and one year later in Venezuela. In Ecuador the first law on bilingual education dates from 1981. Peru passed a law in 1984. Bolivia is still waiting a law, even though it is the country with the largest indigenous population in South America. For an overview of bilingual education programs in the Andes, see Adelaar and Muysken (2004: 606ff).

³⁶ To acknowledge the rights of language minorities and remedy the chaotic situation of indigenous education (motivated by the large number of speech communities and their divergent interests), the Mexican Senate passed a law on the linguistic rights of indigenous

The Hispanicization of native peoples in Mexico went hand in hand with the colonization of its large territory. Spanish colonization of Mexico began relatively early in comparison to South American countries. The conquest of the Aztec Empire commenced in 1519 and ended two years later with the final takeover of Tenochtitlan. From 1521 a series of successful enterprises allowed the Spaniards to seize a vast extension of land extending from the north of today's Texas in the United States through the Mexican plateau and Central America down to Panama City. These lands formed the Voce-royalty of New Spain in 1535. The process of Hispanicization was not that rapid. The cultural and linguistic isolation of ethnic groups under the protection of missionary orders which promoted the use of native languages slowed down the expansion of Spanish in the first century. To clear the way for their language, Spaniards used native institutions to their advantage. They knew well that any project of Hispanicization in a stratified society would meet with failure unless the top of the societal ladder becomes involved. In these terms, the role of the indigenous elites was decisive for colonization. Spaniards took children of the local elites to special education centers in which they could be immersed in the language and culture of the conquerors. The best known of these centers was *Colegio de Santa Cruz de Tlatelolco*, founded in Mexico in 1536, which received 60 pupils of the Indian nobility only in the first year. After a long period of intensive instruction, Indian trainees were sent back to their communities to serve as brokers between the Crown and their people, and agents of cultural and linguistic change. This strategy maximized the efforts of the Spaniards to homogenize the enormous linguistic and cultural diversity they found in Mexico. In this process of homogenization a few Amerindian languages were used as *lingua franca* for evangelization and interethnic communication. The most important of these languages was Nahuatl. It became so widespread in certain areas that it ended up by replacing vernacular languages in few decades.³⁷ These vehicular languages leveled linguistic variation in early colonial times. When the Bourbon reforms banished their use from administration and education in 1770, these and other minor languages had been almost replaced by Spanish. The process varied from one area

peoples in 2003. In addition to giving official status to the indigenous languages in their respective areas of influence and creating a special agency for the protection of ethnolinguistic rights (*Instituto Nacional de Lenguas Indígenas-INALI*), the law amended two articles of the Constitution about the multiethnic composition of the Mexican society and the obligatory nature of bilingual education for indigenous children of elementary school. Although it is too early for an evaluation, similar experiences in other Hispanic countries show that the recognition of ethnolinguistic minorities must be accompanied by concurrent processes of language revitalization, production of pedagogical materials and extensive training of bilingual instructors unless the law is to remain in writing.

³⁷ A similar replacement of vernacular varieties by a *lingua franca* (Quichua) fostered by missionaries is found in the Ecuadorian Andes. For a discussion of this replacement in Imbabura, see Gómez Rendón and Adelaar (forthcoming).

to another, depending on a number of geographic, demographic and sociopolitical factors. But language homogenization did not attain its ultimate goals of shift and leveling at a national level. Over sixty Indian languages exist in present Mexico, and their influence on local Spanish is not unimportant. Indian varieties of Spanish continue to emerge at the rate of Hispanicization and crystallize as native varieties once their speakers are acculturated. Many features of regional Spanish in Mexico are the product of substratum and adstratum influence from Amerindian languages.

5.4.1. Dialects of Spanish in Mexico

Three authors have proposed different classifications of Spanish dialects in Mexico, mainly on the basis of phonetic features. The first classification was presented by Henríquez Ureña (1934). He identifies five dialectal areas, excluding the southwest of the United States. These areas include 1) central Mexico; 2) northern Mexico; 3) the lowlands of the Mexican Gulf which connect to the Pacific lowlands through the Isthmus of Tehuantepec; 4) Yucatan; 5) Chiapas (and most Central America). Rona (1964) is the author of another classification. He makes a fundamental distinction between the Spanish spoken in the states of Chiapas, Tabasco, Yucatan and Quintana Roo (corresponding in Henríquez' division to the third, fourth and fifth areas, respectively) and the dialect spoken in the rest of the country, which he considers rather homogeneous. On the other hand, the classification of Zamora and Guitart (1982) classifies Mexican dialects in three distinct areas: eastern coast of Mexico; central and northern Mexico; and the southern Mexican states on the Guatemalan border. These classifications differ mainly in the number of dialects but agree in other respects, namely: a) central Mexico has a clearly identified dialect; b) the eastern lowlands make also a well-defined dialectal area; and c) the dialects of southern and southeastern Mexico are different from the rest of the country. These distinctions are preserved to some extent in the latest classification of the *Atlas Lingüístico de México* (1990-2001), which upgraded the methodological criteria of previous attempts and obtained more accurate results.

In each of the major dialectal zones (central Mexico; eastern Coast and southern Mexico) the Amerindian influence has its own contribution. Thus, while some Amerindian substratum is present in central and southern Mexico, this is far more important in the south. In contrast, the Amerindian substratum in the eastern Coast is comparatively minor. The Amerindian contribution is not decisive either in central Mexico, where a moderate Nahuatl substratum combines with other factors (e.g. geographical isolation) to explain the present configuration of dialects. The following table summarizes the main phonetic characteristics of the three dialectal areas according to phonetic and morphosyntactic parameters:

Table 5.4. Phonetic and morphological features of Mexican Spanish dialects

Dialect zone	/s/	/x/	Voseo
Central Mexico	fricative alveolar	Velar	+
Eastern Coast	aspirated or dropped	Glottal	-
Southern Mexico	dropped or aspirated	Glottal	+

The first criterion separates the dialects that aspirate or drop /s/ in coda position (eastern Coast and southern Mexico) from the dialects that maintain /s/ as a fricative alveolar. The second criterion concerns the velar realization of /x/ in central Mexico as opposed to the glottal realization of the same phoneme in eastern Coast and southern Mexico. A third criterion distinguishes the areas that use pronominal and verbal *voseo* (central Mexico and southern Mexico) from the areas that do not use *voseo* in any form (eastern Coast). Other distinctions are less clear. For instance, all the areas are characterized as *yeistas* (/ll/ → [y]), the realization of /y/ is diatopical, with a fricative alveolar [ʝ] in Oaxaca and a relaxed open fricative /yⁱ/ in the rest of the country (Moreno de Alba 2002: 111). Similarly, vowel relaxation is characteristic of the central plateau, though exclusively, because others (e.g. the Pacific Coast) tend to relax /i/ and /o/ as well. In the same way, the velarized realization of /n/, not attested in central Mexico or the eastern Coast, is not typical of the entire southern area but only of Yucatan and Chiapas (Moreno de Alba 2004: 216).

Since the early years of the conquest the central plateau became the scenario of intensive contact between Spanish and Amerindian languages of Uto-Aztecan, Otomangue and Tarascan families. The most important of these languages in terms of vitality and lexical contribution to Spanish is Nahuatl. Nevertheless, from an extensive study of lexicography Lope Blanch (1969) showed that Amerindian loanwords are less numerous than often assumed:

“Podríamos afirmar que los indigenismos de uso general en el español de México ascienden, en total, a la cantidad de 156 vocablos, correspondientes a 121 lexemas; sumando a ellos las voces de uso o conocimiento parcial se llegaría a 245 vocablos y 186 lexemas. Cantidades no despreciables, por cierto, pero tampoco tan elevadas como para suponer que su desaparición “produciría un caos verdaderamente horrible” en el habla mexicana, según creía D. Rubio (cf. *supra*, n. 15)” [We may state that Indian loanwords of widespread use in Mexican Spanish amount to 156 items corresponding to 121 lexemes. Add to this number the loanwords of partial use, and we have 245 words and 186 lexemes. While these numbers are not unimportant, they are not so important that their fall in disuse would produce a really disturbing effect in Mexican Spanish, as D. Rubio thought] (Lope Blanch 1969: 49; my translation).

Of the 156 items identified by Lope Blanch, 141 come from Nahuatl. Other Amerindian language families represented are Maya (9), Tarascan (5) and Otomí (1). Only 95 of all the loanwords are known in all Mexico while the rest are used by smaller sectors of the population in the cities or the countryside. Despite the rigorous analysis of the data by Lope Blanch, two remarks need to be made:

- 1) The sample studied by Lope Blanch was collected in the capital, and consequently his findings have a limited scope;
- 2) Amerindian loanwords will continue to enter local and regional varieties of Spanish as the process of Hispanicization advances in different areas of the country. These areas include Spanish varieties which are the product of imperfect learning by Amerindian speakers³⁸ and therefore show the crystallization of lexical items from their first languages. These areas were not considered by the study of Nahuatl loanwords by Lope Blanch. His findings reflect only the use of Spanish monolinguals in the capital city.

5.4.2. Spanish in Querétaro

The state of Querétaro is located at the heart of the central plateau (*Mesa Central de Anahuac*). It has an extension of 11,499 square kilometers. The state capital (Querétaro) is located 211 kilometers north of Mexico City. The capital is home to 615,850 people representing two thirds of the state population (962,470). The great majority of people are Spanish monolingual and only a small part (22,000) bilingual in Spanish and Otomí (Ortiz Álvarez 2005: 55).

There are two studies on the Spanish spoken in Querétaro. One is due to Muñoz Ledo y Mena (1934). The other is the aforementioned *Atlas Lingüístico de México* (1990-2001), which includes an individual section for this area. From a comparison of the data of both sources, it is clear that the phonetic features of Querétaro Spanish have remained intact. Because there is only fragmentary information in both studies about the lexicon and other aspects of grammar, the following discussion deals exclusively with the phonetics.

Two of the most salient features of the central plateau are present in Querétaro: the non-aspiration of the fricative alveolar /s/ and the velar realization of /x/. Five additional features are reported for this variety, as shown in the following table.

³⁸ The same applies to the phonetic and grammatical influences of Amerindian languages on Mexican Spanish. Lope Blanch (1972) states that excepting the morpheme *-eco*, extensively used in toponyms, Nahuatl influence on Mexican Spanish is not certain because all of the often adduced Amerindian traits were present in previous stages of Peninsular Spanish and occur in other parts of Hispanic America where Nahuatl is not present, such as the Antilles and the Andes. For a realistic evaluation of these statements, it is necessary to conduct research on local varieties of Spanish heavily influenced by Amerindian substratum which are often spoken in areas far from the principal urban centers.

Table 5.5. Phonetic features of Spanish in Querétaro

Feature	Example
relaxation or elision of unstressed vowels: /e/ → [ə] ~ [Ø] /o/ → [ə] ~ [Ø]	<pues> [pəs] ~ [ps] <oscuro> [əskuro] ~ [skuro]*
average devoicing or elision of intervocalic /d/ /d/ → [d] ~ [Ø] if V_V	<cerrado> [seʔaðo] ~ [seʔao] <enchilada>[enčɪlada]~[enčɪla:]
merger of intervocalic /ll/ and its eventual elision /ll/ → [y ⁱ] if V_V /ll/ → [Ø] if V_V	<ardilla> [ardiy ⁱ a] ~ [ardía]* <tortilla> [tortiy ⁱ a] ~ [tortía]*
sibilant fricativization of /r/ in coda position /r/ → [r̥] if V_#	<comer> [komeɾ̥]
full realization of cluster /kt/ with sonorization of /k/ /kt/ → [gt] ~ [kt]	<actor> [aktoɾ̃] ~ [agtoɾ̃]
* Occurrence restricted to bilinguals	

These features are not exclusive of Querétaro Spanish. Most of them occur all over the high plateau, even though their realization is particularly marked in Querétaro. As regards the relaxation of unstressed /e/ and /o/, Moreno de Alba (1989: 41) suggests that this feature is regular and perceptible in the entire central plateau while occurring occasionally in the eastern Coast (states of Veracruz, San Luis Potosí and Tamaulipas). About the generally assumed Amerindian substratum of this phonetic feature, Lope Blanch states that no clear link may be traced between the relaxation of unstressed vowels and the influence of Nahuatl because a similar relaxation is not reported for Nahuatl itself while it occurs in several areas of the Andes.

The second feature (devoicing or eventual elision of intervocalic /d/) is more intriguing for a dialectal classification, because it is typical of the Eastern coast and has a clear peninsular origin. Its occurrence might be explained as a phonetic ‘leftover’ of a set of Andalusian features present in early colonial times. However, even if right, this interpretation does not answer why the phonetic feature in question took root in Querétaro and not in other areas.

The third feature is interesting from a language contact perspective. The realization of /ll/ in most dialects of Mexican Spanish is [yⁱ] (Moreno de Alba: 2002: 113). If /ll/ is preceded by /i/ like in *tortilla* and *ardilla*, two phenomena may occur: either the vowel merges with the approximant to produce the segment [y] or the approximant coalesces with the vowel to produce the segment [i]. Rosenblatt (1967: 117) notes that the relaxation and eventual elision of /y/ in intervocalic environment is typical of lowland dialects. However, Querétaro Spanish is spoken in the highlands. We are therefore left with two possible explanations for this

phenomenon: the elision is either an Andalusian remnant or an outcome of imperfect learning of Spanish by Amerindian speakers. A piece of evidence for the second explanation is provided by Muñoz Ledo (1934: 105; 106), who found the coalescence of the vowel with the approximant in the Spanish speech of Otomí Indians from Querétaro. Additional support for this interpretation is the fact that similar mergers occur in the Spanish of indigenous speakers from other areas of the high plateau.

Another feature which deserves some comment here is the sibilant-like fricativization of /r/ in coda position. According to Moreno de Alba (2004: 130ff) the distribution of this feature corresponds roughly to the area of central Mexico, which is characterized by a strong articulation of consonants (Sp. *fuerte consonantismo*) as opposed to the eastern Coast. He supports this view with data from the *Atlas* and points out that sibilant-like fricativization of /r/ does not occur in any of the places where /s/ is aspirated or elided (a feature associated with prominent vowels and dark obstruents). While some authors like Malmberg (1948) put forward a Nahuatl substratum for this feature on the basis of its highland distribution, Lope Blanch (1972) considers this hypothesis erroneous for several reasons. First, the Amerindian language before contact had neither trills nor flaps. Second, the fricativization of /r/ is only one of several realizations of this phoneme in Mexico. And third, the frequency of fricativized /r/ is comparatively lower than the frequency of other realizations. For Lope Blanch the fricativization of the vibrant is not dialect-specific but associated with certain emphatic registers. However, the contact hypothesis cannot be overlooked. For its widespread distribution the fricativization of /r/ may be due to substratum influence from an indigenous language, not necessarily Nahuatl. According to Suárez (1983: 46), twelve languages of a sample of 38 from different Mesoamerican families do have vibrants. Tarascan has a set of two vibrants while Otomí has one vibrant phoneme (Hekking 1995: 31).³⁹ Both are spoken in the central plateau, specifically in areas where the fricativization of /r/ shows the highest frequency.

I am specific in the description of several phonetic phenomena of Querétaro Spanish in order to show their possible Amerindian substratum, but also because they influence the borrowing and accommodation of Spanish loanwords in Otomí (cf. Chapter 10). Thus, the relaxation of unstressed vowels (e.g. *escuela*) may produce target forms in Otomí with vowel elision (e.g. [skuela]).

³⁹ Hekking notes, however, that the trill /rr/ occurs only in Spanish borrowings, e.g. *burro*. No mention is made about whether the vibrant is fricativized or not.

5.5. Spanish: a typological characterization

This section takes as its point of departure the premise that Spanish is one and the same across continents and countries, regardless of dialectal, sociolectal and other variations. The evolution of Spanish in the Americas did not modify its typological nature but enriched the language in ways nobody could imagine before 1492. Even if Spanish is not a monolithic, indivisible entity – not even within the sociopolitical space of Latin America – Spanish remains to date one of main agglutinating factors in the continent. The present section is thought as a linguistic complement to the social and historical events described in section 5.1.1. To the extent that Spanish is the source language in the three borrowing situations studied in this book, its linguistic description will be a solid basis for the analysis of typological constraints on borrowing. The following characterization focuses on morphological typology and parts of speech but includes other parameters such syllable structure, type of affixation, order of constituents, alignment and so forth.

Genetically, Spanish is a language of the Romance branch of the Indo-European family, akin to Portuguese, French and Italian, all of them direct heirs of Latin, with which they share a number of typological characteristics. Spanish originated in the Castilian plains. From there it expanded first throughout the Peninsula since the twelfth century, during the Christian Re-conquest. Later, since 1492, it spread to the five continents in the context of the Spanish colonization.

Phonetically, Castilian Spanish has twenty-three distinct segments, of which eighteen are consonant and five vowels. While the number of vowels is the same for all varieties of Spanish – with certain differences such as vowel relaxation or elision, for example, in Ecuador and Mexico – the number of consonants in Andalusian and American varieties goes down to sixteen. The reason is the loss, in these dialects, of the dental-alveolar distinction /θ/-/s/ and the lateral-approximant distinction /ʎ/-/y/. Both phenomena are characteristic of the vast majority of American varieties of Spanish and were amply discussed in section 5.1.1. The typical Spanish syllable is open (CV), though several other sequences are possible too, the most typical being CVC (e.g. *tan-to*) and CCV (e.g. *tra-bajo*). Diphthongal syllabic nuclei are also very frequent in Spanish, especially those of rising type (second segment stressed). Onsets may be simple or complex. Simple syllabic onsets show no restriction while complex ones are only of the type occlusive-plus-flap. In contrast, codas show a larger number of restrictions, especially in word-final position (e.g. stops are not permitted in coda position at the end of a word).

As regards morphological typology, Spanish is a typical – though not prototypical – example of a fusional language. It shares this characteristic with other languages of the Romance branch due to their common origin in Latin, a highly fusional language. Spanish words usually contain more than one morpheme. However, morphemes in a word do not correspond to the linear sequence of morphs

in this word (Crystal 2006: 194). As a result, the identification of morphological segments is often unfeasible by the fusion of features in one single morph.⁴⁰ This is all the more evident in the Spanish verb phrase. Let us consider the verbs in the following sentence.

- 5) *él* *no* *quiso* *que* *vinieran*
 3S.MASC NEG want.PST.PRF.3S that come.PST.PRF.SBJ.3PL
 ‘He did not want them to come’

The verbs of the main clause (*quiso*) and the subordinate clause (*vinieran*) are morphologically complex: they contain several morphemes which indicate number, person, tense, aspect and mood. Plural number in the subordinate verb is expressed by /-n/, the lack of which in the main verb indicates singular. Person is not indicated by separate morphs, but the same morph for number serves this purpose, i.e. one morph stands for two features. Tense, aspect and mood are even more difficult to assign in morphological terms. The only possible way to know that *quiso* is a perfective form is by comparing it to the verb stem (*quer-* ‘want’) and assigning the former to a specific paradigm of aspect. The same procedure applies to the verb *vinieran*, the stem of which is *ven-* ‘come’.⁴¹ Notice that aspect and tense are closely related in Spanish, so that *quis-* and *vini-* indicate also past tense. The morphological identification of mood is not less complex. The subjunctive in (5) can be roughly assigned to the bound morph *-era*, but this assignment depends on the tense of verb. The morpheme is different when the verb is in present tense (e.g. *veng-a-n*, come-PRS.SUBJ-3PL).

The rich verb morphology of Spanish allows the optional suppression of the pronoun subject in a sentence. This is typical of pro-drop or null-subject languages. Because verbal endings usually enable the identification of subjects without further marking, personal pronouns are used mainly for emphasis and contrast.

These examples show the intricacies of Spanish verb morphology and illustrate the fusional character of this language. Fusion is present in other word classes such as articles and pronouns (plus pronominal clitics). The Spanish article deserves some description for it occurs as a grammatical borrowing in Guaraní. Developed from Latin demonstratives, Spanish articles not only indicate definiteness and number but also grammatical gender, and must concord with nouns in these features. Consider the following examples:

⁴⁰ Notice my use of the terms ‘morpheme’ and ‘morph’. The former refers to the form and the semantic feature together, while the latter refers only to the form.

⁴¹ In fact, both *quis-* and *vini-* are suppletive forms in their corresponding paradigms.

- | | | | | |
|----|-------------|---------------|-------------|--------------|
| 6) | <i>el</i> | <i>niño;</i> | <i>la</i> | <i>niña</i> |
| | DEF.S.MASC | boy | DEF.S.FEM | girl |
| | 'the boy'; | | 'the girl' | |
| | | | | |
| 7) | <i>los</i> | <i>niños;</i> | <i>las</i> | <i>niñas</i> |
| | DEF.PL.MASC | boy.PL | DEF.PL.FEM | girl.PL |
| | 'the boys' | | 'the girls' | |

Notice that only the plural morpheme /-s/ can be segmented, even though the grammatical features involved are three (definiteness, gender and number). Because the Spanish article is a deictic itself, it cannot be preceded or followed by deictic forms such as demonstratives. This makes constructions like (8) ungrammatical.

- | | | | |
|----|----------------|------------|----------------|
| 8) | <i>Ese</i> | <i>el</i> | <i>hombre*</i> |
| | DEM.DIST | DEF.S.MASC | man |
| | 'That the man' | | |

Possessive adjectives cannot precede or follow articles either. This is shown by the ungrammatical noun phrase in (9). To indicate possession, either a possessive adjective follows the noun (9a), or a possessive adjective precedes it (9b):

- | | | | |
|-----|-------------------|-------------|--------------|
| 9) | <i>La</i> | <i>tu</i> | <i>casa*</i> |
| | DEF.S.FEM | ADJ.POSS.2S | house |
| | 'That your house' | | |
| | | | |
| 9a) | <i>La</i> | <i>casa</i> | <i>tuya</i> |
| | DEF.S.FEM | house | PRO.POSS.2S |
| | 'Your house' | | |
| | | | |
| 9b) | <i>Tu</i> | <i>casa</i> | |
| | PRO.POSS.2S | house | |
| | 'Your house' | | |

According to the Principle of System Compatibility (section 3.6.1) Spanish as a fusional language may borrow practically any form-meaning unit from any type of language, since no restrictions exist to morphological compatibility. In contrast, languages of other types (e.g. agglutinating or isolating) can borrow from Spanish only certain types of items depending on their compatibility: agglutinating languages can borrow independent words, roots and one-meaning affixes; isolating languages can borrow independent words and roots. These restrictions will be incorporated in

formulating the language-specific hypotheses on borrowing for the three languages of this study.

The preceding examples showed that Spanish verbal morphology is based on suffixation (i.e. bound forms attach to the end of stems or roots). Indeed, the entire Spanish inflectional morphology is made up of suffixes. Prefixes belong mostly to derivational morphology. Notice that the Spanish noun phrase shows less morphological fusion than the verb phrase. Compared to verb morphology, noun morphology is rather simple in this language. The reason for this simplification is the lack of morphological cases and the replacement thereof with a rich set of prepositions.

Spanish has a considerable number of simple and complex prepositions. These are prepositional periphrases in which basic prepositions combine with nouns to form a prepositional constituent (e.g. *de acuerdo con* 'in accordance with'). In certain grammatical frameworks, the half-open nature of prepositions in Spanish supports their treatment as function words of lexical nature, i.e. items positioned in between lexicon and grammar. In Spanish, prepositions are a salient typological feature determining its degree of analyticity.

In the noun phrase, possession is indicated either by a set of possessive adjectives inflected for person and number, or by the use of preposition *de*. The following examples illustrate both types of possessive constructions:

- 10) *Mi patria es tu hogar*
 ADJ.POS.1S fatherland be.PRS.IND.3S ADJ.POSS.2S home
 'My fatherland is your home'
- 11) *El nieto de Antonio*
 DEF.S.MASC grandson of Antonio
 'Antonio's grandson'
- 12) *Gente del campo*
 People of:DEF.S.MASC countryside
 'People from the countryside'
- 13) *Muros de piedra*
 Wall.PL of stone
 'Stone walls'

Preposition *de* is used to link non-possessive modifiers in the noun phrase. In (12) and (13) the nouns headed by the preposition refer to origin and material, respectively. Clausal modifiers in Spanish are linked by a number of relative pronouns (e.g. *que*, *cuyo*) and adverbial conjunctions (e.g. *donde*, *cuando*).

Relativization in Spanish is the most frequent clause-linking strategy. Coordination and subordination are accomplished by a series of connectives including simple conjunctions (e.g. *y, o, si, como, porque*) and a closed set of simple and complex adverbial conjuncts (e.g. *así, ya que, desde que*). The extensive use of connectives reinforces the Spanish preference for hypotactic constructions. Indeed, Spanish hypotaxis is diametrically opposed to the parataxis characteristic of languages like Guaraní.

Word order in Spanish is rather flexible. (S)VO is the unmarked word order in declarative sentences (14). Other orders are used with pragmatic value. Further mechanisms of clause dislocation include topic fronting (15) and cleft sentences (16).

- 14) *El campesino trabaja la tierra*
 DEF.S.MASC peasant work.PRS.3S DEF.S.FEM land
 'The peasant works the land'
- 15) *La tierra la trabaja el campesino*
 DEF.S.FEM land PRO.3S.ACC work.PRS.3S DEF.S.MASC peasant
 'The land is worked by the peasant'
- 16) *Es la tierra la que trabaja el campesino*
 be.PRS.3S DEF.S.FEM land DEF.S.FEM.ACC REL work.PRS.3S
 DEF.3.MASC peasant
 'It is the land which the peasant works'

The fronted topic in (15) does not have any marker indicating this function. Instead, the speaker uses an accusative clitic pronoun concordant with the fronted noun in number and gender. The strategy in (16) consists of a complex relative structure made up of the article and the relative pronoun in accusative case. These examples show clearly that alignment in Spanish distinguishes accusative arguments either morphologically (through pronominal clitics and prepositions) or syntactically (post-verbal position in declarative sentences) while subjects and agents are both unmarked.

The System of Parts of Speech in Spanish

I base the following description of the parts of speech in Spanish on the typology proposed by Hengeveld (1992) and Hengeveld *et al* (2004). The identification of Spanish along the scale of parts of speech is crucial to the analysis of lexical

borrowings and the formulation of language-specific hypotheses about the type of lexical classes borrowed and their use in the recipient language.

Spanish is a language with a differentiated parts-of-speech system (Type 4). It has individual lexical classes for each of the syntactic slots in predicate and referential phrases. Lexical classes in Spanish include verbs (used as heads of predicate phrases), nouns (used as heads of referential phrases), adjectives (used as modifiers of referential phrases), and manner adverbs (used as modifiers of predicate phrases).

The first distinction in the system of parts of speech of Spanish is based on morphological criteria. Spanish nouns and adjectives are usually marked for number (e.g. plural *-(e)s*) and gender (e.g. *-a* feminine, *-o* masculine), as shown in (17) and (18). In turn, verbs are marked for number but *not* for gender, and most importantly, finite verbs are *always* marked for tense, aspect and mood, as illustrated in (19). None of the latter markers occurs on nouns and adjectives.

- | | | |
|-----|--|---|
| 17) | a) <i>comprador</i>
buyer.MASC-S
‘seller (man)’ | b) <i>comprador-es</i>
buyer.MASC-PL
‘sellers (men)’ |
| 18) | a) <i>solitari-o</i>
lonely-MASC.S
lonely (man) | b) <i>solitari-a-s</i>
lonely-FEM-PL
‘lonely (women)’ |
| 19) | <i>compr-aba-n</i>
buy-PST.IMPF-3P
‘They (men or women) used to buy’ | |

A second distinction, based on morphological criteria as well, separates nouns, verbs and adjectives from manner adverbs. The majority of manner adverbs originate in adjectives, being derived from them by the suffix *mente*: e.g. *casual* ‘coincidental’ > *casual-mente* ‘coincidental-ly’; *serio* ‘serious’ > *seria-mente* ‘seriously’.

On the other hand, nouns and adjectives share a good part of morphology but they are different in two important aspects: first, nouns have intrinsic gender while adjectives do not; second, only adjectives can modify referential phrases while the great majority of nouns cannot.

As regards compounding, a few nouns can form compounds with verbs and other nouns, as illustrated by examples (20a) and (20b):

- | | | |
|-----|--|--|
| 20) | a) <i>pinta-labios</i>
paint-lips
‘lipstick’ | b) <i>casa-cuna</i>
crib-house
‘nursery’ |
|-----|--|--|

However, noun-noun compounding is highly restricted, because not any noun can form a compound with any verb or any noun. This is illustrated by the ungrammaticality of (21), where two nouns cannot stand independently in the same referential phrase. In this case a prepositional connective is required (21b) in between the two nouns.

- | | | | | |
|-----|----|-------------------|----|-----------------------|
| 21) | a) | casa piedra* | b) | casa de piedra |
| | | house stone | | house OF stone |
| | | ‘the stone house’ | | ‘the house of stone’ |

The fact that noun-noun compounds are only few while the number of verb-noun compounds is much larger is further evidence of the non-inherent modifying function of nouns. Nouns can stand alone own in the noun phrase, that is, without any further modifier. This is their defining characteristic. Interestingly, adjectives can also stand alone in the noun phrase, that is, without an explicit noun head. This feature typical of Spanish adjectives is not evidence, however, that adjectives can occupy both the syntactic slots of heads and modifier of the referential phrase. In fact, noun heads are implicit and can be most of the times retrieved from discourse if required, as shown in (22) below. Exceptional are cases of nominalization of adjectives, as in *los Rojos* ‘the Red’ (the communists).

- | | | | | |
|-----|-------------------------------|----------------|---------------|---------------|
| 22) | <i>¿Te</i> | <i>gustan</i> | <i>los</i> | <i>rojos?</i> |
| | You.ACC | like.PRS.PL | DEF.PL.MASC | red.PL.MASC |
| | - ‘Do you like the red ones?’ | | | |
| | <i>¿los</i> | <i>zapatos</i> | <i>rojos?</i> | |
| | DEF.PL.MASC | shoe-PL | red.PL.MASC | |
| | - ‘The red shoes?’ | | | |

A final issue concerns the relative flexibility of Spanish adjectives, according to which they may be used also as modifiers of predicate phrases (adverbs) without further measures. However, this flexibility is restricted to a small subclass of adjectives. Members of this subclass can modify nouns and verbs alike by filling the syntactic slots modifier of referential phrase and modifier of predicate phrase. Consider the adjective *rápido* ‘fast’ in (23) below. The ambiguity of the sentence is caused by the fact that *rápido* can modify the head noun *tren* ‘train’ (interpretation A) or the verb *tomar* ‘take’ (interpretation B). Still, *rápido* can become a full-fledged manner adverb by taking the adverbial ending (*mente*), in which case it produces the second interpretation of (23).

- 23) *No pude tomar el tren rápido*
 NEG could take ART train fast
 Meaning A: 'I could not take the fast train'
 Meaning B: 'I could not take the train quickly'

- 24) *No pude tomar el tren rápidamente*
 NEG could take ART train quickly
 'I could not take the train quickly'

The number of adjectives that can be used also as predicate phrase modifiers is small. Most adjectives cannot be used adverbially. This is shown in (25). The adjective *sincero* 'honest' cannot modify the verb *decir* 'tell' unless it takes the adverbial ending. Finally, there are a few adjectives that cannot take the adverbial ending and are used therefore as manner adverbs in their adjective form (26).

- 25) *dime la verdad sinceramente / sincero**
 tell:1.DAT ART truth honestly / honest
 'Tell me the truth honestly'
- 26) *no corras, ve despacio/despaciamente**
 NEG run:PRS.IND.2S go.IMP.2S slow/slowly
 'I left (the place) running'

The above discussion confirms the classification of Spanish is a type-4 language: a language with individual lexical classes for every syntactic slot. Verbs differ from nouns, adjectives and manner adverbs in that they take markers of tense, aspect and mood while the others not. In turn, nouns are different from adjectives in that they have intrinsic gender and cannot modify other nouns except in compounding. Finally, adjectives differ from manner adverbs in that the vast majority of them cannot act as predicate modifiers without further measures. While a closed class of adjectives behaves flexibly as modifiers of both types of phrases, prototypical adjectives modify only referential phrases. All of this makes Spanish a differentiated language.

Chapter 6

Ecuadorian Quechua

Ecuadorian Quechua (henceforth Quichua)¹ belongs to the northern branch of the Quechua family. Therefore, it is part of Quechua IIB in Torero's classification (1964), which is the one I follow here. Quechua II B includes the Ecuadorian dialects spoken in the Andean Highlands and the Amazon Lowlands plus several Peruvian dialects such as Chachapoyas or Loreto spoken also in the Amazon basin. Ecuadorian Quichua is broadly divided in Highland Quichua (*Quichua de la Sierra*) and Lowland Quichua (*Quichua del Oriente*). According to Knapp (1991), the Highland Quichua population includes all the speakers of Quichua with the exception of those who live at less than 2000 meters above the sea level. The number of Highland Quichua speakers is considerably larger than the number of Lowland Quichua speakers. If the geographical distribution of both dialects is considered however, their respective *spheres of influence*² are closely similar (cf. Map 6.1).

Highland Quichua is spoken in nine provinces of the Ecuadorian Andes, namely: Imbabura, Pichincha, Cotopaxi, Tungurahua, Chimborazo, Bolívar, Cañar, Azuay and Loja.³ Although many studies report that the province of Carchi is Spanish monolingual, the 1990 census showed a number of Quichua speakers scattered in few parishes, who use the language exclusively in domestic settings (Buttner 1993: 23). Quichua speakers are unevenly distributed in the aforementioned provinces.⁴ Central Cotopaxi, Tungurahua, Chimborazo and Bolívar represent two

¹ Proto-Quechua vowels included: unrounded front /i/, rounded back /u/ and low central /a/. The first two were realized as [e] and [o] when preceded by uvular /q/, and as [i] and [u] when preceded by velar /k/ (Adelaar with Muysken 2004: 197). Unlike Peruvian and Bolivian varieties, which preserve the uvular-velar distinction, Ecuadorian Quechua keeps only the velar consonant. In this way the unrounded front and rounded back vowels are realized always as [i] and [u]. Accordingly, Quechua speakers in Ecuador refer to their language as 'Quichua' and *not* 'Quechua' as in Peru and Bolivia.

² Knapp identifies three types of social space in the following terms: "El **núcleo** en el caso típico es el área donde una cultura tiene su más densa población e instituciones políticas claves, culturales y económicas. El **dominio** es donde la cultura presenta predominio numérico. La **esfera** es donde la cultura ejerce alguna influencia sin llegar a ser dominante" (Knapp 1991: 16) This classification can be viewed as an extension of our definition of 'social space' in section 2.1.

³ Notice that Ecuadorian dialects hardly overlap with district borders. According to SIL (2005) there are only four dialects in the Highlands vis-à-vis nine Quichua-speaking provinces. Further explanation is provided in the dialect section.

⁴ Lowland Quichua speakers live mainly in the provinces of Napo and Pastaza. Their number is even harder to estimate since there is a large number of second-language speakers who identify themselves ethnically as Quichua. For a discussion of Quichua-centered ethnogenetic

thirds of the entire Highland Quichua population. Because censuses do not provide specific ethnolinguistic information,⁵ it is difficult to know the exact number of Quichua speakers and their overall percentage of the national population. Sociopolitical interests from national and local governments and Indian organizations prevent an impartial consideration of the actual indigenous population.

The first census (1950) gave a number of 320,056 speakers of Highland Quichua out of a total population of 3,202,757 inhabitants. Highland Quichua speakers represented 10% of the country's population. Buttner (1993: 19f) notes that the census had a number of shortcomings which influenced its output. The most important of them was that nearly 20% of the censused population did not specify their linguistic background. Knapp introduced some corrections to these figures and gave a number of 440,994 speakers of Highland Quichua representing 14% of the country's population in 1950. Forty years later, the 1990 census gave an approximate number of 340,000 Quichua speakers in the Highlands, representing 3.5% of the national population (9,696,979). Again, deficiencies in the collection of data influenced decisively the output (Buttner 1993: 22f). The last national census conducted in 2001 gave a number of 595,798 indigenous speakers⁶ for the Highlands, who represent nearly 5% of the national population (12,156,608). For the same census indigenous speakers in the country represented 7% (830,418). This percentage differs considerably from those provided by ONGs and several Indian organizations including CONAIE. The latter considers that only the speakers of Highland Quichua count above one million. Similarly, the *Instituto Indigenista Interamericano* gave an estimate of 2,634,494 speakers of indigenous languages in Ecuador for 1993. This number represented one quarter of the total population (cf. Adelaar 1999: 11). Taking the percentage from the 2001 census for the entire indigenous population (7%) as a baseline and the percentage from *Instituto Indigenista* (25%) as a threshold, we calculate a reasonable estimate of Amerindian speakers in the country around 16%. Considering that Ecuador's population in 2005 was estimated in 13,363,593 (CEPAL 2005), the above percentage corresponds to some 2,100,000 indigenous speakers. Excluding the population of other ethnolinguistic groups in the Pacific Coast and the Amazon Lowlands, which represent 29% in the last census, we have an approximate number of 1,500,000

processes experienced by several native peoples in the Eastern Lowlands of Ecuador, see Hudelson (1988), Reeve (2002) and Gómez Rendón (2006d).

⁵ On the one hand, most censuses do not include questions about the ethnolinguistic background of respondents. On the other hand, indigenous speakers usually hide their ethnolinguistic identity for sociocultural reasons. Moreover, Indian organizations have boycotted censuses in response to the neglect of administrations in the hard conditions of Quichua speakers.

⁶ The number corresponds to respondents who gave any Indian language as their mother tongue. The vast majority of those who do not speak Spanish in the Highlands as their first language are Quichua speakers.

speakers of Highland Quichua. Without any reliable figures available, this number is a reasonable estimate of the overall number of Highland Quichua speakers.

The issue of the size of the Highland Quichua population is closely related to the issue of its linguistic vitality. Studies on the vitality of Quichua in Ecuador are scarce. Buttner (1993) is the most important sociolinguistic survey of Quichua. Different from most censuses, the main conclusion of this survey is that Quichua is still vital in Highland Ecuador, even though vitality is not uniform across provinces. Based on the data from Buttner (1993: 48), the following table gives the percentages of native Quichua speakers, native Spanish speakers and bilingual speakers among the indigenous population. These percentages are an average of the number for individual communities in each province.

Table 6.1 Quichua vitality according to percentage of native speakers

Province	Quichua	Spanish	Both	No Answer
Imbabura	81.9	15.5	2.3	0.3
Pichincha	13.7	83.7	2.3	0.4
Cotopaxi	80.1	12.5	3.1	4.3
Tungurahua	76.6	20.6	2.3	0.5
Chimborazo	91.9	6.0	2.1	0.0
Bolívar	78.5	18.2	3.3	0.0
Cañar	73.5	23.2	3.3	0.0
Azuay	43.0	55.0	2.0	0.0
Loja	26.1	71.0	2.0	0.0

Source: Buttner 1993: 48

The central Highlands and the province of Imbabura are the strongest Quichua areas. The southern Highlands are less vigorous in terms of native speakers and language maintenance. The most densely populated provinces in the Highlands (Pichincha and Azuay) show the highest percentage of Spanish as a first language. Loja in the southern Highlands and Chimborazo in the central Highlands mark the sharpest contrast between traditional Quichua areas: the former shows the highest degree of Hispanicization while the latter shows the highest degree of Quichua maintenance. As for the number of people who have Quichua *and* Spanish as their first languages (coordinate bilinguals) percentages do not differ significantly across provinces. This does not mean that levels of bilingualism are uniform however. A fine-grained classification of bilingualism (Buttner 1993) shows crucial differences. The survey identified nine levels of monolingualism-bilingualism, namely: Quichua monolingualism; Spanish monolingualism; rudimentary Quichua-Spanish bilingualism (where Quichua is dominant); rudimentary Spanish-Quichua bilingualism (where Spanish is dominant); advanced Quichua-Spanish bilingualism

I; advanced Quichua-Spanish bilingualism II; advanced Spanish-Quichua bilingualism I; advanced Spanish Quichua bilingualism II; coordinate bilingualism. The following table shows the distribution of ethnic monolinguals and bilinguals in the Quichua communities of the nine highland provinces.⁷

Table 6.2 Levels of monolingualism and bilingualism

Level	%
Quichua monolingualism	8.7
Spanish monolingualism	10.4
Rudimentary Q-Sp bilingualism	24.0
Rudimentary Sp-Q bilingualism	6.3
Advanced Q-Sp bilingualism I	28.4
Advanced Q-Sp bilingualism II	4.1
Advanced Sp-Q bilingualism I	11.6
Advanced Sp-Q bilingualism II	0.8
Coordinate Spanish-Quichua bilingualism	0.6
n.a.	5.0

Source: Buttner 1993: 69

Quichua native speakers with an advanced level of bilingualism (I) represent the largest group. Rudimentary bilinguals whose mother tongue is Quichua are the second group in size. The percentages of rudimentary and advanced bilinguals are not considerably different from each other. Quichua monolingualism (8.7%) is slightly lower than Spanish monolingualism (10.4%). The number of coordinate bilinguals is very small if compared to the other groups. In general, the figures show two parallel processes: on the one hand, a steady process of Hispanicization; on the other, the maintenance of Quichua. While these percentages suggest that a complete shift to Spanish will not take place in the medium term in the Highlands, at a local level there are communities which were bilingual thirty years ago but now are Spanish monolingual (e.g. González Suárez, Imbabura). Table 6.3 shows the percentages of Quichua monolingualism and illiteracy according to age groups:

⁷ There are cases of communities which identify themselves ethnically as Quichua but some or most of their members are not native speakers of the language or speak Spanish only (e.g. San Isidro, Loja).

Table 6.3 Quichua monolingualism and illiteracy by age groups

Age group	Illiteracy (%)	Q-monolingualism (%)
14-20	6.9	3.5
21-30	20.0	6.2
31-40	41.8	8.5
41-50	55.4	11.7
51-60	66.8	15.9
60 >	72.4	14.3

Source: Buttner 1993: 54

Two straightforward correlations are observed: one between age group and illiteracy (the younger the speaker, the less monolingual in Quichua); and another between age group and Quichua monolingualism (the older the speaker, the more illiterate). Both correlations demonstrate that Hispanicization goes hand in hand with formal schooling. Buttner (1993: 34) notices two gaps in schooling levels: one between the last two generations (14-30) and the rest; the other between the last three generations (41>) and the younger ones. The first gap results from a wider access by speakers up to thirty years to secondary education and a limited access to it by the rest of speakers. The second gap results from the lack of access by speakers from forty years onwards to elementary education and the access to it by younger speakers. Retrospectively, these tendencies correspond to two major developments in the social structure of the countryside in the last fifty years. The first is the Agrarian Reform initiated in the early 1960s, which resulted in a more extensive coverage of elementary education in rural areas. The second is the application of Bilingual Education Programs and the extension of secondary education to rural areas since the late 1980s. The great majority of Quichua monolinguals above fifty years did not go to elementary school. In addition, it is possible to trace a further correlation between Quichua monolingualism and gender. Women make up the largest group of Quichua monolinguals and the smallest group of bilinguals. Thus, the higher the level of bilingualism, the lower the percentage of bilingual women. Table 6.4 shows this correlation with respect to the types of Quichua-dominant bilingualism.

Table 6.4 Levels of bilingualism by gender

Gender	Quichua Monolingualism	Rudimentary Quichua-Spanish Bilingualism	Advanced (I) Quichua-Spanish Bilingualism	Advanced (II) Quichua-Spanish Bilingualism
Men	21.5	48.6	64.9	72.7
Women	78.0	51.1	34.5	27.3

The process of Hispanicization intensified in Ecuador during the second half of the twentieth century through a wider coverage of formal schooling in rural areas and higher rates of Indian migration to the cities. The last enclaves of Quichua monolingualism in the Highlands are isolated communities in the central provinces located on bleak plateaus or *páramos* at an altitude higher than 3,200 meters. These communities usually do not have access to schooling and their contact with the mainstream society is limited by a lack of roads and transportation. In general, however, the advanced process of Hispanicization reflected in higher levels of bilingualism does not imply the loss of the native language at individual or collective levels. Quichua often coexists with Spanish in different social spaces, albeit one language is preferred to the other depending on the situation.

The survey provides additional information about the differential use of Quichua and Spanish in various socio-communicative spaces. Given the prevailing diglossic situation in the Highlands, both languages are expected to be in complementary distribution across social spaces. Let us see if this is the case. As regards the preferred language in the family, 78% of those whose first language is Quichua prefer this language at home while 15.6% prefer both Quichua and Spanish. On the contrary, 87% of the indigenous speakers whose first language is Spanish prefer this language at home, but only 9.8% prefer both languages. This distribution differs from province to province but is consistent with the data in Table 6.1 about the vitality of Quichua. For example, 66% of the indigenous families from Chimborazo – a traditionally Quichua-speaking province – use exclusively Quichua at home while only 5.5% use Spanish. Conversely, 78% of the indigenous families from Pichincha – a traditionally Spanish-speaking province – use Spanish at home while only 2.7% use Quichua. The distribution of the preferred language in the socio-communicative space of the community is closely similar to the distribution in the domestic space. Quichua is preferred in 81.6% of the indigenous communities of Chimborazo while Spanish is preferred in 87.8% of the indigenous communities of Pichincha (Buttner 1993: 63). Other socio-communicative spaces include collective work parties (*mingas*), community meetings (*asambleas*), parish centers (*cabeceras parroquiales*), open-air markets (*ferias*) and churches. Table 6.5 below shows the usage of Quichua in these spaces for the nine provinces.

Chimborazo is the province with the highest percentage of Quichua usage in the five settings. Pichincha and Azuay are the provinces with the lowest percentage. The percentages of Quichua usage are remarkably lower in parish centers and open-air markets. While these spaces are traditionally Spanish speaking, the dominance of this language is less important in Chimborazo, if compared to Azuay or Pichincha. In contrast, *mingas* and meetings are predominantly Quichua-speaking settings because they are located in the broader space of the community. Finally, the socio-communicative setting of the church is a public, originally Spanish speaking space. However, it has been increasingly appropriated by Quichua speakers in the central

Highlands and Imbabura. On the contrary, Pichincha, Azuay and Loja still prefer the use of Spanish in the church. This is explained by the higher levels of Hispanicization among indigenous speakers in these provinces and the closeness of mestizo churches of towns and cities.

Table 6.5 Uses of Quichua by socio-communicative per settings and province

Province	Minga	Meetings	Church	Urban center	Market
Imbabura	76.4	73.3	31.9	17.6	16.3
Pichincha	4.9	3.4	1.1	0.8	0.8
Cotopaxi	66.9	62.0	52.0	11.5	10.5
Tungurahua	62.3	56.0	29.6	15.5	11.1
Chimborazo	79.4	78.9	72.3	33.2	36.4
Bolívar	67.0	61.5	30.9	16.4	20.0
Cañar	60.2	55.2	20.4	16.0	14.9
Azuay	20.0	10.0	4.0	5.0	2.0
Loja	24.6	21.7	10.1	8.7	11.6

In sum, the status of the Quichua language in Ecuador is one of relative maintenance accompanied with higher levels of bilingualism. In the central provinces the native language remains strong in most communicative spaces, especially in the household and the community. In the rest of the Highlands, Quichua speakers are rapidly shifting to Spanish, and this shift will be complete in one or two generations.

6.1. The History of Quichua in Ecuador

One puzzling question for those who study the history of Quechua in the northern Andes is how this language – which became the official language of the Inca Empire - managed to take firm root in this part of the Cordillera in scarcely sixty years of Inca domination from the conquest by Huaina Capac around 1470 to the fall of the Inca Empire in 1532. The question becomes even more problematic, because the Incas never sought to replace the vernacular languages of their conquered territories nor reduce the linguistic variety of the Empire by imposing Quechua (Mannheim 1991: 36ff). In trying to answer this question, Torero proposed that Quechua was spoken in the present territory of Ecuador well before the Inca invasion: long-distance traders or *mindaláes* introduced Chinchay Quechua in the late fourteenth century and speakers of different linguistic backgrounds began to use it as a lingua franca (Torero 2003: 93-105).⁸ From Jijón y Caamaño (1940, 1941) and Paz y Miño

⁸ Others have explained the early presence of Quechua in the northern Andes by assuming that the language originated in the Ecuadorian Amazon Lowlands. Lathrap (1970: 176ff) and

(1940, 1941, 1942) we know that Pre-Inca languages in the northern Andes included Pasto, Cara, Panzaleo, Puruhá, Cañari and several Jivaroan languages spoken in the southern part of the highlands and the eastern slopes. Nothing of these languages is known except for the lists of toponyms and anthroponyms collected by Jijón y Caamaño and Paz y Miño. The Quito Synod of 1593 ordered the preparation of catechisms and confessionaries in these languages but they are reported lost to date (Adelaar and Muysken 2004: 392). Residual evidence of pre-Inca languages can be traced in the form of substrata in the variety of Ecuadorian Quichua spoken today in their former area of influence. Adelaar and Muysken (2004) suggest, for example, “that a possible substratum from Cara preserved in modern Imbabura Quechua is the use of a labial f” (Adelaar with Muysken 2004: 394). It is generally assumed that pre-Inca languages survived throughout the sixteenth century to be finally replaced by Quichua around the second half of the seventeenth century. There are no available sources that help us establish an exact date, but it is evident that the shift to Quichua did not take place overnight.

The Quichuization of native populations in the northern Andes was *not* the direct result of Inca occupation. It was the Spaniards who realized the potential of Quechua as a lingua franca not only for inter-ethnic communication but, most importantly, for the evangelization of indigenous peoples. Given the large number of vernaculars spoken in the Inca Empire at the time of the Spanish conquest, the evangelization of ethnic groups in their own languages was in principle unfeasible. While pre-Inca languages were promoted as a means of instruction in the first decades of colonization, it became soon obvious that using them for evangelization was rather unrealistic.

But the adoption of Quechua for indoctrination was not exempt from disagreement. Crown officials and members of the clergy viewed the use of an indigenous language with suspicion and resistance. For them Quechua could not transmit theological concepts and therefore could not function properly as an effective vehicle for evangelization. This position was politically motivated. The new rulers were afraid that Quechua might become an agglutinating factor in the promotion of ethnic awareness. The ambivalent position between the use of Quechua and the use of Spanish in evangelization continued until the Bourbon reforms in the second half of the eighteenth century. As a result, a coherent language policy could not be implemented in colonial times. Mannheim (1991) explains the ups and downs of language planning in the Spanish colonies in the following terms:

Hartmann (1979: 287ff) are the advocates of this theory, which is not accepted in today's scholarly circles however.

“Language policy in the Spanish empire was molded by competing interest groups, each of which staked its claim before the Council of the Indies (Heath 1976: 50; Rivarola 1985: 26-27). As a result, the council frequently shifted back and forth between radically different approaches depending upon which pressure group was able to gain its attention. The extent to which the council’s policies were actually implemented was similarly determined by competing interests, this time at a local level” (Mannheim 1991: 64)

The use of Quechua for evangelization received partial support from the language policies of three Councils held in Lima between 1551 and 1583. Clergymen implemented these policies in different areas of the Empire, including the eastern Lowlands. From the last quarter of the sixteenth century a number of clergymen studied Quechua and wrote grammars (*Artes*), dictionaries (*Vocabularios*) and primers for catechization (*Cartillas*) in different varieties of the language. Several courses opened for this purpose in Lima (1550) and Quito (1570). Similarly, efforts were made to standardize the language in order to make its learning easier for priests and facilitate the printing of materials in Quechua. The basis for the standardization was Cuzco Quechua, a variety directly associated with the Inca. Cuzco Quechua presented several phonetic intricacies such as the velar-uvular distinction and the ejective-aspirated distinction of stops (Mannheim 1991: 142). These particularities were eventually omitted in the standardized version and resulted in a language closely resembling the Quechua variety spoken in the northern Andes because of its simplified phonetics. This standard was used until the first half of the seventeenth century (Adelaar and Muysken 2004: 183). Some scholars maintain that the use of standardized Quechua by missionaries influenced decisively the development of Ecuadorian Quichua, particularly in the Amazon Lowlands (Oberem and Hartmann 1971; but see Muysken 2000 for an evaluation of this hypothesis). Still, the influence of standardized Quechua may have not been as decisive as generally assumed, but its use by missionaries fostered the expansion of the language in the northern Andes at the expense of pre-Inca languages.

Because Quechua was not the mother tongue of the peoples of the northern Andes until their native languages were eventually replaced, it is not possible to speak of Ecuadorian Quechua as a distinct variety before the end of the seventeenth century. It is only from the moment that these native peoples abandoned their pre-Inca language (Cara) and adopted Quechua that something like an Ecuadorian variety of Quechua emerged. The historical record shows that the replacement of pre-Inca languages with Quechua was a gradual process that lasted over a hundred years. The question is what Quechua dialect became the basis for Ecuadorian Quechua. By studying early grammatical descriptions Muysken (forthcoming) shows that Quechua in Ecuador kept many features of Peruvian dialects in the

seventeenth century⁹, but that these features were replaced by those typical of present-day Quichua over the next centuries. The following is a summary of Muysken's findings about the process of formation of Ecuadorian Quichua.

The first known source of Ecuadorian Quichua is an anonymous grammar dating back to the late seventeenth century – the exact date is unknown. The manuscript was published by Dedenbach-Salazar (1993), who calculated its time of writing through the loss of phonological distinctions and the replacement of markers characteristic of Quechua-I varieties (central Peru and Bolivia). The most important of these changes are: the lack of distinction between inclusive and exclusive pronouns; the loss of possessive pronominal forms and their replacement by pronoun-genitive constructions; and the loss of verb-object agreement markers. Posterior to this anonymous grammar is the catechism prepared by the Quitoian bishop Francisco Romero in 1725. Romero does not make any use whatsoever of inclusive and exclusive pronouns. He does not use Peruvian benefactive *-pu* either. Furthermore, Romero uses very frequently the subordinating suffixes *-cpi*, *-spa*, *-ngapa* and impersonal *-ri* in intransitive verbs, as is characteristic of present-day Ecuadorian Quichua. From the lower occurrence of the reflexive *-ku* and the use of comparative constructions with *yalli-*, Muysken concludes that Romero's catechism was written after the anonymous manuscript analyzed by Dedenbach-Salazar. Another early source of Ecuadorian Quichua is the grammatical sketch by Nieto Polo (1753). The author says explicitly in the title of his sketch that it deals specifically with the Quechua language spoken in the Province of Quito. This implies that the differences between this variety and Peruvian dialects were not unimportant. By the time Nieto Polo wrote his sketch the Peruvian inclusive-exclusive forms and the suffixes for the benefactive and dative cases had almost completely disappeared from Ecuadorian Quichua, while others such as the transitional pronominal form (1>2) and nominal possessive markers were disappearing gradually. This suggests that Nieto Polo is dealing already with a distinct Ecuadorian variety, and indeed he often compares this variety with Peruvian dialects in order to point out differences. The brief grammatical remarks in Velasco's *Vocabulario de la Lengua Indica*, published in the same year as the work by Nieto Polo (1753), report the same morphological developments in Quichua. More than a century later, in 1884, Cordero published a Quichua grammar in which the transitional pronominal form *-wa* is preserved as an optional marker of first and second person object. Cordero also has as optional the use of possessive marking on nouns instead of genitive constructions – in full use by the end of the nineteenth

⁹ Muysken assumes that Peruvian Quechua (specifically, Cuzco Quechua) was, in the early decades of colonization, a model for the description of Ecuadorian Quichua, and that reference was often made in grammars to (Peruvian) forms that were not used in Ecuador anymore.

century. Cordero's grammar proves archaic in some respects – a feature that becomes clear in his later dictionary (1905), where the occurrence of obsolete forms is frequent – but it certainly describes a distinct Ecuadorian variety of Quechua. In all, the aforementioned changes reflect a gradual restructuring of Ecuadorian Quichua in the lapse of two centuries, from the late seventeenth to the late nineteenth. The coexistence of alternative Ecuadorian and Peruvian forms that disappeared over the years confirms the gradual nature of the process. According to Muysken, this process continues today with the loss of the remaining transitional marker (-*wa*) in modern dialects of the central Highlands.

In fact, changes in Ecuadorian Quichua continue to date, but they are largely motivated by language contact with Spanish¹⁰. The influence of Spanish on the lexicon of Quichua includes basic and non-basic vocabulary and involves practically every semantic field, from kinship and household to religion, education and administration (Gómez Rendón and Adelaar, forthcoming). It should be noted, however, that borrowing is not the same across dialects and idiolects. Generally speaking, the dialects of distant areas with less contact with urban centers show lower levels of borrowing. In a similar way, the idiolects of older generations show much less influence than those of younger bilinguals.¹¹ On the other hand, an increasing bilingualism among Quichua speakers and the use of Quichua in atypical communicative settings such as radio broadcasting have induced a number of structural changes in the language (Fauchois 1988). Grammatical changes in Ecuadorian Quichua as a result of contact with Spanish have been analyzed elsewhere (Gómez Rendón 2007a). As a result, strongly Hispanicized varieties of Quichua continue to emerge as adaptations to the new communicative settings of modern society. Contemporary Quichua is a living language after four centuries of contact because it succeeded in making a compromise between the communicative needs imposed by the official language and the cultural needs of their speakers to preserve their linguistic identity.

6.2. The dialects of Quichua in Ecuador

Ecuadorian Quichua can be classified in two distinct varieties, Highland Quichua and Lowland Quichua. Each corresponds to a specific geographical area: the Andean

¹⁰ While the Quichua-Spanish contact increased dramatically in the last century as a result of the expanding national society and the diffusion of media in rural areas, the contact itself dates back to the early years of the European conquest. The existence of loanwords in Quichua which have long disappeared from modern Spanish dialects (e.g. *parlar* 'speak') demonstrates how old the Spanish-Quichua contact is in the northern Andes.

¹¹ In some cases borrowing became massive and produced mixed lects (Media Lengua) as described by Muysken (1978; 1996) and Gómez Rendón (2005; 2008b) for the Ecuadorian Andes of Ecuador.

Highlands and the Amazon lowlands. Further divisions can be made on the basis of phonetic, morphosyntactic and lexical criteria. For example, Highlands Quichua distinguishes between northern varieties and central-southern varieties. A similar distinction for Lowland Quichua separates northern (Napo) from southern (Pastaza) varieties. Northern Highland Quichua includes the varieties of Imbabura and Pichincha.¹² Central-Southern Highland Quichua include the varieties of Cotopaxi, Tungurahua, Chimborazo, Bolívar, Cañar, Azuay and Loja. Quichua dialectal areas are shown in Map 6.1.

Map 6.1 Map of Quichua Dialects in Ecuador



Source: FEDEPI. © R. Aschmann 2006

¹² While in Imbabura the indigenous language is widely distributed over the whole province, in Pichincha it is spoken only in suburban Calderon (inside the metropolitan area of Quito).

Dialectal differences in Quichua involve regions and provinces alike. From the information of individual entries (Haboud de Ortega *et al* 1982) it is possible to trace these differences. For example, the verb ‘to collapse’ is *tuñurina* in highland and lowland varieties but *tularina* for Pichincha. In this case both words are phonologically similar. In other cases words under the same entry have different phonetic forms, although their origin is still Quichua. The verb ‘to chew’ is *kashuna* in the Highlands but *mukcuna* in the Lowlands. The same holds for ‘old’, which is *rucu* ‘old’ in Highland Quichua but *paya* in Lowland Quichua. Most lexical differences noted in the dictionary make a distinction between highland and lowland dialects. But highland varieties, too, show differences in the lexicon. The word *chamcha* ‘flavorless’ occurs in Chimborazo and Tunguragua (central) but its equivalent is *chamuk* in Imbabura (northern), Cañar and Azuay (southern). In a few cases words occur only in one province. This is the case of *ñusta* ‘princess’ in Imbabura or *zacziquina* ‘spread, said of lianas or pumpkins’ in Bolívar. Finally, there is a large number of local names for endogamic animals plants (e.g. *tauri* ‘lupin’, only in the Highlands; or *sicu* ‘kind of rodent’, only in the Lowlands).

The most noticeable differences in the Ecuadorian dialects are phonetic in nature. To give an idea of dialect variation in Quichua, consider the different pronunciations of *patsac* ‘hundred’ which, according to the aforementioned dictionary, has as many as sixteen different pronunciations: [patsax, patsak, patsag, patsa, p^hatsak, p^hatsa, patsux, patsuk, patsu, patsix, patsik, patsug, patsi, fatsax, fatsak, fatsa]. This variation is not restricted to lexical items. The most important differences concern the morpho-phonological processes that have affected the affixes (Adelaar and Muysken 2004: 237). The case of the genitive-benefactive suffix *-pac* is illustrative in this respect. According to Caimi *Ñucanchic Shimi-yuc-Panga* (1982), this bound morpheme may be realized in nineteenth different allomorphs depending on the dialect: [-pak, -pax, -pa, -bak, -bax, -bag, -ba, -buk, -bux, -bug, -bu, -wak, -wax, -wag, -wa, -k, -x, -g, -w]. Similarly, the affirmative *-tac* has as many as 26 allomorphs. In broad terms, northern and central dialects are more innovative than southern ones (e.g. Cañar) in that the latter have not undergone the aforementioned morpho-phonological processes nor other developments that affect the syntax of the language.¹³ In what concerns the dialects of Bolívar and Imbabura, they show differences as well, with the former being generally more conservative. Bolívar Quichua is the least innovative of the central dialects and therefore is much closer to Cañar Quichua than to the Cotopaxi or Chimborazo dialects. In the

¹³ Muysken’s study *Syntactic developments in the Verb Phrase of Ecuadorian Quichua* (1977) is illuminating in this respect.

following I discuss some of the phonetic particularities of Imbabura Quichua in contrast with other highland dialects.

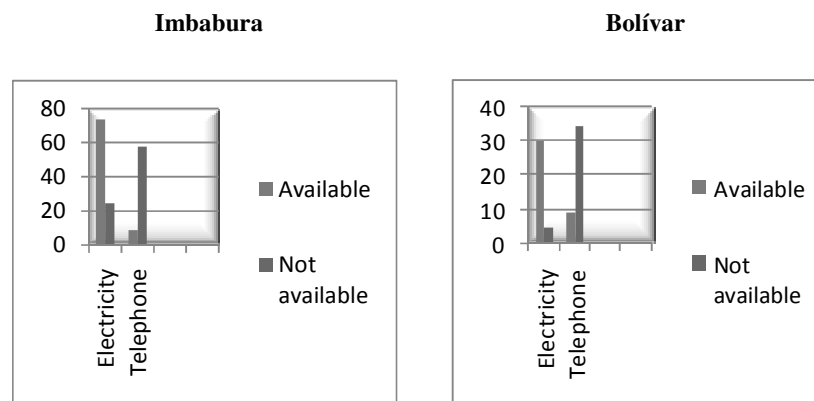
In phonetics, the main difference between Imbabura Quichua and other highland (but also lowland) dialects is the fricativization of stops /p/ and /k/ in all positions except after a nasal. The resulting allophones [f] and [j] differ in word initial position from their counterparts in the rest of Ecuadorian dialects, be they aspirated ([p^h] and [k^h]), or non-aspirated ([p] and [k]). Imbabura fricativized occlusives differ also from non-aspirated realizations in word-medial or word-final position in most dialects. Some examples are *pucuna* ‘to blow’, realized as [fukuna] in Imbabura but [p^hukuna] in central dialects (e.g. Bolivar) and [pukuna] in southern dialects (e.g. Loja); *upiana* ‘to drink’, realized as [ufiana] in Imbabura but [up^hiana] in Cotopaxi and Tungurahua (central dialects) and [upiana] in Azuay (southern dialect); *cari* ‘male’, [jari] in Imbabura, but [k^hari] in Chimborazo (central). Grammatically, the reciprocal *-naku-* is realized as [-naju-] in Imbabura but [naku] in all other provinces. A further phonetic process in Imbabura is the voicing of [t] after nasals and other environments (Fauchois 1988: 62), although a similar phenomenon is reported also for Salasaca (central). In contrast, the voicing of stops is less widespread in Imbabura. According to Adelaar and Muysken (2004: 242) Salasaca and central dialects show the following processes, which are very infrequent or nonexistent in Imbabura: the raising of /a/ to [i] or [u]; the frequent deletion of the final stop or the nasal in affixes such as *-pak* or *-man*; consonant cluster simplification in the affixes *-kpi* (> -ki) or *-špa* (> -ša); vowel cluster simplification (e.g. *wira* > *ira*); and pronunciation of palatal /ž/ as a palatal affricate [č] before voiceless stops (e.g. *kužki* > *kučki*). In general, these morpho-phonological processes make central dialects distinct from their northern and southern counterparts. Explanations for the above distribution of features are largely language-internal, but several nonlinguistic motivations can be identified as well. These have to do with geographical and demographical factors and the related sociolinguistic conditions. I discuss some of them for Imbabura Quichua and Bolivar Quichua.

In 2001 the number of Imbabura Quichua speakers was twice as large as the number of Bolivar Quichua speakers (86.986 vs. 40.094).¹⁴ The distribution of Quichua speakers in both provinces is different too. Quichua is spoken in four of the six cantons of the Imbabura province. In contrast, Quichua is spoken only in one of the six cantons of the Bolivar province. Many Quichua communities in Imbabura are

¹⁴ These numbers are very probably underestimated, as explained in the introduction to section 6.2. A more reasonable number is around 150.000 speakers. Ethnologue gives a number of 300.000 speakers in 1977, which is obviously an exaggeration considering that the whole population of Imbabura (i.e. Mestizos and Indians) reached only 250.000 in 1982 (INEC 2001). There being no census information for the Quichua-speaking population of Imbabura other than the 2001 census, I have preferred to use these figures.

close to urban centers, in particular Otavalo, Cotacachi and Ibarra (province capital). Consequently, the access to the communities is easier thanks to an extensive system of roads. Differently, most Quichua communities in Bolívar are distant from the only urban center of the province (Guaranda) and therefore less accessible. Differences in accessibility are reflected in the supply of facilities such as drinking water, electricity and telephone (cf. Figure 6.1). Nearly three quarters of the indigenous population in Imbabura have access to electricity, while only thirty percent have this supply in Bolívar. Notice that electricity implies access to radio and television broadcasting in Spanish and better communication with the mainstream society. In general, Bolívar Quichua communities are more isolated and more conservative.

Figure 6.1 Access to basic facilities in two Quichua-speaking highland provinces



As for the number of Quichua native speakers both provinces do not differ much (81.9% in Imbabura vs. 78.5% in Bolívar, see Table 6.1). There is a noticeable difference between both provinces with respect to the use of Quichua in socio-communicative settings (cf. Table 6.5). Imbabura Quichua is used with higher frequency in all settings except for the market. These data contradict the above statement about the higher degree of isolation and conservatism of Bolívar Quichua. However, other factors must be considered for a comprehensive evaluation. Quichua is stronger in Imbabura not only on account of its larger number of speakers but also because their attitude is one of deep ethnic awareness and positive identification with the native language. Positive attitudes towards Quichua and the use of the indigenous language in public spaces are less noticeable in Bolívar. These facts explain why despite being relatively more conservative, Bolívar Quichua is less used across social spaces. In support of this explanation two anecdotic but illuminating

facts can be added. In 2005 the municipal council of Otavalo – the second biggest city of Imbabura and the one with the largest number of Quichua speakers in the province – ordered their officials to take courses in Quichua so as to provide better services to Quichua users. No similar decision is reported for Bolivar. Also, Imbabura Quichua has been often a model for other dialects and comparatively numerous materials for teaching have been used this variety, with the opposition of other Quichua communities (cf. Buttner 1993: 195).

In short, Imbabura Quichua speakers are increasingly bilingualism but maintain their language while Bolivar Quichua speakers are less bilingual and tend to shift to Spanish rather rapidly. As a result, Imbabura Quichua is more innovative and most of its lexical and structural changes are induced by contact with Spanish. In contrast, Bolivar Quichua is generally more conservative and therefore less hispanicized.

6.3. Quichua: a typological characterization

Quechua has differently evolved as a result of its geographical expansion. Quechua dialectalization has produced remarkable differences across varieties such that “many linguists now prefer to speak of ‘Quechuan languages’ rather than of ‘Quechua dialects’” (Adelaar and Muysken 2004: 168). Still, Quechua varieties – including Ecuadorian Quichua – remain essentially uniform in their typological character. The following typological description of Quichua is based on the assumptions that Quichua is typologically similar to other Quechua languages, and that Ecuadorian varieties show similar typological features.

Quichua belongs to Quechua IIB in Torero’s classification (Torero 1964). The branch covers an extensive area including “the dialects of the Ecuadorian highlands and *Oriente* (the eastern lowlands); the Colombian Quichua dialect usually called *Inga* or *Ingano* (Caquetá, Nariño, Putumayo); the dialects spoken in the Peruvian department of Loreto in the Amazonian lowlands (which are, in fact, extensions of the varieties spoken in the Amazonian region of Ecuador); the *Lamista* dialect spoken in the area of Lamas (department of San Martín, Peru); and that of Chachapoyas and Luya (department of Amazonas, Peru)” (Adelaar and Muysken 2004: 186f). Quechua IIB differs from the Quechua I spoken in central Peru but show certain resemblance to varieties outside this area (e.g. Santiagueño Quichua in Argentina). Notwithstanding the existence of a dialectal continuum between Quechua varieties, a major split exists between central Quechua and the other varieties. For Adelaar and Muysken (2004: 188) this split reflects an initial diversification in Proto-Quechua.

The phonological inventory of Ecuadorian Quichua includes three vowels (/a/, /i/, /u/) and sixteen consonants (/p/, /t/, /k/, /ts/, /tʃ/, /ʃ/, /s/, /x/, /ʒ/, /m/, /n/, /l/, /r/, /ɾ/, /w/, /y/). These segments occur phonologically in Ecuadorian Quichua but their

realization differs across dialects.¹⁵ Differences consist mainly in the aspiration or glottalization of stops (cf. *supra*). Thus, /p/ is realized as aspirated in most central dialects ([p^h]) but non-aspirated in southern dialects ([p]) and fricativized in Imbabura ([Φ]). This variation may be understood more clearly by assuming that varieties of Ecuadorian Quichua form a diasystem. In this perspective, the phonological inventory consists of diaphones condensing equivalences between sounds of different dialects. For the stop /p/ the equivalences are represented as follows:

Imbabura:	p ~ /Φ/ ----- > /P/
Bolívar:	p ~ p ^h

In this representation the sign ~ stands for phonetic contrast at dialect level while upper-case /P/ represents the diaphone. Other phonetic details of Imbabura and Bolívar Quichua were discussed in section 6.2. Syllables in Ecuadorian Quichua are basically open (CV) but CVC syllables are common as well. Consonant clusters are permitted in onsets but not in coda position. Word-final clusters are absent. Stress is assigned to the penultimate syllable by default, with only few instances of last-syllable stress (Adelaar and Muysken 2004: 206).

The entrance of Spanish loanwords to the core vocabulary of Quichua has enriched the native inventory with the consonants /b/, /d/, /g/, /β/, /z/ and the medial vowels /e/ and /o/. These sounds show a high degree of integration in Quichua and may be considered part of the phonological inventory of the language (Cole 1982: 199). The integration of these sounds has been facilitated in part by the fact that except for /β/, they all have counterparts in native allophones: [b] is an allophone of /p/, and so is [d] of /t/ and [g] of /k/, in nasal environments. The same holds for [e] and [o] in relation to /i/ and /u/.

From a morphological point of view, Quichua is a typically agglutinative language, with a rich and very regular morphology. Compared to Peruvian Quechua, Ecuadorian varieties have experienced a simplification involving two changes: the loss of verb-object agreement and the loss of possessive nominal suffixes. For comparison Cole (1982:6) gives the following examples of second person object agreement in San Martín Quechua (Peruvian) and Ecuadorian Quichua.

¹⁵ This inventory differs from the one given by Cole for Imbabura Quichua (Cole 1982: 199) which consists of five vowels including Spanish-borrowed /e/ and /o/ plus twenty-two consonants including Spanish-borrowed /b/, /d/, /g/, /β/, /z/ but also /j/, /Φ/ and /ɲ/. In addition, Cole postulates the phoneme /t/ instead of two phonemes /t/ and /t̚/. The inclusion of the Spanish-borrowed vowels and consonants is fully justified by the abundance of loanwords with these sounds. In contrast, the inclusion of /j/, /Φ/ and /ɲ/ is less substantiated, because they are allophones of /x/, /p/ and /n/ in Imbabura and do *not* occur in other Ecuadorian varieties.

	<u>San Martín Quechua</u>		<u>Ecuadorian Quichua</u>		
1)	a. <i>Ñuka-ka</i>	<i>maka-yki</i>	b. <i>Ñuka-ka</i>	<i>kan-ta</i>	<i>maka-ni</i>
	1SG-TOP	hit-1S.2OBJ	1S-TOP	2S-ACC	hit-1S
	‘I hit you’		‘I hit you’		

The gradual loss of person agreement markers in Quichua started in the late sixteenth century (cf. 6.2). The second change consists in the loss of possessive nominal suffixes. Consider the following examples from Cerrón-Palomino (1987:200).

	<u>Junin Quechua</u>	<u>Ecuadorian Quechua</u>
2)	a. <i>maki-yki</i>	b. <i>kan-pak</i> <i>maki</i>
	hand-2S.POSS	2S.GEN hand
	‘your hand’	‘your hand’

The loss of possessive nominal suffixes has encouraged a tendency present in Quichua towards higher levels of analyticity. Such tendency contrasts with the great degree of synthesis of Peruvian dialects. Additionally, the extensive use of pronominal roots in Quichua seem to have induced a more frequent use of personal pronouns in subordinate and main clauses, where other varieties use them basically for emphasis.

Contact with Spanish has induced further changes. One of them involves the use of Quichua *kikin* 'proper' as a polite second-person pronoun on the model of the Spanish polite form *usted* '2S.HON'. Arguably, this form developed as a pronoun relatively early, when social relations between Spaniards and Indians were based on caste hierarchies. Nowadays *kikin* is falling into disuse, being preserved only in conservative varieties. But the influence of Spanish on the pronominal paradigm extends to the subset of interrogative pronouns in most Hispanicized varieties. Also, Ecuadorian dialects have incorporated a few bound Spanish morphemes through the borrowing of words from this language. These belong to nominal morphology and include agentive *-dur* as in *ñau-pa-dur* 'spokesman', diminutive *-itu* as in *waw-it* 'little child'. In case marking contact-induced changes include (i) the lack of distinction between inalienable and alienable possession; (ii) the loss of different forms for comitative and instrumental cases; (iii) the frequent drop of the obligatory accusative marker on direct objects; (iv) the increasing tendency to use the plural marker on nouns after numeral modifiers; and (v) the use of Spanish lexical borrowings to express local or spatial relations.¹⁶ These changes have been discussed elsewhere (Gómez Rendón 2007a).

¹⁶ Other developments in Quichua verbal morphology are less clearly assigned to contact-induced change. These include a) the use of reciprocal *-naku-* as a verbal plural marker on

Morphologically, Quichua do not have gender markers nor definiteness markers such as articles. In general terms, Quichua morphology is based on inflectional and derivational suffixes. The only exception is the prefix *la-* ‘in-law’, albeit it does not occur in all Ecuadorian varieties. Case suffixes like *-manta* in the following example have been considered by some linguists as true postpositions.

- 3) *Otavalo-manta* *shamu-ni*
 Otavalo-ABL come-1S
 ‘I come from Otavallo’

The order of constituents in the noun phrase is fixed in Quichua, with modifiers preceding noun heads (30a-b). The order in possessive constructions is possessor-possessed, even if pronominals are involved (cf. 28b).

- 4) a. *turu* *pilchi* b. *sumac* *huasi*
 clay bowl nice house
 ‘bowl of clay’ ‘nice house’

The Quichua noun phrase has experienced two noticeable changes as a result of contact with Spanish: the use of determiners *shuk* ‘one’ and *kay* ‘this’ to replace the native topicalizer *-ka*; and the occurrence of Spanish diminutive and augmentative endings in borrowed and native lexemes (cf. *supra*). Apart from Spanish numerals, which are ubiquitous, Ecuadorian Quichua has borrowed several Spanish quantifiers. Unlike numerals, Spanish quantifiers have not replaced their native counterparts but co-occur with them in duplets with emphatic purposes. The most frequent are *tuditu* ‘all’ and *algunu* ‘some’.

The influence of Spanish extends also to the verb phrase. TMA structures influenced by contact with Spanish include the replacing of *-ngapak* with purposive *-chun* in co-referential constructions on the model of the Spanish subjunctive; (ii) the use of Spanish *dizi-* ‘say’ in reportatives and quotatives; and (iii) the use of Spanish modal verbs. Spanish loan verbs are borrowed as verb roots without their infinitive endings and any extra marking. Certain verbal roots from Spanish are subject to further morpho-phonological changes such as syllabic elision (cf. Chapter 10). At sentence level Quichua shows a fixed (S)OV word order, as shown in (5).

- 5) *ñuka-ka* *ishkai* *churi* *chari-ni*
 1S-TOP two son have-1S
 ‘I have two sons’

intransitive verbs; b) the extension of reflexive *-ri-* to cover reciprocal meaning (cf. *supra*); and c) the use of reflexive *-ri-* on the model of Spanish impersonal passive *se*. For an argumentation of these changes as induced by contact, see Gómez Rendón (2007a).

Traditionally Quichua uses a nominalization strategy for clausal subordination (32). However, there is a tendency towards the replacement of nominalization with finite subordination on the model of Spanish subordinate clauses. The results are independent clauses linked by Spanish connectors. The subordination strategy makes use of Spanish subordinators such as the relativizer *que* ‘that’ (after *verba dicendi*), the relative pronoun *lo que* ‘that (which)’, and a few conjunctions such as *purki* ‘because’ or *si* ‘if’. Compare the nominalization in (6) with the subordination by means of Spanish *lo que* in (7).

- | | | | | |
|----|--|----------------|---------------------|-------------------|
| 6) | <i>chaya-shpa</i> | <i>paikuna</i> | <i>muna-shka-ta</i> | <i>apa-shka-n</i> |
| | arrive-GER | 3.PL | want-PTCP-ACC | take-PRF-3 |
| 7) | <i>chayashpa</i> | <i>paykuna</i> | <i>apa-shka-n</i> | <i>lo-que</i> |
| | arrive-GER | 3.PL | take-PRF-3 | that-which |
| | ‘Upon their arrival, they took what they wanted’ | | | |

The affluence of Spanish vocabulary in Quichua goes hand in hand with less visible changes at clausal and sentential levels. The co-occurrence of Spanish loanwords and syntactic calquing on the model of Spanish suggest some kind of correlation between both phenomena, so that a cause-effect chain between lexical borrowing and syntactic borrowing may be hypothesized. Other, less frequent syntactic changes induced by contact with Spanish include: i) Spanish SVO word order in declarative sentences; ii) Spanish SVO word order in non-verbal predicative constructions involving a copula; iii) shift from RelN to NRel with native interrogative pronouns used as relative pronouns; iv) question formation through Spanish interrogative intonation contours on unmarked declarative sentences; and v) replacement of nominalized clauses with adverbial subordinated clauses by means of Spanish subordinators. In addition, several Spanish connectors are used in Quichua, such as additive *y* ‘and’, contrastives *o* ‘or’ and *díno* ‘if not’ (Sp. *de no*) and disjunctive *pero* ‘but’. Spanish time adverbs are used as discourse markers (e.g. *aura* ‘nowadays’ (< Sp. *ahora*); *intonses* ‘then’ (< Sp. *entonces*); and *siympre* ‘always’ (< Sp. *siempre*). The days of the week and the times of the day are Spanish but co-occur with native items in loan blends.

The System of Parts of Speech in Quechua

In this section I substantiate my classification of Quechua as a flexible language (type-2) in Hengeveld’s classification. Type-2 languages have two lexical classes. One lexical class corresponds to heads of predicate phrases (verbs). The other class includes items that may be used in any syntactic position except as head of predicate phrases. The following examples from Peruvian Quechua (Schachter 1985: 17) support this claim.

- 8) a. *chay hatun runa* b. *rikashaka: hatun-ta*
 DEM big man see:PST:1S big-ACC
 ‘that big man’ ‘I saw the big one’
- 9) a. *chay alkalde runa* b. *rikashaka: alkalde-ta*
 DEM mayor man see:PST:1S mayor-ACC
 ‘That man who is mayor’ ‘I saw the mayor’

In these examples *hatun* is both a referential-phrase modifier (8a) and a referential-phrase head (8b). Similarly, *alkalde* is both a modifier in (9a) and a head in (9b). Evidence against this classification has been presented by Beck (2002: 144ff). According to Beck, the lack of a noun-adjective distinction in Quechua is not thorough because only the adjectives can be modified by adverbs like *maymi* ‘very’, as shown in (10):

- 10) *chay warmi maymi sumak-mi*
 DEM woman very pretty-FOC
 ‘That woman is very pretty’

Noun modification with *maymi* is ungrammatical, as illustrated by Cole (1985: 99-100) in (11) below. Still, when asked about the grammaticality of this sentence, however, several Quichua speakers in Imbabura and Bolivar considered (11) perfectly possible:

- 11) *chay warmi maymi duktur-mi*
 DEM woman very doctor-FOC
 ‘That woman is a real doctor’

While a semantic distinction between property concepts and entity concepts underlies the argument, Beck himself admits that “the existence of a semantic distinction of this type is in itself not enough to establish that there is a parts-of-speech distinction between nouns and adjectives in the lexicon” (2002:144). The theory of parts of speech by Hengeveld *et al* (2004) does not exclude such a distinction. Rather, it argues for the existence of a non-specialized lexical class for both concepts.

Beck’s second argument maintains that noun-noun constructions should be treated as compounds because nouns acting as attributes of other nouns cannot occur more than once in the same noun phrase – as opposed to adjectival modifiers which may be recursive. As additional evidence, Beck mentions that noun-noun constructions may be attributives of other nouns, as shown in example (12) from Cerrón-Palomino (1987: 300).

- 12) *hara chakra rumi*
 corn field stone
 ‘stone of the cornfield’

Readings of (12) such as ‘field stone of corn’ or ‘cornfield of stone’ are not possible. In fact, *hara* and *chakra* form an attributive compound that modifies *rumi*. The resulting interpretation of (12) is closely similar to English ‘stone of cornfield’. According to Quichua phonology, *hara chakra* should be realized as a compound if the main stress falls on the first syllable of *rumi*. However, I have not conducted a phonological analysis of these structures in order to know which stress pattern obtains.

Beck’s third argument against the typological classification of Quichua as a type-2 language states that property-concept words used as heads of referential phrases (13) are actually adjectives standing for deleted heads in elliptical constructions. A conclusive proof of this would be, for Beck, “their reliance on context to supply the identity of a nominal head” (2002: 145). Therefore, sentences like (39) are ungrammatical if out of context.

- 13) *puka-ta ri-ka*
 red-ACC see-PST
 ‘he sees the red one’

The claim that contextual reference is required for the correct interpretation of (13) is not conclusive either. Color terms are universally associated to objects and do not exist independently, being to this extent context-dependent in any human language and irrelevant for a noun-adjective distinction.¹⁷ On the other hand, it is not relevant that lexemes like *puka* in (13) are context-dependent, but that they occupy the syntactic position of referential heads without further measures and take nominal morphology (e.g. accusative markers).

Additional evidence for the classification of Quichua as a type-2 language comes from the fact that many lexical items used as referential-phrase heads and referential-phrase modifiers can be used as predicate-phrase modifiers too. *Yanka* ‘useless’ occupies the position of referential phrase modifier in (14) but also the position of predicate phrase modifier in (15) without any morphological derivation.

- 14) *kai-ka yanka yura ka-n-mi*
 that-TOP useless plant be-3-AFF
 ‘That is a useless plant’

¹⁷ Typically, dictionaries define colors with reference to physical objects. The Webster Dictionary, for example, defines ‘white’ according to metaphors such as intensity of light, racial groups and the like.

- 15) *kaina chaupi tuta-kaman yanka shuya-ku-rka-ni*
 yesterday middle night-up.to useless wait-DUR-PST-1S
 ‘Yesterday I waited until midnight in vain’

Similarly, *utka* ‘speed’ functions as head of the referential phrase in (16) and modifier of the predicate *shamui* ‘come’ in (17).

- 16) *utka-ka rura-shpa alli-mi ka-n*
 speed-TOP work-GER good-AFF be.3
 ‘quickness is good in working’
- 17) *utka huasi-man shamui-i*
 speed house-ALL come-IMP
 ‘come home quickly’

The arguments against the classification of Quichua as a flexible language are insufficient. I propose therefore to classify Quichua as a language which makes no distinction between nouns, adjectives and adverbs. This classification is corroborated in part by Adelaar and Muysken (2004), who state that “adjectives are similar to nouns in their syntactic behavior [and] it is not always easy to distinguish between the two categories” (2004: 208), although for both authors adjectives are different from nouns in that they cannot stand alone as subjects except if followed by *ka-q* ‘the one that is’. While this condition holds for subjects, it does not for adjectives in other syntactic positions, as shown by (13) above.

From the examples of Ecuadorian and Peruvian Quechua discussed above it is clear that lexical flexibility is characteristic of Quechua and thus must be considered one of its intrinsic typological features. The functional adaptation of Spanish borrowings provides additional support to this classification (cf. Chapter 10).

Ecuadorian Quichua remains typologically identical to other Quechua languages in agglutination, suffixation and verb-final word order, but differ from them in (pro)nominal morphology due to simplification. Spanish influence on Quichua consists in the addition of sound segments to the phonological inventory, the incorporation of function words such as determiners, quantifiers, connectors and adverbs, and the replacement of nominalization by relativization on the model of Spanish subordinate sentences. Albeit prominent, these changes have not modified the inherent typological character of Quichua.

6.4. Borrowing Hypotheses for Quichua

The language-specific hypotheses developed in the following will be tested in Chapters 10 and 11 on the corpus of Imbabura and Bolívar. The hypotheses involve

predictions about frequencies, types and functions of Spanish borrowings in the corpus. They are based on the hierarchies discussed in section 4.3 concerning a) the principle of functional explanation; b) the principle of system compatibility; c) the scales of borrowability; and d) the theory of parts of speech. The numeration corresponds to that followed in section 4.3.

Hypothesis from the Principle of Functional Explanation

H.1 Quichua will borrow Spanish discourse items easier than non-discourse items.¹⁸

H.1.1 Quichua will borrow Spanish discourse elements such as topic and focus markers but also evidentials and connectors.

Predictions from the principle of system compatibility

H.2 Quichua (agglutinative) will borrow from Spanish (fusional) free words and roots, but less likely clitics (e.g. pronominal proclitics) and bound morphemes (e.g. plural markers, gender markers, etc.).

Predictions from the scales of borrowability

H.3 Quichua will borrow Spanish lexical items easier than grammatical ones.

H.3.1 Quichua will borrow items from open lexical classes (e.g. nouns) easier than items from half-open classes (e.g. prepositions) and closed classes (e.g. pronouns).

H.3.2 Quichua will borrow Spanish lexical items in the following order of frequency: nouns, verbs, adjectives and adverbs. Adpositions (i.e. prepositions) will be borrowed, if at all, with less easily because Quichua (postpositional) does not have a syntactic slot for them. Function words such as conjunctions and pronouns will be borrowed only seldom. The pro-drop character of Spanish will disfavor the borrowing of Spanish pronouns in Quichua. Articles will not be borrowed at all.

Predictions from the theory of parts of speech

H.4 The typological distance between Spanish (source language) and Quichua (recipient language) is bridged in the borrowing process following the hierarchy of parts of speech: head of predicate phrase > head of referential phrase > modifier of referential phrase > modifier of predicate phrase.

H.4.1 Accordingly, Spanish forms that function as heads of phrases (i.e. verbs and nouns) will be borrowed easier than forms that function as modifiers (i.e. adjectives and adverbs). Also, Spanish forms that function as heads of predicate phrases (i.e. verbs) will be the most easily borrowed lexical class; forms that function as modifiers of predicate phrases (i.e. manner adverbs) will be the hardest class to be borrowed. While H.4.1 contrasts with H.3.2 above, both hypotheses will be tested.

¹⁸ Easier borrowability implies precedence in time and dominance in frequency.

- H.4.2** If Quichua borrows items from one lexical class, it will borrow items from previous lexical classes in the hierarchy. Accordingly, if Quichua borrows modifiers of referential phrases (Spanish adjectives), it will borrow heads of referential and predicate phrases (Spanish nouns and verbs) but not necessarily modifiers of predicate phrases (Spanish manner adverbs).
- H.4.3** As a flexible language, Quichua will borrow more easily lexemes from the lexical class immediately following the last differentiated lexical class in its system of parts of speech. Therefore, Quichua will borrow nouns more easily, because nouns are the lexical class that follows the last differentiated class (verbs) in its system.
- H.5** The syntactic distribution of borrowed lexemes in Quichua will follow the same distribution of native lexical classes (functional adaptation hypothesis). Accordingly, if Quichua borrows Spanish adjectives, it will use them as heads of referential phrases but also as modifiers of referential and predicate phrases, which corresponds to the distribution of native Quichua non-verbs). Therefore, Spanish borrowing will thus not modify the system of parts of speech in Quichua.
- H.6** The distribution of borrowed lexemes will follow the same distribution of their lexical classes in Spanish (functional specialization hypothesis). Accordingly, if Quichua borrows Spanish adjectives and adverbs, it will use them only in their original positions of modifiers of referential and predicate phrases but not interchangeably as if they formed one lexical class. The functional specialization of Spanish borrowings will thus result in a gradual differentiation of the parts-of-speech system of Quichua. While H5 and H6 make opposite predictions, both hypotheses will be tested.
- H.7** No predictions can be made from the lexicalization hypothesis because it applies only to rigid languages, and Quichua is flexible (cf. *supra*).

The foregoing hypotheses will be tested systematically on the Quichua corpus of Imbabura and Bolivar in the light of linguistic and nonlinguistic factors influencing the borrowing process (Chapters 10 and 11).

Chapter 7

Paraguayan Guaraní

Paraguayan Guaraní¹ or *avañe'ẽ* (people's language) is a Tupi-Guaraní language of the Tupi family spoken in Paraguay, the Argentinean provinces of Corrientes, Misiones, Formosa, Chaco and the north of Santa Fe, as well as in the southern part of the Brazilian states of Mato Grosso do Sul and Paraná (Dietrich 2002: 32). Tupi languages are spoken over an extensive area in South America, "approximately from 4° in the North to 30° in the South" (Gregores and Suárez 1967: 13). While Tupi languages keep a close resemblance to each other, similarity is notably reduced between Paraguayan Guaraní and other languages of the same family. Furthermore, *avañe'ẽ* is distinct from languages of the subfamily Tupi-Guaraní spoken in present Paraguay. Gregores and Suárez (1967) quote a statement from Cadogan in which he maintains that

"For a Paraguayan not used to have dealings with the Mbya [another Tupi-Guaraní language], it would be practically impossible to reach an understanding with, for example, a Mbya woman not used to contacts with Paraguayans. As to men, most of them have come into contact with Paraguayans and learnt how to express themselves, more or less, in 'Paraguayan Guaraní'. But the differences between both languages are great and even for me it is difficult to follow a conversation between two Indians when they are speaking in their own language" (Gregores and Suárez 1967: 16)

Other languages of the Tupi-Guaraní family spoken in Paraguay are Paĩ Tavyterã, Mbya, Chiripá, Ache, Chiripá, Tapieté and Chiriguano, the last two spoken also in Bolivia. Scholars (e.g. Dietrich 1996) have called these languages 'ethnic Guaraní' to distinguish them from Paraguayan Guaraní (called by some 'Mestizo Guaraní')² and Classical Guaraní (also called 'Jesuitic Guaraní'). It is generally assumed that Paraguayan Guaraní originated in one variety of 'ethnic Guaraní' once spoken in

¹ I refer to 'Paraguayan Guaraní' as Paraguay's national language to distinguish it from other Guaraní languages. Paraguayan Guaraní includes *Guaraníete* 'pure Guaraní' and *jopara* 'mixed Guaraní.' Both terms are not well differentiated in the literature. For some, Paraguayan Guaraní is *not* equivalent to *jopara*. These terms are discussed in section 7.2.

² Dietrich (2002: 31) notes that no strict equivalence exists between 'Guaraní criollo' (Mestizo Guaraní) and 'Guaraní paraguayó' (Paraguayan Guaraní) but he does not explain the difference. His use of *Guaraní criollo* is similar to my use of *Paraguayan Guaraní*, which refers to the national language of Paraguay spoken by people of non-Indian descent in this country. The use of the adjective 'mestizo' makes explicit reference to the mixture characteristic of contemporary *avañe'ẽ*. Various glottonyms are used in the literature, and debates persist around which term is the correct one.

Paraguay at the time of the Spanish conquest, but there is no way to establish with certainty which dialect contributed to the formation of present-day Guaraní. Neither do we know how Jesuitic Guaraní is related to present-day Guaraní nor how it contributed to its present form. From an extensive investigation (Thun, *Atlas Lingüístico Guaraní-Románico – Sociología*, 2002) it is clear that variation in present Guaraní is less dialectal than sociolectal and idiolectal.

In 1992 the percentage of Guaraní monolinguals (39.30%) was considerably higher than the percentage of Spanish monolinguals (6.40%), particularly in rural areas (MEC 1999). Also, the percentage of bilinguals (49%) was less than half of the country's population (4,152,588 in 1992). By 2002 bilinguals above five years increased to 59% (2,655,423 speakers) while Guaraní monolinguals decreased to 27% (776,092 speakers). By the same year the percentage of bilinguals from rural areas had increased to 17.62%, with a similar decrease in the percentage of Guaraní monolinguals in the same areas. Guaraní speakers including bilinguals and monolinguals above five years of age counted 3,946,904 people, according to the 2002 census³.

Other European speech communities in Paraguay include Portuguese, German and English. The majority of their members speak Spanish but only a small number speak Guaraní. On the other hand, the total population of speech communities of Tupi-Guaraní and other families amounted to 89,169 in 2002.

Table 7.1 Speakers above 5 years by area, sex and language (2002)

Languages	Country			Urban			Rural		
	Both	Men	Women	Both	Men	Women	Both	Men	Women
Guaraní	3,946,904	2,008,237	1,938,667	2,165,630	1,058,223	1,107,407	1,781,274	950,014	831,260
Spanish	3,170,812	1,552,319	1,618,493	2,285,301	1,092,874	1,192,427	885,511	459,445	426,066
Portuguese	326,496	177,504	148,992	205,977	110,161	95,816	120,519	67,343	53,176
Guaraní/ Spanish	2,655,423	1,312,980	1,342,443	1,862,561	901,305	961,256	792,862	411,675	381,187
Spanish/ Portuguese	264,706	145,361	119,345	191,338	103,057	88,281	73,368	42,304	31,064
Guaraní/ Portuguese	196,716	111,513	85,203	157,830	87,460	70,370	38,886	24,053	14,833

Source: Dgeec. *Resultados Finales Censo Nacional de Población y Vivienda 2002*

³ Census data must be taken with caution, however, because no census so far has measured the levels of bilingualism and the use of language in social spaces.

Table 7.1 show two major facts. Firstly, if societal bilingualism means the use of two languages by the majority of a country's population, Paraguay is *not* a bilingual country. While societal bilingualism may be attained in the near future thanks to the regular increment in the number of bilinguals over the last years, it is unlikely that Spanish and Guaraní become spoken on equal grounds in Paraguay,⁴ given the dominant position of Spanish in relation to Guaraní. Secondly, the geographical distribution of languages in urban areas (Spanish) and rural areas (Guaraní) is gradually disappearing as a result of the dissemination of bilingualism in detriment of (Guaraní) monolingualism. Still, Paraguay continues to be a unique case in the continent, but this uniqueness is founded less on the assumed bilingualism of the Paraguayan society, than on the fact that Guaraní is the only indigenous language in Latin America spoken by non-Indians as their mother tongue.

Fishman (1967) described Paraguay as a case of diglossic bilingualism, where people speak two languages but use them in different social contexts. The concept of diglossic bilingualism points out different uses but fails to recognize the existing linguistic conflict. This conflict is modeled by important sociopolitical factors. Rubin (1968) studied these factors in terms of socio-communicative contexts and variables of language choice. He was the first sociolinguist who linked language usage to political, social and economic power and described the influence of social conditions such as literacy, migration, social mobility and group cohesion on language choice. Rubin (1973 141-156) considers four major factors ('variables' or 'dimensions' in his terminology) influencing language choice in Paraguay:

- a) The first factor is geographic. It distinguishes rural from urban areas. The typical space of Guaraní is the countryside while that of Spanish are the cities.
- b) The second factor is the formality of the speech event. Rubin defines formality as "a limited set of expected behaviors" and informality as "the normal set of behaviors allowed within one group". Spanish is typically associated with formal speech events (especially if one of the interlocutors is socially dominant) while Guaraní is related to informal ones. Formality is partly determined by the physical setting of the speech event and the topic of the verbal exchange. Situations that are not strictly formal may develop gradually into informal ones.
- c) The third factor is the degree of intimacy. Spanish is generally associated with a lower degree of intimacy (and formality). In contrast, Guaraní is usually associated with a higher degree of intimacy. Solidarity and group identity may increase or foster intimacy. The relation between Guaraní and intimacy is not

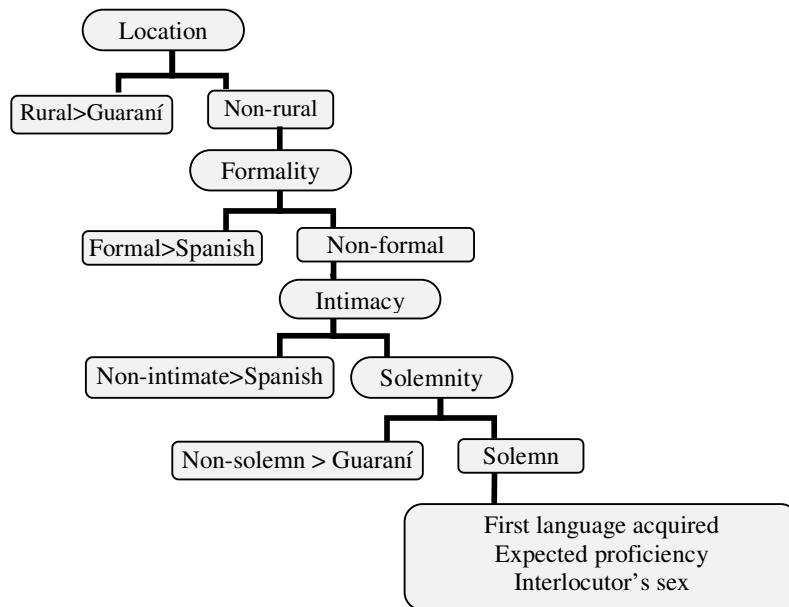
⁴ Unfortunately we do not have a sociolinguistic survey similar to that of Quichua (cf. section 5.2.), which provides quantifiable data on the uses of Spanish and Guaraní in different socio-communicative spaces. The most comprehensive – though by now outdated – sociolinguistic study of Paraguay is that by Rubin (1968).

predetermined however, because language choice is influenced very often by the speakers' mother tongue.

- d) The fourth factor is the degree of solemnity or seriousness of the speech event. Solemn speech belongs to Spanish while humorous speech is exclusive of Guaraní. Notice that a lower level of solemnity usually corresponds to a lower level of formality and intimacy.

Also, Rubin identifies minor factors that may cause deviations from the expected speech behaviors of interlocutors. These include the pressure by educational institutions on teachers, parents and students; the linguistic proficiency and the gender of the interlocutor; and the first language of acquisition. According to Rubin, major and minor factors change as society develops over time. As formal education and bilingualism increase, speakers are freer to choose one language or the other under less sociocultural pressure. This dynamic view of bilingualism would explain the changing language usage in today's Paraguay. Figure 7.1 shows schematically the interplay of factors determining language choice. The model assumes that higher dimensions determine lower ones. From top to bottom, language choices become specific and the sociolinguistic dimensions multiply.

Figure 7.1 Dimensions ordered by language choice



Based on Rubin (1968)

Rubin referred explicitly to the sociopolitical conflict of the Paraguayan society as a factor unbalancing the status of bilingualism. Many of his readers have downplayed the sociopolitical conflict by assuming wrongly that Paraguay is a model of stable bilingualism. Quite the opposite, the sociolinguistic situation in Paraguay is *not* stable but changing. Moreover, the complementary distribution of Guaraní and Spanish shows that Paraguay is typically diglossic⁵ even though bilingualism is becoming the rule. Rubin noticed that language shift was in progress already in the sixties. Interestingly, the shift did not favor one language over the other but fostered bilingual proficiency and the mixing of both languages (Rubin 1973: 126f). But what is the limit of Paraguayan bilingualism? If the use of one language remains mutually exclusive of the other, full bilingualism can be attained, other things being equal. Considering that no society is completely bilingual, Rubin suggests that societal bilingualism in Paraguay can be approached but never accomplished.

The idea of an unfinished bilingualism is shared by Melià (1973), for whom “real bilinguals (to be specific, coordinate bilinguals) are unviable in Paraguay, as it is not possible to master both languages, Spanish and Guaraní, with equal, or nearly equal, proficiency, not because of structural deficiencies of Guaraní but because of its specific developments” (Melià 1973: 26; my translation). These developments result from the coexistence of languages over a long period of time and eventually converge in the emergence of a third language genetically different from Spanish and Guaraní. The new language will grow out of a complex process of language mixing. This process was surmised by Rubin as the following quote shows:

“Este aumento en la habilidad bilingüe se refleja en el pueblo debido a un alto grado de ‘codeswitching’. Frecuentemente, cuando se les preguntaba a los informantes qué idioma usaban en situaciones específicas contestaban ‘dzopará’ [jopará] (mezcla) refiriéndose tanto al cambio de un idioma a otro entre frases como también a la mezcla más íntima dentro de las frases mismas. Las observaciones de los informantes revelan que un gran porcentaje de las conversaciones informales consiste en un equilibrio entre los idiomas que se utilizan en un mismo discurso. El cambio a un aumento en la habilidad bilingüe produce también un alto grado de ‘codeswitching’. Aunque no tengo datos históricos que documenten esto, se dice que hace veinte años, la gente hablaba más el Guaraní” [This increase in bilingualism is reflected on a high degree of ‘codeswitching’. When asked about what language they use in specific situations, informants usually answered ‘dzopará’ (mixture) in reference not only to the switch of languages from one phrase to another but also to the mixing within

⁵ For Melià (1973) the situation in Paraguay is neither bilingualism nor diglossic bilingualism (in Fishman’s sense) but diglossia. For Melià, only this term unmasks the true dimension of Paraguay’s linguistic conflict. In a similar way, von Gleich (1993) insists that nobody in Paraguay is bilingual, and Paraguayan bilingualism is more mythical than real.

phrases themselves. My informants' remarks show that a large percentage of informal conversations correspond to a balance between languages mixed in the same discourse.] (Rubin 1973: 127; my translation).

Melià's third language has a proper name. It is called *jopara* 'mixed Guaraní', and it is viewed as the opposite of *Guaraníete* 'pure Guaraní'. The specific characteristics of both varieties are discussed in section 7.2 in the context of linguistic variation in Paraguay. While *jopara* is typically associated with urban settings, *Guaraníete* is identified with the countryside. A clear-cut geographical division cannot be assumed however, since higher levels of mixture are naturally expected as rural speakers become bilingual (cf. *supra*). For many *jopara* is equivalent to colloquial Guaraní while *Guaraníete* is used only in literary works and textbooks. The current bilingual programs promote *jopara* as the language of schooling. Over the last decades many efforts have been made to 'cleanse' Guaraní *jopara* by producing prescriptive grammars and dictionaries that fill lexical gaps through neologisms and other equally fruitless strategies. The debate about which language should be used in education continues today.

The distance between *jopara* and *Guaraníete* is increasing day by day. According to Granda, "if the gap between actual language use and language reference models continues or increases, it might give rise not only to an unwanted collective complex of linguistic inferiority but also to an increasingly dangerous state of double internal diglossia" (1981: 134; my translation).

7.1. The history of Guaraní in Paraguay

Juan Díaz de Solís (1516) and Alejo García (1524) explored the territory of Paraguay with the purpose of finding an easier route to the Inca Empire. By 1525 the news of a silver booty seized by Garcia in the eastern slopes of the Andes encouraged Sebastian Gaboto to lead an exploration party along the Paraná and Paraguay rivers. A few years later Sebastián de Mendoza set up a large expedition to the estuary of Río de la Plata, where he founded the city of Buenos Aires in 1536. Sent by Mendoza to rescue the exploration party of Juan de Ayolas and Domingo Martínez de Irala in the Chaco, Juan de Salazar and Gonzalo de Mendoza founded the outpost of Asunción on the eastern banks of the Paraguay River in 1537. Only Irala and his party survived to the bellicose Indians of the Chaco and were forced to return to Asunción. There he met a strong resistance from Cario Indians (Guaraní), whose leader Lambaré was defeated in the outskirts of Asunción one year later.

The Guaraní-speaking groups inhabiting the territory of present Paraguay at the moment of the Spanish invasion were "the Carios, whose territory was limited by the Paraguay, Tebicuary and Jejuí rivers, on the one hand, and the highlands of

Ybyturuzú; the Tapé, who inhabited the highlands of the same name on the Grande del Sur river; the Chandules or Islander Guaraníes; the Itatines in northeastern Paraguay; the Chiriguanos in the eastern slopes of the Andes; and the Guarayos in Santa Cruz de la Sierra [Bolivia]" (Corvalán 1992: 2; my translation). The Spaniards realized very early that the linguistic homogeneity of the Spaniards could help their colonization enterprise. The contact was different in each case, not only because the Guaraní groups were many but also because the methods and goals of the conquerors changed according to the situation (Melià 1988: 18). While the Carios were contacted by military parties in the late 1530s (cf. *supra*), the Tapes were contacted by Jesuit missionaries in 1628.⁶ These opposite situations suggest that there were two different conquests.⁷ Each conquest had its own agents, its own place and time, but above all, its own methods.

The first Conquest of Guaraní

Few years after their defeat by Irala, the Carios entered into marriage alliances with the Spaniards. Indian women were given to the conquerors as domestic workforce and became mothers to a large numbers of mixed-blood children whom they cared for and raised alone. For some authors, this explains the spreading of Guaraní among the mestizo population of Paraguay unlike the events in other Spanish colonies where mestizos did not maintain the Indian language but assimilated linguistically to the dominant Spanish society. While miscegenation was decisive for the configuration of the linguistic situation in Paraguay, ethnic mixing did not encourage an equal use of Spanish and Guaraní. The assumption that miscegenation allowed bilingualism is deeply rooted in historiography and veils the sociopolitical domination of colonial Paraguay, where both languages coexisted *but* were used in mutually exclusive settings (cf. Melià 1988: 216).

Before the Spanish conquest the Carios and other Guaraní groups lived in settlements scattered over a vast area of land that covered present Paraguay, the southernmost part of Brazil and the north of Argentina. According to Clastres (1974: 79f) Guaraní territory covered an area of about 500,000 square kilometers. A conservative estimate of the Guaraní population at the time of the Spanish conquest gives an approximate of 200,000 people (Melià 1988: 239). Still, the scattered

⁶ Cario, Tovañi and Guarambaré Indians were the first Guaraní to be contacted by the Spaniards and were eventually absorbed in the mixing process. Tapé, Itatí and Paranaguá Indians had their first contact with Europeans only through Jesuit missionaries. Mbya Indians, in contrast, had only occasional contact with Europeans and preserved their culture and language to a great extent without Spanish influence well into the twentieth century (cf. Trinidad Sanabria 2002; Melià 1988).

⁷ The idea of the different conquests is inspired in the work of Melià (1988) about the four different types of 'reductions' to which Guaraní was subject during the last four hundred or so years (cf. *infra*).

settlement pattern of Guaraní peoples prevented the Spaniards from using their workforce at a large scale as they did in the Andes.

In Asuncion, the major settlement in Guaraní territory, Spaniards were easily outnumbered by Indians. This situation prevailed into the first half of the sixteenth century, even after the decimation of the Guaraní by epidemic outbreaks. In 1617 the number of Indians in Asuncion was 28,200 vis-à-vis 350 Spanish colonists (Necker 1975: 145). With such a demographic unbalance, intermarriage became the best strategy for Spaniards to create long-lasting bonds that encourage pacific coexistence. Spaniards used to marry several Indian women, and polygamy became a common practice in the district. As a result, miscegenation spread rapidly along with the Indian language, which was transmitted from the Indian mothers to their mestizo offspring. Mestizos became more numerous over the years but were eventually absorbed into the Spanish enclaves of the area. The easy incorporation of mestizos to the colonial society was facilitated in part by the absence of a strong caste system, which left space for interracial and intercultural practices. This does not mean however that Mestizos were not discriminated. They were indeed, but unlike Indians, Mestizos could mask their descent and became more 'Spanish-like' in a process of socio-psychological "whitening" (Maeder 1975: 82).

The initial demographic situation in Paraguay had a strong impact on the languages involved. Not only Guaraní spread all over Paraguay but its linguistic structure experienced noticeable changes as a result of usage in contexts different from those of pre-contact times. At the same time, Spanish continued to be used in all official transactions and was associated with the ruling elites. Still, Crown officials protested that Guaraní was displacing Spanish in the area to the point that even the few unmixed descendants of Spaniards who remained in Paraguay preferred to speak the native language with each other instead of speaking Spanish. The use of Guaraní by culturally indigenous mestizos who were part of the colonial society and by Spaniards who learned the language from their close contact with the overwhelming number of Mestizos and Indians required a series of adaptive strategies from the indigenous language which eventually shaped present Guaraní.

The second Conquest of Guaraní

The Guaraní peoples who remained outside Spanish influence were the object of evangelization enterprises, first by Franciscan and later by Jesuit missionaries. While Franciscan missions developed at the heart of the colonial matrix, Jesuitic missions (1610-1768) attained a higher degree of autonomy and self-support without parallel in Hispanic America. *Reducciones* or missions were villages formed by people from different ethnic groups under the rule of the Jesuits. Entry was prohibited to everyone except missionaries. This policy prevented any type of mixing with the Spanish population and saved Indians from the numerous epidemic

outbreaks that assailed towns and cities. In 1760, eight years before the Jesuits were expelled from the Spanish colonies, the number of Guaraní Indians had increased to 104,184 in the seven *reducciones* that existed in Paraguay⁸, while the population of Spaniards and Mestizos in urban centers counted 39,739 citizens (Maeder 1975: 81). A remarkable characteristic of *reducciones* was the considerable degree of independence granted to Indians for their own cultural and economic ways. The Guaraní language was one of the most cherished cultural traditions which flourished in the space of *reducciones*. No other language was used in the missions for oral and written communication in daily life, both by Indians and Jesuits. The latter made their best efforts to standardize the language by providing it with a phonological spelling, grammars, dictionaries and all kind of materials for religious indoctrination. Melià (1988: 249) calls this process *reducción* 'reduction'.⁹ He offers a detailed description of the standardization the Jesuits made of the language so that it can express all the concepts considered useful for the everyday and religious life of Indians. The effects of standardization in the structure of the language are discussed in the next section. For the time being suffice it to say that the Guaraní 'created' by the Jesuits sought to erase the dialectal differences present in *reducciones*. Arguably, a process of dialect leveling took place during the one hundred and fifty years of Jesuitic missionary administration and resulted in the creation of a *Guaraní koine* (Lustig 1996: 23).

After the missions were dissolved by the Crown under suspicion of creating an autonomous regime independent from Spain, their Indians either fled to the wilderness or integrated into the colonial society. Only a very small number of them remained in the surroundings of the former missions. They organized themselves in communities with a relative autonomy and fewer contacts with the outside world, thus surviving until 1848 (cf. *infra*) when a pro-Spanish nationalist regime ordered their dissolution (Plà 1970: 17).

⁸ A total of 29 *reducciones* were scattered in Paraguay (7), northern Argentina (15) and southern Brazil (7).

⁹ Melià identifies four 'reductions' of Guaraní since the Spanish Conquest: the Hispanic reduction; the Jesuit reduction; the National-Indigenous reduction after the Independence from Spain and the formation of the Paraguayan state; and the anthropological reduction by academic researchers (Melià 1988: 260).

Map 7.1 Jesuitical missions in the seventeenth and eighteenth centuries

Source: *Historia del Paraguay* (2003)

The differences between the Guaraní spoken by Indians in the *reducciones* and the Guaraní spoken by Mestizos and Spaniards in the urban centers were numerous, but most of them consist in changes induced in the language by contact with Spanish. The mutual influence between Jesuitic Guaraní and urban Guaraní remains unknown to date (Dietrich 1995: 204). From the present configuration of Paraguayan Guaraní it is obvious that the contribution of the Jesuitic Guaraní spoken by Indians from former *reducciones* who came to live in the towns became dissolved over the years until their eventual assimilation in the colonial society. Melià (1988) expresses this view in the following terms:

“Con la migración de los indios misioneros fuera de sus reducciones, movimiento que se irá prolongando durante el siglo XIX, y la mayor interferencia de la población criolla en aquellos mismos pueblos, es muy probable que la distancia dialectal entre ambas formas de la lengua Guaraní haya disminuido, pero en el sentido de una mayor criollización. Los factores que mantenían al Guaraní como “variedad alta” con escritura y literatura y con su relativa autonomía dentro de la reducción desaparecen, mientras se acentúan los factores que actuaban dentro del colonialismo criollo”. [With the migration of the Indians out of the *reducciones* – a process that extended well into the nineteenth century – and the increasing presence of the Mestizo population in the same town, it is very likely that the dialectal distance

between both varieties of the Guaraní language diminished, but in the direction of an increasing creolization. The factors disappeared that made Guaraní the “higher variety”, i.e. the variety with a writing system and literature tradition but also with relative autonomy inside the *reducciones*; at the same time, the factors playing a role in mestizo colonialism became important] (Melià 1988: 243; my translation).

Guaraní after the Independence

When Paraguay declared its independence from Spain in 1811, it was an isolated district of the Viceroyalty of Rio de la Plata, with minor economic and political relevance for the Crown. Guaraní was then spoken all over Paraguay by people of Indian and Spanish descent. But the new establishment did not create an auspicious context for the use of Guaraní in education and administration. The diglossic character of the colonial society remained unchanged. After their independence from Spain, Paraguayans continued to show ambivalent attitudes towards Guaraní: on the one hand, Guaraní is seen as the greatest symbol of Paraguayan identity; on the other, Guaraní is associated with backwardness and primitivism. This ambivalence explains why political leaders often took contradictory stands and made divergent decisions about the use of Guaraní.

The dictator Gaspar Rodríguez de Francia ruled the country during 26 years, from 1814 to his death in 1840. Apart from isolating Paraguay from the rest of the world and promoting an autarchic system based on economic self-support and agrarian communalism, Francia gave a new impetus to the old educational establishment by opening schools across the country and ordering that education be provided only in Spanish. Contradictorily, Francia himself used Guaraní for all administrative and political issues and considered it a distinctive trait of Paraguayan identity. Francia's successor, Carlos Antonio López, a European-grown progressist, was a blatant detractor of Guaraní. In 1848 López launched a campaign for the replacement of Guaraní family names with Spanish names. About the same year he ordered the dissolution of the few Indian communities that remained from the former Jesuit missions. However, this measure indirectly strengthened the use of Guaraní, because many Indians who were Guaraní monolinguals became incorporated to the mainstream Paraguayan society (Zajícová 2002: 4). The effects of such incorporation have not received yet any special attention from historians and linguists.

With López' death in 1862, his son Francisco Solano López became president of Paraguay and had to face a bloody war with Argentina, Brazil and Uruguay. While the aftermath of the war was incommensurable in demographic and economic terms, it strengthened Guaraní as a Paraguayan symbol. Guaraní was used in the battlefield and the trenches as a secret code or in folk songs, many of which became part of Paraguayan oral tradition. Francisco López realized the agglutinating

potential of Guaraní and organized during the war a Congress of Spelling (*Congreso de Grafía*) which set the first rules for the orthography of Paraguayan Guaraní. The first literary works written in Paraguayan Guaraní date from this time. Journals like *Cabichu'í* and *Cacique Lambaré* reported the events of war and mocked the warlike skills of the enemy.

Six years of war left a decimated nation at the mercy of the winning powers. And the winners did not have any interest in promoting Guaraní. A puppet regime was appointed on 15 August 1869 during a mass celebrated at the Cathedral by the General Vicar of the Argentinean Army, who preached about “the need to regenerate the Paraguayan people in order to promote their development” and the need to expel Guaraní from Paraguay for being “a dreadful creation of ignorance and backwardness” (García Mellid 1988: 34). Shortly afterwards, on the 7th of March 1870, the Minister of the Interior Cirilo Antonio Rivarola decreed on behalf of the provisional government that “schoolteachers and pupils are explicitly prohibited to use Guaraní in the classrooms, the only language of which shall be Spanish” (Zajícová 2002: 5). A few years later the Argentinean political writer and activist Domingo F. Sarmiento embarked on an educational reform for Paraguay following his ideological tenant of a new American civilization, in which “languages of wild men” had no place (Trinidad Sanabria 2002). For the next fifty years a linguistic policy that favored Spanish monolingualism at the cost of Guaraní prevailed in education and administration. War tested the agglutinating power of Guaraní once more in the 1930s. The Chaco War between Paraguay and Bolivia (1932-1935) motivated the recognition of Guaraní at all spheres of administration. Politicians, military and religious leaders usually gave their speeches and harangues in Guaraní while poets used to compose popular songs in the language.

Most of the liberal, revolutionary and dictatorial administrations of the twentieth century did nothing for the promotion of Guaraní. Winds of change came with the nationalization of the language on the 15th of August 1967, when the Constitution granted Guaraní the status of a ‘national language’. The new status did not have any practical consequence however. Only with the fall of Stroessner’s dictatorship in 1989 and the passing of a new Constitution in 1992, Guaraní obtained its official status on a par with Spanish. The Ministry of Education began to implement a bilingual education program in which every child must be taught in his/her own mother tongue. But the implementation of bilingual programs did not escape controversy: some people consider that bilingual programs perpetuate the same structures of oppression on Guaraní by giving too much space to Spanish borrowings; others consider that Guaraní and Spanish are related by their common history, so that it is inevitable that Guaraní carries all kinds of traces from Spanish, just like Paraguayan Spanish carries numerous traces from Guaraní.

Because Paraguayan Guaraní was used only in oral form until recently, the greatest challenge faced by language policy makers in Paraguay is the design of a

large-scale adaptation of the language to the contents of modern education. This adaptation, which Melià calls ‘the national-indigenous reduction of Guaraní’ (Melià 1988: 260), includes the following tasks: the construction of a specialized Guaraní lexicon for social and physical sciences; a new description of Guaraní according to its own categories; the normativization of Guaraní for its use in public spaces; the promotion of literature and the preparation of teaching materials in Guaraní; and the training of teachers in the implementation of bilingual education programs (Pereira Jacquet 2003). Of these goals the creation of a modern lexicon has absorbed the efforts of linguists and teachers during the last decade. Every year dictionaries appear to bridge the lexical gap between Guaraní and Spanish. Some of these dictionaries have been criticized by linguists on account of the flawed and arbitrary mechanisms used to create a mere magnum of neologisms that Guaraní speakers never use in daily communication (De Guaranía 1998; Melià 1998; Trinidad Sanabria 2002;). The writing of good descriptive grammars to facilitate the teaching of Guaraní according to its own linguistic categories has been completely neglected so far. In the new millennium language planning has become the arena of political disputes. Far from solved, controversies about education models proliferate. In a recent article about the ideologies behind the debate of Guaraní in bilingual education programs, Mortimer (2006) summarizes in very clear terms what seems to be the actual motivation of the dispute:

“More than being over the kind of Guaraní being used in schools, the current struggle seems to be over the degree to which the language is incorporated into the curriculum and the degree to which this incorporation might represent an additional academic challenge for students who have traditionally spoken the language of greater access and power—that is to say, the degree to which the incorporation of Guaraní into school challenges the advantages Spanish speaking students have traditionally enjoyed. The formal incorporation of both languages into public education undoubtedly represents improvement in access for Guaraní dominant children to both literacy and knowledge” (Mortimer 2006: 68).

The core of the controversies about the use of Guaraní in education seems therefore the confrontation of views about linguistic mixture. Underlying this confrontation are long-established relations of power between different sectors of a culturally and racially mixed society like the Paraguayan.

7.2. Language variation and language mixing in Paraguay

Paraguay is described as more homogeneous in linguistic terms than any other country in the Americas. The basis for this statement is the allegedly stable

bilingualism of Paraguay and the high degree of racial mixture at all levels of the Paraguayan society. While homogeneity is certainly prominent in Paraguay, the fact is that Paraguayan bilingualism is neither stable nor societal, but changing and diglossic.

The homogeneity of Paraguayan Guaraní is based on a narrow dialectal variation. Guaraní dialectal differences are visible between the variety spoken in Paraguay and the one spoken by Paraguayan immigrants in the Argentinean Province of Corrientes. The Guaraní of Corrientes is not simply “an extension of Paraguayan Guaraní but an independent variety of Guaraní which has evolved since the late 1800s” (Dietrich 2002: 34f; my translation). Two opposite tendencies are observed in Corrientes Guaraní: one is the archaic realization of certain sounds (e.g. the first person pronoun *che* is realized as [če] in Corrientes but as [še] in Paraguay); the other is the loss of several distinctive features of phonology (e.g. nasalization) and the lexicon (e.g. kinship terms). Gregores and Suárez (1968) report that their informants were aware of the differences between their (Paraguayan) dialects of Guaraní and the dialect spoken in Northeast Argentina, “but that they [the differences] were never so great as to impair communication in any serious way” (Gregores and Suárez 1968: 16). These authors notice that linguistic borders do not match political borders between Paraguay and Argentina. The same applies to the dialects of Guaraní spoken by Paraguayan immigrants in Brazil, even if, in this case, the contact language is Portuguese instead of Spanish. The findings of *Atlas Lingüístico Guaraní-Románico* (2002) point in the same direction.

Variation in Paraguayan Guaraní is more visible, in a diatopic perspective, between urban and rural varieties. However, the gap between both varieties is being gradually bridged by an increasing bilingualism in rural areas (cf. 5.3). In principle, the urban-rural split is correlated to differences in social class, economic position, education and age. Therefore, variation in Paraguayan Guaraní is sociolectal rather than dialectal. In other words, Paraguayan Guaraní is diastatically heterogeneous, but dialectally homogeneous.

In the literature rural Guaraní is associated with *Guaraníete* ‘true Guaraní’ while urban Guaraní is sometimes identified with *jopara* ‘mixed Guaraní’. However, there is no exact correspondence between these varieties and their assigned areas. It is perfectly possible to find *jopara* in rural areas – in fact some of our rural informants spoke more *jopara* than *Guaraníete*. Still, the degree and range of mixture is visibly lower in the countryside. *Guaraníete* has been also identified with academic Guaraní, i.e. the language created by scholars through a systematic ‘cleansing’ of the Spanish lexicon (Mortimer 2006: 2). Academic Guaraní is used only by a small number of educated Paraguayans in formal settings (Lustig 1996: 20; Rodríguez-Alcalá 2002: 79). The other side of the coin is *jopara*, i.e. the colloquial variety of Guaraní spoken by most Paraguayans. *Jopara* carries numerous lexical and grammatical imprints from Spanish. In general, there is no consensus

about which variety is referred to by one term or the other. Let us see some definitions of *Guaraníete* and *jopara* in the literature.

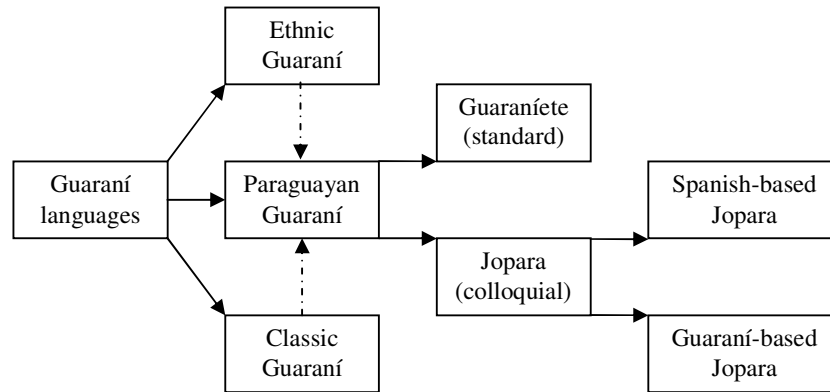
Dietrich equates *Guaraníete* with the ‘educated standard variety’ of Guaraní (Sp. *norma culta*) as opposed to *jopara*, the mixed language “characterized by many lexical and syntactic influences from Spanish” (2002: 40). In similar terms, Lustig (2000: 2) associates *Guaraníete* with a Guaraní purged of Spanish items. Neither Dietrich nor Lustig are specific about whether *Guaraníete* is equivalent to Paraguayan Guaraní or not. Still, the assumption is implicit in both authors that *Guaraníete* is one form of Paraguayan Guaraní. The definitions of *jopara* in the literature are more numerous. Lustig provides a definition of *jopara* as “a mixed language from Spanish and Guaraní in which most of the Paraguayans communicate in their daily life” (Lustig 1996: 1; my emphasis), but he adds that *jopara* is better described as a ‘mixture of languages’ rather than a ‘mixed language’, since it is *not* a language in strict terms, because it has no rules. A similar view is held by Dietrich (1993: 18) who considers *jopara* a non-stabilized mixed language on its way either to normativization or replacement by Spanish. In terms of use and distribution Lustig defines *jopara* as “a diastatic and diasituational variety of Paraguayan Guaraní which occupies an intermediate position in a continuum of different degrees of Hispanicization or Guaranítization, from ethnic Guaraní through pure academic Guaraní to Paraguayan Spanish and Standard Spanish” (Lustig 1996: 3; my translation). That is, *jopara* is the variety in which Paraguayan Guaraní is realized in daily communication. A rather different definition of *jopara* appears in an official document on the Paraguayan educative reform prepared by the Paraguayan Ministry of Education (MEC). This document defines *jopara* first as the lexical borrowing which is not integrated into Guaraní phonology and morphosyntax, and by extension as the variety which uses non-integrated forms. *Jopara* is the opposite of *jehe’a*, i.e. the lexical borrowing which, by virtue of its integration to the structure of the language, is part and parcel of the Guaraní lexicon and follows its orthographic rules (MEC 2004). A further distinction is made according to the possible types of *jopara*: the use of unincorporated single words (lexical *jopara*); the mixing of Spanish and Guaraní within one syntactic unit (syntactic *jopara*); and the use of Spanish and Guaraní in alternated form within one text (discursive *jopara*). Both *jopara* and *jehe’a* are different from Paraguayan Guaraní, the language used by most Paraguayans in daily communication.

In sum, there are several definitions of *Guaraníete* and *jopara*, some subtler than others but all used more or less interchangeably. The list of terms referring to either variety can be long, as noticed by Mortimer (2006: 59), who found as many as twenty-six terms for *Guaraníete* and no fewer than twenty-three for *jopara*. Add to the list two terms associated to *jopara* which are increasingly used in specialized and non-specialized circles: one term is *guarañol*, a hybrid from *Guaraní* and *español*, coined by Melià (1988: 247); the other is *castení*, a hybrid from *castellano* and

Guaraní. While both terms refer to language mixing, they are not interchangeable and mean two different types of mixture. *Castení* refers to the *jopara* whose matrix language is Spanish; *guarañol* refers to the *jopara* whose matrix language is Guaraní. In the following I provide a characterization of *jopara guarañol* and the place it occupies in the continuum discussed in section 5.1.3.

A classification of Guaraní varieties is represented in Figure 7.2 below. The Guaraní branch of the Tupi family distinguishes three sub-branches: ethnic Guaraní, spoken by indigenous peoples of Paraguay; Classical Guaraní, developed in the Jesuitic missions during the seventeenth and eighteenth centuries¹⁰; and Paraguayan Guaraní, the language spoken by Paraguayan Mestizos. Paraguayan Guaraní is further divided into *Guaraníete* (standard Guaraní used in written form and formal speech) and *jopara* (colloquial Guaraní). Paraguayan Guaraní received two contributions: from Classical Guaraní when the Jesuitic missions were dissolved and their Indians came to live in the towns; and from ethnic Guaraní, through the ongoing migration of members of Guaraní ethnic groups to the cities. Finally, *jopara* distinguishes two varieties according to the matrix language involved: Spanish-based *jopara* or *guarañol* and Guaraní-based *jopara* or *castení*.¹¹

Figure 7.2. Language varieties within the branch of Guaraní languages



¹⁰ Notice that Classical Guaraní is different from *Guaraníete*, even if both resemble each other in several ways, mainly in the minimum influence of Spanish in their respective systems.

¹¹ Strictly speaking, though, *castení* should not belong to Guaraní, because its matrix language is Spanish.

Borrowing and language mixing

Nobody knows exactly when Spanish borrowing in Guaraní started. From other colonial settings in the Americas we assume that it was in the early years of colonization. In fact, Spanish borrowings in Amerindian languages are as old as the Spanish presence in the continent. For example, many aboriginal languages preserve Spanish words long disappeared from modern Spanish or words which are pronounced as in old Spanish: e.g. *parlana* 'to speak' in Quichua, from old Spanish *parlar* 'to speak'; *obexa* 'sheep' [obešá] in Guaraní, pronounced [obexa] in modern Spanish; *domi* 'money' in Otomí, from *tomín*, a type of Spanish currency in the sixteenth century. These words and many others are conspicuous evidence of early contacts with Spanish. In the communicative setting of Spanish colonization, lexical borrowing was a common practice. Of course, the wide gap between borrowing and mixing proper has been bridged only by few languages.

Mixing as defined here is the massive entry of foreign elements in a language which ends by changing its original configuration. Some authors call this process 'relexification' (e.g. Muysken 1981): a term from Creole linguistics to describe the massive lexical replacement occurred in certain non-Creole languages. Two well-known cases of mixing in Amerindian languages come from Nahuatl (Hill and Hill 1977, 1986) and Media Lengua (Muysken 1979, 1985, 1997; Gómez Rendón 2005, 2008b). I propose to include jopara as another case of language mixing for a number of reasons to be explained later in this section. The few linguistic studies of mixing in *jopara* are Domínguez (1982), Armatto de Welti (1982), Lustig (1996) and Gómez Rendón (forthcoming/a).

Language mixing in diachronic perspective

The social and cultural conquest of Guaraní peoples and their language had two different settings, as explained in the previous section. One took place in the colonial urban centers founded by the Spaniards since 1537 and involved intense miscegenation (*mestizaje*) in cultural and racial terms. The other took place in the Jesuitic missions and involved the isolation of indigenous people from the colonial society, which enabled them to preserve a great part of their former life style and their own language. Neither setting was harmless however. Spanish settlers and missionaries, each in their own ways, undertook the 'reduction' of Guaraní peoples. This reduction was far more systematic in the case of the Jesuitic missions. The Spanish settlers did not make any effort to provide the indigenous language with an orthographic system of graphemes and rules while the Jesuits did so. In the urban centers Guaraní was spoken alongside Spanish while in the missions it was the only language for communication. Guaraní experienced an intense contact with Spanish in colonial towns by a steady increase in the number of bilinguals among Mestizos

and Spaniards. Still, the linguistic processes undergone by Guaraní in both settings resemble each other very closely. They involved the Hispanicization of the lexicon and certain grammatical categories of the indigenous language.

In the lexicon the process involved three strategies: the borrowing of Spanish lexical and grammatical items; the use of native Guaraní words with Spanish meanings; and the formation of new words on the basis of Spanish semantic structures.

From early religious works and the proceedings of indigenous councils it is possible to trace the use of Spanish borrowings back to the late sixteenth century. In the first Guaraní catechism written by Fray Luis de Bolaños (1583), besides religious names (e.g. Jesus, Maria) and formulae (e.g. amen) we find a few Spanish words including *padre* 'father', *gracia* 'grace' and *cruz* 'cross'. In a 1753 document, written by the president of the Indian council Nicolás Ñeengyrú to the governor of Buenos Aires, we find Spanish borrowings of administrative character (e.g. *rey* 'king', *cabildo* 'council', *Corregidor* 'royal representative') but also a few words from basic vocabulary (e.g. *señor* 'sir', *nombre* 'name'). If Spanish was present in the written language of monolingual clergymen and mission Indians, it is not unrealistic to assume its presence in the oral language of bilingual Mestizos and Spaniards in towns. A series of official documents written in Guaraní by Mestizo leaders of the Independence wars show an abundance of lexical and grammatical borrowings from Spanish (cf. Romero 1992).

The mechanisms of semantic calquing and word formation had far-reaching consequences for the development of the language. They consisted in the mapping of Spanish semantic units onto native forms and in the creation of neologisms based on Spanish word-formation rules. Both practices were familiar in the colonial period, especially among the Jesuits. Like in other areas of the Spanish Empire, the missionaries used native words to express religious concepts. In this case the advantage was that missionaries find no resistance from religious officials, who criticized vehemently the use of native languages for religious indoctrination in the case of Quechua, for example (Mannheim: 1991: 65). The word *tupã* is illustrative in this respect. Originally, *Tupã* was one of the highest Tupi-Guaraní divinities. Jesuits used this word to mean 'God'. The same word served to create neologisms such as *Tupã-sy* 'God-mother' for the Virgin Mary, or *Tupa-o* 'God-house' for the church. Later on *Tupã* was replaced by *Ñande-jára*, literally 'Our Lord.'

In the grammar, Guaraní lost several grammatical categories. According to Zajícová (2002: 3), because it was women, not (Spanish) men, who transmitted Guaraní to their offspring, the language lost certain gender-based categories. One of them was the use of the affirmative adverb, which had originally two forms, *ta* in men's speech and *heẽ* in women's speech. Nowadays we find only the second form in Paraguayan Guaraní while the other is preserved in ethnic varieties. Similar changes affected the kinship system. Pre-contact Guaraní used a gender- and age-

based categorization of kinship, according to which the speaker's sex and age determined the use of the reference term. Thus, a woman addressed her brother as *kyvy* regardless of age while a man called his brother *ryke'y* (if he was older than his brother) or *ryvy* (if he was younger). Similar distinctions were made by sisters and brothers when referring to older or younger siblings. Of this fine-grained classificatory system only a few terms remain while most have been replaced by Spanish kinship terms which do not make similar distinctions (Dietrich 2002: 33f).¹²

These glimpses into the contact-induced changes that occurred in Paraguayan Guaraní give us an idea of how different the language is from pre-contact Guaraní – or from ethnic Guaraní for that matter. The study of language mixing in Paraguay is largely limited by the absence of written documents or primary sources that show the evolution of mixing through the aforementioned strategies. Fragmentary evidence comes from a few testimonial narratives of Crown administrators and visitors during the colonial period. The following is one of the earliest references to language mixing in Paraguay.

“Todo el vulgo, aun las mujeres de rango, niños y niñas, hablan el Guaraní como su lengua natal, aunque los más hablen bastante bien el español. A decir verdad, mezclan ambas lenguas y no entienden bien ninguna...Así nació una tercera o sea la que usan hoy en día”
[all people, even elite women, boys and girls, speak Guaraní as their mother tongue, even though most of them speak Spanish as well. Actually, they mix both languages and do not understand either properly...In this way a third language emerged, which is the one they use nowadays] (Dobrizhoffer 1783, quoted in Melià 1974: 59).

The first written documents in *jopara* date from the War of the Triple Alliance. As mentioned above, the Guaraní language became then an agglutinating symbol of Paraguayan identity. Two journals written in *jopara* under the name of *Cacique Lambare* and *Cabichu'i* appeared between 1867 and 1868. These early pieces of Paraguayan folk literature have been analyzed by Lustig (2002). *Cacique Lambare* and *Cabichu'i* were Guaraní monolingual publications for soldiers on the front. Considering this readership, the editors used colloquial Guaraní only. The following excerpt is taken from *Cabichu'i* (quoted in Lustig 2002: 4).

¹² For a similar development in Ecuadorian Quichua, see Gómez Rendón 2007a.

Extract from *Cabichu'i* (*jopara* literature from the late nineteenth century)

Paraguayan Guaraní (<i>jopara</i>)	English Translation
Toikove 21 de octubre tres de noviembre ndive mburuvicha ha'e soldados umi ára javeve. Taimarã'e'ÿ entero ikatu haguã ogosa ñande Karai Guasu ome'ëva condecoración eta. Enterove por parejo jafelicita chupekuéra, ñande Mariscal remime'ë jarohory hendivekuéra, Toikove ñane Retã ñande Mariscal ndive, ha'e umi Mburuvicha eta oascende ramo va'ekue.	On the 21st of October and the 3rd of November our officials and soldiers shall celebrate. Everybody shall be party in honor of the many decorations gained by our Great General. Each and all of us shall congratulate them, and with them we shall show our joy as a gift for our General. That our Land may live and so live our General and the promoted officials.

The text contains sixteen different Spanish items among numerals, nouns, verbs and adverbs. Spanish items are adapted to the morphosyntactic structure of Guaraní. Interestingly, the editors of this journal state explicitly that they are using 'pure' Guaraní. Lustig quotes another text in *jopara* from the Chaco War against Bolivia, composed by a famous Paraguayan folk singer, in which its author boasts his use of 'pure Guaraní' but uses a large number of Spanish borrowings or code switches. The entire *jopara* literature of the late nineteenth and early twentieth centuries takes as its point of departure the events of war and depicts Guaraní as *the* symbol of the nation, regardless of any concern about mixture.

The historical record suggests that *jopara* has been spoken in Paraguay at least since the seventeenth century, and since then it has been strongly associated with ethnic (Paraguayan) identity. Speaking *jopara* always implied loyalty to this identity. Only recently *jopara* has been negatively associated with impurity. Clearly, certain changes in the linguistic ideology of the Paraguayan society occurred in the second half of the twentieth century. The attacks on *jopara* became harder since the early seventies, some years after the declaration of Guaraní as a national language, and have reached their peak in the last decade, after the promotion of Guaraní to the status of official language and the implementation of the Paraguayan Educative Reform. Ironically, the new status of Guaraní has encouraged the underestimation of colloquial Guaraní. While there are many influencing factors to be consider, those of ideology and politics are among the most crucial, as explained by Mortimer (2006: 68).

But what type of mixture is *jopara*? As explained above, *jopara* includes two varieties ways: one takes Spanish as its morphosyntactic matrix (section 5.1.3.1); the other takes Guaraní as the matrix. The first variety was discussed in section 5.1.3.1. Here I discuss the variety of *jopara* based on Guaraní morphosyntax. Part of the following discussion was presented elsewhere (Gómez Rendón, forthcoming/a).

Assuming that language mixing is determined by a combination of the lexicon and the grammar of the languages in contact, four different combinations of Spanish and Guaraní are possible. This is shown in Figure 7.3 below.

First, any combination of lexicon and grammar from two dialectal varieties of one language produces another dialectal variety of the same language, one that is typologically similar to the dialects which contributed to its emergence. Because *jopara* is not a mixture of dialects but one of different languages, the mixtures which combine lexicons and grammars of either Guaraní (G) or Spanish (Sp) are not considered *jopara*.

Figure 7.3 Combinatory possibilities of Spanish-Guaraní language mixing

		GRAMMAR	
		G	Sp
L E X I C O N	G	G	J _{Sp} (CASTENÍ)
	Sp	J _G (GUARAÑOL)	Sp

We are left thus with two possible *jopara* mixtures: one (J_{Sp}) whose matrix language is Spanish and (most of) its lexicon Guaraní; and another (J_G) whose matrix language is Guaraní and (most of) its lexicon Spanish. I have proposed to call the first variety *casteni* and the second *guarañol*. The first variety (J_{Sp}) is not reported in the literature because linguists and sociolinguists refer to *jopara* only as ‘Hispanicized Guaraní’. However, this variety (J_{Sp}) exists indeed in the form of Guaraníticized Spanish, as discussed in section 5.1.3.1. Castení speakers usually live in the cities, particularly in Asunción; their socioeconomic status is low and their education level is usually elementary. For the second type of mixture (J_G) the evidence is abundant. It includes part of the corpus collected for this investigation. The text corresponds to a speaker of Ciudad del Este. Spanish loanwords are italicized and switches appear in square brackets.

Example from Jopara

Upei aju agotyó amba'apo [compra venta de mercaderías, principalmente mercaderías]_{CS1} ojeguerúva Brasilgui, [comestibles, ropa, un poco de todo upei otra actividad]_{CS2} ajapova'ekue avei are *avende*, *avende plástico* [bolsas de plástico]_{CS3} upea ajogua Paraguaygui, ha *la* ante koa ko *trabajo* ajapova hina ko'agã, aha avei *coloniape* aha aykyti yvyra, ambo'i ajapo chugui peteĩ *corte pisora* oipuruva umi *europeo* {...} ha che agueru [de que como paraguayo]_{CS4} ñande jaguereko *la* ñande *identidad*, ha amogotyó *lado* ouramo peteĩ [cultura diferente]_{CS5} omoingue ñandeve ko'ape, [dificultad algunas veces]_{CS6} *la* ñane ñe'ẽ Guaraníme, *porque* omoingue ñandéve [otra cultura]_{CS7} ha avei upekuevo ñande ñacompara ñande *cultura* iculturakuéra ndive, ha noipe'a ñande hegui michimi pe ñande [identidad nacional]_{CS8}, pe ñande rekoteve, ha'e avei heta ñande *paisano*, ñane retaygua, *aveces* opreferi [la otra cultura]_{CS9} *entonces* oipe'a ñandehegui [un poquitito]_{CS10} *la* ñande *paraguayo* reko. Péicha che *atopa* [la influencia del portugues]_{CS11} (FC, Paraguay, 2004)

Given the strong presence of code switching, one major question is in what proportion code-switching and borrowing occur in this variety of *jopara*. From the statistical analysis of texts (Gómez Rendón, forthcoming/a) I have shown that code-switching tends to be more frequent than borrowing (1.37 to 1) and inter-sentential switches more numerous than intra-sentential ones. A detailed analysis of the morphosyntactic matrix allows to conclude that: 1) Guaraní is the matrix language of this variety of *jopara* because it provides most of the system morphemes; 2) word order is Guaraní and Spanish, even though syntactic calquing from the latter is prolific; 3) constituent order in the noun phrase is Guaraní; 4) the order of morphemes in derivation and inflection is Guaraní, even though *jopara* is less complex, both morphologically and syntactically, than traditional Guaraní (cf. *infra*).

The next question is whether this variety is still Guaraní in typological terms. To answer this question, we need first to describe the typological features of traditional varieties of Guaraní and compare them to those of innovative varieties such as *jopara*. This is done in the next section.

7.3. Paraguayan Guaraní: a typological characterization

The following typological description is based on traditional Guaraní as spoken in rural areas. Notice that rural Guaraní is not necessarily equivalent to *Guaraníete* or 'pure Guaraní'. Rural Guaraní is a variety used in daily communication and characterized by a lesser influence from Spanish; *Guaraníete* refers to a standardized

variety purged of Spanish elements. The typology of traditional Guaraní¹³ will be compared with that of *jopara* Guaraní. The term ‘Paraguayan Guaraní’ is a cover term for traditional Guaraní and *jopara* Guaraní.

Paraguayan Guaraní is one of several Guaraní languages of the Tupi Guaraní family (cf. Figure 7.2). The Guaraní branch extends over several countries, including Bolivia, Paraguay and Argentina. The Tupi-Guaraní family covers a wider area, thereby representing the largest language family in South America in geographical distribution (cf. Dietrich 1990). Guaraní is spoken also in the northeast of Argentina and the south of Brazil by Paraguayan immigrants.

The phonological inventory of traditional Guaraní includes twenty-six sounds: fourteen consonants (/p/, /t/, /k/, /s/, /ʃ/, /h/, /m/, /n/, /ŋ/, /v/, /y/, /ɣ/, /r/, /ʀ/); twelve vowels among oral and nasal (/a/, /ä/, /e/, /ẽ/, /i/, /ĩ/, /i/, /ĩ/, /o/, /õ/, /u/, /ũ/). This inventory differs from the one presented by Gregores and Suárez (1968) in two respects. On the one hand, it does not include the lateral /l/, the voiceless labiovelar stop /kʷ/, the voiced labiovelar nasal /ŋʷ/, or the voiced fricative labiovelar velar /ɣʷ/. Lateral /l/ shows a low frequency and occurs mostly in Spanish and other borrowings. Segments /kʷ/, /ŋʷ/ and /ɣʷ/ are allophonic realizations of the non-labiovelar phonemes /k/, /ŋ/ and /ɣ/ before and after vowels and after /i/. On the other hand, the inventory includes nasals as distinct phonemes and *not* as allophones of oral vowels (Gregores and Suárez 1968: 82f).

In addition to the aforementioned sounds, the phonological inventory of *jopara* Guaraní includes six sounds /Φ/, /č/, /ð/, /ř/, /l/ and /ʎ/. With the exception of /l/ and /ʎ/, which may come from another indigenous language (cf. Gregores and Suárez 1967: 89), the occurrence of these phonemes is limited mostly to Spanish loanwords. On occasion these sounds appear in native items, especially in the speech of younger bilinguals. Segments /Φ/, /č/, /ř/ show the same primary articulation as native phonemes /p/, /š/, /r/ but differ from them in their secondary articulation. Laterals /l/ and /ʎ/ have no native counterparts in the place and manner of articulation and thus may be considered exclusive of *jopara* Guaraní. A significant degree of free variation is found across *jopara* idiolects between /č/-/š/ and /l/-/r/. The vowel inventory of *jopara* Guaraní has remained virtually untouched by Spanish, except for the tendency observed in bilingual children and young adults to either relax the high central vowel /i/ to produce [ɪ], or pronounce it like the fricative velar [ɣ]. Because this phenomenon is limited to urban lects, it is possible to state that the six-vowel set of traditional Guaraní is preserved in the vast majority of *jopara* speakers.

Nasal harmony and spreading nasalization are two salient features of the suprasegmental phonology of traditional Guaraní. Both features show the effects

¹³ Traditional Guaraní not only has fewer Spanish loanwords but the existing ones are assimilated to its phonological system. *Jopara* has many more Spanish loanwords, most of which are unassimilated.

from contact in *jopara* varieties. Bilingual children and young adults do not (fully) nasalize affixes attached to nasal roots (e.g. reciprocal *jajo* instead of *ñaño*). Regressive and progressive nasalization does not occur either in the speech of all *jopara* speakers (e.g. *mitānguéra* ‘children’ is sometimes realized as [mita’ngwera] instead of [mitā’gwera]).

Stress in traditional Guaraní typically falls on the last syllable. Loanwords are assimilated to this pattern, i.e. “with stress in the last syllable, no matter in which syllable the stress originally fell” (Gregores and Suárez 1967: 91). Here *jopara* Guaraní makes a difference once again, because the majority of loans occurs unassimilated and preserves primary stress in the same syllable as in the source language. An example is Spanish /késol/ ‘cheese’, which occurs assimilated as [kesú] in traditional Guaraní but unassimilated as [késol] in *jopara* Guaraní. A thorough description of assimilation of Spanish loanwords in Guaraní is presented in section 10.1.2.

The main syllabic pattern of traditional Guaraní is CV, although CVC syllables are not infrequent. Onsets and codas consist are always monophonemic (Gregores and Suárez 1968: 61). There is no restriction in onsets. Codas may be only /m/, /n/ or /ŋ/. Spanish loanwords with consonant clusters in onsets are assimilated by adding a syllable with the same vowel as in the original syllable. One of the earliest Spanish loanwords assimilated along this pattern is *cruz* ‘cross’, pronounced as *kurusu* in traditional Guaraní. Because Spanish loanwords in *jopara* Guaraní usually occur unassimilated, there are no restrictions for onsets and codas. Accordingly, one finds clusters formed by a plosive and a flap (e.g. /tr/, /pr/) and sibilants in coda position – especially in plural words borrowed as frozen expressions (e.g. *kosa-s-kuéra* ‘things’).

Morphologically, traditional Guaraní is defined as agglutinative and polysynthetic. It has prefixes, suffixes and circumfixes. Consider the affirmative sentence in (1) and its negative counterpart in (2):

- 1) *ne-mo-memby-jevý-ta*
 3_{subj}2_{obj}-CAUS-have.son-again-FUT
 ‘He will make you have a son again’
- 2) *no-ne-mo-memby-jevý-ta-i*
 NEG-3_{subj}2_{obj}-CAUS-have.son-again-FUT
 ‘He will not make you have a son again’

(Lustig 1996: 19)

Both prefixes (*ne-*, *mo-*) and suffixes (*-jevý*, *-ta*) are added to the root *memby* ‘son’ in (1). The circumfix *no-i* indicates negation in (2). A larger number of affixes may

be attached to the root. Interestingly, example (1) comes from a *jopara* speaker (Lustig 1996). *Jopara* preserves the morphosyntactic structure of Guaraní, despite the plethora of Spanish lexical borrowings. Still, there is an increasing tendency in this variety to depart from traditional polysynthesis towards a higher degree of analyticity. Consider the answer given by a *jopara* speaker to the question whether knowing a second language is good for monolinguals:

- 3) a. *chéve guarã nda-i-perhudisial-r-i*,
 1.OBJ for NEG-3.PRS-detrimental-EUPH-NEG
 b. *re-mombarete-ve-hína pene arandu*
 2S-strengthen-MORE-PROG 2.POSS knowledge
 c. *a-medida-que la ñe'ẽ rei-kuaa*
 inasmuch.as DEM speak 2S-know
 'For me it is not bad, because you strengthen your knowledge to the extent you know the language'

The example contains three Spanish borrowings: the adjective *perhudisial* 'detrimental'; the complex conjunction *a medida que*; and the article *la*. Let us focus on the linking strategies in (3). Although clause (b) is semantically dependent on (a), the causal relation made explicit in the English translation through the conjunction 'because' is only implicit in (3) in so far as both clauses are linked by simple juxtaposition. On the other hand, (c) is linked to (b) by Spanish *a medida que* 'to the extent that', thereby indicating a 'fulfilled condition' and subordinating (c) to (b). While both linking mechanisms coexist in *jopara*, traditional Guaraní shows a strong preference for the use of parataxis and postpositions – instead of connectives. This makes *jopara* somewhat less synthetic than traditional Guaraní. Compared to an equivalent construction in traditional Guaraní, clause (c) is a syntactic calque from Spanish. In (c) the verb head *kuaa* 'to know' has two arguments, the second-person subject expressed by the prefix *rei-* and the object *la ñe'ẽ* 'the language'. The construction is fully grammatical in *jopara* and yet syntactically different from (4), in which noun incorporation has taken place, thus leaving one explicit argument:

- 4) *Re-ñe'ẽ-kuaa*
 2S-speak-know
 'you know how to speak (it)'

In fact, the increasing replacement of noun incorporation with phrasal constructions is a strong evidence of the greater degree of analyticity in *jopara*. In general, incorporated constructions are more frequent in traditional Guaraní. Compare examples (5a-b).

- 5) a. *A-johei-ta* *che-juru* b. *A-je-juru-hei-ta*
 1S-wash-FUT my-mouth 1S-REFL-mouth-wash-FUT
 'I will wash my mouth' 'I will wash my mouth'

As it seems, polysynthesis is *not* the rule in *jopara*. It may be hypothesized that Spanish connectives influence decisively the degree of synthesis, but further analysis is required. Loan connectives are discussed in Chapter 11.

Pre-contact Guaraní did not make gender distinctions in nouns while number marking was optional. Traditional and mixed varieties of Paraguayan Guaraní still lack gender marking in nouns, but *jopara* Guaraní tends to mark number with more frequency. Also, pre-contact Guaraní did not have articles to express definiteness, but traditional Guaraní and *jopara* Guaraní use Spanish articles *la* for singular and *lo*¹⁴ for plural (Gregores and Suarez 1967: 144). Spanish articles are used somewhat differently in Paraguayan Guaraní. In (6) *la* precedes the possessive adjective, which is ungrammatical in Spanish:

- 6) *ij-apyte-pe-kuéra* *o-u* ***la*** *che* *tio*
 3.POSS-middle-LOC-PL 3S-come DEM 1.POSS uncle
ha o-henoi la iñ-ermano-kuéra
 and 3-call DEM 3.POSS-sibling-pl
 'My uncle came with them and then called his brothers and sisters'

Articles in Paraguayan Guaraní perform a demonstrative (deictic) function (Lustig 1996: 10). They are used to mark definiteness, reference and cohesion in discourse. The use of Spanish articles in Paraguayan Guaraní is described in Chapter 11.

The order of constituents in the Guaraní noun phrase is head-modifier in attributive constructions and modifier-head in possessive constructions. Traditional Guaraní has basically two ways of expressing possession: the juxtaposition of nouns in the order possessor-possessed as in (7); and the composite postposition *-pegua* attached to possessor noun as in (8).

- 7) *umi* *organización* *dirihente-kuéra* *ndive* *ro-ñe'ẽ*
 some organization leader-PL with 1PL.EXC-speak
 'we speak with some leaders of the organization'
- 8) *mbyja* *ára-pegua* *o-mombe'u* *Ñandejara* *i-pu'aka-ha*
 star sky-ABL 3-tell Our.Lord 3-be.poweul-REL
 'The stars from the sky tell the power of God'

¹⁴ The form *lo* comes from the masculine plural article *los* after the elision of the sibilant.

The juxtaposed construction in (7) is similar in meaning to the possessor-ablative construction in (8). Ablative constructions are often used as fixed expressions and show relatively low frequency (cf. Guasch 1997: 62). Neither plus human nor minus alienable are determining factors in possession marking. Alternatively, *jopara* uses the Spanish preposition *de* between possessed and possessor. Consider the following example:

- 9) *Oĩ-há-pe guive o-je-gueraha preso padre-de-familia*
 3.be.REL-LOC FROM 3-PASS-take imprisoned parent-of-family
 ‘Since then, parents of families were imprisoned’

Jopara constructions with *de* are restricted to Spanish loanwords. This suggests that they should be analyzed rather as phrasal borrowings (cf. Chapter 10). Other Spanish prepositions are not borrowed into *jopara*, and the language remains postpositional.

While the word order of traditional Guaraní is SOV, there is a tendency in Paraguayan Guaraní to SVO due to Spanish influence (Gregores and Suarez 1967: 182). Still, word order in Paraguayan Guaraní remains relatively free. Personal pronouns are dropped with frequency. If explicit, they serve emphatic and contrastive purposes. On the other hand, Paraguayan Guaraní is an active-stative language in which case marking is based on active-inactive distinctions (Velázquez-Castillo 2002).

The foregoing discussion shows that *jopara* resembles traditional Guaraní in features such as affixation and word order, but it differs in others like clause linking, articles and connectors.

The System of Parts of Speech in Paraguayan Guaraní

Paraguayan Guaraní is a flexible type-2 language in Hengeveld’s classification. Accordingly, it has only two lexical classes: verbs and non-verbs. The class of verbs is clearly identified by the existence of two morphological paradigms (*areal* verbs and *aireal* verbs) as shown in the following examples.

- | | |
|---|--|
| 10) a. (Che) <i>a-guata</i>
1S 1S-go
‘I walk’ | b. (Nde) <i>re-mba’apo</i>
2S 2S-work
‘You work’ |
| 11) a. (Che) <i>ai-pota</i>
1S 1S-want
‘I want’ | b. (Nde) <i>rei-pota</i>
2S 2S-want
‘You want’ |

Non-verbs occupy any of the following syntactic positions without further measures: head of referential phrase, modifier of referential phrase, and modifier of predicate phrase. The following examples illustrate the syntactic flexibility of non-verbs:

- 12) a. *Ko karai tuja* b. *Che tuva tuja*
 DEM man old 1S father old
 ‘This old man’ ‘The oldness of my father’
- 13) a. *Che ro-hayhu asy* b. *Nde rayhu asy*
 1S 2.OBJ-love intense 2S love intense
 ‘I love you passionately’ ‘Your passionate love’

The same lexeme, *tuja*, modifies a referential phrase in (12a) and heads a referential phrase in (12b). Likewise, *asy* modifies a predicate phrase in (13a) and a referential phrase in (13b). Another feature typical of Paraguayan Guaraní is the capacity of most lexemes to be used predicatively. This feature is most visible in the case of quality-attributive verbs (Gregores and Suárez 1967: 138), which may be used as heads of predicate and referential phrases, as shown in (14a-b):

- 14) a. *a-vy’á ne-recha-rehe* b. *a-vy’á ne-recha-vo*
 1S-happiness 2S-see-by 1S-happiness 2S-see-when
 ‘My happiness of seeing you’ ‘I am happy to see you’

In similar terms, a predicative reading of *tuja* in (12a) is “to be old”. The predicative use of nouns, adjectives and manner adverbs is illustrated in the following examples:

- 15) a. *Pe kyse puku* b. *Che che-kyse*
 DEM knife red 1S 1S.POSS-knife
 ‘That red knife’ ‘I have a knife’
- 16) a. *A-jahe’o pochy-rehe* b. *Che che-pochy*
 1S-cry angry-by 1S 1S-anger
 ‘I cry from anger’ ‘I am angry’
- 17) a. *o-mbohovai mbarete* b. *o-mo-mbarete*
 3-react strongly 3-CAUS-strongly
 ‘He reacts strongly’ ‘He strengthens [it]’

Despite the extensive predicative use of most lexemes, the existence of a clear-cut lexical class of ‘pure’ verbs identified on the basis of their morphology (cf. Nordhoff 2004) prevents a classification of Paraguayan Guaraní as a type-1 language.

The examples given in support of my classification of Paraguayan Guaraní as a type-2 flexible language come from traditional Guaraní. Neither loanwords from Spanish nor syntactic calquing occur in these examples. The question is whether this classification is valid also for *jopara*. To answer this question I analyze examples from colloquial Guaraní collected in the field. The assumption to be confirmed is that *jopara* Guaraní maintains the same distribution of parts of speech as traditional Guaraní. In the following examples Spanish borrowings appear underlined.

- 18) *o-ñe'ẽ* la *Guaraní-me*
 3-speak ART Guaraní-LOC
 si *ha'ekuéra* *oi-pota* la kampesino vóto
 if 3.PL 3-want ART peasant vote
 'They speak in Guaraní if they want to get the peasant vote'
- 19) porque *pe* *nde* *mitã-ramoguare* *reĩ-ramo-guare*
 because DEM 2S child-WHEN.PST 2S.be-WHEN.NMLZ.PST
 nde servicio-marina-pe entero *re-gueruka* *cheve*,
 2S navy-LOC all 2S-send 1S.ACC
 'Because you sent me your photos since your childhood till you entered the navy'
- 20) *o-ñe'ẽ* atravesado la *Guaraní*
 3-speak crossed ART Guaraní
 'They speak Guaraní in a confusing manner'

From a flexible language which makes no distinction between nouns, adjectives and adverbs, it is expected that (1) Spanish nouns may be used as adjectives, (2) adjectives as nouns, and (3) adjectives and nouns as adverbs. The first hypothesis is met by (18), where the Spanish noun *campesino* 'peasant' modifies the noun head *voto* 'vote'. The fact that *campesino* can be used also as adjective in Spanish contributes to a similar use in *jopara*. The second prediction is confirmed by (19), where the Spanish adjective *entero* 'entire' is used as a noun, with the meaning of 'everything'. Finally, the prediction about the behavior of adjectives as adverbs is confirmed by (20), where the Spanish adjective *atravesado* 'crossed, mixed up' modifies the Guaraní verb *ñe'ẽ* 'speak'. Notice that these examples do not involve a process of derivation. But *jopara* shows also the extended predicative use of most lexemes, as illustrate by the following examples.

- 21) *nda-che-tiempo-i* la *a-japo* *haguã* otra-cosa
 NEG-1S-time-NEG PRO_(x) 1S-do for other-thing_(x)
 'I don't have time to do other things'

- 22) *i-conocido-iterei* *nd-o-guereko-i-ha* *la* *culpa*
 3-known-very NEG-3S-have-NEG-REL ART blame
 ‘It is well known that he is not guilty’
- 23) *Nda-i-deprovecho-mo’ãi* *chupe* *la* *Guaraní*
 NEG-3-useful-FUT 3.ACC ART Guaraní
 ‘Guaraní will not be useful for him/her’

Examples (21) and (22) include the Spanish noun *tiempo* ‘time’ and the adjective *conocido* ‘known’. Both lexemes carry verbal morphology: the verbal prefix *che-* (first-person singular) attached to the loanword *tiempo* in (21) makes the verb ‘to have time’; the pronominal prefix *i-* (third-person) in (22) promotes the loan adjective *conocido* to the category of verb. Notice that the prefixes in (21) and (22) are *not* derivational but inflectional morphemes. Finally, example (23) shows the prepositional phrase *de provecho* ‘of use’ used as a predicate with the meaning of ‘to be useful’. The bulk of the evidence attests *jopara* as a flexible language. This flexibility does not imply however that Spanish loanwords are *always* used in non-prototypical functions. Examples of borrowed lexemes used in their original lexical classes are numerous. Thus, Spanish verbs are *always* used as heads of predicate phrases (24) and manner adverbs always used as modifiers of predicate phrases (25).

- 24) *jamás* *na-ñe-komunika-mo’ãi*
 never NEG-REC-communicate-FUT
la *ña-ñe-komunika* *háicha* *la* *Guaraní-me*
 PRO 1PL-REC-communicate like PRO Guaraní-LOC
 ‘They will never communicate in the way we communicate in Guaraní’
- 25) *Che* *nd-ai-kuaá-i-nte* *exactamente*
 1S NEG-1S-know-NEG-only exactly
la *mba’e* *parte-pa* *la* *nde* *róga* *oĩ*
 ART what place-INT ART 2S.POSS house 3.be
 ‘I just don’t know exactly in what place your house is’

Certainly, not all lexemes are equally likely to be used as verbs. This means that a semantic constraint operates for native elements of open classes too. Notwithstanding this restriction, the fact that Spanish loanwords may be used also in non-prototypical positions confirms that *jopara* maintains the system of parts of speech of traditional Guaraní.

7.4. Borrowing hypotheses for Paraguayan Guaraní

The language-specific hypotheses presented in this section are tested in Chapters 10 and 11 on the Guaraní corpus collected in Paraguay. The hypotheses involve predictions about frequencies, types and functions of Spanish borrowings in the corpus. They are based on the hierarchies discussed in section 4.3 concerning a) the principle of functional explanation; b) the principle of system compatibility; c) the scales of borrowability; and d) the theory of parts of speech. The numbers correspond to those in section 4.3.

Predictions from the Principle of Functional Explanation

- H.1** Paraguayan Guaraní will borrow Spanish discourse elements easier than non-discourse elements.
- H.1.1** Paraguayan Guaraní will borrow Spanish discourse elements such as topic and focus markers but evidentials and connectors.

Predictions from the principle of system compatibility

- H.2** Paraguayan Guaraní (agglutinative) will borrow from Spanish (fusional) free words and roots, but less likely clitics (e.g. pronominal proclitics) and bound morphemes (e.g. plural markers, gender markers, etc.)

Predictions from the scales of borrowability

- H.3** Paraguayan Guaraní will borrow Spanish lexical elements easier than grammatical ones.
- H.3.1** Paraguayan Guaraní will borrow items from open lexical classes (e.g. nouns) easier than items from half-open (e.g. prepositions) and closed classes (e.g. articles).
- H.3.2** Paraguayan Guaraní will borrow Spanish lexical items in the following order of frequency: nouns, verbs, adjectives and adverbs. Adpositions (i.e. prepositions) will be borrowed, if at all, less easily because Paraguayan Guaraní (postpositional) does not have a syntactic slot for them. In contrast, articles may be borrowed to the extent that a syntactic slot for them is available in Guaraní. In turn, pronoun borrowing will be disfavored by the pro-drop character of Spanish will disfavor the borrowing of Spanish pronouns. Conjunct borrowing is not expected, other things being equal.

Predictions from the theory of parts of speech

- H.4** The typological distance between Spanish (source language) and Paraguayan Guaraní (recipient language) is bridged in the borrowing process following the hierarchy of parts of speech: head of predicate phrase > head of referential phrase > modifier of referential phrase > modifier of predicate phrase.
- H.4.1** Accordingly, Spanish forms that function as heads of phrases (i.e. verbs and nouns) will be borrowed easier than forms that function as modifiers (i.e.

adjectives and adverbs). Also, Spanish forms that function as heads of predicate phrases (i.e. verbs) will be the most easily borrowed lexical class; forms that function as modifiers of predicate phrases (i.e. manner adverbs) will be the hardest class to be borrowed. While H.4.1 contrasts with H.3.2 above, both hypotheses will be tested.

- H.4.2** If Paraguayan Guaraní borrows items from one lexical class, it borrows items from previous lexical classes in the hierarchy. Accordingly, if Paraguayan Guaraní borrows modifiers of referential phrases (Spanish adjectives), it will borrow heads of referential and predicate phrases too (Spanish nouns and verbs) but not necessarily modifiers of predicate phrases (Spanish manner adverbs).
- H.4.3** As a flexible language, Paraguayan Guaraní will borrow more easily lexemes from the lexical class immediately following the last differentiated lexical class in its parts-of-speech system. Therefore, Paraguayan will borrow nouns more easily, because nouns are the lexical class that follows the last differentiated class (verbs) in its system
- H.5** The syntactic distribution of borrowed lexemes in Paraguayan Guaraní will follow the same distribution of native lexical classes (functional adaptation hypothesis). Accordingly, if Paraguayan Guaraní borrows Spanish adjectives, it will use them as heads of referential phrases but also as modifiers of referential and predicate phrases, which corresponds to the distribution of native Guaraní non-verbs. In addition, all Spanish borrowings might be used alternatively as predicates given the same use of lexical classes in Paraguayan Guaraní. Therefore, Spanish borrowing will thus not modify the system of parts of speech in Guaraní.
- H.6** The distribution of borrowed lexemes will follow the same distribution of their lexical classes in Spanish (functional specialization hypothesis). Accordingly, if Paraguayan Guaraní borrows Spanish adjectives and adverbs, it will use them only in their original positions of modifiers of referential and predicate phrases but not interchangeably as if they formed one lexical class. The functional specialization of Spanish borrowings will thus result in a gradual differentiation of the parts-of-speech system of Paraguayan Guaraní. While H5 and H6 make opposite predictions, both hypotheses will be tested.
- H.7** No predictions can be made from the lexicalization hypothesis because it applies only to rigid languages and Paraguayan Guaraní is flexible (cf. *supra*).

The foregoing hypotheses will be tested systematically on the Guaraní corpus in the light of linguistic and nonlinguistic factors influencing the borrowing process (Chapters 10 and 11).

Chapter 8

Otomí

Otomí or *hñäñho*¹ is spoken in the Mexican states of Hidalgo, México and Querétaro, with some speakers also in Puebla and Veracruz. It belongs to the Otopamean branch of the Otomanguean language family, along with Pame, Chichimeca, Mazahua, Matlatzinca and Ocuilteco. The Otomanguean family ranks second in geographical distribution after the Uto-Aztecan (Ortiz Alvarez 2005: 37). The Otomí varieties studied here are spoken in the state of Querétaro. In addition to traditional Otomí areas in central Mexico, speakers of *hñäñho* are present in a few towns in Guanajuato and Michoacán as well as in the town of Ixtenco in the Nahuatl-dominant state of Tlaxcala. Otomí is spoken in these enclaves only by a handful of speakers, even if people still consider themselves ethnically as Otomí. At the same time, an ever-increasing number of Otomí speakers have migrated to Mexico City, Monterrey, Guadalajara and Mazatlán.

A marked process of dialectalization is observed in present-day Otomí as a result of a pattern of scattered settlement across states and the lack of contact among Otomí areas. Tlaxcala Otomí is the most deviant variety in comparison to Querétaro Otomí, with no contact among speakers of these varieties and a lower degree of mutual intelligibility. Dialectal variation in Otomí and other issues of genetic classification are discussed in section 8.3. In addition to dialectalization, many Otomí communities are experiencing a rapid shift to Spanish, particularly in the state of Mexico,² where the highest levels of migration are attested (Barrientos López 2004: 6).

The 1970 census gave a total number of 221,080 speakers of *hñäñho*, unevenly distributed over eight states. The 2000 census show rather similar numbers. In that year the states with the largest number of Otomí speakers (Hidalgo, Mexico and Querétaro) counted a total of 241,496 speakers. The Otomí population

¹ The word 'Otomí' is an ethnic denomination considered negative by speakers of this language, who prefer to call themselves *ñäñho* and their language *hñäñho*. Following Hekking (1998: 8) the etymology of the word *hñäñho* means 'well-spoken language', and *ñäñho*, 'those who speak well'. Another hypothesis has it that the root *-ñho* is a derivation of the word *xĩu* 'nose' in reference to the nasal character of Otomí. Hekking (1995: 8) grants no credibility to this hypothesis because it is unlikely that such a term is used for self-identification.

² This process becomes evident if we consider the total population of Otomí households in the state of Mexico for 2000. Of 279,036 individuals living in Otomí households, only 104,579 (37.5%) reported to be speakers of Otomí. This noticeable gap shows not only that a rapid process of language shift is taking place in most Otomí communities as indicated above, but also that a large number of these communities remain ethnically loyal after losing their language.

of Veracruz, Puebla, Guanajuato and Tlaxcala is minimal. However, these numbers do not coincide with those given by *Comisión Nacional para el Desarrollo de los Pueblos Indígenas*, a state agency in charge of promoting the development of indigenous peoples in Mexico, according to which the number of Otomí speakers was 327,318 for the same year. Censual data should be read with caution because all censuses so far have failed to account for two relevant facts. One is the migration of indigenous people to the cities, where they usually report as non-speakers of indigenous languages even if they actually are, for reasons of low linguistic and ethnic self-esteem (Hekking 1998: 21). The other is ethnic identification, according to which non-speakers of Otomí report themselves as speakers on the basis of their identification with the Otomí culture, especially in communities where the language is widely spoken. In view of these factors, it is more likely that the total number of Otomí speakers should rise above 300,000, and this is in fact the figure used by most linguistic and anthropological studies.

Otomí is the strongest indigenous language in the state of Querétaro. There are very small numbers of speakers of Mazahua and Pame on the state border with the states of Mexico and San Luis Potosí.³ Otomí is spoken along with Nahuatl (Uto-Aztec) and Tepehua (Totonaco-Tepehua) in the state of Hidalgo, and along with Nahuatl and other Otomanguean languages including Mazahua, Matlatzinca and Ocuilteco in the state of Mexico. In all the states, however, Otomí is in permanent contact with Spanish. Mexico shows an ongoing process of Hispanicization of its indigenous peoples. This process is especially visible among Otomí speakers, as it becomes clear from a comparison of the percentages of indigenous people who speak native languages in Mexico (Table 8.1 below). Unlike other indigenous peoples with high percentages of language maintenance –Tzeltal and Tzotzil are the most remarkable – speakers of *hñāñho* represent only half of the Otomí ethnic population. A similar degree of language loss is shown by Mazahua, another language of the Otopamean branch.

³ Other newcomers are speakers of Mixe (Mixe-Zoque) and Nahuatl (Uto-Aztecan), though in small numbers.

Table 8.1 Total population and number of speakers per indigenous language

Ethnic group	Population	Native Speakers
Tzeltal	384,074	87.6%
Tzotzil	406,962	86.9%
Mazateco	305,836	80.5%
Huasteco	226,447	76.5%
Mixteco	726,601	70.3%
Náhuatl	2,445,969	67.5%
Totonaca	411,266	66.1%
Zapoteco	777,253	65.1%
Maya	1,475,575	60.5%
Otomí	646,875	50.6%
Mazahua	326,660	46.5%

Source: Comisión Nacional para el Desarrollo de los Pueblos Indígenas

Table 8.2 shows the percentage of speakers of the languages of the Otopamean branch from the total ethnic population. The Otopamean language with the highest degree of language loss is Ocuilteco, spoken in Mexico State, where there are speakers of Mazahua and Otomí as well.

Table 8.2 Total population and number of speakers in the Otopamean branch

Ethnic group	Population	Native speakers
Pame	12,572	77.7%
Chichimeca Jonaz	3,169	62.7%
Otomí	646,875	50.6%
Matlatzinca	3,005	47.9%
Mazahua	326,660	46.5%
Ocuilteco	1,759	29.7%

These data suggest that the linguistic vitality of Otopamean languages is severely endangered, especially if compared to languages from other families such as Uto-Aztec (Nahuatl) or Mayan (Tzeltal, Tzotzil), but also to Otomanguean languages such as Zapotec or Mixtec.

Relevant factors influencing language shift among Otomí speakers in Querétaro include the following, according to Hekking (1998: 19-21): the lower socioeconomic status of Otomí speakers; a traditional association of Otomí language and culture with negative features; the comparatively small number of Otomí speakers in relation to other ethnic groups; and the lack of contact among speakers of different Otomí varieties. While each of these factors contributes differently to the process of language shift and loss, the first of them is the most influencing one in my opinion. The small size of the Otomí population with respect to other ethnolinguistic groups is only a secondary factor, because the number of Otomí speakers is larger than the number of speakers of other indigenous languages in Mexico. From a dynamic demographical perspective, the non-contact among Otomí speakers is also decisive. At a community scale isolation gives the impression that Otomí speakers are few, thereby reinforcing a linguistic ideology of 'minority group'.⁴

From an examination of different sources I conclude that no agreement exists about the number of speakers of Querétaro Otomí. Ortiz Álvarez (2005: 55) gives 22,077 speakers in 2000. However, the sum of Otomí speakers from the highlands and the semi-desert given by Mendoza *et al* (2006: 10) amounts to 19,321 speakers in the same year. Still, both figures are lower than those for the states of Hidalgo and Mexico, with 110,043 and 104,357 speakers, respectively. According to Mendoza *et al* (2006) speakers of Querétaro Otomí were concentrated in the municipality of Amealco (13,007), while their number in the municipalities of Colón, Cadereyta, Peñamiller, Tolimán, Pedro Escobedo, Enrique Montes and Tequisquiapan was only 6,314. A demographic report prepared by SEDESU (2006) on the basis of *II Censo de Población y Vivienda 2005* show that these numbers do not account for all the Otomí population in the state, because a large number of Otomí speakers are settled today in the capital Querétaro.

⁴ Isolation is being reduced nowadays, because Otomí groups of different states are supporting organizational and political initiatives for integration. Hekking (1998: 21) noticed an ethnic mobilization in the mid nineties.

Table 8.3 Otomí Population in indigenous households⁵ by municipality in 2005

Municipio	Población total	Población indígena	% respecto al municipio
Querétaro Arteaga	1,598,139	41,091	2.6
Amealco de Bonfil	56,457	18,261	32.3
Tolimán	23,963	8,529	35.6
Querétaro	734,139	7,229	1.0
Cadereyta de Montes	57,204	1,995	3.5
San Juan del Río	208,462	1,936	0.9
Corregidora	104,218	692	0.7
Ezequiel Montes	34,729	420	1.2
Jalpan de Serra	22,025	402	1.8
Tequisquiapan	54,929	302	0.5
El Marqués	79,743	265	0.3
Pedro Escobedo	56,553	265	0.5
Colón	51,625	218	0.4
Arroyo Seco	12,493	148	1.2
Landa de Matamoros	18,905	127	0.7
Peñamiller	17,007	103	0.6
Pinal de Amoles	25,325	95	0.4
Huimilpan	32,728	57	0.2
San Joaquín	7,634	47	0.6

Source: SEDESU, *Anuario Económico*, 2006: 80

As shown in Table 8.2, the major Otomí areas are the highlands in the municipality of Amealco (*Sierra Queretana*) and the semi-desert in the municipalities of Tolimán, Cadereyta, Colón and Peñamiller. Excluding the urban center of Querétaro, which is not a traditional Otomí area, the population of the highlands and the semi-desert represents around 70% of the entire Otomí population in the state. The following map shows the geographic distribution of Otomí speakers in the state of Querétaro plus that of other minority groups such as Pame and Huastec settled in a few villages in northern *Sierra Gorda*, with a population of 1035 speakers in 2005.

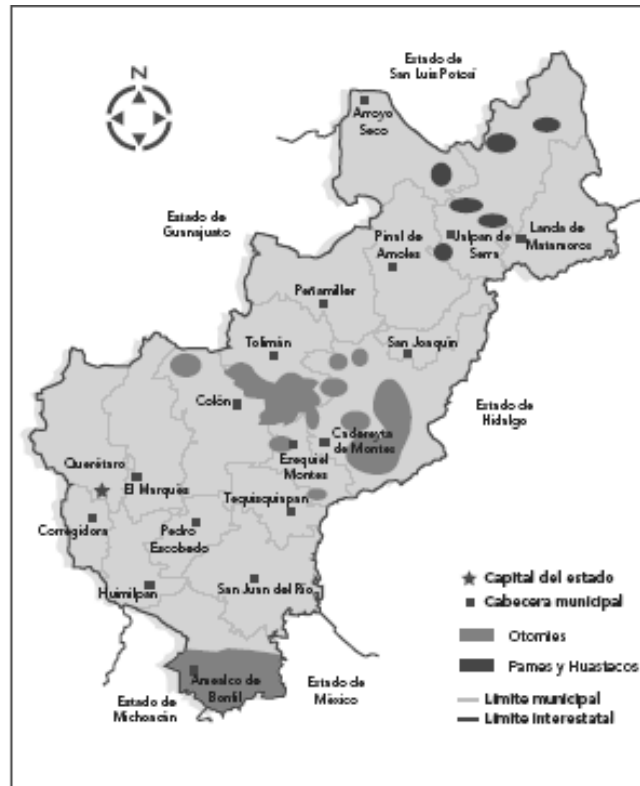
The Otomí dialects of Santiago Mexquititlán and Tolimán are spoken in the municipalities of Amealco (highlands) and Tolimán (semi-desert), respectively. These communities concentrate the majority of Otomí speakers in their respective municipalities: 18,261⁶ and 8,529 speakers, respectively. The Otomí population in these municipalities is distributed over 142 *barrios* or counties. Otomí speakers in rural communities show a traditional pattern of scattered settlement, especially in Tolimán. In the semi-desert area 9,055 speakers live in 72 *barrios* with less than 250

⁵ The table includes people from households whose head and/or his/her partner speak the indigenous language. It includes small numbers of Pame and Huastec speakers representing 3% of the population.

⁶ For the Otomí population of Santiago Mexquititlán between 1970 and 1986, see Hekking 1998: 21.

inhabitants (SEDESU 2005: 80). This type of settlement is common to most Otomí communities in central Mexico and exerts a major influence on the process of dialectalization.

Map 8.1 Languages spoken in the state of Querétaro, central Mexico.



Source: SEDESU, *Anuario Económico*, 2006: 79

The overall number of Querétaro Otomí speakers corresponds to 0.4% of all indigenous speakers in the country (Ortiz Álvarez 2005: 61). While this percentage supports Hekking's idea about the size of the speech community as a factor influencing language shift, it is clear that Otomí speakers represent the overwhelming majority of indigenous people in Querétaro. Comparatively, speakers of Hidalgo Otomí are more numerous (110,043) but coexist with a larger Nahuatl population (221,684). Also, the number of Otomí speakers in the state of Mexico is five times (104,357) larger than the number of Otomí speakers in Querétaro but the former live together with a medium-size population of Nahuatl speakers (55,802)

and a large-size population of Mazahua speakers (113,424). The degree of language loss is higher in Querétaro even though Otomí is the only indigenous language spoken in the state. In neighboring states where Otomí is spoken along with other indigenous languages, the degree of language loss is lower. The statistics prove that it is not the size of the speech community which co-determines the loss of the indigenous language but the influence of nonlinguistic factors such as socioeconomic status and lower ethnic self-identification⁷.

Language shift and loss in the Otomí population is reflected in the rates of monolingualism and bilingualism. In the last sixty years the bilingual indigenous population of Mexico has shown a steady increase in absolute figures, with a corresponding decrease in the number of monolingual speakers. The percentage of monolinguals from the total population of indigenous speakers was 52% in 1930 but only 16% in 2000. The percentage of bilinguals increased for the same period from 48% to 84%. According to Ortiz Álvarez (2005: 66), the indigenous languages with the largest number of monolinguals belong to the Mayan family (31% of the total ethnic population). In contrast, Otopamean languages show the lowest numbers of monolinguals (e.g. 5.4% Otomí monolinguals). At the same time, the bilingual population of Otomanguean languages showed the highest percentage (26.8%) in 2000. Bilingual indigenous speakers of Otomí and Spanish were 267,409 for the same year. In general, the number of bilingual men is two times larger than the number of bilingual women, and the number of monolinguals increases with age.

These figures can be correlated to the geographical mobilization of speakers (migration). In this case the migration of Otomanguean speakers is one of the largest. Still, Ortiz Álvarez (2005: 90) identifies a double tendency for Otomí speakers. These continue to be concentrated in the states of Hidalgo, Mexico and Querétaro but also migrate in small numbers to non-traditional areas in the states of Baja California Sur (3.2%), Zacatecas (2.1%), Yucatan (1.4%) and Chihuahua (1.2). While these percentages show that Otomí migration is comparatively low across states, they do not include the large numbers of Otomí immigrants to the capital cities of their respective states. The cities that attract seasonal Otomí immigrants in the state of Querétaro are Cadereyta, Ezequiel Montes, San Juan del Río and Querétaro City. Furthermore, migration to several destinations in the United States is important nowadays, especially among the Otomí speakers of Hidalgo. Otomí speakers of Querétaro and Mexico prefer regional migration over international migration (Barrientos López 2004: 14).

⁷ That demographical factors do not necessarily influence language maintenance is demonstrated by the average yearly growth rate of the indigenous population in Querétaro (2.2%) for the ten-year period between 1990 and 2000. This rate is much higher than the rates for Hidalgo (0.7%) and Mexico (1.4%) (Ortiz Álvarez (2005: 46f).

Interestingly, the effects of regional and international migration on the maintenance of the indigenous language are different. While regional migration implies shift to Spanish and the eventual loss of the indigenous language – if migration is permanent – international migration has encouraged Otomí speakers to agglutinate in political, organizational and interest groups which promote the use of the indigenous language as a symbol of ethnic identity (Alcántara Beatriz 2006: 27f). Paradoxically, international migration contributes to language maintenance. This is explained by the diglossic situation of indigenous languages and Spanish in Mexico. Outside their communities, Otomí people in Mexico speak their language only in domestic settings and prefer Spanish in all other socio-communicative spaces. In contrast, the third language used in non-Spanish speaking countries like the United States does not lead to shift but leaves the door open to the use of the group's language as an agglutinating symbol of identity. This use is not viable in Mexico, where the indigenous language is considered culturally alien and marginalized.

While Otomí is still widely spoken in community spaces such as religious services, meetings, schools and households, the number of speakers from the total Otomí population has decreased dramatically over the last years as a result of an interrupted transmission of the language from older to younger generations and the side effects of formal schooling and literacy. The Valley of Mezquital in Hidalgo and the Semi-Desert in Querétaro report a growing number of young people who do not speak Otomí or speak it only in domestic spaces. This has caused a functional reduction of the indigenous language and its reduction to fewer spheres. The expected result is that passive speakers become semi-speakers and eventually Spanish monolinguals (Mendoza Rico *et al* 2004: 9ff; Alcántara *et al* 2004: 27f). Schooling and literacy also influence language practices and lead to language shift. For the states of Hidalgo, Mexico and Querétaro, the literacy rates are over 75% (cf. Moreno Alcántara *et al* 2004: 51; Barrientos López 2006: 31; Mendoza Rico *et al* 2004: 47). There exist programs of bilingual education in Otomí and Spanish, especially in the state of Hidalgo, whose bilingual schools have become leaders in the field and a model for the Otomí communities from other states. Nevertheless, the presence of Spanish in the schooling system remains dominant, and the inclusion of the indigenous language is justified inasmuch as it facilitates the learning of Spanish and literacy in this language. Besides, most parents are unwilling that their children be taught in Otomí, because of the low prestige of this language and the idea that its use hinders the learning of Spanish. In this context, the best efforts of bilingual teachers fall on fruitless soil. In the Otomí villages of the semi-desert area and the highlands of Querétaro several efforts have been made in order to set up a bilingual education program for elementary school, but resistance from parents themselves has influenced decisively the success of the program.

Two additional problems related to bilingual education can be identified for Querétaro Otomí. One is that teachers who speak other dialects (mainly Mezquital Otomí) have been hired to solve the lack of well-trained bilingual teachers in Querétaro, with the expected result that dialectal differences interfere in the teaching-learning process. The interference issue is even more problematic because there is no standardized spelling for all Otomí dialects, and the differences between the spelling systems are numerous.⁸ This lack of normativity in writing makes Otomí literacy a real challenge for pupils and teachers.

It is necessary to stress the fact that in spite of a comparatively large number of Otomí speakers, the language shows clear signs of a decreasing vitality, accompanied by shift and loss in several Otomí communities, which remain ethnically self-identified as Otomí after the demise of their language. A widespread process of linguistic borrowing is accompanied by higher levels of bilingualism in the Otomí population.

8.1. The history of Otomí

Otomí history has been obscured by a historiographical tradition that depicts Otomí people as savage and backward in comparison with the major civilizations of central Mexico (e.g. Toltecs, Aztecs).⁹ Many historical events in which the Otomí people played a central role have been obliterated from the historical record due to a Nahuatl-centered historiographic tradition that tends to downplay the contributions of other ethnic groups of pre-Columbian Mexico. Today, scholars from different fields such as linguistics, history, archeology and anthropology begin to recognize the central role played by the Otomí in the history of Mesoamerica. The following account of Otomí history is therefore based on the works of scholars who have made an effort in each of their disciplines to unveil the Otomí past (e.g. Soustelle 1937; van de Fliert 1988; Galinier 1990; Wright Carr 1997; Hekking 1998; Lastra 2007).

According to Soustelle (1937: 470) the Otomí are associated to the oldest demographical strata in pre-Columbian Mexico. The ancestors of the Otomí and other Otopamean peoples migrated to central Mexico from either of two areas: 1) the territory of the present state of Oaxaca, with the largest concentration of languages of the Otomanguean family; or 2) the territory of today's Veracruz on the Mexican Gulf, known by chroniclers as *Nonoualco* and associated to the first historical culture in Mexico, the Olmecs (Soustelle 1937: 448). Neither hypothesis has been thoroughly demonstrated by archeological facts however. Still, historical records and

⁸ When writing this section, the author was informed that a standardized spelling system for all Otomí dialects had been approved and waited for a regulatory framework for its implementation (Hekking and Ángeles Gonzáles, p.c.)

⁹ A similar bias prevails in Andean historiography in favor of the Inca and in detriment of other ethnic groups.

glotto-chronological evidence show clearly that the present Otomí territory was not populated originally by speakers of this language, and that Otomí presence is explained by migration waves to the central plateau from the south or the east of Mexico in the first century of the Christian era.

Otomí speakers played a decisive role in the development of the multi-ethnic city-state of Teotihuacán during the five hundred years from its inception ca. 300 A.D. to its fall in the eighth century (Lastra 2007; Wright Carr 1997). The end of Teotihuacán implied the ethnic re-organization of the social space in central México. As a result, some groups re-settled uninhabited areas in the central valleys while others gathered to form independent chiefdoms, the most important of which was the Toltec city-state of Tula in the present state of Hidalgo. The area in and around Tula was populated by Otomí peoples before the entry of Nahuatl-speaking Toltecs (Soustelle 1937: 451). The presence of Nahuatl speakers in central Mexico is late in comparison to that of other ethnic groups, especially Otomanguean. The successive migrations of Nahuatl-speaking groups from the north of Mexico to the central plateau unchained a process of acculturation in which the newcomers became gradually integrated into the Mesoamerican culture and adopted many of its material and scientific developments (van de Fliert 1988: 43). The archeological record shows that the Otomí played a crucial role in the development of the urban center of Tula and the building of an extensive trade and ritual network in the area. With the fall of Tula around the late twelfth century, the influence of Nahuatl-speaking groups in central Mexico increased gradually until the end of the fourteenth century. The Otomí chiefdoms of Chapa de Mota and Jilotepec flourished in the northwestern part of Mexico State, the southern part of Hidalgo and the southern portion of Querétaro (van de Fliert 1988: 44). Nicknamed by chronicles as the “Otomí kidney”, this area concentrated the largest part of the Otomí population before the Spanish Conquest. Today it remains the core area of Otomí influence.

With the birth of the Aztec empire around 1376, all ethnic groups inhabiting the valley of Mexico and neighboring areas came under its rule. The Nahuatl rulers of Texcoco found no resistance from the Otomí centers of Otumba, Tepotzotlan and Tulancingo, but the Nahuatl kingdom of Azcapotzalco annexed the Otomí chiefdom of Jilotepec only after several battles. As Soustelle (1937: 463) explains, the attitude of the Nahuatl invaders towards the Otomí population was not the same in all cases. Texcoco rulers maintained good relations with their Otomí subjects and let them remain in their areas of occupation. On the contrary, the Aztecs of Azcapotzalco imposed hard taxing conditions on their vassals and expelled Otomí groups from their traditional territories. By the first half of the fifteenth century most Otomí cities and chiefdoms were under Aztec rule. The only exceptions were the Otomí who lived in the highlands of today's state of Veracruz and the Otomí people of Tlaxcala, who preserved their independence in exchange of military services to the Aztecs. The harsh subjection to which most Otomí peoples fell victim through the

dispossession of their traditional lands may explain their support to the Spaniards during the conquest and during the first century of colonization.

The sociopolitical events on the central plateau since the emergence of the Nahuatl kingdoms in the fourteenth century resulted in new patterns of settlement among ethnic groups, with important consequences for the ethnolinguistic configuration of the area. The increasing political presence of Nahuatl chiefdoms resulted in the expansion of their language over central Mexico. In turn, the effect of the forced displacement of Otomí peoples was the emergence of two discontinuous Otomí-speaking areas separated by a Nahuatl-speaking land. The reshaping of the linguistic landscape of central Mexico did not result from shift but from military occupation and expelling of former inhabitants. Where Nahuatl peoples coexisted peacefully with speakers of other languages, bilingualism was the rule and the expansion of Nahuatl did not occur at the expense of other languages. Nahuatl-Otomí bilingualism was widespread in the central and northern areas of the present state of Mexico (Soustelle 1937: 477), where most Otomí speakers were concentrated at the time of the Spanish Conquest. Unlike the eastern and southeastern areas of the plateau, the central area was continuous and not interrupted by Nahuatl, although speakers of this language were scattered all over the area and most Otomí were competent in Nahuatl as well. A similar multilingual situation was that of the Toluca valley (the western part of the central plateau), where Otomí coexisted with other Otopamean languages and with Nahuatl in some villages of the southern valley (e.g. Coatepec, Texcaliacac). The northern part of the central plateau including most of Querétaro state, northern Hidalgo, and Guanajuato was beyond Aztec influence, being the home land of nomadic groups of the Otopamean family (i.e. Pame and Chichimec). The role played by Otomí peoples in the conquest and the colonization of the northern part of the central plateau deserves special attention.

Long years of hard taxing and the uprooting from their homelands strengthened in most Otomí groups the hatred towards Nahuatl-speaking invaders. These feelings were rapidly noticed by the Spaniards and used for their own purposes. Even the Otomí people of Tlaxcala, unconditional allies of the Aztecs, after a few battles with the Spaniards, realized that they could use the newcomers to make the Aztec rulers pay off old debts. The Otomí became thus the best allies of the Spaniards in their conquest of Mexico, providing them not only with soldiers but also with all kinds of supplies even in the hardest moments. Moreover, with Otomí assistance the Spaniards initiated the colonization of the silver-rich area to the north of the Mexico valley (the present states of Guanajuato and Zacatecas). On account of the strategic position of their territory, which connects the valley of Mexico to the northern area dominated by the bellicose Chichimecs, the Otomí were the most helpful allies of the colonizers. Their position was even more strategic because the Otomí shared with the Chichimecs a number of cultural traits originated in their common ancestry (their languages belong to the Otopamean family) and old relations of trade. Unlike

most Spanish towns which later became large cities, Querétaro was founded by a Christianized Otomí Indian, who worked also as a peace-maker for the Spaniards in northern Hidalgo. Similarly, Otomí speakers were present in the foundation of the oldest towns in the state of Querétaro (e.g. Tolimán in 1532)¹⁰ but also in neighboring Guanajuato (e.g. San Miguel Allende in 1547). Moreover, Otomí leaders and their people participated actively in the colonization of the Chichimec territory for one hundred years, albeit their success was partial. The territory came under full Spanish control only in the first half of the eighteenth century, when the Spaniards carried out the systematic extermination of Chichimec and Pame Indians. The survivors were grouped in towns for the Spaniards to benefit from their workforce.

The immediate effect of the conquest of the Chichimec territory was the expansion of the Otomí language to the north of its traditional area, that is, to the present state of Querétaro and to northern Hidalgo. A further effect was the emergence of bilingual towns in which Otomí was spoken along with Chichimec or Pame (Tolimán was one of these multilingual centers). In contrast, a simultaneous recession of the Otomí language from the core of the central plateau occurred as a result of three factors: 1) the use by most missionaries of Nahuatl in the evangelization of indigenous peoples; 2) the moving of Otomí people from their traditional area in the central plateau to the north for the colonization of the Chichimec territory; and 3) the moving of Nahuatl Indians from different parts of central Mexico to former Otomí areas to work in agriculture and mining activities. The 'Nahuatlization' of the central plateau was further encouraged by the traditional Otomí-Nahuatl bilingualism of the area before the Spanish conquest. Many of the existing Nahuatl-speaking towns in the Valley of Mexico were originally Otomí three or four hundred years ago. A recent case of Nahuatlization among Otomí speakers is the enclave of Ixtenco (Tlaxcala) where the shift to Nahuatl is virtually completed.

The Otomí migration to territories north of the valley of Mexico resulted in the dislocation of the once compact Otomí area in the central plateau and the following dialectalization. In the state of Querétaro (but also in a large portion of Hidalgo) the process of dialectalization speeded up since the late seventeenth century through the progressive encroachment of Otomí lands by an increasing number of Spanish *haciendas* formed in the fertile valleys at the expense of Indian territory. When the Otomí failed to defeat the Chichimec, they became 'useless' for the Spanish Crown and lost many of their benefits. The conditions were then set for the expropriation of Otomí lands, which were taken over by ranchers, miners and *hacienda* owners

¹⁰ Santiago Mexquititlán was founded around 1520 by Spanish settlers who sought to facilitate the trade of land staples and the improvement of tax collection in the area (van de Fliert 1988: 53).

(Prieto and Utrilla 1997: 32), and the displacement of Otomí elites from the urban centers. The outcome of these events was the recession of Otomí to their present areas of in the semi-desert region in northern Querétaro (Tolimán) and the southern highlands (Amealco) of the state. Both areas became niches of refuge in which the indigenous language could survive after the Otomí were expelled from the cities and their lands taken over by the Spaniards. The semi-desert and the highlands had been previously colonized by the Otomí through different processes: while the semi-desert was settled during the colonization of the Chichimec territory, the southern highlands were populated as an extension of the Otomí traditional area in the northern part of Mexico State (Jilotepec). The dialects differences between these areas result from distinct demographic compositions (e.g. the presence of non-Otomí indigenous groups), the urbanization process leaded by nearby cities, and the urban migration of Otomí speakers. Prieto and Utrilla (1997: 33ff) maintain that the Otomí of Amealco (southern highlands) is closer to the variant spoken in northern Mexico state while the Otomí of Tolimán and Cadereyta (semi-desert) is similar to the variant spoken in Hidalgo (Valley of Mezquital) with some Chichimec substratum. Finally, the Otomí of San Idelfonso is similar to the varieties of Tolimán and Cadereyta, but it lacks Chichimec substratum (Hekking, p.c.).

The historical events just described suggest that the uniformity of the Otomí language has considerably diminished in the last centuries, but that there are ethnic Otomí groups that remain culturally distinct even if they have lost their language in favor of Spanish (cf. *supra*). The corollary is that present Otomí groups are more culturally than linguistically homogeneous. Nevertheless, we should recall that Otomí groups have received cultural influences from Otopamean and Nahuatl speech communities as a result of their coexistence in the cultural *sprachbund* of Mesoamerica. This influence led Wright Carr (1997: 2) to pose the question of the Otomí cultural unity in the following terms:

“Conviene preguntarnos si los Otomíes han sido, en diferentes momentos de su historia, un grupo lingüístico, una cultura o una etnia, o bien una combinación de estas variables. Su identidad lingüística es evidente: los Otomíes son los hablantes de un conjunto de lenguas estrechamente emparentadas que descienden de un idioma proto-Otomí, hablado hace varios siglos en el centro de México. La existencia de una cultura Otomí es menos evidente, ya que desde tiempos remotos los hablantes de Otomí han habitado entornos geográficos diversos, entremezclados con otras comunidades lingüísticas. En tiempos recientes se ha tratado de fomentar, con base en la semejanza de sus hablas, la integración étnica de los Otomíes” [It is wise to ask ourselves whether the Otomís have been, in different moments of their history, a linguistic group, a cultural group, an ethnic group, or a mixture of these variables. Their linguistic identity is obvious: the Otomís are the speakers of a set of closely related

languages which come from a proto-Otomí language spoken several centuries ago in central Mexico. The existence of an Otomí culture is less obvious: since long time ago Otomí speakers have inhabited various geographical settings and mixed with other linguistic groups. In modern times, the ethnic integration of the Otomí people has been fostered on the basis of their language] (Wright Carr 1997: 2; my translation).

8.2. The dialects of Otomí

The dialectal diversification of Otomí and the fact that intelligibility is seriously reduced between certain varieties has led some authors to consider Otomí a diasystem composed of different Otomí languages (e.g. Suárez 1983: xvi; Palancar 2006: 325). Positions in this respect vary from those who sustain the aforementioned view to those who consider Otomí one single language composed of a number of dialects. While case studies deal with a individual Otomí dialects, there is no comprehensive description of the Otomí dialectal variation apart from the study presented by Soustelle (1937) seventy years ago. For this reason and for the systematic treatment of data in this work, the following discussion of Otomí dialects is based on Soustelle, with additional information proposed by several authors in the last years. Notice, however, a number of communities in which Otomí was still vital in the early thirties are today Spanish monolingual while others have changed as a result of urban migration.

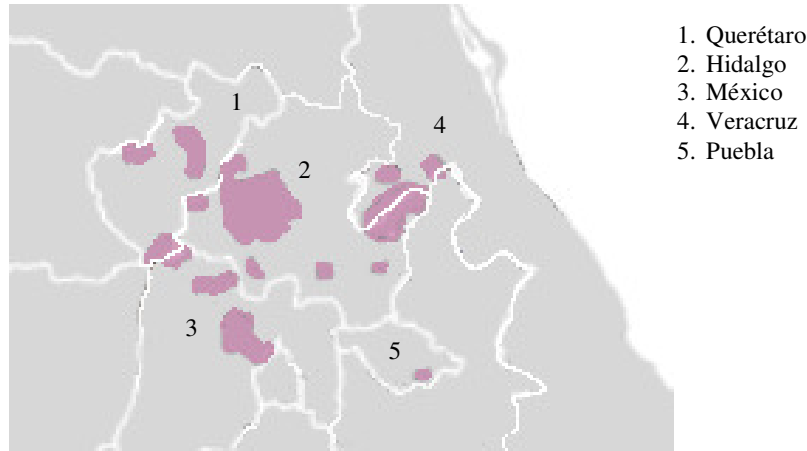
Soustelle investigated the dialects spoken in 33 villages and their neighboring areas in seven different states, including Querétaro, Hidalgo, Mexico, Tlaxcala, Veracruz, Puebla, Michoacán and Guanajuato. His work is based primarily on the analysis of phonetic-phonological variation, and secondarily on the analysis of lexical variation. Morphosyntactic variation plays no role in Soustelle's classification because the language shows a high degree of uniformity in this field (Soustelle 1937: 212ff) and because of his assumption that changes are harder to occur in the field of grammar.

At the phonological level Soustelle identifies thirty-two parameters of dialectal variation (1937: 191f). The most important are: a) the voiced-voiceless distinction in stops; b) the full or partial occlusivity of stops; c) the fricativization and glottalization of stops; d) vowel alternation, especially the variation between /a/ and /o/ and between /e/ and /i/. For Soustelle, the parameters of voicing, occlusivity and glottalization are the primary criteria for dialectal classification. In contrast, vowel alternation is less consistent and should be considered a secondary factor of variation. In a similar way, distributional criteria such as the syllable-initial position of phonemes do not vary systematically across dialects. From his analysis of the foregoing criteria in thirty-three localities, Soustelle identifies seven distinct dialectal areas (Soustelle 1937: 203). Geographically, these areas are:

- I. The state of Querétaro and part of Guanajuato
- II. The central valleys of Hidalgo and southeastern Sierra Gorda
- III. The area of Jilotepec in northern Mexico, and the Otomí enclave of Michoacán
- IV. The plateau of Ixtlahuanca, the Otomí enclave of Amanalco in the highlands astride Mexico and Michoacán
- V. The south portion of Sierra de las Cruces and the adjoining plateau
- VI. The eastern slopes of the central plateau in Hidalgo, and the southern Toluca Valley
- VII. The plateau of Tlaxcala on the slopes of the Malinche volcano

Querétaro Otomí – including the dialects of Santiago Mexquititlán and Tolimán – is part of the first dialectal area (Group I). A number of Otomí-speaking localities mentioned by Soustelle are today monolingual in Spanish. For example, the formerly Otomí community of La Cañada in the vicinity of Querétaro City has shifted to Spanish, and so have most communities in the municipalities of San Joaquín and Peñamiller. Today, only the communities located in the municipalities of Amealco and Tolimán have an important number of Otomí speakers. Other areas outside the state of Querétaro such as Ixtenco (VII) in Tlaxcala and San Felipe los Alzate in Michoacán have almost completely shifted to Spanish. In general, Otomí is spoken with different degrees of vitality in Querétaro, Hidalgo, (northern) Mexico and Puebla.

Wright Carr (1997: 2) proposed another classification, based on Soustelle (1937), Manrique (1969), Galinier (1987) and Lastra (1993). According to Wright Carr, the Otomí language includes four dialectal areas: 1) Western Otomí, spoken from the Valley of Toluca through the Valley of Mezquital up to Sierra Gorda, corresponding in Soustelle's classification to groups I, II III and V; 2) Eastern Otomí, spoken in the eastern mountains of Sierra Madre, corresponding to Group IV; 3) Tilapa Otomí, in the southeastern portion of the Toluca Valley, part of Group VI; and 4) Ixtenco Otomí, spoken in Tlaxcala and classified as Group VII in Soustelle's division. Wright's classification fails to make a distinction between western varieties (I, III, IV, V) central varieties (II) and eastern varieties (VI). Neither does Wright Carr account for the similarities between the varieties of the southern valley of Toluca and the eastern highlands of Sierra Madre. Wright Carr's proposal is too broad to allow further distinctions at lower levels. The present distribution of Otomí is shown in Map. 8.2. Notice that no Otomí areas are identified in Guanajuato and Michoacán because of the dying status of the language in these states.

Map 8.2 Present-day distribution of Otomí dialectal areas per state

The specificity of Soustelle's classification allows further divisions. While the seven groups described by Soustelle represent dialectal areas, the varieties of the thirty-three localities he studied are considered distinct dialects. An intermediate category between dialectal areas and dialects are 'subgroups'. Soustelle finds three subgroups in Group I:

- A) the Valley of Laja and the highlands of Tierra Blanca in the state of Guanajuato, where few Otomí speakers exist to date;
- B) the central and southern part of Querétaro state, including the nowadays Spanish monolingual community of La Cañada and the municipality of Amealco;
- C) the Sierra Gorda dialects, spoken in the northern part of Querétaro state, including the highlands from Tolimán to Jalpan and Pinal de Amoles (today Spanish monolingual), the communities of San Miguel, Tolimán, San Antonio, Higuera (Spanish monolingual) and Tetillas in the area of Cadereyta, Boyé and Sombrerete, where few Otomí speakers are reported today.

The phonological characteristics of subgroup B –including the dialect of Santiago Mexquititlán – are the absence of the fricative /θ/, the occurrence of the fricative bilabial /ɸ/, the nasal vowels /ã/, /õ/ and /ɔ̃/, and the collapse of variants /a/, /ɔ/ and /o/ in the latter. The main features of subgroup C –including the dialect of Tolimán – are those of subgroup B plus the occurrence of the fricative dental /ð/.

Table 8.4 Otomí dialects per locality, state and linguistic features

	Representative Localities	States of concentration	Phonetically relevant Phenomena
I	Zimapán, Tolimán, Amealco	Querétaro	Voicing, occlusivity, fricativization in equal proportions; wide vocalic variation; mixed dentilabial and palatal consonants
II	El Cardonal	Hidalgo	High occurrence of fricativized stops and labials, mixed palatals, partial occlusivity, predominant vowels ô, â, o
III	San Andrés Timilpan	Mexico	Low occurrence of fricativization; no fricativized palatals; partial occlusivity in dentilabial and labial sounds
IV	San José del Sitio	Mexico	No fricatives, partial occlusivity in bilabial stops; predominant vowels <u>o</u> , e, i
V	Ameyalco	Mexico	High frequency of fricativization, especially in bilabial consonants; predominant vowels <u>o</u> , e, i
VI	Santa Ana Hueytlálpan	Hidalgo	Higher occurrence of devoicing and fricativization; wide vocalic variation
VII	Ixtenco	Tlaxcala	Low occurrence of fricativization; no fricativized palatals; diphthongization of i and e; partial occlusivity in dentilabial and labial sounds

Excluding the first subgroup (Guanajuato), Querétaro Otomí is classified in two subdialects: the varieties spoken in the semi-desert and part of Sierra Gorda (northern part of the state) and the varieties spoken in the highlands of Amealco (southern part of the state). In roughly similar terms, Hekking *et al* (forthcoming) identifies four varieties of Querétaro Otomí grouped in two distinct areas: the

varieties of Santiago Mexquititlán and San Idelfonso Tultepec¹¹ in the municipality of Amealco; and the varieties of Tolimán and Cadereyta in the semi-desert area. According to Hekking, the variety of Santiago Mexquititlán is closer to the Otomí spoken in northern Mexico State whereas the varieties of San Idelfonso, Tolimán and Cadereyta are closer to the Otomí spoken in the Valley of Mezquital in Hidalgo state. Another classification of Querétaro Otomí is presented by Mendoza Rico *et al* (2006: 11), who identify the variants of Amealco, Tolimán and Cadereyta while tracing linkages to the neighboring dialects in the states of Mexico and Hidalgo (cf. *supra*). These linkages are somewhat different from those proposed by Soustelle (1937: 204). In Soustelle's words, "the diversification is much more noticeable in southern dialects than in northern dialects; Groups I and II [Querétaro and Hidalgo] show a high degree of homogeneity in contrast to the extreme diversity found in the southern part of the Otomí area" (Soustelle 1937: 204; my translation). In this perspective, the dialect of Tolimán is closer to the dialect of Mezquital than the dialect of Santiago Mexquititlán to the dialect of northern Mexico.

The foregoing discussion is related to the issue of the Otomí spelling systems. As mentioned above, there is no standardized spelling for all Otomí dialects. This is not surprising given the marked dialectalization of the language. Fifteen different spelling systems had been proposed for Otomí until 1999 (Zimmerman 1999: 157). For the dialects of Santiago Mexquititlán, San Idelfonso, Tolimán and Cadereyta in the state of Querétaro there is a unified system since 1999. The major features of this spelling concern the vocalic system: the underlining of vowels to represent openness (a, e, i, o, u); the use of dieresis to represent nasalization (ä, ö, ä, ö); the marking of contrastive high, low and ascending tones with {`}, {˘} and vowel duplication {vv}, respectively. As noticed by Hekking *et al* (forthcoming), the spelling conventions of Querétaro Otomí are similar to those developed in the state of Hidalgo. Empirical evidence of the cross-dialectal applicability of this spelling is that teaching materials have been developed with this spelling since 1999 outside the state of Querétaro. The use of a spelling system for various dialects has received additional support from linguists working on the Highland Otomí of Hidalgo (Group V), who use the same spelling as the one of the Valley of Mezquital (Group II). These developments show an emerging consciousness among Otomí speakers but also the fundamental unity of the language across its dialects.

¹¹ For Soustelle (1937: 184, 199) the Otomí variety of San Idelfonso is a "rather particular dialect" spoken only in this village but understood by speakers of Huichapan, Chantantongo and San Luis in Hidalgo state. He groups San Idelfonso in the dialectal area of Jilotepec corresponding to northern Mexico state.

8.3. (Querétaro) Otomí: a typological characterization

The following typological description assumes that Otomí dialects belong to one and the same language. Accordingly, it is expected that most, if not all, of the typological features discussed hereunder apply to any Otomí variety regardless of its geographical distribution or sociolinguistic situation. At the same time, it is necessary to insist that this fundamental unity does not obliterate dialectal differences, which occur at all levels of the language. On the basis of this assumption the typological characterization that follows is based largely on the dialect of Santiago Mexquititlán. Three reasons substantiate this choice: first, half of the Otomí corpus for this investigation was collected in this locality; second, the existence of grammatical descriptions and dictionaries for this dialect helps us provide a trustworthy account of the language; and third, it is clear from the previous dialectal discussion that the dialect of Santiago Mexquititlán is representative of Querétaro Otomí. In order to highlight differences from, and similarities to, neighboring dialects, I make use of grammatical descriptions available for other varieties. In the following I refer to the language as ‘Otomí’ in general and use the terms ‘Querétaro Otomí’ when making claims about particular features of this dialect. The following description assumes the language as spoken in the present and do not presuppose any pre-contact situation unless otherwise specified.

Otomí belongs to the Otopamean branch of the Otomanguean family. According to Suárez (1983: xvi), of twenty-four languages that make up the Otomanguean family, seven form the Otopamean branch. The geographical distribution of the Otomanguean family is limited to central and southern Mexico, but the internal differentiation is the largest of all Mesoamerican families. Accordingly, Suárez (1983: 26) considers Otomanguean not a family itself but something like a “hyper-family” or “stock”. The differentiation within the Otopamean branch is just as great: Otopamean languages have a range of differentiation similar to the one attested in the Mayan language family (Suárez 1983: 26). In fact, several authors consider Otomí not a language but a group of languages forming a diasystem.

The phonological inventory of Otomí is rather complex, and it is there that Otomí dialects differ most from each other. Querétaro Otomí has thirty-four phonemes, including ten vowels, two semi-consonants, and twenty-two consonants. In addition to the five vowels of Spanish (/a/, /e/, /i/, /o/, /u/), the language has two central vowels (/ə/, /i/), two open-mid vowels (/ɛ/, /ɔ/) and one nasal vowel in allophonic variation (/ã/ ~ /õ/). Hekking (1995: 30) notes that other dialects show a larger number of nasal vowels. Highland Otomí (Hidalgo state), for example, has five nasal vowels (Voigtlander and Echegoyen 1985) while the Otomí dialect of San José del Sitio (Mexico state), has nine nasal vowels, one for each oral vowel (Soustelle 1937: 129-181). Bartholomew (1968) points out that nasalization across

Otomí dialects is irregular in high vowels in comparison to low vowels. The corollary is that every Otomí dialect shows at least one nasal vowel for any of the following oral segments /a/, /e/, /o/, /ə/, /ɛ/, /ɔ/. In contrast, nasal high vowels vary across dialects and is absent in certain varieties (e.g. Santiago Mexquititlán). The inventory of consonant sounds in Querétaro Otomí includes: sixteen phonemes similar to their Spanish counterparts (/p/, /t/, /k/, /b/, /d/, /g/, /f/, /č/, /s/, /x/, /m/, /n/, /ɲ/, /r/, /ʃ/, /l/), two semi-consonants (/w/, /y/), the glottal stop /ʔ/, the glottal fricative /h/, three apical sibilants /z/, /ts/, /tʰ/, and one palatal sibilant /š/. Like other neighboring dialects, Querétaro Otomí does not show vowel harmony.

In Soustelle's classification the Otomí of Santiago Mexquititlán (henceforth Santiago Otomí) belongs to subgroup B (Amealco) of the dialectal area I (Querétaro). Soustelle described several features for this area (cf. Table 5.15) and the respective subgroup (cf. *supra*). Let us see now whether these are confirmed by the aforementioned phonological inventory. The occurrence of voiced and voiceless stops in Santiago Otomí where other dialects show only voiceless segments confirms the voicing tendency identified by Soustelle for group I (1937: 198f). The glottalization of stops and the realization of full occlusivity described as typical of group I are also confirmed for Santiago Otomí. In contrast, the wide range of variation in the vowel system is not attested, because vowels show fixed phonetic values. Of the features proposed for subgroup B, the occurrence of the fricative bilabial /ɸ/ and the absence of the fricative dental /θ/ are confirmed for Santiago Otomí. On the contrary, the existence of three nasal vowels and the collapse of /a/, /o/ and /ɔ/ into /o/ are disconfirmed: this dialect has only one nasal vowel /õ/ in allophonic variation with /ã/, and only the first two vowels occur as allophonic realizations.

Three sounds did not occur in Classical Otomí: the alveopalatal affricate /č/, the trill /ʀ/, and the lateral /l/. While these sounds occur all in Spanish loanwords, they occur in native forms too (Hekking 1995: 31). For example, *ts'aki* is realized as [č'aki] in Santiago Otomí. The existence of the alveopalatal affricate and the lateral in Nahuatl suggest also the origin of these sounds in the contact of Otomí with this language.

The most salient feature of Otomí suprasegmental phonology is the tonal system. The language has three tones: one high, one low, and one ascending. Tones are marked in writing only if contrastive. Non-contrastive tonal realizations depend on style or register as well as on phonetic environments. Nasalization is generally considered a distinctive segmental feature on vowels but suprasegmental processes involving nasalization are well known. Soustelle identified a widespread phenomenon of prenasalization involving segments /t/ and /d/ in the dialect of San Jose del Sitio. Both sounds become [nt] and [nd] in word-initial position. Grammatical descriptions of Santiago Otomí do not refer to prenasalization, but clusters [nt] and [nd] occur in native forms and assimilated loanwords with /t/ and

/d/ in word-initial position. Hekking and Bakker (2007) find no evidence of contact-induced changes in tone, vowel and consonant harmony, but no reference is made to suprasegmental phonology.

Syllables in Otomí are typically open (CV) but other patterns are frequent as well. I have found no explicit reference to syllable number for Santiago Otomí. Nonetheless, an analysis of the corpus shows that the description of San Jose del Sitio (Soustelle 1937: 135f) is fully applicable to Querétaro Otomí: accordingly, the frequency of consonant-vowel monosyllables is high, but this does not mean that Otomí is a typical monosyllabic language – in fact, most words are disyllabic (CVCV). Consonant clusters are frequent in onsets but not permitted in coda position. Onsets of type NCC result from prenasalization as explained above (e.g. *nt'udi*). Alien clusters have been introduced in Otomí from Spanish, especially the stop-flap onsets /tr/, /pr/ and /kr/. Similarly, restrictions across syllabic boundaries have been changed by Spanish loanwords: e.g. *ektarya* ‘hectare’, *septyembre* ‘September’, with non-Otomí clusters /kt/ and /pt/ (Hekking and Bakker 2007). The stability of alien clusters depends on the age of the loanword and the degree of bilingualism of the speaker.

The introduction of new sounds and syllabic patterns through loanwords has not provoked major changes in the phonological inventory of Otomí, simply because a large number of loanwords are accommodated to the native system (cf. section 10.1.3). Comparing the phonological inventory of present-day Otomí with the inventories of Paraguayan Guaraní (5.3.3) and Quichua (5.2.3) provides further evidence for this claim. Given that contact-induced changes in the phonological system of a recipient language are directly related to the number of unintegrated loanwords, Otomí is the least influenced of the three languages not only in terms of number of loanwords but also of frequency of assimilation. The next section tests this correlation on the corpus of each language.

Morphologically, Otomí shows a split typology consisting in a mixture of synthetic and analytic structures (Hekking 1995: 5; Hekking and Bakker 2007). The split morphology of Otomí corresponds to the types of morphemes in the language. Querétaro Otomí has two types of bound morphemes: proclitics and affixes. Affixes, the great majority of which are suffixes, are part of verbal morphology while proclitics fit either in verbal or nominal paradigms. Other authors (Soustelle 1937: 143ff; Andrews 1993) consider proclitics true prefixes on the basis of their cross-syllabic coalescence in all Otomí dialects. According to Soustelle (1937: 138) any monosyllable preceding or following a polysyllabic word merges with the latter in pronunciation. Since proclitics are monosyllabic, they are expected to lose their phonetic shape by merger. Neither Hekking (1995) nor Hekking and Bakker (2007) provide counterevidence to Soustelle’s claim. I prefer the term ‘proclitic’ because

most grammatical sketches of Querétaro Otomí use this term and the current spelling writes proclitics as separate forms.¹²

Hekking and Bakker (2007) maintain that Otomí shows a synthetic structure at the level of the phrase but an analytic structure at the level of the sentence. A few examples from (Hekking 1995) illustrate this. Consider the following noun phrases.

- 1) *Ár=ngú* *ar=Xuwa*
 3.POSS=house DEF.S=Juan
 ‘Juan’s house’

- 2) *Ma=ngú-hu*
 1PL.POSS=house-INCL
 ‘Our house’ (first person inclusive)

- 3) *Yá=wa* *ar=tsa’yo*
 3PL.POSS=foot DEF.S=dog
 ‘The dog’s feet’

- 4) *Da=r=nxutsi-ga*
 PRS.1=DEF.S=girl-EMPH.S1
 ‘I am a woman’

- 5) *Hin=d=ar* *‘behñä*
 NEG=PRS.1=DEF.S woman
 ‘I am not a woman’

Proclitics are ubiquitous in the noun phrase. They indicate definiteness and number, but also person, negation, tense and aspect. Number marking is made exclusively through proclitics, since there are no plural markers. Gender is not grammaticalized in Otomí but signaled lexical, when necessary, through the nouns *tsu* ‘male’ and *ndq* ‘female.’ Possessive proclitics are another type of adnominal particles. Possession is the only syntactic relation that can be marked in the noun phrase (1-3). If a noun is used predicatively as in (4-5), the noun phrase carries the same tense and aspect markers of verbs. Otomí is not a head-marking language in the noun phrase: all markers are attached to the proclitics while the noun head usually occurs bare. Exceptions are the clusivity markers attached to the head noun when preceded by possessive proclitics (2).

¹² Further evidence for the analysis of these forms as proclitics is that modifiers occur between proclitics and nouns: e.g. *ar=dütä*, DEF.S=tall, ‘the tall one’, *ar=na-data*, DEF.S=very=tall, ‘the very tall one, the giant’.

In the following examples noun phrases in subject position appear in square brackets.

- 6) [Ar=*Māndo*] *mi=ñä-wi* *ár=nänä*
 DEF.S=Armando IMPF.3=speak-DUAL 3.POSS=mother
 ‘Armando spoke with his mother’
- 7) *Bi=pä-hya* *da=ot-’ya*
 PRS.3=know-EMPH.3PL.PROX FUT.3=write-EMPH.3PL.PROX
 ‘They know how to write’
- 8) *Di=ne* *ga=fax-’i* *ar=xudi*
 PRS.1=want FUT.1=help-OBJ.2 DEF.S=tomorrow
 ‘Tomorrow I want to help you’
- 9) [Ya=*mepute*] *um-bi* *ar=nhñuni* *ya=mbane*
 DEF.PL=porter give-BEN DEF.S=mole DEF.PL=godfather
 ‘The porters give mole to their godfathers’
- 10) [Ma=*txu*] *bu-se* *j=ár* *ñäni* *ar=hñe*
 POSS.1=grandmother live-REFL LOC=POSS.3 side DEF.S=river
 ‘My grandmother lives by herself at the riverside’
- 11) [Nugö] *di=bu-kwa*
 PRO.1 PRS.1=be.LOC-PROX
 ‘I am here’

The marking of syntactic relations between arguments of the predicate is done through proclitics, suffixes and a few prefixes. Proclitics play also a major role in the verb phrase: they mark person, tense and aspect (Hekking 1995: 47). There are seven types of verbal suffixes: markers of number and clusivity (6); emphatic markers (7); markers of direct object (8); 4) markers of indirect object (9); markers of reflexivity-limitativity (10); and markers of location (11). In principle all of these suffixes can be attached to a verb root at the same time. Hekking (1995: 50) notes however that suffixes usually are not more than two. Hekking quotes the following example (12) as an extreme case of agglutination in Otomí: four suffixes attached in a predetermined order to the verb root *hongí* ‘to look for’.

- 12) *Bi=hong-a-wi-tho-wa*
 PRS.3=look.for-OBJ.1-DUAL-LIM-LOC.PROX
 ‘He/she looks for us only (around) here’

The verb phrase in traditional Otomí is relatively complex in morphological terms. It may include several inflectional affixes, as illustrated in (80). It differs therefore from the noun phrase, which shows a slightly higher degree of analyticity. While the structure of the verb phrase is similar to Classical Otomí, deviant cases such as (13) and (14) are reported too.

- 13) *Ya=mep̄te* *un=ar*
 DEF.PL=porter give=DEF.S
nhñuni *ne* *ar=sei* *yá=mbane*
 mole and DEF.S=pulque POSS.3PL=godfather
 ‘The porters gave mole and pulque to their godfathers’
- 14) *Yá=meni* *xi=‘yot’u* ‘*nar=mixa* *pa* *ya=hkwete*
 POSS.3PL=relative PRF.3=make INDEF.S=mass for DEF.PL=forebears
 ‘Their relatives gave a mass to the forebears’

Hekking (1995: 37f) mentions three ways to express the relation between the predicate and the indirect object: 1) through the verbal suffix *-pi* or any of its variants; 2) through simple juxtaposition, if the semantic relation is implicit in the meaning of the verb; or 3) through the Spanish preposition *pa* (from *para* ‘for’). The first alternative is illustrated in (9) above. The second alternative is exemplified in (13). Finally, the third alternative is illustrated in (14). The use of Spanish prepositions in Otomí is prolific and has changed the native ways in which phrasal constituents are related. Example (15) shows the use of the Spanish preposition *con* ‘with’, instead of the Otomí instrumental marker.

- 15) *Ma=tada* *bí=daki* *ar=m̄ti* ***ko*** *ár=ndojwai*
 POSS.1=father PST.3=attack DEF.S=animal with POSS.3=machete
 ‘My father attacked the animal with his machete’

Summing up, Otomí shows a split morphological type according to which verb phrases are more complex than noun phrases. Relations between arguments often remain implicit. Arguments are traditionally juxtaposed in a fixed order. The use of Spanish prepositions is a recent development in the Otomí verb phrase.

The next issue has to do with the type of Otomí at the level of the sentence. Otomí shows the greatest level of analyticity at this level. This is shown in its tendency to asyndetism and juxtaposition. Hekking and Bakker (2007) summarize the sentence structure of Otomí in the following terms:

“At the sentence level the structure is more analytical, and it is not uncommon to find asyndetic compounding and bare juxtaposition of

The following examples include complex-verb, subordinated, coordinated and relativized constructions, characterized all by asyndetism. Let us begin with complex-verb constructions.

- Juxtaposition is the preferred strategy in complex-verb constructions. The verbs 'try' and 'want' are auxiliaries and their relation to the main predicate is *not* mediated by connectors. Consider now the following examples of clausal coordination.

- In (18) the clauses are coordinated by *ne* ‘and’. Other particles used for coordination are ‘*nehe* ‘too’ and *ne‘nehe* ‘in addition’. In contrast, the clauses are coordinated by simple juxtaposition in (19). Hekking (1995) does not mention which strategy is preferred, but his statement that complex sentences *usually* lack connectors suggests that asyndetic constructions are the unmarked choice. Subordination constructions are exemplified below.

- 20) *Ar=bätsi* *bí=nzoni* *bí=ntsär'i* *na* *nts'editho*
 DEF.S=child PST.3=cry PST.3=burn SUPL hard-LIM
 'The child cried because he burned painfully'

- 21) *M-tada-ga* *mbi=xoka=r* *goxthi*
 POSS.1-father-EMPH.1 PST.3=open=DEF.S door
bí=um-bu 'nar na ndutse ko=r ntsu
 PST.3=give-BEN INDF.S SUP chill with=DEF.S fear
 'As soon as my father opened the door, he got a chill because of fear'
- 22) *Ya=bädi* *mpefi* *xi=hño*
 DEF.PL=knewer work PRF.3=well
nuya 'ñete kat=ya jä'i
 DEM-DST.3PL sorcerer cheat=DEF.PL person
 'While medicine-men work well, sorcerers cheat people'

Examples (20) to (22) illustrate different types of subordination: (20) indicates a causal relation between two events; (21) refers to the anteriority of one event with respect to another; and (22) signals simultaneity of events. None of the above constructions makes use of connectors to link the subordinate clause to the main clause. Instead, (20) and (22) use simple juxtaposition while (21) has a proclitic of tense in the subordinated clause to indicate anteriority to the main clause. Notice in (21) the fusion of the Spanish preposition *con* 'with' and the proclitic of definiteness and singular number *-r*. To the foregoing strategies for subordination Hekking adds a number of connective particles to express causality (*ngetho*, *jange*), comparison (*tengu*, *ngu*, *jangu*), simultaneity (*nä'ä*) or finality (*ma*). The clauses headed by these particles are all adverbial. The next examples illustrate subordinate relative clauses with and without connectors. Relative clauses are bracketed.

- 23) *Bui* *xingu* *ya=ngú*
 be.3 much DEF.PL=house
 [*hinti* *pets'i* *ya=nsogi* *pa=r* *dehe*]
 nothing have DEF.PL=key for=DEF.S water
 'There are many houses which do not have stopcock for water'
- 24) *Ya=ts'udi* *tsi* *ya=mänsanä*
 DEF.PL=pig eat DEF.PL=apple
 [*nu'u* *tagi* *ndezu* *ja=r* *zá*]
 DEM-PL.PROX fall from LOC=DEF.S tree
 'The pigs eat the apples that fall from the trees'
- 25) *M-besinu-ga* *xi-ku-ga* *enä*
 POSS.1-neighbor-EMPH.1 say-OBJ.1-EMPH.1 say
 [*hinda* 'wä'y=ar *njeña*]
 NEG-FUT.3 rain=DEF.S year
 'My neighbor told me there will be no rain this summer'

Adjectival subordinate clauses, equivalent to relative clauses in many Indo-European languages, are not linked to the main clause by connectives in traditional Otomí.¹³ Known as “the gapping strategy” (Comrie 1989: 147f), this mechanism of juxtaposition makes no reference to the antecedent in the relative clause. This is illustrated in example (23). Adjectival subordinate clauses are headed also by particles for deictic reference (24) which include *nu’ä*, *nä’ä*, *ge’ä*, *nu’u* and *ge’u* as well as interrogative *to*. A further strategy is the use of a Spanish preposition. Subordinate clauses in reported or indirect speech are not linked to main clauses by connectives. Instead, they use finite verb forms such as *embi* ‘say.3S.DAT’ or *enä* ‘say.3S’ (25). The following examples show complex sentences in which the subordinate clause indicates the purpose of the main clause.

- 26) *Kä* *j=ar* *nijä* *q’u-w=ar* *rosaryo*
 walk.down LOC=DEF.S church make-DUAL.INCL=DEF.S rosary
 ‘They walk down to pray the rosary’
- 27) *Ngötho=r* *pa* *pong=arXuwa*
 all-DEF.S day leave-DEF.S Juan
ma *bi=qx=ya* *jwä*
 for PRS.3=hunt=DEF.PL fish
 ‘Juan left (home) the whole day to fish’

Again, juxtaposition (26) and connective particles (27) are the typical strategies for subordination. According to Hekking (1995: 45), classical Ottoman does not mark final clauses (purpose) if their subject is co-referential with the main clause, but marking is obligatory if otherwise.

The influence of Spanish is changing the typological structure of Otomí considerably through the increasing use of prepositions and conjunctions. A few examples of this use in hypotactic constructions demonstrate this sufficiently.

- 28) *Ga=eh-e* *j=ar* *nijä*,
 FUT.1=come-PL.EXCL LOC=DEF.S church
 [*pa* *ge* *da=nä-w=ar* *majä*]
 for that FUT.2=speaker-DUAL.INCL=DEF.S priest
 ‘We will come to the Church for him to speak to the priest’

¹³ Notice that adjectival clauses in Otomí are always post-nominal. See Hekking and Bakker (2007) for further explanations.

- 29) *Ar=bätsi b́i=nzoni [porke b́i=ntsät'i na nts'edí]*
 DEF.S=child PST.3=cry because PST.3=burn SUPL hard
 'The child cried because he burned painfully'
- 30) *När=jä'i [ke xka xi-kí]*
 DEF.S=person that PRF.3 say-ACC.1
ge m-tyo-ga-nu
 Npd POSS.1-uncle-EMPH.1-EMPH.EXO.3S
 'The one who said it to me is my uncle'

Spanish connectors are varied in Otomí. Hekking identifies twenty-two different Spanish connectors in his corpus, some of which occur more frequently than others. The most frequent by far are *pa* (short form of Spanish *para* 'for') and its compound forms (e.g. *pa ge*) as illustrated in (28), followed by others like *como* 'as' and *porque* 'because' (29). Less frequent is the Spanish conjunction *que* 'that', which heads dependent (adjectival) clauses (30). The occurrence of Spanish connectors in everyday speech has modified the way Otomí marks syntactic relations in the sentence.

Compared to the noun phrase, the sentence shows more analytical structures. Syntactic relations are expressed asyndetically by means of juxtaposition or syndetically by deictics, proclitics, adverbial particles, Spanish prepositions and conjunctions. The ongoing shift from juxtaposition to connectivity through native particles or borrowed prepositions makes contemporary Otomí more hypotactic than classical Otomí. The final outcome of this shift might be the loss of verbal suffixes from colloquial speech (Hekking 1995: 155ff). Chapter 11 gives a detailed analysis of prepositional and conjunctive connectivity in contemporary Otomí.

What about constituent order in Otomí? Possession in Otomí follows a possessed-possessor order while attributive modification a modifier-head order. The order of adjectival (relative) clauses in complex noun phrases is post-nominal. Compared to the fixed VOS word order of classical Otomí, the modern language shows other alternatives, in particular a tendency towards SVO. For the Otomí of Santiago Mexquititlán, Hekking (1995: 36) identifies SVO as the basic word order while the same order is prevalent in the Otomí of San Andrés Cuexcontitlán in the state of México (Lastra 1994). In contrast, Suárez (1983: 95) identifies Otomí as a VOS language on the basis of Highland Otomí (Hidalgo). Soustelle, in turn, classifies Otomí as a typical VSO language on the basis of the Otomí spoken in San Jose del Sitio (Mexico State). Because there is no comprehensive study of syntactic variation across dialects, we cannot make any generalization upon a solid empirical basis. Still, one tendency is clear in Querétaro Otomí: the increasing frequency of SVO order as compared to VOS or VSO orders. Compare SVO examples (7), (13), (16), (20), (21) and (24), with VSO examples (23) and (27).

The System of Parts of Speech in (Querétaro) Otomí

Few issues in Amerindian linguistics prove as controversial as the identification of parts of speech. The classification of parts of speech in Ecuadorian Quichua and Paraguayan Guaraní showed this clearly. The reasons for the failure of most grammatical descriptions to properly identify parts of speech lie on a long tradition that makes use of linguistic categories proper to Western European languages. In addition, there is the influence of other factors such as a) a process of dialectalization which makes invalid for one dialect what is valid for another; b) the influence of Spanish at the lexical and grammatical levels; and c) the fact that lexical categories in some Amerindian languages make subtler distinctions than those used in most European languages. For Otomí all these factors conspire intricately and make conclusive statements unsustainable. Therefore, the typological classification of parts of speech elaborated in the following should be considered a tentative proposal awaiting further study. Most of what is said here is not new, except for the way it is said. The analysis is based on previous work on the topic by several authors (cf. Soustelle 1937; Voigtlander and Echegoyen 1985; Lastra 1992; Hekking 1995; Palancar 2006; Bakker and Hekking 2007; Bakker *et al* 2008). Of these sources, particular attention will be paid to Palancar (2006), who deals specifically with parts of speech in Otomí.

From the start it is hard to establish a clear-cut division between verbs and nouns in Otomí. Most nouns can be used predicatively without any mechanism of derivation. Soustelle explains this special feature of Otomí in the following terms:

“The distinction between nouns and verbs is very uncertain and hard to capture. As far as form is concerned, we should point out that most words might be both nouns and verbs. Therefore, a large number of words do not tell us whether they are nouns or verbs only by their form. In fact there is only one much-reduced class of nouns that can never be used as verbs. These are the nouns carrying the nominalizer prefix *t-*. [...] However, it is hard to make a clear noun-verb distinction even in the case an allegedly nominal prefix is present” (Soustelle 1937: 165; my translation).

I maintain that it is perfectly possible to make a distinction between nouns and verbs in Otomí on the basis of morphological distribution.¹⁴ The following arguments support this view.

Otomí verbal morphology consists basically of proclitics and suffixes. Proclitics mark person, tense and aspect. Suffixes mark number, inclusive-exclusive

¹⁴ Purely syntactic criteria are less helpful to identify lexical classes in Otomí, for word order patterns vary across dialects and a number of pragmatic and discourse factors intervene.

distinctions, emphasis, location, comitativity, direct object and indirect object. Verbal proclitics are distinct from nominal proclitics in that the latter indicate definiteness and number. Both types of proclitics are not interchangeable. Nominal proclitics do not precede verbs just like verbal proclitics do not precede nouns. In contrast, verbal suffixes occur on verbs but also on nouns. They include the markers of number, clusivity, location and emphasis. Nouns in predicative function are always marked by one of these suffixes, which they share with verbs. On the contrary, suffixes marking comitativity, locativity, direct and indirect object do not occur on nouns (Bakker *et al* 2008). The following examples from Hekking (1995) illustrate the aforementioned distribution of proclitics and affixes. Verbal proclitics and suffixes appear in bold.

- 31) *Ar=xudi* ***ga=pa*** *ma=xoro*
 DEF.S=tomorrow FUT.1=sell POSS.1=guajotole [turkey]
 ‘Tomorrow I will sell my guajolote’
- 32) *Ar=ts’unt’u* ***da=’yapi*** *ár=’ye* *ar=nxutsi*
 DEF.S=bridegroom FUT.3=ask.OI.3 POSS.3=hand DEF.S=bride
 ‘The bridegroom will ask the bride’s hand for marriage’
- 33) ***Bí=mānda-wi*** *’nar=hé’mi* *ár=amigo* *Enrike*
 PST.3=send-DUAL INDEF.S=letter POSS.3=friend Enrique
 ‘He sent a letter to his friend Enrique’
- 34) ***Di=kut’a-hu***
 PRS=five-INCL.PL
 ‘We are five’
- 35) *Ar=Xuwa* *mi=ña-wi* *ár=to*
 DEF.S=Juan IMPF.3=speak-DUAL POSS.3=mother-in-law
 ‘Juan was talking with his mother-in-law’

In sum, while most nouns can be used predicatively, they still make a class of lexical elements different from verbs according to the distribution of morphemes. In other words, nouns and verbs in Otomí cannot be grouped in one indistinct class of flexible elements.¹⁵ Still, both lexical classes are open to the extent that new members enter through borrowing or compounding (cf. *infra*).

¹⁵ Notice that a similar distribution of parts of speech was found in Guaraní, where nouns make a lexical class separate from verbs but still can be used predicatively.

Notwithstanding the relevance of a noun-verb distinction for any classification of parts of speech, several authors maintain that the major issue in Otomí concerns the existence of adjectives (Soustelle 1937: 165; Palancar 2006: 28; Bakker *et al* 2008). The remaining part of this section focuses on the discussion of this lexical class.

According to recent studies, lexical items classified as ‘adjectives’ in most European languages belong to either nouns or verbs in Otomí, depending on morphosyntactic criteria. Implicit in this proposal is the aforementioned distinction between nouns and verbs.¹⁶ Let us first have a look at noun-like adjectives, i.e. adjectives showing nominal morphology. This morphology includes not only the proclitics of definiteness and number, which some authors call ‘articles’ (e.g. Hekking and Andrés de Jesús 1984; Hekking 1995)¹⁷ but also the verbal proclitics used on nouns with predicative function. In the following examples the proclitics *gar*, *ar* and *ya* accompany lexemes encoding property concepts such as *nduxte* ‘naughty’ and *junt’ei* ‘jealous’:

- 36) a. *g=ar* *nduxte* b. *ar=nduxte*
 PRS.2=DEF.S naughty DEF.S=naughty
 ‘Your are naughty’ ‘The naughty person’
- 37) *Ya=junt’ei* *mi=tsa* *m=ar* *’ñu*
 DEF.PL=jealousy IMPF.3=feel IMPF.3=DEF.S pain
 ‘Jealousy hurt him’
- 38) *Ar=’ñohō* *i=bi=zu* *ni* *’nar= pa*
 DEF.S=mister PRF.3=PST.3=scare nor INDEF.S=day
bi=mengi *bi=’ñeme* *ya=junt’ei*.
 PST.3=return PST.3=believe DEF.PL=jealous
 ‘The guy was so scared that he never felt jealous again’

Noun-like adjectives are accompanied with proclitics indicating person and tense (*gar*), number and definiteness (*ar*, *ya*). According to Palancar (2006: 347), there are twenty different noun-like adjectives in Otomí. Most, if not all, of these lexemes refer to properties attributable to human beings. Property-concept nouns behave like other nouns in that they are *not* linked to their subject noun phrases by a copula if used predicatively. Because the language does not use copulas for non-verbal

¹⁶ Soustelle maintains that it is difficult to distinguish nouns from verbs in Otomí but insists that property concepts are encoded either as nouns or verbs (Soustelle 1937: 165).

¹⁷ No equivalence exists. The most important difference between these nominal proclitics and articles is that nominal proclitics are obligatory with determiners, quantifiers or interrogative pronouns (Hekking 1995: 57f).

predication, property-concept nouns should be thus considered denominal verbs rather than adjectives per se. Consider the following example from San Idelfonso Otomí (Palancar 2006: 349):

- 39) *No=r* *já'í* *ar=günt'ei*
 DEF=S person DEF.S=jealous
 'The man is jealous'
- 40) *Ga'tho* *nu* *ma=míxi* *ya=nduxte*
 all DEF.PL POSS.1=cat PRS.3PL=naughty
 'All my cats are naughty'

A parallel class of property-concept lexemes in Otomí is that of verb-like adjectives. Palancar classifies this class as part of a larger class of stative verbs distinct from active verbs according to morphological parameters. For example, the third-person imperfect proclitic for active verbs is *mí*, but the same proclitic for stative verbs is *már* (Palancar 2006: 333f). This partition applies to all Otomí dialects, although it is not clear what members make up the class. Palancar mentions that “the verb *jóhyá* ‘be glad’ is one of a very few active verbs in Otomí that depict PCs [property concepts]” (Palancar 2006: 336). In contrast, Hekking and Bakker (2007) assign the same lexeme to the class of intransitive verbs along with others like *dáthi* ‘be ill’ or *txutxu'lo* ‘small’. In the following examples of *jóhya* a resultative-state reading is obligatory for (41) whereas a present reading is required for (42). Notice also the different use of tenses: past in the first sentence, present in the second.

- 41) *Nú* *ma=nõno* *xa* *bi=n-jóhyá*
 DEF POSS.1=mother Int PST.3=NI-be.glad
 'My mother got very glad'
(Palancar 2006: 336)
- 42) *Di=johya-he*
 PRS.1=be.happy-EXCL.1
 'We are happy' (and not 'we become happy')

In consideration of additional morphological criteria, Palancar makes a further distinction of verb-like adjectives in two subclasses. The first subclass is characterized by its overlap with active verbs as regards inflection. Palancar lists eleven of such verbs: e.g. *dõtá* ‘be big’, *tx'úlo* ‘be small’, or *tse^htho* ‘be strong’, *nzátho* ‘be beautiful’, *rá'yo* ‘be new’, *txu* ‘be old (for a man)’, etc. The great majority of lexemes from this class refer to human characteristics, like noun-like adjectives. Similarly, the members of this subclass vary from dialect to dialect. For instance,

Hekking classified *rá'yo* 'be new' (cf. *supra*) rather as a property-concept noun. Consider the following example:

- 43) *Nuya* *ya='bets'i* *hingi* *ya='ra'yo*
 DEM.PROX.PL DEF.PL=thing NEG-PRS.3 DEF.PL=new
 'These things are not new'

Differences in classification are observed in other lexemes such as *dötá* (San Idelfonso) or *dätä* (Santiago Mexquititlán) 'be big'. Palancar classified *dötá* as a property-concept verb whereas Hekking classifies the same lexeme as a noun.

The second class of verb-like adjectives is an open class. Adjectives borrowed from Spanish become members of this class. Palancar (2006: 337) characterizes this class on the basis of four morphosyntactic peculiarities: a) their argument is encoded with object morphology; b) they receive a morphologically conditioned nasal prefix¹⁸; c) they use a special set of function words; and d) they lack a morphosyntactic bound form.

The coding of arguments with object morphology is the most salient feature of verb-like adjectives. In Otomí, verbal suffixes marking patients (direct objects) in transitive verbs mark experiencers in intransitive verbs. The same suffixes occur on verb-like adjectives. According to Palancar, this feature makes verb-like adjectives similar to stative verbs and "reveals that Otomí has an active/stative split involving intransitive verbs" (Palancar 2006: 338).¹⁹ The following examples from Hekking and Bakker (2007) and Palancar (2006) illustrate this morphological feature of Otomí. Verbal prefixes appear in bold.

- 44) a. *Xi=nts'**ut**'i-gi* b. *Xi=nts'**ut**'i-'i*
 PRF.3=thin-OBJ.1 PRF.3=thin-OBJ.2
 'I am thin' 'Your are thin'
- 45) *Xi=**ñh**ets'i-'i*
 PRF.3=be.tall-OBJ.2
 'You are tall'

The second feature of verb-like adjectives consists in the occurrence of a morphologically conditioned nasal prefix: /n-/ in (44a-b) and /ñ-/ in (45). Palancar

¹⁸ This nasal prefix has the same form as the corresponding prefix of nasal intransitive verbs. Compare examples (112), (113) and (109).

¹⁹ Otomí would be, therefore, similar to Guaraní in this respect (cf. section 7.3).

considers this prefix a verbal marker of stativity, which does not occur on verbs of the active type.²⁰

The third feature consists in the occurrence of the verbal proclitic *xi* before verb-like adjectives in present tense. Notice that this proclitic encodes perfectivity on other verbs. This means that verb-like adjectives describing a present state of affairs require perfect morphology instead of null morphology as other verbs. The fourth feature of verb-like adjectives acting as stative verbs is the lack of a bound form. The absence of such form in the vast majority of stative verbs draws a divide between them and the rest of verbs, which always have two forms, free and bound (cf. Palancar 2004).

Verb-like adjectives are used also as modifiers of referential phrases. Hekking and Bakker (2007) give some examples of this use. Compare the following examples:

- 46) *Ar=hets'i* *'ñoho*
 DEF.S=tall man
 'The tall man'
- 47) *Ar=ts'ut'i* *nxutsi*
 DEF.S=thin girl
 'The thin girl'

For authors like Voigtlander and Echegoyen (1985), Lastra (1992), Andrews (1993), and Hekking (1995), these examples are instances of nominal modification, according to which the lexemes attributing a quality or property to the head noun should be considered adjectives. Palancar parts company with these authors because he considers constructions like (46) and (47) instances of nominal compounding. The first (dependent) element of these compounds is a property-concept verb and the second element a noun. Other compounds are formed only by nouns. The difference between noun-noun compounds and verb-noun compounds lies on the semantics of the dependent element. In the first case this element specifies the function or the source indicated by the second element, whereas in the second case the dependent element expresses a property of the entity referred to by the nominal element. Palancar shows that both types of compounds have similar characteristics: a) they are head nouns in nominal predication; b) they have morphologically adjusted forms; c) they are new lexemes; d) they occur in lexical pairs; and e) they show restrictions concerning internal modification. (Palancar 2006: 353). In example (48)

²⁰ An alternative interpretation is that they are denominal verbs, i.e. nouns derived into transitive verbs by the nasal prefix. The resulting verb form would have two arguments: an impersonal zero subject and a recipient (Dik Bakker, p.c.).

below the noun-noun compound '*bṓts'e-hmé* basket-tortilla' or 'basket for tortillas' is a head noun in nominal predication. The same status is given to verb-noun compound '*bó-míxi* be.black-cat or 'black cat' in (49):

- 48) *Nú ná=r 'bṓts'e ar='bṓts'e-hmé*
 DEF DEM=S basket DEF.S=basket-tortilla
 'This basket (here) is a tortilla-basket'
- 49) *No ma=míxi hínge ar='bó-míxi*
 DEF.S POSS.1=cat NEG DEF.S=be.black-cat
 'My cat is not black' (literally, 'my cat is not a black cat')

Verb-noun compounds usually insert a nasal infix between the dependent element and the head. In addition, they have a suppletive bound form which occurs exclusively in compounds.²¹ In principle, all property-concept lexemes, be they verbs or nouns, may form compounds with other lexemes and create novel words. However, not all lexical combinations are possible, which, according to Palancar, "serves as another important piece of evidence that such lexical combinations should be treated as compounds, and not as adjectives, in syntactic attribution" (Palancar 2006: 357). Verb-noun compounding is the most productive type and includes property-concept verbs of the first class. In contrast, because stative verbs are limited to lexical conventions, their compounding is less productive. On the other hand, Palancar notices that the combination of more than one dependent member in verb-noun compounds is grammatical but hardly found in colloquial speech. Further restrictions on compounds concern internal modification: compounds may be modified internally only by intensifier *rá-*, a prefix attached to the whole compound, *not* to either of its elements; similarly, limitative *-tho* modifies property-concept verbs but not stative verbs in a verb-noun compound.

In view of the foregoing arguments, Palancar concludes that 1) nouns and stative verbs encode property concepts in Otomí, and 2) nouns and verbs referring to property concepts form compounds with other nouns and produce novel lexical items. An inspection of the Otomí corpus collected for this investigation demonstrates that similar constructions occur in Santiago Mexquititlán and Tolimán dialect and that differences consist in the different membership of some lexemes to

²¹ Also, noun-noun compounds have suppletive bound forms different from their free counterparts. The free form of *déhe* 'water' contrasts with its bound form *-thé* in compounds such as *dōthe* 'river'. The difference is that suppletive bound forms may also occur in constructions other than compounds.

one or another class. In sum, Otomí has no adjectives, only rigid verbs and a number of flexible and inflexible nouns.²²

A lexical class left aside so far is the class of (manner) adverbs. Of the Otomí adverbs listed by Hekking (1995: 54), only two function as modifiers of predicate phrases: *nts'edi* 'strongly' and *'nihi* 'quickly'. Interestingly, they can also modify referential phrases. This is illustrated in the following examples.

- 50) *'Nar=nts'edi* *uñä* *mi=tekwe* *ár=mfeni*
 INDEF.S=strong headache IMPF.3=waste POSS.3=brain
 'A strong headache exhausted his brain'
- 51) *Di=ne* *ga=pehni* *nuna* *ar=hé'mi*
 PRS.1=want FUT.1=send DEM DEF.S=paper
'mehni *j-a=r* *'mehni* *ngut'ä*
 send.PTCP LOC=DEF.S post quick
 'I want to send this letter by express post' (lit. by quick post)
- 52) *Pente* *ar=ndähi* *bí=ndui* *bí=jwihni*
 Suddenly DEF.S=wind PST.3=begin PST.3=blow
nts'edi *j=ar* *'rani*
 strongly LOC=DEF.S bridge
 'Suddenly the wind began to blow strongly over the bridge'
- 53) *Ba=ehe* *ngut'ä*
 IMP.2.EGO=come quickly
 'Come quickly over here'

Clearly, the same lexeme can be used in adjectival and adverbial function²³ without any kind of derivation. Lexemes of this type are few and form a closed class. They cannot be used as heads of referential phrases (nouns), but they can be used as heads

²² Palancar postulates the existence of a small class of 'acategorical lexemes'. These are bound forms expressing property concepts and occurring only in verb-noun compounds. These bound forms include only *t'úlo-* 'small', *dä*, big, and *m'ó* 'blue'. The strong resemblance between these forms and property-concept verbs points to a diachronic relation between both classes. Palancar maintains that Classical Otomí have a closed lexical class of adjectives, the remnants of which are the aforementioned bound forms (Palancar 2006: 360).

²³ Notice however the slight difference between the adjectival uses of *nts'edi* 'strong' in (104) and *ngut'ä* 'quick' in (106). The first lexeme occurs pre-nominally when used as an adjective while the second occurs post-nominally when used in the same function. Syntactically speaking, while *nts'edi* can be used both pre-nominally and post-nominally depending on its adjectival or adverbial function, *ngut'ä* is used only post-nominally regardless of its function. Arguably, this difference in syntactic behavior may be ascribed to the different subclasses of property-concept items.

of predicate phrases (verbs). Accordingly, they could be classified as instances of property-concept verbs, with an additional adverbial function. The use of stative verbs as modifiers of predicate phrases supports this classification. Consider the following examples:

- 54) *tsa xi=hño*
 feel PRF.3=good
 'it feels good'
- 55) *Hmä ar=apyo pɛts'i xingu ya=nzaki xi=hño*
 IMPF:say DEF.S=celery have much DEF.PL=life PRF.3=good
pa da=ts'i ne hmä ge ar='ñithi
 for FUT.3=IMPF.eat and IMPF.say DEM DEF.S=medicine
 'It is said that celery has a lot of good nutritional substances and is medicinal' (lit. 'It is said that celery has a lot of good life for eating...')

In (54) the stative verb *hño* modifies the predicate *tsa* 'feel'. The verb form modifies the noun *nzaki* 'life' in (55). The position of this lexeme like that of *ngut'ä* in (51) is post-nominal. This position suggests that the adjectival function of these lexemes is not prototypical but an extension of their predicative function. Alternatively, *xi hño* can be interpreted as a subordinate clause, hence its post-nominal position.

Summing up, Otomí distinguishes nouns from verbs but lacks adjectives and adverbs while using morphosyntactic strategies instead. Property concepts are encoded either by nouns or verbs and form compounds with other nouns. Stative verbs are used as modifiers of predicate phrases.

According to the theory of parts of speech proposed by Hengeveld (1992) and Hengeveld *et al* (2004), languages that distinguish two contiguous lexical classes may be flexible (Type 2) or rigid (Type 6). Flexible languages show one class of verbs and other of non-verbs, the last class encompassing nouns, adjectives and adverbs. Quichua and Guaraní are this type of languages. Otomí is different from them in several aspects. First, the use of nouns as modifiers of referential phrases is limited to a closed class of nouns in noun-noun compounds. Second, nouns cannot be used as modifiers of predicate phrases. Third, adjectives correspond to a subclass of verbs (stative verbs).²⁴ Fourth, while nouns can be used as heads of predicate phrases, they make a lexical class different from verbs. And fifth, the role played by verbs in the modification of phrases in Otomí suggests a clear resource to morphosyntactic strategies. All this demonstrates that Otomí is a rigid language of

²⁴ The existence of a small number of adjectival (bound) forms - remnants of a former lexical class of adjectives, according to Palancar – is insufficient to hypothesize the existence of a closed class of adjectives in Otomí, as typical of rigid languages with an intermediate parts-of-speech system (Type 5/6).

type 6, i.e. it distinguishes nouns and verbs as separate lexical classes while using morphosyntactic mechanisms for nominal and verbal modification.

Two caveats are required however. One is that the classification of parts of speech elaborated in this section is based mainly on dialects of the Querétaro area and should be restricted only to Querétaro Otomí. The other is that the above classification of parts of speech describes present-day Otomí. This stipulation is important since the language has experienced changes as a result of contact with Spanish in the four last centuries and these changes may increase with bilingualism.

8.4. Borrowing hypotheses for (Querétaro) Otomí

The language-specific hypotheses presented in this section are tested in Chapters 10 and 11 on the Otomí corpus collected in Santiago Mexquititlán and Tolimán. The hypotheses involve predictions about frequencies, types and functions of Spanish borrowings in the corpus. They are based on the hierarchies discussed in section 4.3 concerning a) the principle of functional explanation; b) the principle of system compatibility; c) the scales of borrowability; and d) the theory of parts of speech. The numbers correspond to those in section 4.3.

Predictions from the Principle of Functional Explanation

- H.1** Querétaro Otomí will borrow Spanish discourse elements easier than non-discourse elements.
- H.1.1** Querétaro Otomí will borrow from Spanish discourse elements such as topic and focus markers but evidentials and connectors.

Predictions from the principle of system compatibility

- H.2** Considering the morphological type of Spanish (inflectional), Querétaro Otomí (synthetic in phrase, analytic in the sentence) will borrow from Spanish (fusional) free words and roots, but neither clitics nor bound morphemes.

Predictions from the scales of borrowability

- H.3** Querétaro Otomí will borrow lexical elements easier than grammatical elements.
- H.3.1** Querétaro Otomí will borrow items from open lexical classes (e.g. nouns) easier than items from half-open (e.g. prepositions) and closed classes (e.g. articles).
- H.3.2** Querétaro Otomí will borrow Spanish lexical items in the following order of frequency: nouns, verbs, adjectives and adverbs. Adpositions (i.e. prepositions) will be borrowed, if at all, less easily because Querétaro Otomí does not have a syntactic slot for them, unless a gap-filling strategy is involved (cf. 2.6.2.2). Pronoun borrowing will be disfavored by the pro-drop character of Spanish. Articles may be borrowed to the extent that a

syntactic slot for them is available in the language. Conjunct borrowing is not expected, other things being equal.

Predictions from the theory of parts of speech

- H.4** The typological distance between Spanish (source language) and Querétaro Otomí (recipient language) is bridged in the borrowing process following the hierarchy of parts of speech: head of predicate phrase > head of referential phrase > modifier of referential phrase > modifier of predicate phrase.
- H.4.1** Accordingly, Spanish forms that function as heads of phrases (i.e. verbs and nouns) will be borrowed easier than forms that function as modifiers (i.e. adjectives and adverbs). Also, Spanish forms that function as heads of predicate phrases (i.e. verbs) will be the most easily borrowed lexical class; forms that function as modifiers of predicate phrases (i.e. manner adverbs) will be the hardest class to be borrowed. While H.4.1 contrasts with H.3.2 above, both hypotheses will be tested.
- H.4.2** If Querétaro Otomí borrows items from one lexical class, it borrows items from previous lexical classes in the hierarchy. Accordingly, if Querétaro Otomí borrows modifiers of referential phrases (Spanish adjectives), it will borrow heads of referential and predicate phrases too (Spanish nouns and verbs) but not necessarily modifiers of predicate phrases (Spanish manner adverbs).
- H.4.3** As a rigid language, Querétaro Otomí will borrow more easily lexemes from the lexical class immediately following the last differentiated lexical class in its parts-of-speech system. Therefore, Querétaro Otomí will borrow adjectives more easily, because adjectives are the lexical class that follows the last differentiated class (nouns) in its system.
- H.5** The syntactic distribution of borrowed lexemes in Querétaro Otomí will follow the same distribution of native lexical classes (functional adaptation hypothesis). Accordingly, if Querétaro Otomí borrows Spanish nouns and verbs, it will use them as heads of referential and predicate phrases, respectively. In turn, if adjectives and adverbs are borrowed, they will be used either as nouns or stative verbs. In addition, Spanish nouns might be used alternatively as verbs given the same use of native nouns in Querétaro Otomí. Therefore, Spanish borrowing will not modify the system of parts of speech in Querétaro Otomí.
- H.6** The distribution of borrowed lexemes will follow the same distribution of their lexical classes in Spanish (functional specialization hypothesis). Accordingly, adjectives and adverbs borrowed from Spanish will be used in Querétaro Otomí only in their original position of modifiers of referential and predicate phrases, even though the language does not have individual lexical classes fulfilling both syntactic functions (cf. *infra*). The functional

specialization of Spanish borrowings will thus result in a gradual differentiation of the parts-of-speech system of Querétaro Otomí. While H5 and H6 make opposite predictions, both hypotheses will be tested.

- H.7** If Querétaro Otomí borrows adjectives and adverbs and uses them in their original syntactic positions, a process of lexicalization will take place, by which the language will gradually replace morphosyntactic strategies with lexical items for the modification of referential and predicate phrases.

The foregoing hypotheses will be tested systematically on the Otomí corpus of Santiago Mexquititlán and Tolimán in the light of linguistic and nonlinguistic factors influencing the borrowing process (Chapters 10 and 11).

Chapter 9

Borrowing hypotheses in comparative perspective

The present investigation seeks to outline how the typologies of the languages in contact determine the outcomes of borrowing. It is therefore of great importance to relate the language-specific predictions made in previous chapters to each other. A cross-linguistic comparison will provide a comprehensive idea of how typology is expected to influence borrowing and how it interplays with nonlinguistic factors.

9.1. Predictions from the Principle of Functional Explanation

While the prediction from the Principle of Functional Explanation (i.e. discourse elements will be borrowed more easily than non-discourse elements) is valid for the three languages, differences are expected depending on: 1) the degree of bilingualism at the level of the speaker and the speech community; 2) the sociolinguistic situation of the recipient language vis-à-vis the source language. Because Guaraní speakers show the highest degree of bilingualism and Paraguayan Guaraní has a higher socio-political position, the discursive pressures exerted by Spanish will be less intense on Guaraní speakers and their need to borrow discourse elements consequently lesser.

Further differences are expected from the discourse structure of the recipient languages. The marking of evidentiality is of primary importance for Quichua discourse but only secondary in Guaraní and Otomí. This is reflected in the rich set of evidentials in Quichua (cf. Gómez Rendón 2006b) as compared to Guaraní and Otomí. Therefore, it is expected that Quichua borrow Spanish evidential forms of lexical (e.g. *dizque*) and periphrastic type (e.g. *se dice*). Similarly, because Quichua is a topic-prominent language, the borrowing of topic markers is expected, if available in the source language. While Spanish lacks topic markers, it makes use of syntax to mark topic and focus. Therefore, it is expected that Quichua calque word orders for encoding pragmatic values. The testing of this hypothesis implies the analysis of syntactic borrowing, but the task goes beyond this study.

9.2. Predictions from the Principle of System Compatibility

These predictions are based on the influence of the morphological type of the languages in contact on the outcomes of borrowing. The morphological type of the source language (Spanish) predicts that free forms will be borrowed more easily than bound forms. The morphological type of the recipient language predicts that the three languages will borrow free forms and roots but not affixes.

9.3. Predictions from the scales of borrowability

While the scales of borrowability predict cross-linguistic preferences in borrowing (lexical over grammatical; open over closed), the typology of the languages might determine different outcomes. Quichua and Guaraní will not borrow prepositions because they are postpositional languages. It is likely, however, that these languages borrow lexical items and use them as postpositions. Spanish prepositions are not expected in Otomí either, because the language lacks a syntactic slot for adpositions in general. Still, preposition borrowing cannot be left out as a way to fill syntactic gaps in Otomí.

The recipient languages do not have a class of conjunctions. Instead, Quichua uses discourse shifters; Guaraní has postpositions; and Otomí makes use of deictic particles or simply leaves clause connections implicit. In this context, these languages are not expected to borrow conjunctions, unless we assume they borrow them to fill syntactic gaps. The recipient languages do not have articles either. However, Otomí and Guaraní have two classes of function words performing similar functions: nominal proclitics and deictic particles, respectively. In contrast, Quichua marks definiteness only by means of a topicalizer. In this context, only Otomí and Guaraní could borrow articles from Spanish, given the functional equivalence of nominal proclitics and deictic particles.

Finally, the recipient languages have a separate class of personal pronouns. Pronoun borrowing is not expected for any of the three languages, other things being equal. The borrowing of other subclasses of pronouns cannot be excluded however. In any case, the fact that Spanish is a pro-drop language and personal pronouns are, therefore, less salient in discourse can be a decisive factor.

9.4. Predictions from the Theory of Parts of Speech

The general predictions from the theory of parts of speech concern the order in which lexical classes are expected to occur in borrowing. This order is broadly determined by the hierarchy of parts of speech and holds for any language. In contrast, the language-specific predictions from the theory of parts of speech hypothesize two possible scenarios for the use of loanwords in the recipient language: one in which loanwords are functionally adapted to the native system of parts of speech, without any typological modification; and another in which loanwords are used according to the system of parts of speech of the source language, with some typological modification. The second scenario has two alternative solutions depending on whether the recipient language is flexible or rigid: if flexible, general lexical classes (e.g. non-verbs) split into specialized classes (e.g. nouns vs. modifiers); if rigid, new lexical classes emerge and replace morphosyntactic elements of the language.