Paraguayan Guarani:

Some considerations about language mixing and an acoustic study of urban and rural vowels

Meritxell Fernández Barrera

MA Linguistics, Leiden University

August 2015

Supervisor: Dr. E.I. (Mily) Crevels

Second reader: Martin Kohlberger

Acknowledgements

This thesis has been at the same time a great challenge and a big opportunity to deepen my knowledge in fieldwork linguistics. I am sincerely grateful to all the people who contributed to making this opportunity become reality.

Firstly, I would like to express my gratitude to my supervisor Mily Crevels. She not only guided me in the selection of a research topic for the thesis, but also accompanied me throughout the whole process of designing and implementing fieldwork in Paraguay: from writing my research proposal, to applying for funding and getting in touch with the speech community, she provided extremely useful advice and support all along the way. She also shared with me unique bibliographical references about Paraguayan Guarani that I would not have been able to find in the library, and she carefully read this thesis, always providing insightful comments to improve it.

My thanks go also to Martin Kohlberger, whose course on Fieldwork Phonetics was a source of inspiration for undertaking phonetic research in the field. The course provided me with invaluable methodological tools and Martin's advice on recording equipment was also precious.

I am highly indebted to all the members of the Literacy and Lifelong Learning Department at the Paraguayan Ministry of Education and Culture. They welcomed me as a member of the team and they made their best to make me feel at home in their country. Furthermore, they allowed me to participate in their monitoring mission in the Department of Concepción and they helped me establish contact with the rural speech community. Without them, I would never have been able to interview rural speakers. I also thank the members of the technical office of the Spanish Agency for International Development Cooperation in Asunción. They provided invaluable logistic support and advice.

Last but not least, I would like to thank all the Guarani speakers that kindly accepted to participate in my research. I thank them for their patience and generosity in sharing their linguistic knowledge with me, and also for leading me to a better understanding of Paraguayan culture.

Financial support for this thesis came from the Leiden University Fund and from the Sustainable Humanities Internship Fund of Leiden University.

Abstract

This thesis tackles two aspects of Paraguayan Guarani. Firstly, it discusses the role of language mixing as a result of contact with Spanish. In this sense, it describes the types of mixed structures reported in previous literature and it reviews contradictory claims considering, on the one hand, that mixed structures are just part of a discursive strategy (code-switching), and on the other hand, that they are part of a fully-fledged new code (a so-called mixed language). The thesis concludes that only by measuring the stability of mixed structures across speakers and across contexts of use will it be possible to reach a solid conclusion with regard to this matter. Secondly, against the prevalent assumption that Paraguayan Guarani is mostly homogeneous across regions, the thesis conducts an acoustic study comparing the formant values of Guarani vowels produced by urban speakers from Asunción and rural speakers from Concepción. Results show that the differences between individual segments are not statistically significant. However, significant differences are revealed in the F2 of two vowels, /ɨ/ and /i/, in specific consonant environments, thus pointing to potential variation in coarticulation patterns.

Table of Contents

List of Tables

List of Figures

Glosses

The following glosses are used in the Guarani and Spanish data presented in the examples.

1SG>2SG	1st person agent argument, 2nd person object argument
3	third person
A1	1st person set A crossreference marker
ART	article
B1	1st person set B crossreference marker
CAUS	causative
CONJ	conjunction
DAT	dative
DEM	demonstrative
FUT	future
INT	interrogative
LOC	locative
NEG	negation
NEUT	neuter
PE	plural exclusive
PFV	perfective
PL	plural
PRES	present
PST	past
REFL	reflexive
REL	relational prefix
SG	singular

Motivation and summary

Paraguayan Guarani was initially suggested by my supervisor Mily Crevels as a research topic for this thesis. My goal was to have a first fieldwork experience in language documentation in Latin America and at the same time I had a keen interest in language contact processes and in the sociolinguistic dynamics that shape contact varieties. As it turned out Paraguayan Guarani allowed all these levels of analysis.

First of all, as will be discussed later on, works documenting Paraguayan Guarani are scarce, especially with regard to the phonetics of the language. Indeed, most phonetic studies have relied on a reduced number of Guarani speakers residing outside the country. Some authors concerned with a morphosyntactic description of the language have compiled oral corpora (e.g. Gregores and Suárez 1967, Tonhauser 2006, Gómez-Rendón 2008, Kalfell 2010). However most of those corpora have been obtained from speakers in urban areas where the degree of Spanish-Guarani bilingualism is high, and where in any case daily exposure to Spanish is considerable.

Secondly, Paraguayan Guarani is well-known for the profound influence Spanish has exerted on it. This influence is most evident in the lexicon, were phonologically non-adapted borrowings from Spanish abound, but Spanish has also left its imprint on phonology (e.g. new segments) and morphosyntax (e.g. metatypy). This convergence with Spanish has led to a growing body of literature discussing the status of Paraguayan Guarani in the continuum of language-mixing phenomena (i.e. from code-switching practices to a new mixed code).

And thirdly, the current linguistic landscape of Paraguay is considered unique from a sociolinguistic perspective: Guarani and Spanish are both recognized as official languages on equal ground, and a big part of the non-indigenous population speaks Guarani. This is often portrayed in the literature as a triumph of the native language over the pressure of the language brought by the colonizers.

Bearing in mind these three particularities of the language, I set out to design a phonetic experiment that would allow throwing light on current processes of mixing between the Spanish and the Guarani phonetic systems. If innovative phonetic features were to be observed in Paraguayan Guarani with regard to the original Spanish and Guarani phonetic systems, then there would be ground to believe that Paraguayan Guarani is emerging as a third fully-fledged linguistic code.

However, as I started to read background literature, I soon run into a major obstacle: there is no consensus in the literature as to what exactly Paraguayan Guarani is. In other words, it is not yet clear whether the reported utterances, in which there is a mixture of Spanish and Guarani, are the expression of a new linguistic code or rather just instances of code-mixing practices. I realized that the term *Jopara*, originally meaning mixture in Guarani, is used repeatedly in the literature to refer to Paraguayan Guarani, but not always with the same meaning. Different works assume different definitions and even some times, the same work, will assume incompatible definitions (e.g. assuming that Jopara is both a new language and a

code-switching practice between two different languages). This led me to slightly reformulate my goals, and to devote a considerable part of the thesis to discuss the available evidence on *what* Paraguayan Guarani is and *who* speaks it. Section 1 attempts to answer the first question, while section 2 deals with the second one.

Thus, section 1 includes a discussion of the terms used to refer to the language (section 1.1), a review of the mixed structures reported by previous works (section 1.2), and a discussion of whether Paraguayan Guarani should be considered a third language or not (section 1.3). The second section describes the origins of language contact in Paraguay (section 2.2), the current bilingual situation in the country as well as the different bilingual profiles (section 2.1). One of the conclusions of section 2 is that in Paraguay, the urban-rural divide and the type of bilingual profile generally conflate into one single variable. In other words, L1 Guarani speakers tend to be located in rural areas, whereas L2 Guarani speakers are more common in urban areas. In the absence of previous comparative acoustic studies, section 3 describes an experiment aimed at contrasting vowel formants in these two groups of speakers. Thus, instead of analyzing the phonetics of mixed sentences, it aims at describing the potential acoustic variation of Paraguayan Guarani vowels, depending on the bilingual profile of the speaker and her geographic location. Fieldwork was carried out in the urban area of Asunción and in two rural areas in the northern Department of Concepción. The elicited items are inserted into sentences where no overt Spanish material is present, so that any observed variation between the two groups of speakers can only be attributed to their bilingual profile, and not to the elements of Spanish origin included in the sentence. Despite the fact that previous literature would suggest differences in the production of non-native phonemes by L2 Guarani speakers (such as /ɨ/), no significant difference was found in the formants of vowels between both groups of speakers. However, a significant difference was observed between the two groups in the production of /ɨ/ when neighboring consonants are taken into account. This finding suggests that differences between L1 and L2 Guarani speakers may lie in coarticulatory patterns rather than in the acoustic properties of individual segments.

The thesis concludes with a discussion of the main findings of the thesis and a suggestion for further lines of research (section 4).

1. Paraguayan Guarani and language mixing

1.1. Paraguayan Guarani: language *vs.* variety, register and discursive strategies

Given the unsettled debate about the characteristics and the genetic classification of the Guarani spoken in Paraguay, it is worth devoting a section to make clear what is understood by *Paraguayan Guarani* in this thesis and why this term has been chosen.

A defining feature of so-called Paraguayan Guarani is that it is spoken by the non-indigenous population. This recalls the distinction between ethnic and non-ethnic Guarani. The former includes the languages spoken by indigenous groups, such as Mbya Guarani or Ache Guarani. The latter is considered to be the modern version of the Guarani spoken by the indigenous peoples who integrated in the Spanish colonial society, although the evolution from this earlier version of Guarani to modern Paraguayan Guarani has yet to be described (Thun 2005: 312).

I prefer to use the term Paraguayan Guarani to the often used label *Jopara* since it is more neutral and thus more adequate for descriptive purposes. This is due to the fact that Jopara, which means 'mixture' in Guarani, has been used with many different meanings and it has also been used too often to support language ideologies in favor of a 'pure' Guarani. Using it to refer to the language often implies making a priori assumptions either about the degree of mixing in the language or about the desirability of such mixing. In the specialized literature, authors range from considering Jopara a conventionalized linguistic code to considering it a discursive practice. The descriptive and the prescriptive meaning of Jopara are usually also mixed up. In other words, in some cases Jopara is used merely to refer to the documented language practices, whereas in some other cases it is used as a label to identify those language practices that are incorrect according to the purist linguistic norm. Furthermore, since in most of the cases no linguistic data are offered to exemplify what Jopara is, the concept seems to be the result of an intellectual exercise ideologically motivated rather than of empirical analysis, as suggested by Penner (2007: 73).

Table 1 provides a summary of the different meanings that the term Jopara acquires both in the specialized literature and in colloquial usage. The sources where the term acquires each meaning are also reported there. The list of sources is provided for illustrative purposes and should not be considered exhaustive. It should further be noted that in the same source sometimes different meanings of Jopara appear throughout the text. Meanings *A* to *D* are descriptive, whereas meaning *E* is prescriptive. Meanings *A* and *B* refer to a linguistic code, language and variety, respectively. Meaning *C* refers to a set of varieties defined by a discursive practice based on code-switching between Guarani and Spanish. Meaning *D* refers to utterances with specific mixed structures. Meaning *E* refers to utterances and by extension to the language in general.

	Meanings of Jopara	Sources
A	The Guarani spoken by the non-indigenous population in Paraguay. Synonymous of Paraguayan Guarani	Bakker *et al.* (2008) Gómez Rendón (2008: 210)
B	Guarani **variety** more heavily influenced by Spanish (≠*guaraniete*)	Cerno (2011: 21) Pottier (1970: 43)
C	**Set of varieties** from a continuum of mixtures ranging from mostly based on Guarani with some Spanish insertions to mostly based on Spanish with some Guarani insertions (≠*guaraniete*)	Johnston (2007: 4) Estigarribia (2015: 210) Speakers' view as reported in Penner 2007: 82
D	**Instances of utterances** where a Spanish root not adapted orthographically is inserted into a Guarani matrix	MEC (2006)
E	**Incorrect** Guarani (and in some cases also incorrect Spanish)	Guasch 1948 Galeano 2011 Speakers' view as reported in Penner 2007: 82

Table 1. Different meanings of Jopara according to the linguistic phenomena it is considered to encompass

Let us first consider the descriptive meanings. Meaning *A* gathers all those conceptions that see Jopara as a fully-fledged linguistic code that is the mother tongue of most Paraguayans and is the result of the still undescribed leveling between the different dialects of Guarani spoken in colonial times. Jopara refers in this meaning to spoken Guarani and does not include *Guaraniete* or academic Guarani which is a written standard created by scholars and purged from Spanish borrowings (Mortimer 2006: 51). This understanding mostly coincides with the definition of Paraguayan Guarani provided above and it does not assume, a priori, that there is a dichotomy in spoken Guarani between a pure variety and a mixed variety. This remains an empirical question to be analyzed on the basis of linguistic data.

According to meaning *B*, Jopara refers to a specific variety of Paraguayan Guarani which is more heavily influenced by Spanish. The difference with regard to meaning *A* is considerable, since if Jopara is a variety more heavily mixed, then there must necessarily be another variety which is pure (or at least less mixed) in the linguistic repertoire of the speakers. In other words, if Jopara is just one of the varieties of the language, this implies that the speakers are able to produce some other variety which is less mixed. So far this fact remains unclear, since *Guaraniete*, understood as academic Guarani, belongs almost exclusively to the written domain. Sometimes *Guaraniete* is understood as referring to rural Guarani, in contrast to urban Guarani. However a systematic description of the differences between urban and rural Guarani is still lacking. Only when this split is solidly documented will it be justified to speak of Jopara *vs.* Guaraniete as two varieties of the language.

It is not rare in the literature to find authors who seem to hesitate between an understanding of Jopara as simply referring to spoken Guarani (meaning *A*) and as a variety of spoken Guarani which is more influenced by Spanish (meaning *B*). This is the case of Dietrich (2010) who refers to Jopara as "[…] much more a style of speaking than a proper language" (Dietrich 2010: 40), but later on he refers to two levels of Guarani, Guaraniete "[…] which is never spoken" and "spoken Guarani, or Jopara" (*Ib.*). These two interpretations are incompatible, since if Jopara is the only Guarani spoken, then it cannot be just a style of speaking characterized by code-switching. This would amount to saying that speakers of Guarani do not have a *proper* language but only a code-switched speech. In a way, this would deny the existence of (Paraguayan) Guarani.

Meaning *C* is similar to meaning *B* in the sense that it also presupposes the existence of less mixed varieties, but in this case Jopara is understood as an umbrella term encompassing a *set* of varieties from a continuum ranging from predominantly Guarani to predominantly Spanish. In this line, Estigarribia (2015: 210) equates Jopara to a *mixed lect* as conceptualized by Backus (2003), namely as a register that displays insertional code-switching and alternational code-switching and in which the grammatical structure can come from Guarani or from Spanish. From this point of view Jopara would include *casteñí*, that is to say, the production of code-switched speech where most of the material is from Spanish origin, and *guarañol*, the code-switched patterns where most of the linguistic material is Guarani. This wide meaning seems to be also the one assumed by Johnston (2007) in a phonetic study analyzing the duration of voiceless stops /p, t, k/ in Spanish loans inflected with Guarani morphology. This is reflected in the sentences the author elicited as Jopara instances. Whereas in (1) the structure is mainly Guarani with the Spanish lexical insertions *juga* and *muñeka*, in (2) the pattern is the opposite: the structure is Spanish with the Guarani insertion *tatu* (Johnston 2007: 20). In the following examples the use of italics indicates that the morpheme is of Spanish origin.

(1) a-*juga*-ta che *muñeka*-re
A1SG-play-FUT 1SG doll-with
'I am going to play with my doll'

(2) *ese* tatu *se* *me* *escap-ó*
DEM armadillo 3SG.REFL 1SG.DAT run.away-3SG.PST
'That armadillo run away from me'

Similarly, the speakers' understanding of Jopara seems to be large enough to include both sentences with a Guarani matrix and sentences with a Spanish matrix. This is illustrated by Penner (2007: 82), who reports a study where 800 speakers were asked to decide for a set of sentences whether they were cases of Guarani, Spanish or Jopara. Almost 75% of the participants considered the sentence in (3) to be Jopara. This proves that the presence of a single Guarani root (in this case the interrogative particle *piko*) in a Spanish matrix is enough for the speakers to consider it an instance of Jopara.

(3) *¿No* *hiciste* todavía piko *lo* *que* *te* *ped-í*?
NEG do.2SG.PST still INT 3SG.NEUT CONJ 2SG.DAT ask.for-1SG.PST
'Didn't you do what I asked you?' Penner (2007: 82)

In any case, meaning *C*, just like meaning *B*, implies the existence of registers of Guarani where there is less Spanish influence, and so far these have not been documented and systematically described. When are these pure varieties and/or registers used? What are their lexical and morphosyntactic differences with regard to Jopara? Before these questions can be answered, using Jopara in any of these two meanings is an aprioristic use of the term, unless it is just used in contrast to Guaraniete, the written variety.

Meaning *D* is more restricted in scope since it refers to specific utterances where a Spanish root not adapted phonologically is inserted into a Guarani matrix. This meaning was the one suggested by the Ministry of Education and Culture (2006) which considered that Jopara included loans not adapted phonologically and inserted into a Guarani matrix. As a counterpart, *Jehe'a* was understood by the Ministry to refer to the loanwords adapted to Guarani phonology. However, as noted by Hauck (2014: 129), this classification confounds orthographic integration with phonological integration since it admits non-native phonological sequences as native provided they are written according to Guarani orthography. The examples of Jehe'a shown in (4) are indeed only adapted orthographically but not phonologically, since consonant clusters such as "dr" ("in kuádrope") are not part of native Guarani phonology. The definition of Jopara could thus be reformulated as instances of utterances where a Spanish root not adapted *orthographically* is inserted into a Guarani matrix.

(4) amongurusu kuádrope […]
a-mo-kurusu kuádro-pe
A1SG-CAUS-cross box-LOC
'Tick the box […]' (Hauck 2014: 128)

As for the prescriptive meaning of Jopara (meaning *E*), this has permeated both the specialized literature and the speakers' metalinguistic representations. From this point of view, Jopara denotes the deviance from a supposed linguistic norm excluding any type of Spanish interference. This negative connotation of Jopara was already present in the literature around the mid-twentieth century, such as in the foreword of the bilingual dictionary published by Guasch (1948) (cited by Penner 2007: 58). It is also a recurrent notion in the works of the most purist grammarians of the Guarani language (e.g. Galeano 2011: 20). Penner (2007: 84) further hypothesizes that according to the speakers' perceptions Jopara may refer to any incorrect utterance (in Spanish or Guarani), independently of whether there are elements from the other code or not.

Given the broad range of linguistic phenomena that Jopara has come to designate and also its prescriptive connotation, the term appears to be ill-suited for descriptive purposes. It is thus not surprising that there is a trend in the most recent literature to use the term Paraguayan Guarani instead of Jopara to refer to the language leaving apart discursive phenomena which involve mixing with Spanish material (e.g. Walker 1999; Tonhauser 2006, 2011; Gómez-Rendón

2008; Estigarribia 2015). An alternative term also used to avoid the label Jopara is *colloquial Guarani*. This was the term used in one of the first descriptive works of the language (Gregores and Suárez 1967). I prefer to keep the term Paraguayan Guarani to refer to the language, since it seems to be broader than colloquial Guarani, which may be understood as covering only certain language registers.

1.2. About mixed structures in Paraguayan Guarani

The amount of different meanings that Jopara has acquired in previous works is indicative of the relevance of mixing processes in Paraguayan Guarani. This section describes them in a more systematic way. The three main processes to take into account are: structural convergence, insertion and alternation.

On the one hand, structural convergence refers to the influence of Spanish on the structures of Paraguayan Guarani without the intervention of any overt linguistic material. Example (6) shows a case of structural convergence towards Spanish, without any Spanish morphemes. In this example possession is expressed with the verb 'to have' instead of with the predicative nominal constructions of Guarani (5).

(5) che che-r-a'y h-eta
1 B1SG-REL-son REL-much
There is multiplicity of my son(s) of me
'I have many sons' (Dietrich 2010: 43)

(6) che a-reko h-eta che-r-a'y
1 A1SG-have REL-much 1-REL-son
'I have multiplicity of my sons'
'I have many sons' (Dietrich 2010: 43)

On the other hand, insertion and alternation refer to processes that lead to the combination of Spanish and Guarani linguistic material in mixed utterances.

According to Muysken (2000:3), insertion consists of the integration of material from one language into the structure of another language; alternation refers to a suite of structures from different languages. Using the notion of matrix as defined by Myers-Scotton (1993: 6), namely, as the language providing the morphosyntactic frame, there is only one matrix language in insertion, whereas in alternation there are two matrix languages. Examples (7) to (9) exemplify some of these structures. Some of the examples come from a volume compiling the transcription of interviews of former members of the *Christian Agrarian Leagues*[1] (Comisión NRDHC 1992).

Example (7) is a clear case of insertion. It illustrates the insertion of Spanish roots into a Guarani matrix. The Spanish verbal root *persegi* 'follow' is inflected with the personal prefix

[1] The *Ligas Agrarias Cristianas* was a political movement that emerged between the 60's and the 80's in rural areas of Paraguay to defend agricultural interests.

ore- 'B1PE' and with the perfective suffix *-ma*. In the same example the Spanish root *vecino* 'neighbor' is inflected with the Guarani plural marker *-kuéra*.

Example (8) is a clear case of alternation: there is a change of grammatical frame from the first sentence, where the verbal morphology is Guarani (the active third person personal prefix *o-* is attached to the root *japo*, which means 'do'), to the second sentence, where the morphology is Spanish.

(7) ha upéi ore-*persegí*-ma *vecino*-kuéra almacenero
and after B1PE-follow-PFV neighbor-PL shopkeeper
'and after (that) we were chased by the shopkeepers of the area'
(Comisión NRDHC 1992: 11)[2]

(8) […] ojapo *primero* ha *segundo grado*, i*firma* ojapomi ha upéantema.// Es muy triste no saber leer, no saber firmar.

'[…] they do the first and the second grade, they barely learn how to sign and that is all. // It is very sad not to know how to read, how to sign. ' (Thun 2005: 334)

Example (9) is less obvious. On the one hand, it could be a case of alternation, since in the second part of the sentence ('la escuelita en el año 1974'), there is no Guarani morphology at all, so a Spanish matrix can be assumed. However, unlike in example (8), the Spanish constituents are in a dependency relation with a Guarani verb, namely *romoĩ*, so it is difficult to tell whether it is a case of alternation or insertion. Some heuristic criteria have been suggested in the literature to assess this type of structures. According to Muysken (1995), when several constituents in a row are switched, which together do not form a constituent, alternation is more likely. The same prediction is made when the switched element is at the periphery of an utterance, as in this case (Muysken 1995: 180).

(9) ha upéi ro- moĩ la escuelita en el año 1974
and after A1PE-put ART school in ART year 1974
'and later we opened the school in 1974' (Comisión NRDHC 1992: 35)

This initial characterization corresponds to a first descriptive level in terms of the structure of the mixed utterances. A second level of analysis addresses the issue of whether the structure is conventionalized or just a discourse strategy. The main difference between the two is that whereas in the first case the mixed utterance is stable in the speech community and thus can be considered part of the code, in the second case the mixed structure is just the result of discourse choices made on the spot by the speakers.

On the one hand, conventionalized mixed structures are the result of borrowings, either of single words or multiword expressions. On the other hand, discourse strategies refer to code-switching, in which the speaker switches from one language to the other. If the switch takes place at the border of two sentences it is considered intersentential, whereas it is intrasentential if the switch occurs within a sentence. The relation between, on the one hand,

[2] In the examples coming from the Agrarian Leagues corpus I keep the orthography used by the transcribers.

the processes of insertion and alternation and, on the other hand, the degree of conventionalization, is summarized in Table 2.

Term	Overt structure	Type of bilingual production
insertion	Spanish root (or constituent) not adapted phonologically inserted into Guarani matrix (e.g. inflectional morphology)	*If conventionalized→ Borrowing*. This possibility does not assume the bilingual competence of the speaker.
		If not conventionalized→ Insertional code-switch. The speaker introduces a Spanish word or constituent in a Guarani utterance. There is a certain bilingual competence and probably there is an alternative Guarani word that refers to the same concept.
bilingual sentence	Set of constituents in Guarani followed by a set of constituents in Spanish (or the opposite order). The Spanish constituents are in a dependency relation with the Guarani matrix.	*If insertion and conventionalized→ Borrowing of multiword expression*. A suite of borrowed multiword expressions which produces the impression of a code-switch. No bilingual competence is assumed.
		If insertion and not conventionalized→ Insertional code-switch. The speaker introduces a suite of Spanish constituents in a Guarani utterance. There is a certain bilingual competence.
		If alternation and not conventionalized→ Alternational code-switch. The speaker switches between the two codes, Spanish and Guarani. Therefore a certain bilingual competence is assumed.
alternation	Full sentence in Guarani followed by a full sentence in Spanish (or the opposite order).	*Alternational code-switch*. The speaker switches between the two codes, Spanish and Guarani. Therefore a certain bilingual competence is assumed.

Table 2. Types of mixed structures and their potential analyses in terms of bilingual productions

In the case of **alternation** between sentences (such as in (8)), since two matrix languages are involved, the categorization as a discourse strategy is rather straightforward. This discourse strategy is usually referred to as *alternational code-switch*. These cases assume a certain bilingual competence of the speaker, even if the competence covers only certain semantic domains, as is typical for cases of diglossia.

In the case of **insertion** (as in (7)), the question is open as to whether the insertion of the foreign element is pervasive, stable and conventionalized across the speech community or whether it is rather a deliberate switch between languages. The question does not have an

easy answer in Paraguayan Guarani, since in the absence of a widely accepted linguistic norm considerable variation is to be suspected across speakers and contexts. If the insertion is conventionalized, it is a case of *borrowing*. This means that the element imported from the donor language is already integrated into the recipient language (although not necessarily adapted phonologically) and thus knowledge of the donor language is not assumed from the speaker. The speaker does not even have to be aware of the foreign origin of the borrowed element. If, on the other hand, the insertion is not conventionalized, the mixed structure is the result of the use of two codes. In other words, the speaker switches from one language to the other and therefore some bilingual competence is assumed. This is referred to as *insertional code-switch*. Insertional code-switch is sometimes signaled with prosodic cues such as emphasis or a preceding pause (Auer 1999: 314), and even if it does not require a proficient knowledge of the source language, it does assume the awareness of the speaker of the foreign origin of the inserted element.

In cases like (9), which are referred to as **'bilingual sentence'** in Table 2, there are three possibilities. If structurally the utterance is considered an alternation, then an alternational code-switch can be diagnosed. If the speaker can productively use the Spanish grammatical elements which appear in the multiword expression (such as the preposition *en*) there is some ground to believe it is an alternation. If, on the contrary, it is considered an insertion, then it still has to be decided whether the insertion is a borrowing or just an insertional code-switch. This depends on stability measures. If the multiword expression is used consistently by the speech community and there is no equivalent Guarani term, then there are more chances that it is a borrowing. If the referent of the multiword expression is a concept rather than a proposition, it is also more probable that it is a borrowing.

In the literature that analyzes mixed structures in Paraguayan Guarani like those reported in (7) to (9), the criteria and the terminology used to identify each type of structure are not always consistent, and thus it is difficult to compare the findings of different studies. Gómez Rendón distinguishes *borrowings*, *complex borrowings* and *code-switching*. He suggests heuristic criteria to distinguish borrowings from code-switching, such as resemanticization, phonological adaptation and frequency of occurrence (Gómez Rendón 2008: 30). However, he notes that for some idiomatic expressions these criteria are not enough to identify them as clear cases of borrowing or code-switching. Thun (2005: 331-332) distinguishes *citations*, *code-switching* and *code-mixing*. According to this author, the main difference between, on the one hand, citations, and on the other hand, code-switching and code-mixing, is that the former are part of a stabilized mixed code, whereas the latter are discourse strategies which rely on the bilingual competence of the speakers. Thus, citations correspond to borrowings (both single-word and complex). An additional criterion characterizes citations according to Thun (2005: 333), namely, that they refer to a semantic domain where the source language, Spanish, is typically dominant, hence the term 'citation', since speakers are somehow 'citing' what they learnt from an environment where Spanish is typically used. Code-mixing seems to refer to insertional code-switching, whereas code-switching denotes exclusively intersentential code-switching, since the term identifies utterances where the switch occurs at the border of two sentences (Thun 2005: 335). According to this criterion, cases like (9) are not code-switching, since there is only a Guarani verb and thus there is no alternation between matrix languages across sentences. This approach seems to restrict alternation to switches between clauses.

However, Muysken (2004: 4) has noted that alternation can occur also within the clause, and this is the approach reflected in Table 2.

To sum up, what these studies have in common is that the terms they use to refer to the mixed utterances do not denote the structure of the utterance (e.g. whether there are one or two matrixes, whether there is a single word insertion or the insertion of several constituents). The terms used both by Gómez Rendón and by Thun refer to the type of bilingual production which underlies the mixed utterances rather than their overt structure. A problem with this approach is that it jumps to conclusions about the conventionalization of the structure before stability measures (consistency across speakers and within the productions of a single speaker) are implemented. Taking into account the fact that in Paraguayan Guarani utterances in which there is an apparent switch within the clause are very common and that these cases are not straightforwardly classified as borrowings or code-switching, it seems desirable to keep both levels of analysis separate.

1.3. Assessing whether Paraguayan Guarani is a mixed language

An open question with regard to Paraguayan Guarani is whether it is a mixed language. In the literature sometimes the expression *third language* has been used (e.g. Melià 1975, 2013), somehow pointing to the possibility that Paraguayan Guarani is a new linguistic code having emerged from the combination of Guarani and Spanish. In other cases, Paraguayan Guarani is referred to as a *mixed code* (e.g Thun 2005: 337), but just in a very general sense indicating that the language has integrated elements of Spanish origin. However, the use of these expressions has rarely been accompanied by a thorough analysis of linguistic data proving the presence of the defining characteristics of a mixed language.

An exception is Estigarribia (2015), who analyses whether the “Jopara” used in the novel Ramona Quebranto (the first novel written completely in colloquial Guarani) is a mixed language. He concludes that it is not, and that it is rather a register that includes both insertional and alternational code-switching. As discussed above, he calls this register a mixed lect, following Backus’ (2003) terminology and thus concludes that it is not a mixed language. The author bases his analysis on the assumption that the language in which Ramona Quebranto was written represents oral Guarani. However, nothing precludes the possibility that the novel uses code-switching as a stylistic device and that in spoken Guarani there are other registers in which code-switching does not play a role. If this was the case, these registers could still be considered as candidates to qualify as a mixed language.

In general terms, mixed languages are defined as bilingual mixtures with a split ancestry and identifiable source languages, having emerged in a situation of community bilingualism (Matras and Bakker 2003: 1; Meakins 2013: 159). From an overview of the proposed theoretical models, mixed languages can be identified on the basis of five dimensions: a) proportion of borrowings; b) structural type of the mixed structure; c) full functionality; d) rapid vs. slow emergence; e) community bilingualism as a condition for emergence; f) conventionalization. These dimensions are presented below, together with a discussion of each of them for Paraguayan Guarani.

a) The **proportion of borrowings** has to be significantly higher than in cases of extreme borrowing. According to Bakker and Mous (1994: 5-6), in cases of extreme borrowing the morphemes borrowed from the donor language never go beyond 45% of the lexicon, whereas in the prototypical cases of mixed languages this proportion can reach 90%. The range between these two percentages is, according to the authors, unattested.

In the case of Paraguayan Guarani, there is no general agreement on the proportion of Spanish-derived elements. The figures vary depending on the source. Whereas Gómez-Rendón (2008: 290-291) found only around 17% of borrowings and 18% of code-switching in an oral corpus containing recordings from urban and rural areas (total of 35% of Spanish-derived elements), there is an average of 48% of loanwords in a set of selected oral corpora used for the development of the Guaraní-Romance Linguistic Atlas (Thun 2002), and a surprisingly higher percentage in the novel Ramona Quebranto, which according to the analysis of Estigarribia (2015: 195) presents as much as 70% of Spanish words. This is probably due to the fact that the novel represents a specific urban register more influenced by Spanish.

In any case, the amount of Spanish-derived elements does not reach the case of prototypical mixed languages where the lexifying language provides 90% of the morphemes. The figures provided by Estigarribia (2015) are closer to this threshold, but those reported by Gómez-Rendón (2008) and Thun (2002) are closer to the threshold for languages with extreme borrowing. Variables such as semantic domain and communicative context should be controlled for, since they may trigger the use of more or less elements of Spanish origin.

b) The **structural type** of mixed languages is typically based either on a lexicon-grammar split or on a noun-verb split. In lexicon-grammar languages there is a clear split between lexicon and grammar (syntax, morphology, phonology) in terms of the source language; in noun-verb languages the split seems to align with the nominal and the verbal system. The former, also known as *intertwined languages* (Bakker 2003: 109), are best exemplified by Media Lengua. Media Lengua is a language spoken in Ecuador that has Quechua as its matrix language and Spanish as the lexifying source (Muysken 1979, 1981, 1997). Michif is the clearest example of the second type, since in this case the lexifying source is determined by whether we are in the noun or in the verbal domain. This way, French provides both noun roots and nominal morphology, and Cree provides verbal roots and verbal morphology (Bakker 1997). The literature mentions also a third type of mixed language, less discussed than the previous two and structurally different from them. This type, which goes by the name of *converted languages* (Bakker 2000, Bakker 2003: 110), refers to languages where all the morphemes come from one language but the semantic, phonological, morphological and syntactic patterns come from another language. The main difference between this case and the two precedent types is that what is borrowed from another language is not the "morphological matter" but only the structural patterns. Structurally this type resembles extreme cases of convergence or metatypy[3], although according to Bakker (2003: 110) the difference lies in the rapidity with

[3] A case of metatypy reported by Ross (1999) concerns Takia, an Austronesian language having undergone structural changes following the model of Waskia, a neighboring Papuan language. The following example, from Ross (1999), shows a word-by-word correspondence between the two languages:

which convergence intertwining takes place (sometimes in just one generation) as opposed to convergence, which is typically a slower process.

Paraguayan Guarani follows mostly the lexicon-grammar split, since most of the Spanish-derived elements are part of the lexicon. Borrowings include nouns, verbs and also discourse markers. However, as shown in (9), Spanish provides also some derivational morphology: the Guarani roots *výro* 'silly' and *jaguar* 'dog' are transformed into verbs with the Spanish derivational suffix *–ear*. To this new root, Guarani inflectional morphology (the 3rd person prefix *o*-) is attached. Thus, Paraguayan Guarani does not align completely with any of the structural types proposed for mixed languages.

(10)

Guarani root	*Derivation with Spanish –ear*	*Inflected Guarani verb*
výro (silly)	*výro+ear→ vyreá* (fool about)	*ovyreá* (he fools about) (Lustig 1996)
jagua (dog)	*jaguar+ear→jaguareá*[4] (denounce)	*ojaguareá* (he denounces)

c) **Full functionality**. A prototypical mixed language is a fully-fledged linguistic code with functionality in all semantic domains and communicative contexts. Given the diglossic distribution of Spanish and Paraguayan Guarani (Melià 2013: 63), it is problematic to assume the full functionality of Paraguayan Guarani. In some professional domains, in the administrative milieu and in education, Paraguayan Guarani still lacks terms for expressing certain concepts and so speakers usually resort to Spanish. The Ministry of Education has created neologisms to express scientific notions to enhance education in Guarani, but the newly created terms are not always accepted by the speech community.

d) **Rapid vs. slow emergence**. Thomason (1995: 17) distinguishes two types of mixed languages depending on the social conditions of their emergence. On the one hand, slow-developing mixed languages emerge in persistent ethnic groups as a result of slow processes of language change. Examples of this type of mixed language are Ma'a, a Southern Cushitic language spoken in Tanzania which has experienced massive borrowing from Bantu; and Caló, spoken in Spain, with Romani lexicon and mostly Spanish grammar. On the other hand, rapidly developing mixed languages emerge in new ethnic groups relatively quickly, sometimes within one generation. Examples of this kind of mixed language are Media Lengua or Michif.

In the case of Paraguayan Guarani, there is some evidence to believe that Spanish influence has taken place gradually and over an extended period of time: some of the loanwords, the most ancient ones, are not adapted phonologically and can be traced back to early colonial society; other loans are not adapted phonologically and are therefore more recent. Thus, it would seem that it belongs to the first category proposed by Thomason. However, the question is still open whether in the last decades there has been a renewed and more accelerated influence of Spanish on Paraguayan Guarani due to demographic changes

---

Takia: tamol an ŋai i-fun-ag=da 'The man is hitting me.'
man DET me he-hit-me=IMPF

Waski: kadi mu aga umo-so
man DET me hit-PRES.he

[4] This additional example was obtained through personal communication with Hedy Penner, Linguistics Professor at the Catholic University of Asunción.

(urbanization) and the improvement of education in rural areas. Furthermore, Paraguayan Guarani does not clearly display the main linguistic feature shared by slowly-developing mixed languages according to Thomason (1995: 20), namely, a profound influence of the dominant language at all levels of language structure and grammar. Paraguayan Guarani does show some influence of Spanish in the structural sphere (e.g. introduction of new phonemes (Gregores and Suárez 1967: 89), of a new stress pattern introduced through unassimilated loanwords (Gregores and Suárez 1967: 91), constructions converging with Spanish such as in examples (5) and (6)), but in no case has the grammatical replacement been as important as in Ma'a, where now most of the grammar is Bantu with few Cushitic features.

e) **Community bilingualism** as a condition of emergence. Prototypical mixed languages are usually assumed to have emerged in a context of widespread bilingualism (Matras and Bakker 2003: 1). According to this scenario, the factor triggering the bilingual mixture would be not so much the communicational need but the intention to perform identity acts. Since speakers are supposed to be functional bilinguals, their decision to switch code would not be motivated by a lack of knowledge in one of the languages, but rather by their will to state their mixed ethnicity.

Estigarribia (2015: 199) claims that Jopara has emerged in a situation of extended (if not full) bilingualism. However, since for his analysis he relies on the novel Ramona Quebranto, it appears that he refers to urban code-switching practices where at least some knowledge of Spanish is required. This condition was not very plausible in the early period of language mixing, since the demographic characteristics of colonial society together with the limited access to education made full bilingualism in Spanish and Guarani highly improbable. In more recent times bilingualism has also been limited. Only in the second half of the 20th century has the number of bilinguals increased considerably (see section 2).

f) **Conventionalization**. In order to be considered a new language, the bilingual mixture has to be stabilized. However, assessing the degree of stability is not always an easy task. The clearest situation is when the speech community no longer speaks the source languages. If the community is monolingual, then it is straightforward to assume that the mixed code has been conventionalized since the speakers do not have the bilingual competence required to practice code-switching. Such a prototypical case is Michif, where speakers do not speak any of the two source languages, Cree and French. Thus the elements of French and Cree that are present in Michif must necessarily be part of an autonomous linguistic code.

When the source languages are still spoken by the community, stability measures are required to tell apart cases of occasional mixing depending on the bilingual competence of the speaker, from cases of mixing which are part of the abstract linguistic code shared by the speech community. In the case of Paraguayan Guarani there are a high percentage of bilinguals and speakers who have at least some competence in Spanish. Stability measures should target: i) intra-speaker and inter-speaker consistency in language choices; and (ii) inter-generational stability. The former refer to measures assessing the degree of lexical and structural variability within the speech of a single individual (intra-speaker measures), and between the speech of different individuals (inter-speaker measures). With a lower amount of variation and a higher consistency in language choices there is a higher the probability of a stabilized mixed code. An

example would be to measure the use of the two types of possessive structures reported in examples (5) and (6) in order to assess whether the structure in (6), calqued from Spanish, is consistently used every time that possession is expressed.

The latter measures explore whether the mixed code is being transmitted to younger generations (see Meakins 2013: 114*ff.* for an assessment of those measures in Gurindji Kriol). Child-directed speech can be in this sense a valuable source of information, since if children are acquiring the mixed code before they acquire Spanish, this would be a clear indicator of its autonomy from Spanish. To my knowledge, this type of studies has barely targeted Paraguayan Guarani until now.

Summing up, according to criteria (a)–(f), the available evidence does not allow to conclude that Paraguayan Guarani is a mixed language. Starting by the basic criterion of the proportion of borrowings, available sources would point rather to a language with extreme borrowing than to a mixed language with a dominant lexifying source. Given the diglossic relation between Spanish and Paraguayan Guarani, it is possible that the proportion of borrowings increases in corpora referring to certain semantic domains and in corpora produced in an informal communicative context. In any case, measures of consistency across speakers would be required to assess the stability of the mixture for these semantic domains and communicative contexts. Without further research, referring to Paraguayan Guarani as a *third language* cannot possibly be anything else than a metaphorical expression denoting the extreme lexical influence of Spanish on Guarani.

2. Spanish-Guarani bilingualism: current and past

Paraguay has been typically considered among (socio)linguists to present a unique situation in the linguistic landscape of the Americas. Indeed, most of the country's population is allegedly bilingual in both Spanish, the language of the colonizers, and Guarani, the language of the original inhabitants of the area. The Paraguayan state also proudly highlights the constitutionally recognized co-officiality of Spanish and Guarani as well as the current efforts for implementing the Law of Languages (such as the creation of a Secretariat of Languages, the creation of the Academy of the Guarani Language and the development of a national plan of languages). However, critical voices denounce that the authorities might be taking for granted Paraguayan bilingualism and thus might be failing to take effective action to ensure its continuity (Melià 2003: 44; Zajícová 2009). This, they claim, might lead in the long run to the crystallization of incipient processes of language shift towards Spanish derived from ongoing socioeconomic changes (such as the increase of urbanization and the improvement of access to education). In the light of these considerations, the Paraguayan *linguistic uniqueness* should be revisited, trying to understand what the actual extent of its bilingualism is, as well as its distribution across regions and across socioeconomic classes. These issues will be discussed in section 2.1.

Section 2.2 introduces a historical perspective on the current situation of Paraguayan Guarani, which seems to run counter to typical processes of native language decline and language shift after the Spanish conquest. Language continuity in Paraguay can be compared to similar processes in other former Spanish colonies, such as Maya in the Yucatan Peninsula. Contrary to other ethnic groups which have kept their life style, culture and also language thanks to their isolation from the main original settlements of Spanish colonizers[5], the speakers of these languages did establish contact with the colonizers from the early moments of the conquest, and later integrated into colonial society. The question arises as to whether the processes leading to language continuity rather than language decline and shift are parallel in these different contexts, or whether the path followed by Paraguayan Guarani has its own peculiarities.

2.1. Current bilingualism in Paraguay and future prospects

One of the elements that single out Paraguay language-wise is the role that bilingualism plays in the construction of national identity in official discourse. In this sense, bilingualism has been understood as the linguistic correlate of the biological interbreeding on which Paraguayans base their identity (Melià 2013: 59). Over the 20th century, census data seem to confirm high levels of bilingualism in Paraguay (50% of the population). This percentage has remained stable until the census of 2002, where an increase of about 10% is reported. However, despite the high levels of bilingualism, there are still fundamental differences between the urban and the rural population. In this sense, in the 2002 Census (Zajícová 2009), the percentage of households where Guarani is the preferred language is 82,5% in the rural areas, while it is only 42,6% in the urban areas. Spanish is the preferred language in 54,9% of urban households, whereas this is the case in only 8,5% of rural households. The linguistic divide is thus still considerable.

[5] E.g. several ethnic groups living in remote areas of Paraguay, which did not come into contact with Paraguayan society until the XXth century (Melià 2010: 94-95).

In any case, it is important to note, as highlighted by Zajícová (2009: 67), that this increase in the percentage of bilinguals is explained through an increase in the number of former monolingual Guarani speakers having acquired Spanish, and not the opposite. The most recent data available (from the permanent household survey from 2013) seem to confirm the trend observed in previous censuses. Figure 1, which reports the percentage of speakers using one of the languages at home or both, is in this sense illustrative. In the period from 2009 to 2013, the number of speakers using exclusively Guarani at home has decreased from over 40% to 35%. Conversely, the number of speakers using both Guarani and Spanish has increased from 30% to over 35%. It would appear that this share of 5% of speakers has been lost by Guarani in favor of a bilingual language use. The percentage of speakers using exclusively Spanish has remained stable over this period at a constant share of 25%.

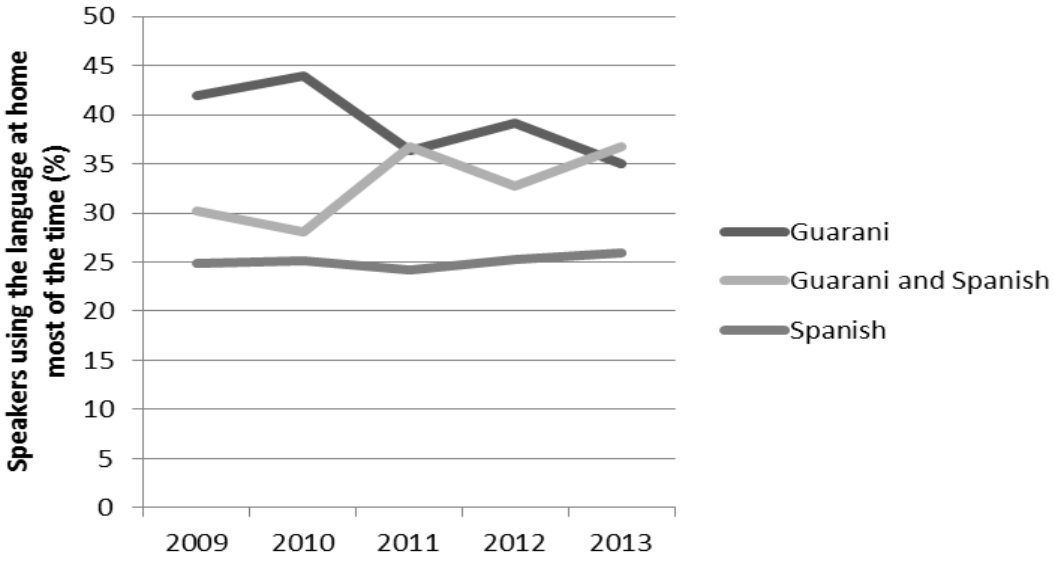

Figure 1. Percentage of speakers using Guarani, Spanish or both most of the time at home. Source (Own elaboration on the basis of data from the Permanent household survey 2013, DGEEC)

Thus, in terms of language continuity, the increase in the number of bilingual speakers could be the intermediate step before language shift. Whether this is the case or whether a situation of stable bilingualism can be foreseen for the near future can only be told from the analysis of further sociolinguistic variables.

Among these sociolinguistic variables, we should consider urbanization trends and access to education. Both variables have been correlated in some cases with shift to dominant language. These factors are for instance mentioned as forces driving the decline of Andean languages (see Pearce and Heggarty 2011: 12). Increased migration to urban centers and changes in the contact between indigenous and non-indigenous peoples have also been identified as potential factors influencing language loss in Guatemala (Yoshioka 2010: 10). In the Mexican Yucatan Peninsula recent internal emigration and urbanization have coincided with an increase of Spanish monolinguals and a decrease of Maya monolinguals (Pfeiler 2014). Similarly to the Paraguayan linguistic situation during the colonial rule, the main language in Yucatan after the

Spanish conquest continued to be Maya. This situation only started to be reversed during the last three decades, due to changes in the economy of the region and to a transformation from a rural to an urban lifestyle. Current bilingualism in this area is interpreted as a transition towards monolingualism, since in many families Maya is no longer transmitted to the younger generations (Pfeiler 2014: 211).

Paraguay has also experienced an increase in urbanization in the last decades. Urban population increased from 35% of the total population in 1950 to almost 60% in 2015 (UN 2015). Rural exodus typically implies that the center of the economic activity shifts towards more technical domains (such as the construction industry) or simply to service activities (commerce), where speakers might feel the urge to borrow more terms from Spanish or they might simply switch to Spanish. This phenomenon has been reported in Yucatan, where nowadays two varieties of Maya coexist: *Hach Maya*, the *pure* variety –mostly spoken by the elders–, and Xe'ek', meaning 'mixture', which is spoken by the younger generations (Pfeiler 2014: 211).

As discussed in section 1.1, this duality between more pure varieties and more mixed ones is also present in Paraguayan Guarani. However, it is an open issue whether the impact of Spanish on Guarani (either in terms of the amount of lexical borrowings or of interference in deeper areas of language structure) has increased over the last decades due to demographic changes. It is an especially controversial question, since it has been suggested that language mixing in Paraguayan Guarani may actually date from the days of the colony (Thun 2005). In this regard, the Paraguayan linguist Domingo Aguilera (personal communication, July 2014) considers that there are two types of influence of Spanish on Guarani. The first one, which he considers to be stabilized and thus homogeneous geographically, is the result of diachronic contact processes between Spanish and Guarani. The second one would be a much more recent phenomenon of the urban areas due to increased contact with Spanish and would be much more instable. Unfortunately, in the absence of oral corpora it is difficult to quantify the impact of this renewed influence of Spanish on urban Guarani. Alternative methodologies, such as documenting the language histories of current Guarani speakers could maybe provide insights in this domain[6].

The second socioeconomic change that may have an effect on Guarani is the improvement of access to education. Rural areas, traditionally deprived from educational facilities, witnessed a dramatic increase of student enrollment in primary school during the 70s (Kluck 1990). Despite

[6] Indeed, the perception of Guarani speakers on the differences between current Guarani and the Guarani spoken by their caregivers can constitute valuable evidence as to recent language change. Some of the speakers interviewed for the elicitation of the data analyzed in this thesis actually mentioned that their caregivers (grandparents) used a variety of Guarani lexically less influenced by Spanish. They also mentioned that the caregivers (who had moved to the capital from their original rural area) would usually try to introduce more elements of Spanish origin to make themselves understood by the children. This would suggest that an increase of Spanish borrowings might have occurred during the socioeconomic transformations of the 20th century.

the limitations of the current educational system[7], the improvement of access to education in Paraguay in the last decades may be correlated to an improvement of the knowledge of Spanish among the rural population. This is probably the factor that explains the previously mentioned increase of bilinguals in the 2002 Census. Paradoxically, if this is not accompanied by effective bilingual programs it may lead in the long term to a decline of Guarani, given the socioeconomic prestige linked to Spanish and the tendency to migrate to urban centers. In this respect it is worth highlighting that existing programs of bilingual education in Paraguay have received extensive criticism (Choi 2004, Mortimer 2006, Caballero 2008). On top of the limitations of current bilingual programs, it should be highlighted that profound socioeconomic inequalities may be at the heart of the failure of bilingual policies. Indeed, Guarani speaking communities in rural areas seem to be those enduring the worst socioeconomic conditions. Rural communities do not only suffer from extremely high levels of poverty (51% of the rural population) but also lack access to fundamental services such as sanitation, electricity and healthcare (Correa *et al.* 2007). Chances of improving the economic conditions of the household seem to be conditioned by the knowledge of Spanish. Within the rural labor market, Guarani monolinguals have a more difficult access to non-farm employment, a sector known to contribute to the mitigation of poverty in rural households, since it provides a stable source of income. In this sense, the percentage of Guarani monolinguals is 82% among the farm workforce, whereas it is only 56,1% among the workforce employed in non-farm activities (*Ib.*). Knowledge of Spanish can thus easily be seen as a commodity allowing for upward social mobility and lead to the failure of current efforts to improve attitudes towards Guarani. This is maybe not surprising, since the affirmation of linguistic rights without the resolution of socioeconomic inequalities is rarely effective[8].

However, there is also a particular trait of Paraguayan bilingualism that may contribute to the continuity of Guarani: the diglossic distribution of Guarani and Spanish. Indeed, whereas so far Spanish seems to monopolize institutional and technical domains, Guarani seems to keep a distinct role in informal contexts, either in humorous uses or as a marker of intimacy or affection. This distribution of functional domains between Spanish and Guarani creates spheres where the two languages are not competitors. This could play a crucial role in ensuring the continuation of a stable diglossic bilingualism (see Melià 2013: 63-64 about the concept of bilingualism with diglossia in Paraguay).

Regarding language choice, in the last decades it seems indeed that this diglossic distribution based on the formality of the situation has gained terrain over the dimension of geographic location. Indeed, if in the beginning of the 60s language choice was highly correlated with the dimension rural *vs.* urban, at the end of the same decade degree of formality was already playing a role in language choice in rural areas. These are the findings of the studies carried out by Rubin in 1960 and in 1965-1967, respectively (Rubin 1985). The author points to urbanization, to the improvement of communications between urban and rural areas and to the introduction of the transistor radio in rural areas as explaining factors of this change. These

[7] Among them the following should be highlighted: shortage of adequately trained teachers (PREAL 2013: 24), the still high percentage of out-of-school children (18,8% in 2012 according to the Unesco Institute for Statistics database) and of dropout in primary education (15,8% in 2011 according to the Unesco Institute for Statistics database).

[8] See Hamel (1994) on neoliberal multiculturalism.

elements probably increased the exposition to Spanish as well as the opportunities to use it. Furthermore, younger generations probably improved their knowledge of Spanish, since they started to have access to higher level schools in the neighboring cities. In spite of the potential role of diglossia in ensuring the continuation of bilingualism, we should also be aware of its consequences on the referential power of the involved languages. Indeed, as Guarani becomes more and more confined to informal settings marked with humorous or affective values, it may as well gradually lose expressive power in other semantic and pragmatic domains. It is in this sense symptomatic that the urban population needs to resort to Spanish loans more often than the rural population[9]. This is consistent with the fact that the diglossic distribution of Spanish and Guarani is older in the capital.

Data from the 2002 Census seemed to indicate that Guarani did indeed keep a distinct role in the privacy of the household, since in 60% of Paraguayan households the preferred language for communication was still Guarani. But this tendency is disproved by more recent data from 2013 (see Figure 1) showing that over 35% of speakers prefer the use of both languages at home[10].

Beyond the diglossic distribution of Spanish and Guarani, other factors may in the long run play in favor of the continuity of Paraguayan Guarani. Among them, one of the most crucial may be the ongoing efforts of the government to standardize and normalize Paraguayan Guarani. The Law of Languages (Law 4251), enacted in 2010, established the creation of the Secretariat of Languages, a permanent institution in charge of language planning. In August 2014 the Secretariat was working on the definition of a "National Plan of Languages" determining the official uses of Guarani and Spanish as well as the steps to take in order to change negative attitudes towards Guarani. The Law attributes the competence of defining the linguistic norm of Guarani to the Academy of the Guarani Language, a private entity partially funded by the state. The norm defined by the Academy will guide the language practices in the public administration and it will inform as well the production of didactic materials by the Ministry of Education. The creation of an agreed standard may contribute to the stabilization of the language and to its use in bureaucratic and professional domains.

There is thus still room to believe that forces towards language shift can be countervailed, provided that the initiatives in the domain of language policy are met by social and economic reforms empowering Guarani speaking communities.

2.2. History of Guarani-Spanish contact

If there is something unique about the linguistic landscape of Paraguay it is the extensive knowledge of an indigenous language by a population that is no longer indigenous. Clearly this

[9] This was observed during the linguistic interviews I carried out in urban and rural areas: in an exercise of picture naming, the participants from Asunción used more often Spanish loans to name animals (i.e. for 'butterfly' they used *mariposa* instead of *panambi*; and for 'turtle' they used *tortuga* instead of *karumbe*).

[10] Since the 2002 Census did not allow more than one answer to the question 'Which language do the members of the household usually use for communication among themselves?' (Zajícová 2009: 63), it is also possible that the relevance of bilingualism in the household went unnoticed.

is not considered to be the default situation in Latin America, where native languages have remained mostly constrained to the ethnic groups that originally spoke them. The key for understanding such a situation lies in the historical processes that shaped language practices in early Paraguayan colonial society. As noted by historians, despite the central role played by Spanish as an administrative language, the population remained mostly monolingual in Guarani. Several factors have been suggested in the literature as potential causes for such a situation, among which: (i) the demography of the region; (ii) limited access to education and to cultural products in Spanish; (iii) the language policy of the Jesuit missionaries in the so-called Jesuit *reductions*.

The demography of the colonial province is indeed one of the most salient factors for an understanding of the linguistic history of Paraguay. The number of Spanish colonizers in the area was always low and the shortage of women among the early Spanish colonizers led to frequent intermarriage and concubinage between Spanish men and Guarani women (Service 1971), so that from the very beginning Guarani must have been part of daily life in the household. The crucial role of Guarani becomes even more plausible if we take into account descriptions of family organization in the first decades after contact. Indeed, it was not rare among Spanish colonizers to establish concubinage with several women; so one can imagine the central role taken by Guarani speaking women in family life. In other Spanish colonies in Latin America, the later increase of female immigration from Spain in the 17th century reduced the rate of interbreeding, since Spanish men opted for marrying Spanish women (Socolow 2000), a fact that must have had an impact on cultural as well as on language practices, renewing Spanish influence on family structure. However, the flow of feminine migration to Paraguay did not increase[11] and, therefore, it can be assumed that there was no counterbalancing factor to the use of Guarani in the family.

Another factor to take into account is the relative homogeneity of the population in the area before Spanish contact (Rubin 1985: 113). The cultural and linguistic homogeneity of the native peoples which entered into socioeconomic relations with the Spanish colonizers probably ensured an advantaged position for Guarani, which was not in direct competition with other native languages in colonial society. Historical sources do refer to the existence of dialectal variation in the Guarani spoken in pre-contact societies, but it is still unknown whether processes of dialect levelling took place in the new social environments created after the Spanish settlement. The situation was probably different in the various settlements of the province, which were of three types: Spanish villages, indigenous villages (either under the control of the Franciscans or under the control of clergymen), and Jesuit reductions.

We know little about the language practices in the first two types of settlement, but it seems clear from the literature that the language policy of the Spanish Crown, which established the need of promoting the learning of Spanish by the indigenous peoples, remained widely unapplied in Paraguay. On the one hand, Spanish colonizers may have been reluctant to empower indigenous peoples with the knowledge of Spanish, as noted by Heath (1972) with regard to Mexico. On the other hand, the means for promoting the knowledge of Spanish were

[11] As noted by Melià (1986: 239), migration to Paraguay had almost completely ceased by the end of the 16th century.

definitely scarce. Indeed, the resources for education in Spanish were limited: access to schools and even cultural products in Spanish, such as books or theater plays, was limited in Paraguayan colonial society. Spanish was an almost exclusive commodity of the upper class, which also run the colonial administration and obviously used Spanish in the communication with the metropolis (Melià 1986: 239). However, unlike other colonial societies, given the poverty of the country and the lack of migration from the metropolis, a clear-cut elite class defined by language, education and economic position did not develop in Paraguay (Rubin 1985: 114).

As for the lower classes, they did not easily have access to the knowledge of Spanish. This may explain why independently of their ethnic affiliation (Spanish or indigenous), the lower classes[12] were Guarani speakers[13]. This linguistic reality is consistent with the fact that the lower classes, either Spanish (including mestizos) or indigenous, shared a rather homogeneous culture (Telesca 2009). Language was thus highly correlated with social class rather than with ethnicity already in the early colony.

Much more is known about language practices in the Jesuit reductions. The linguistic policy of the Jesuits enhanced the maintenance of Guarani in the reductions and avoided the teaching of Spanish. Guarani was the language used in the administration of the reductions, but it was also the language used for evangelization and education. Through the Jesuits, thus, Guarani acquired a literary tradition unknown before. There was also an effort to create neologisms, so that borrowing from Spanish was not required to name the changing cultural reality of the Indians in the reductions. As to the absence of instruction in Spanish, this may have been related to the social and cultural project of the Jesuits: by not teaching Spanish to the Indians of the Reductions the Jesuits avoided their contact with the Spanish of the colonial society, and with the vices related to them, which might have been a negative influence on the evangelization project of the Jesuits (Melià 1986: 128). Contact with the Spanish language was very limited, since whites or mestizos were not allowed in the reductions. The fact that the Indians of the reductions did not learn Spanish may also have contributed to the continuation of the predominance of Guarani after the expulsion of the Jesuit missionaries, when the Indians of the reductions mixed with colonial society. One should not forget, however, that the lower classes of this society were already made up by Guarani monolinguals, although data allowing to tell the differences between the Guarani spoken in the reductions and the Guarani spoken in colonial villages is lacking. In this sense, the impact of the Guarani spoken in the Jesuit reductions on Paraguayan Guarani still remains unknown.

[12] In this sense, however, it should be noted that the 'Spanish' category also included mestizos (Melià 1986: 239 and Telesca 2009: 169-170; 198; 202-206); so from census data it is difficult to differentiate creoles from mestizos. It looks like from the beginning of colonial society there was a type of social ethnicity rather than a biological ethnicity.

[13] This does not seem to be an exclusive situation of Paraguay, since the use of the native language not only by mestizos but also by the lower Spanish classes has been reported in other colonial societies, such as in Yucatán (Gabbert 2004).

3. Empirical study

3.1. Phonology and phonetics of Paraguayan Guarani

3.1.1. Phonology

There are 12 vowels in the Guarani phonemic inventory. Six of them are oral vowels and the rest are their nasal counterparts. Three positions of tongue-advancement are distinguished, as well as two contrasting heights. According to Gregores and Suárez (1967: 49) it may even be possible to speak of a three-way contrast in the tongue-height dimension, since the low vowel /a/ has higher variants in certain contexts.

		front	*central*	*back*
	high	i	ɨ	u
oral	*mid*	e		o
	low		a	
nasal	*high*	ĩ	ɨ̃	ũ
	mid	ẽ		õ
	low		ã	

Table 3. Paraguayan Guarani vowel inventory

As for the consonant inventory, a decision to be made is whether the sounds borrowed from Spanish should be included. A purist perspective would certainly exclude them and limit the inventory to the phonemes of pre-contact Guarani. However, as discussed in section 1, mixing practices are very common in current Paraguayan Guarani and Spanish loanwords are not adapted any more. Excluding these phonemes provides thus a limited perspective of the actual inventory used by the speakers. This way, some works list both sets of phonemes, although separately. This is the strategy followed by Gregores and Suárez who analyze separately the phonology of pre-contact Guarani (Gregores and Suárez 1967: 80 ff.) and the phonology of a corpus with frequent Spanish loanwords not adapted phonologically (Gregores and Suárez 1967: 91 ff.). Conversely, Krivoshein de Canese and Corvalán (1987) and Tonhauser (2006: 129), include only the phonemes of Guarani prior to Spanish influence.

Table 4 includes the phonemes derived from Spanish against a shaded background. With regard to the phoneme /l/, which has a low frequency of occurrence in the language, it is not clear whether it was introduced by Spanish or by some other indigenous language. Indeed, according to Gregores and Suárez (1967: 89), some of the words in which it appears are not Spanish loans nor do they have an unambiguous Guarani etymology (e.g. *lemu* ‘bug’; *lepiju* ‘old man’).

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Labiovelar	Glottal
Plosive/ Nasal	p ^{m}b|m			t ^{n}d|n			k ^{ŋ}g| ŋ	k^{w} ^{ŋ}g^{w}| ŋ^{w}	ʔ
Affricate /Nasal						d͡ʒ |ɲ			
Trill				r					
Tap				ɾ| ɾ̃					
Fricative		f v|ṽ	ð	s	ʃ		ɣ|ɣ̃	ɣ^{w}|ɣ̃^{w}	h
Affricate					t͡ʃ				
Lateral				l | l̃		ʎ			

Table 4. Paraguayan Guarani consonant inventory

A subset of the sounds reported in Table 4 are in complementary distribution. Some of them appear only in oral contexts and have a nasal counterpart that appears in nasal contexts. Oral and nasal allophones are presented in pairs in Table 5 (e.g. ^{m}b|m). A more detailed description of processes of nasalization is provided below.

Previous works are not completely consistent in their inclusion of these pairs of allophones in the consonant inventory. Gregores and Suárez (1967: 51), for instance, do not include the prenasalized stops as phonemes, since they are considered allophones of the nasal phonemes /m/, /n/, /ŋ/. According to these authors the prenasalized stops are just allophones with an oral release used in non-nasal contexts (Gregores and Suárez 1967: 49, 67). However, when it comes to the pair /j - ɲ/, Gregores and Suárez include the oral allophone, namely /j/, as a phoneme, instead of the nasal phoneme /ɲ/[14]. To be consistent, they would have had to leave out /j/ from the phoneme inventory as an oral allophone of the nasal /ɲ/. On the other hand, Tonhauser (2006: 129) includes the oral and the nasal allophones for the pairs /^{m}b, m/, /^{n}d, n/ and /j, ɲ/, but not for the pair /^{ŋ}g, ŋ/. Indeed, in this last case, only the oral allophone, namely, the prenasalized stop /^{ŋ}g/, is included. Only Walker (1999: 68) includes all the allophones in the phoneme chart.

The oral allophones cannot be excluded from the phoneme inventory, since they are the ones postulated in the underlying form (see (11)). However, the nasal allophones are also underlying phonemes in some cases (see (12)), so they have to be included as well. This decision is reflected in Table 4, which follows the approach suggested by Walker (1999).

(11) /kujã/ → [kũ'ɲã] 'woman'(Tonhauser 2006: 129)
(12) /pohano/→ [pohã'no] 'medicate'

Apart from the treatment of the pairs of nasal and oral allophones, several discrepancies can be noted in the literature regarding the Guarani phoneme inventory.

Firstly, whereas some works include the palatal approximant /j/ in the inventory, some other works include the palatal affricate /d͡ʒ/ (e.g. Tonhauser 2006). These two sounds seem to be allophones which occur in free variation.

[14] /j/ is transcribed by the authors as /y/.

Secondly, Gregores and Suárez (1967) and Walker (1999: 183) include three additional labialized phonemes, namely, /kʷ/, /ŋʷ/ and /ɣʷ/, which are missing from the inventories suggested by Tonhauser (2006) and by Krivoshein de Canese and Corvalán (1982). These three phonemes are included in Table 4. Following Gregores and Suárez, the labial-velar approximant /w/ is not included as a phoneme in its own right, given the limited contexts in which it occurs, namely, only after /k/, /ŋ/ or /ɣ/. Other authors consider that /w/ does occur independently, as a dialectal variant of the labialized fricative /ɣʷ/ (Kaiser 2008: 286).

A further difference between the various inventories proposed in the literature has to do with the specification of the constriction location for the posterior fricative. Gregores and Suárez (1967: 51) consider it to be a velar fricative (/x/), whereas Krivoshein de Canese and Corvalán (1987) and Tonhauser (2006: 129) classify it as glottal (/h/). In Gregores and Suárez (1967: 81) the glottal fricative (/h/) is considered as an allophone which occurs in free variation with /x/. On the basis of my recordings, I have only included as a phoneme the glottal fricative.

Some didactic materials of Paraguayan Guarani include an additional phoneme, namely, the prenasalized voiceless stop /ⁿt/ (e.g. Krivoshein de Canese and Acosta Alcaraz 2007: 18; Centurión Cervin and Davalos Arce 2009: 9). However, this choice is controversial, given the very limited distribution of this sound, which occurs exclusively in the suffix *–nte* 'only'. Besides that, according to the rule that foresees the resyllabification of onset clusters (Kaiser 2008: 191), the group /ⁿt/ is expected to split into two sounds when it occurs in medial position: a nasal belonging to the coda of the preceding syllable, and a stop acting as the onset of the following syllable (see (13)). If one adds to these considerations the cross-linguistic markedness of prenasalized voiceless sounds vis-à-vis prenasalized voiced sounds (Herbert 1986: 67), there are reasons to believe that /ⁿt/ is just a sequence of phonemes. In any case, a definite answer will only be possible by taking into consideration phonetic measures such as timing and relative duration (Herbert 1986: 10). If the duration of the group /ⁿt/ approximates the duration of individual consonants in the language, then there would be ground to consider its status as a single phoneme.

(13) chénte → [chen.te] 'only me'

With regard to the phonological processes of Guarani, nasal harmony is without doubt the most characteristic one. A basic unit of analysis in the description of nasal harmony in Guarani is the domain of nasalization, termed *nasal span* by Gregores and Suárez (1967: 67). Within the nasal span, it is possible to predict the nasalization of phonemes. Given its predictability, this nasality will not be considered phonemic. In general terms, it can be said that there is nasality when the word contains a nasal consonant or a phonemic nasal vowel and that the nasal span extends to the whole word (Kaiser 2008: 290). An example of allophonic variation due to nasality is presented in (14) (examples come from Kaiser 2008: 291). In (14), since there is no nasal consonant or phonemic nasal vowel, the reflexive prefix surfaces as /d͡ʒe/. Conversely, in ((14) it surfaces as /ɲe/, due to the presence of /ũ/.

(14) a) /o- d͡ʒe- prepara/ → [o. d͡ʒe.pre.pa'ra]
3 -REFL- prepare
'He prepares himself.'

b) /o-d͡ʒe- pɨrũ/ → [õ. ɲẽ.p ɾ̃.' ɾ̃ũ]
3 -REFL- step
'He takes himself.'

It should be noted that nasality spreads from /ũ/ not only to /d͡ʒ/ making it surface as /ɲ/, but also to all the vowels in the word and to the continuant consonant /r/, which is realized with a certain amount of nasality. The only phoneme that is not affected is the voiceless stop /p/. Indeed, only the voiceless phonemes (/p, t, k, s, ʃ, ʔ/) are transparent to nasality, namely, they allow the nasal feature to spread to neighboring phonemes without being affected by it. This is why in Table 5 no nasal allophones are associated with these phonemes. Sonorant consonants such as /h/, /r/ or /v/ are affected, and are thus realized as the nasal allophones /h̃/, /r̃/ and /ṽ/, respectively. This type of spread is named *long-distance continuous nasal spreading* by Kaiser (2008: 292). It is only triggered by phonemically nasal vowels and it affects all the segments in the word except the voiceless segments.

Another type of nasal spread identified by Kaiser (2008: 292) is the so-called *discontinuous nasal spreading*, which is triggered by nasal consonants and which only affects the prenasalized stops /^{m}b/, /^{n}d/, /^{ŋ}g/, the affricate /d͡ʒ/, and the vowels adjacent to the nasal consonant. In example (15), from Kaiser (2008: 291), the rightmost prenasalized stop (/^{n}d/ in *hendu*) transmits the nasality feature to the only other prenasalized stop (/^{n}d/ in *ndo*) which is realized as/n/:

(15) /^{n}do ro he^{n}du i / → [no.ro.hẽn.'dui]
NEG 1SG-2SG hear NEG
'I don't hear you.'

The neighboring vowel is also nasalized ([hẽn.'dui]), as a result of *local nasal spread*. This third type of nasal spread is also exemplified in (16), from Kaiser (2008: 292). In this case /n/ causes the preceding vowel to nasalize. All the other phonemes are not affected.

(16) pohanohára → [po.hã.no.'ha.ra] 'doctor'

Nasal spread in Guarani has been shown to operate not only leftward, but also rightward, as shown in example (17), from Kaiser (2008: 296), where the nasality spreads from the nasal vowel at the end of *tĩ* ('nose') to the first syllable of the next morpheme, *ruguy* ('hair'). It should be noted, though, that nasal spread seems to be blocked by the stressed oral vowel at the end of *ruguy*. However, Kaiser (2008: 296) considers that available data are still scarce to tell with certainty whether nasal spread is blocked by the oral vowel, or whether rightward spread is limited to just one syllable.

(17) tĩruguy→ [tĩ.ɾ̃ũ.'ɣ^{w}ɨ] [15] 'nose bleed'

[15] In order to ensure the consistency with the phoneme inventory proposed above, I have slightly altered the transcription of the word proposed by Kaiser by using [ɣ^{w}] instead of [w].

Blocking is also observed in leftward nasal spread. See for instance (18), from (Kaiser 2008: 295), where nasality does not affect the initial prefix *o-*. It is believed that initial vowels are opaque to nasal spreading, but Kaiser argues that the morpheme boundary may be the actual cause of the blocking.

(18) o-pyryrỹi → [o.pɨ̃.rɨ̃. 'rɨ̃ĩ]
A3-jump
'He jumps.'

Syllable structure is primarily CV in Guarani, although there are some vowel-only (V) syllables (e.g. *óga* 'house') and some closed syllables (CVC) (Kaiser 2008: 283) (e.g. *tembi'u* 'food'). Consonant clusters have entered the language through non-adapted loans from Spanish (e.g. *kuátro* 'four', *kuádro* 'square'). Stress falls predominantly on the last syllable. Therefore, the official orthography requires accent marks only to signal exceptions to this pattern (e.g. *eíra* 'honey').

3.1.2. Previous works on the phonetics of Paraguayan Guarani

Whereas the morphosyntax of the language and to a certain extent also its phonology have been analyzed from multiple angles, to date the phonetics of Paraguayan Guarani remains one of the least researched areas. Indeed, only a handful of studies provide a phonetic description of the phonemes of the language. Interestingly, one of the pioneer descriptive works of Paraguayan Guarani was a phonetic description of vowels and consonants. The study, by Ramon Caballero, a bilingual speaker from Asunción, was carried out at the Laboratory of Experimental Phonetics of Collège de France, in Paris, under the supervision of Jean-Pierre Rousselot, considered the founder of experimental phonetics. The study was published in 1911 (Caballero 1911) in the *Revue de Phonétique*, a journal devoted to experimental phonetics studies and co-edited by Jean-Pierre Rousselot. On the basis of the techniques available at the time (e.g. the artificial palate), the study describes the articulation of vowels and consonants, taking into account the different articulatory contexts and the possible allophones.

More recent works have focused on the study of phonological processes such as nasal harmony, and have provided acoustic data leading to a deeper understanding of those processes. Walker (1999; 2013: 208 ff.), for instance, studies how the spread of nasality affects voiceless stops, which according to phonological descriptions of the language are supposed to be transparent to flanking nasality. The study confirms that voiceless stops do not show evidence of nasal airflow. However, the nasal context does have an effect on the duration properties of the voiceless stops. More concretely, in a nasal context (e.g. [ãkỹ] 'wet') they show a longer voicing time during closure and a longer VOT than in an oral context (e.g. [aky] 'green, unripe').

Price and Stewart (2013) zoom in on another aspect of nasal spread in Paraguayan Guarani, namely, its gradual nature. By measuring the degree of nasality of segments at different distances from the nasal sound, the authors prove both that nasality is anticipatory and that it builds up gradually as the nasal sound approaches. Four types of segments were measured in

this study, namely vowels, voiceless stops, prenasalized stops and fricatives. Distance from the nasal sound was found to be a significant predictor of the degree of nasality of vowels, voiceless stops and prenasalized stops. More concretely, according to this study, phonetically the nasal span does not go beyond two syllables from the nasal sound. Indeed, in the third syllable from the nasal sound the degree of nasality was found to be not significantly different from an oral span. It would thus appear that some segments which are considered phonologically nasal, such as the syllable *ku-* in example (19), do not display the phonetic properties of nasality any more than oral segments.

(19) /kurapepẽ / → [kũr̃apẽpẽ] 'pumpkin'

Another set of studies has taken as a main point of interest Paraguayan bilingualism and its effects on the phonetics of the language. This way, Johnston (2007) compared the production of vowels and voiceless stops (/p, t, k/) in Guarani, in Spanish[16] and in Jopara[17] by two groups of bilinguals. One of the groups was primarily Guarani speaking, whereas the other group was identified as having Spanish as its primary language. The study did not find significant differences between the Spanish and the Guarani vowels, neither between the productions of the two groups (Johnston 2007: 15). Conversely, VOT ratio measures of voiceless stops revealed significant differences between Guarani subjects and Spanish subjects (Johnston 2007: 16-17). Primarily Guarani speaking subjects presented longer VOT than the primarily Spanish speaking subjects in both languages and in all three stops, except for Jopara /t/ which was produced with a shorter VOT by the Guarani speakers.

Both groups of speakers were found to adjust their VOT when speaking in their non-primary language. This way, Spanish speakers, produced a considerably longer VOT in Guarani than in Spanish, going from a mean VOT ratio in Guarani of 7.2s, 7.8s and 13.5s for /p/, /t/ and /k/, respectively, to 8.4s, 12.8s and 16.4s in Guarani. The Spanish speakers increased their VOT also in Jopara, but without reaching the average VOT produced by Guarani subjects. In the case of Guarani speakers, they also adjusted their VOT when speaking in Spanish and in Jopara. In this case, however, they seemed to compensate in the wrong direction, since the VOT in their production in Spanish was longer than in their production in Guarani. As for their production in Jopara, Guarani subjects again increased heir VOT, except for /t/ where there was a decrease.

The study by Johnston thus provides evidence for the existence of a certain phonetic variation in the production of segments by different types of bilinguals.

In a similarly comparative vein, Jiquilin-Ramirez and Albano (2014) contrast the co-articulatory patterns of vowels and posterior fricatives in Spanish and Guarani as produced by bilingual speakers. Measures of fricatives included the four first spectral moments, namely, the center of gravity, the standard deviation, the skewness and the kurtosis. Cross-linguistic differences were expected since, according to Krivoshein de Canese and Corvalán (1987), the back fricative of Paraguayan Guarani is glottal, whereas the back fricative of Spanish is velar. Within-language differences were also expected, due to the sensitivity of the constriction location of

[16] Of course the set of vowels elicited in Guarani and Jopara was larger than in Spanish, since it included the nasal vowels absent from the Spanish phonemic inventory.

[17] Johnston considers Jopara as referring to mixed words where a Spanish root has been inflected with Guarani morphology.

fricatives to vowel context. Both hypotheses were confirmed. Firstly, the Spanish fricative was found to have a significantly higher center of gravity in the context of /i/ than the Guarani fricative. These results would be compatible with a fronter articulation of the Spanish fricative in the context of /i/, since it has been shown that there is a correlation between frontness of the constriction and center of gravity due to the smaller cavity in front of the constriction (Gordon *et al.* 2002). In the rest of vowel contexts the difference was not significant. Secondly, within both languages the measures of the back fricative differed significantly between the vowel contexts /i/ and /u/. Thus, the constriction location of the fricative seems to be affected by surrounding vowels in a similar way in both languages.

3.2. Hypothesis: acoustic differences related to different bilingual profiles

As shown in section 2, even if the literature does not acknowledge diatopic variation in Paraguayan Guarani, it is well-known that there is a diversity of bilingual profiles. These profiles are usually correlated with geographic location. More concretely, Guarani monolinguals or incipient bilinguals with a passive knowledge of Spanish are in general located in rural areas, whereas Spanish monolinguals or Spanish-Guarani bilinguals with Spanish as L1 are mostly located in the urban areas. Factors such as the L2 status of Guarani in the urban areas, its lower usage and its acquisition at a later age, make it reasonable to expect acoustic differences in the production by urban speakers of Guarani sounds which are not part of the Spanish inventory.

In the following I describe the literature that supports that prediction and I conclude with the formulation of the specific hypotheses to be tested in this study.

3.2.1. Factors impacting native-like production of L2 phonetic categories

The literature on the phonetics of bilingual speakers (e.g. Flege *et al.* 2003) predicts that the phonetic categories produced by L2 speakers will differ from those produced by L1 speakers as a function of several factors. Among them three types can be especially highlighted:

(i) factors related to age and to the state of development of the L1 when the L2 started to be learned;
(ii) factors related to the environment during L2 acquisition (i.e. exposure to and usage of the L2 and the L1);
(iii) factors related to the interaction between the L1 and the L2 phonemic subsystems.

Regarding the first set of factors, several studies have reported a correlation between age at onset of L2 acquisition and proficiency in L2 phonetic categories. Age at onset of L2 acquisition is frequently indexed by the time of arrival to the host country. In a study of native Italians having migrated to Canada, Munro *et al.* (1996) observed that the perceived foreign accent in the production of 11 English vowels became higher as age of arrival increased. Similarly Piske *et al.* (2001) found that early Italian-English bilinguals received significantly lower scores for accentedness in the production of English sentences than late bilinguals. These findings seem to confirm the so-called Critical Period Hypothesis, according to which native-like proficiency in

the L2 is no longer possible after the critical period has ended due to the loss of neural plasticity as the L1 matures (e.g. Lenneberg 1967). There is a discrepancy in the literature as to when the critical period ends. Scovel (1988) placed it at 12 years, Patkowski (1990) at 15 years and Long (1990) considered that the acquisition of an L2 before the age of 6 was a good predictor for the absence of a foreign accent. However, several studies cast doubt on the existence of a clear-cut critical period for L2 acquisition. In this sense, for instance, in the study by Flege, Frieda and Nozawa (1997) early Italian-English bilinguals with a mean age of onset of acquisition of 5.8 years were found to have a perceivable foreign accent. Conversely, evidence has also been reported that it is possible to acquire a free-accented L2 speech when age at onset of acquisition is above 12 years. In this sense, Bongaerts *et al.* (1997) found that among a group of native Dutch speakers having acquired English after age 12, some individuals obtained scores similar to the control group of native English speakers. One potential explanation for the inconsistency of these findings with the critical period hypothesis is the high correlation of age of acquisition with other variables, such as the amount and quality of L2 input (Flege and Mackay 2011). From this point of view, age of acquisition may be seen as a macro-variable (*Ib.*) that co-varies with factors from the second group, namely, with the exposure to the L2 and its usage. Indeed, speakers with a lower age of acquisition have in most cases probably been exposed to the L2 for a longer period than speakers with a higher age of acquisition. Furthermore, in a migration context, individuals with a lower age of acquisition may have benefited from an insertion in particular social settings enabling more frequent and intense interactions in the L2. For instance, they may have been schooled in the L2 and they may even have married a native speaker of the L2 (Flege *et al.* 2003: 468).

Several studies have confirmed the impact of environment variables on ultimate L2 proficiency. Flege, Bohn and Jang (1997), for instance, found that English-language experience (measured as length of residence in the United States) was positively correlated with accuracy in the production of English vowels by non-native speakers. Conversely, the study by Flege, Frieda and Nozawa (1997) revealed that the degree of activation of the L1 may have a negative effect on the accuracy in the production of the L2 sounds. The authors compared two groups of native Italian speakers who were matched for length of residence in Canada (34 years on average) and found that those who reported a frequent use of Italian received significantly higher foreign accent ratings than those who used Italian seldom.

As for the third set of factors, they play a crucial role in the Speech Learning Model (Flege 1995, Flege *et al.* 2003), an explanatory framework conceived to account for the different mechanisms that may lead to non-native pronunciation in L2 speakers. One of the main assumptions of the model is that the phonetic subsystems of the L1 and the L2 coexist in a shared phonological space (Flege 1995: 239) and therefore can have an influence on each other. The model puts a special emphasis on the role that perception has on the ability to produce L2 sounds in a native-like fashion. This way, it predicts two main scenarios leading to non-native pronunciation, depending on whether the speaker is able to discern the differences between an L2 sound and an L1 sound which is close to it in the acoustic space, or whether she fails to do so.

The first scenario, also termed *phonetic category assimilation*, refers to the equation of an L2 sound to a pre-existing L1 phonetic category despite the existence of acoustic differences

between the two categories. This will usually be the case when the speaker fails to perceive those differences and is thus unable to reproduce them. Similarly, a speaker may be unable to perceive the differences between two very similar L2 sounds and thus map them onto a single L1 sound. The failure to perceive these particular L2 sounds will lead to the inability to accurately produce them and will unavoidably block category formation. In contrast, when the speaker is able to perceive the differences between the new L2 sound and the closest L1 sound, a new phonetic category is more likely to be created (Flege *et al.* 2003). Therefore, sound similarity is not the main determinant of accurate production of L2 sounds. This contradicts the previously proposed contrastive analysis hypothesis (e.g. Lado 1957), according to which L2 sounds similar to L1 sounds would be produced more easily, whereas L2 sounds different from L1 sounds would be produced with more difficulty. Phonetic category assimilation is also in consonance with some studies having reported better performance in the production of L2 sounds which were completely different from pre-existing L1 categories, than in the production of L2 sounds which were similar to L1 sounds. Flege (1987), for instance, found that native English speakers produced French /u/ significantly different from native speakers (they produced higher F2 values than French monolinguals). However, their production of /y/ was not significantly different from that of native speakers. This is attributed to a possible interference of the pre-existing English /u/, whereas no previous L1 vowel was similar enough to French /y/ to cause category assimilation.

The second scenario, known as *phonetic category dissimilation*, occurs when a new category has been established for an L2 sound that is close to an L1 sound in the acoustic space. In this case, the speaker is aware of the difference between the two sounds, and will accordingly try to keep them separate by producing an exaggerated version of the L2 sound. Thus, by this mechanism, the similar L1 and L2 sounds will drift away in the acoustic space. In this sense, Flege *et al.* (2003) found that early Italian-English bilinguals who used their L1 infrequently produced English /eᴵ/ (as in *bade)* with more tongue movement than native English speakers. This articulatory pattern was explained by the authors as an attempt by the early bilinguals to dissimilate their native Italian /e/ category from the newly created English /eᴵ/. In the same experiment, late bilinguals (both those using Italian frequently and infrequently) were shown to produce less tongue movement than native English speakers. This was attributed to their failure to form a new category for English /eᴵ/, thus assimilating it to the phonetic properties of Italian /e/ (*Ib.*).

3.2.2. Hypotheses about acoustic differences between urban and rural Guarani vowels

The differences between the bilingual profiles of urban and rural speakers of Paraguayan Guarani give reason to expect acoustic differences in the production of Guarani sounds that are not present in the Spanish phonemic inventory.

Firstly, regarding age of acquisition, the general trend in urban settings is towards a later acquisition of Guarani. Urban speakers having Spanish as their L1 typically report having started learning Guarani at school in the interactions with other children in the playground, therefore around the age of 6. By that time the phonetic categories of the L1 must have been already well developed.

Secondly, regarding environmental factors, urban speakers report using Guarani less frequently than Spanish, which is the most used language in daily routine. This is unsurprising, since urban life is dominated by contexts in which Spanish is the default language (e.g. professional milieu, commercial activity, public administration, media such as radio and TV). The amount of input and practice of Guarani is thus lower in an urban setting and thus the influence of Spanish phonetic categories potentially stronger.

The vowel system of Spanish is smaller than that of Guarani, with only five oral vowels. Apart from the nasal vowels, the Spanish vowel inventory does not include the high central vowel /ɨ/. With regard to this new L2 category, urban speakers may display one of the two processes described by the Speech Learning Model:

- phonetic category assimilation: urban speakers will assimilate /ɨ/ to one of the two neighboring Spanish categories, namely, /i/ or /u/, and therefore produce /ɨ/ with very similar F1 and F2 to one of these two vowels.
- phonetic category dissimilation: if they perceive the acoustic differences between /ɨ/ and the neighboring vowels, urban speakers will overcompensate to maintain the distance between Spanish /i/ and /u/ and the Guarani sound /ɨ/. Depending on the whether speakers try to dissimilate /ɨ/ from /i/ or from /u/, /ɨ/ F2 will be lower (if dissimilated from /i/) or higher (if dissimilated from /u/).

Given the fact that most urban speakers have at least some exposure to Guarani at a young age, it is to be expected that they are able to perceive the differences between the Guarani high central vowel and the high vowels of Spanish /i/ and /u/. If this is the case, the probability for the creation of a new category for /ɨ/ is high, according to the Speech Learning Model. Phonetic category dissimilation is thus more likely than phonetic category assimilation.

There is also the possibility that the acoustic characteristics of urban /ɨ/ do not differ significantly from /ɨ/ produced by rural speakers. This would imply that urban speakers manage to prevent their Spanish phonetic subsystem to interfere with their Guarani phonetic categories. This does not mean that urban speakers keep two completely separated phonemic spaces, a possibility that has been rejected by several authors (e.g. Grosjean 1989, 1999). It would rather mean that the acquisition took place at an early age when the interaction between the L2 and the L1 was still not dominated by the phonetic categories of the L1, which become more powerful attractors of L2 sounds as they consolidate through childhood (Flege *et al.* 2003: 469). It would also probably mean that the Guarani input to which children are exposed in their first years of age is not negligible. An additional explanation would be that Guarani /ɨ/ is sufficiently different from Spanish /i/ and /u/ to be categorized as a new category by L2 learners, thus avoiding L1 interference.

3.3. Acoustic measures of urban and rural Guarani

This section has a twofold goal. On the one hand, it aims at providing a general comparison of urban and rural Guarani vowels. On the other hand, it aims at testing the hypothesis that L2 Guarani speakers produce a significantly different /ɨ̃/ phoneme, given that their L1 (Spanish)

lacks this phoneme. On the basis of previous studies on the phonetics of the L2 (see Section 3.2), acoustic differences are to be expected given the different vowel systems of Spanish and Guarani.

3.3.1. Experimental design and dataset

The experimental design was aimed at contrasting the Guarani vowels in different phonemic environments in the speech of urban and rural speakers. The target vowels appear always in stressed position in words following the pattern (C)V'CV, where the last vowel is stressed. From a much larger set of stimuli, only the words listed in Table 5 were included in the final dataset. This is due to the fact that the urban speakers did not know the Guarani translation for several concepts (e.g. *petɨ̃*, 'tobacco'; *akɨ̃*, 'ripe', akaŋɨ̃, 'lost') and only the words produced by both the urban and the rural speakers were retained for the comparison. A consequence of the exclusion of several words is that the phonological contexts are not always balanced across stimuli for different vowels. Formants across vowels will thus not be comparable and it will be impossible to define the vowel space of the language. Instead the analysis will focus on the comparison of formants for each vowel between the urban and the rural group.

a	**e**	**i**	**o**	**u**	**ɨ**
tupa ita pyta juka vaka	oke	avati panambi	poko	hetu juru tanimbu	pɨ kotɨ ɨvotɨ jukɨ hasɨ jasɨ pochɨ
ã	**ẽ**	**ĩ**	**õ**	**ũ**	
tupã mitã mokã	okẽ	morotĩ potĩ tĩ	kasõ merõ	kũ	

Table 5. Elicited items

The recordings were made with a Zoom H4n recorder at a sampling rate of 44.1 Hz, and a head-worn dynamic microphone Shure WH20. In Asunción the recordings were made in a quiet room, usually at the workplace of the participants. In the rural areas recordings were made outdoors, usually outside the house or the chapel that served as school for the adult literacy courses.

The choice of the elicitation method was motivated by several factors. Firstly, many of the participants were enrolled in literacy courses and therefore were still learning how to read. Thus it was not an option to make them read carrier sentences embedding the target items, since this would have yielded unnatural productions. Secondly, even for many of the literate participants Guarani remains mostly an oral language and thus they are not used to reading text in Guarani. Therefore the methodology chosen consisted in picture naming and sentence translation. Each item was produced twice. In the first part of the interview the participants were shown a series of pictures that they had to name using a single word. In the second part of the interview I would read sentences in Spanish that the participant had to translate into

Guarani. The sentences contained the target words. Participants were encouraged to take some time to think about the most natural translation before giving an answer. In both cases they were asked to give their answer twice. Participants were interviewed in pairs, so that when a doubt about the most suitable translation arose they could discuss it among them. By encouraging them to discuss doubts the goal was to elicit more natural productions than if they had just been interacting with me. This is also why in the beginning of the interview I made several questions targeting sociolinguistic information (i.e. age, place of residence, occupation, contexts of language use) in Guarani. In so doing the aim was to mitigate the accommodation of the participants to my Spanish accent.

3.3.2. Participants

During the 6 weeks that my stay in Paraguay lasted, I interviewed a total of 42 speakers, from Asunción and from two rural areas in the Northern Department of Concepción, namely, Horqueta and Yby Ya'u. Horqueta and Yby Ya'u are located at over 400 km from Asunción, and they have a population of over 50.000 and over 20.000, respectively. The participants did not live in these two cities, but in communities located in neighboring rural areas. 18 of the speakers were from Asunción, 11 from Yby Ya'u, 10 from Horqueta, and 3 minors from the correctional institute of Itaugua (at about 30 km from Asunción). All the participants were females, except for the 3 minors of the correctional institute. A sample of 12 female speakers was selected for the analysis carried out in this thesis: 4 speakers from Asunción, 4 from Horqueta and 4 from Yby Ya'u. The selection was made on the basis of the quality of the recordings (in some of the recordings made outdoors there was considerable background noise due to the wind) and the completeness of the interview (some speakers were able to complete only one part of the interview due to time constraints).

The 4 speakers from Asunción were officials at the Ministry of Education, had Spanish as their L1 and this is the language that they use in most contexts of their daily routine, although they do use Guarani in certain environments (i.e. in local shops and markets) and in certain registers (i.e. jokes, emotionally charged speech). Their age ranges from 36 to 48, with a mean age of 42.75.

The 8 speakers from Horqueta and from Yby Ya'u almost exclusively devoted themselves to housework and family care and their partners were occupied in the agricultural sector. Their age ranges from 26 to 74, with a mean age of 48.13. They lived in rural areas some kilometers away from the city and thus did not have regular contact with urban society. All of them were following the second course of the literacy program led by the Paraguayan Ministry of Culture. Their L1 was Guarani and that is also the language that they used in most contexts. However they had all at least a passive understanding of Spanish and interestingly some of them reported having started to speak some Spanish to their grandchildren who were living in the city, or to their children having migrated to neighboring countries such as Argentina. Their active competence in Spanish seemed to be correlated with age (younger speakers were more fluent in Spanish than older speakers). They also had exposure to Spanish through the radio and television.

Speaker	Group	Age	Place of residence	Previous places of residence
a2	Asunción	48	Asunción	-
a9	Asunción	36	Asunción	-
a13	Asunción	49	Luque	Asunción (Chacarita)
a14	Asunción	38	Capiata	Asunción
h1	Horqueta	74	Horqueta, km30	-
h5	Horqueta	29	Horqueta, km31	Arroyito
h6	Horqueta	48	Horqueta, km31	-
h7	Horqueta	49	Horqueta, km31	-
y1	Yby Ya'u	61	Yby Ya'u (Hechapyra)	-
y3	Yby Ya'u	26	Yby Ya'u (Hechapyra)	-
y4	Yby Ya'u	45	Yby Ya'u (Hechapyra)	-
y8	Yby Ya'u	53	Yby Ya'u	-

Table 6. List or participants

3.3.3. Measurements

3.3.3.1. Acoustic landmarks

In order to ensure consistency in the annotation of the target vowel segments, the following acoustic landmarks were used to identify the beginning of the vowel:

- Following a stop: the beginning of the vowel is marked by the start of periodicity in the sound wave. The beginning of voicing is also signaled through the vertical striation in the spectrogram. See Figure 2.

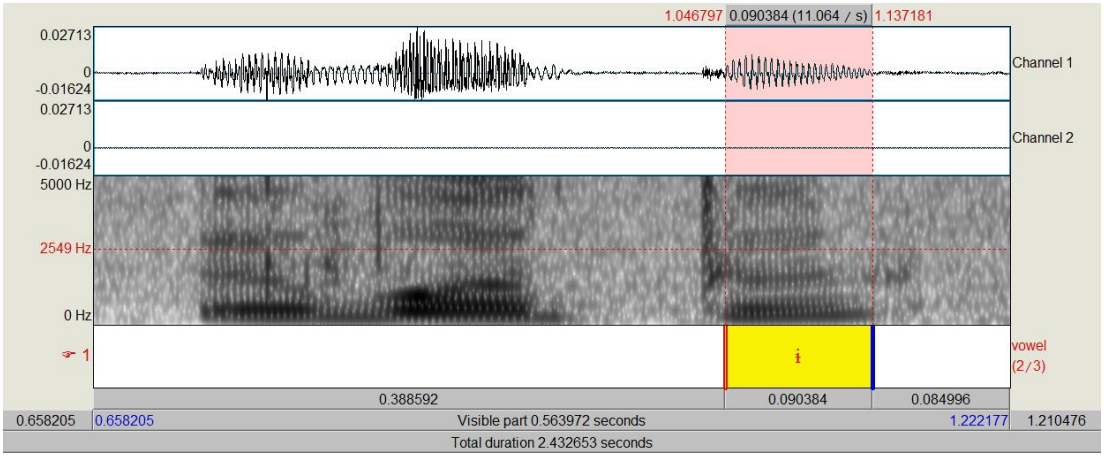

Figure 2. Praat annotation of /ĩ/ in /ɲvotĩ/

- Following a fricative: the beginning of the vowel is marked by the beginning of periodicity in the sound wave and by the presence of formants in the spectrogram after the turbulence of the fricative.

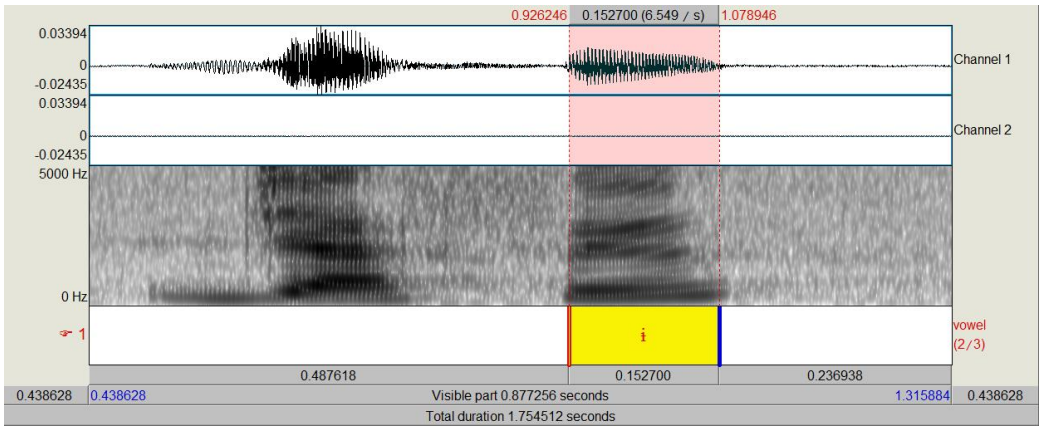

Figure 3. Praat annotation of /ĩ/ in /jasĩ/

- Following a flap: the flap is seen in the spectrogram as a dip between vowels. The vowel begins with the increase in amplitude in the periodic sound wave and with the presence of formants in the spectrogram, especially F2 and F3 which are very weak during the flap.

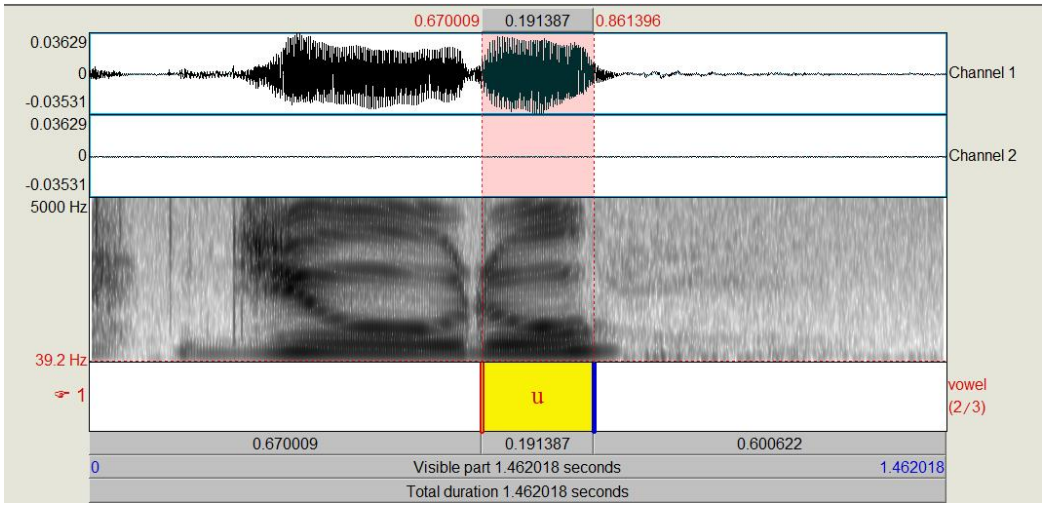

Figure 4. Praat annotation of /u/ in /juru/

The vowel was considered to last until F2 was visible in the spectrogram.

3.3.3.2. Formant measures

The .wav files were segmented with the aid of Audacity and subsequently the target vowels were manually annotated in Praat TextGrids. Formant measures at midpoint were automatically extracted with Praat via LPC (linear predictive coding) with a script having a fixed ceiling identifying the first five formants up to 5500 Hz. Measurements were taken on wide-band Fourrier spectrograms with a Gaussian window shape (window length 5ms, dynamic range 70dB). Midpoint values instead of average measures were chosen because it was considered that midpoint was a more stable measure avoiding vowel onset and offset transitions.

3.3.4. Data analysis

This section describes the two methods that were used to analyze and compare the acoustic measurements between the urban and the rural group: (i) a first exploratory analysis based on descriptive statistics; and (ii) an inferential analysis based on linear mixed-effects models.

3.3.4.1. Exploratory analysis

This section reports the trends observed in formant values for the different vowels and across the two groups of speakers.

3.3.4.1.1. Vowel normalization

Vowel normalization techniques aim at factoring out variation in formant resonances due to physiological differences between the speakers, namely, mouth and vocal fold size. Prior to normalization, formant values are not directly comparable across speakers and therefore it is difficult to explore the main trends in the data.

There are several methods for vowel normalization. One basic distinction has to do with the type of information that is used in the normalization procedure. *Vowel-intrinsic* methods rely only on acoustic information which is contained within a single vowel token to normalize that same vowel token. *Vowel-extrinsic* methods use information that is contained in several vowel tokens produced by the speaker. Similarly, there are *formant-intrinsic* methods, which use information contained in one single formant, and *formant-extrinsic* methods, which use information across several formants.

Adank *et al.*(2004) evaluated several vowel-intrinsic and vowel-extrinsic methods and came to the conclusion that vowel-extrinsic methods performed better at preserving phonemic variation, reducing physiological variation and preserving sociolinguistic variation. They also found that vowel-extrinsic methods that are formant-extrinsic perform worse than methods relying exclusively on data from one formant (Adank *et al.2004*: 3106).

Therefore, in this study formant values were normalized using a vowel-extrinsic method that includes only information within formants, namely Lobanov's (1971) z-score transformation, represented in the following equation:

$$F_{ti}^{N} = \frac{F_{ti} - \mu_{ti}}{\delta_{ti}}$$

where F_{ti}^{N} stands for normalized formant of vowel *i* and speaker *t*; μ_{ti} is the mean formant frequency across all the vowels produced by the speaker, and δ_{ti} is the standard deviation for average μ_{ti}.

Lobanov's z-score transformation yields data that is not in Hertz values and this makes it difficult to relate the normalized values to the actual vowel quality. A scaling algorithm translates back the normalized values to Hertz, so that, when plotted, F1 and F2 are more easily related to actual frequency values. Both normalization and scaling where done through NORM, a web-based interface to the vowels R package, a piece of software designed to normalize, scale and plot vowel formant data (Thomas and Kendall 2007).

3.3.4.1.2. Observed trends

The data in the following plots are based on normalized values. Two measures of centrality are explored: median (shown in boxplots) and mean values (shown in the scatterplot). Figure 5 shows boxplots for each vowel comparing F1 in the urban and the rural group, whereas Figure 6 shows boxplots for each vowel comparing F2 in both groups. Tick marks for the y axis have been set at steps of 100Hz in both figures so that the magnitude of the differences in Hz can be easily inspected visually. The black dot signals the median, namely the number separating the lower half of the data form the higher half. Comparing the median in the urban and the rural group provides a first approximation to the differences between the two groups.

The first observation to make is that differences between the urban and the rural medians seem to be more important in F2 than in F1. Furthermore, given the smaller size of the boxes plotting F1 data, it becomes clear that there is much less variation in F1 values than in F2 values.

Whereas F2 values of /a/, /e/ and /o/ and their nasal counterparts are quite similar, /ɨ/, /ĩ/ and /ũ/ show some differences between both groups. F2 values are lower for /ɨ/ in the urban group, whereas they are higher for /ĩ/, /ũ/. Since F2 is positively correlated with vowel fronting, an initial hypothesis is that /ĩ/ and /ũ/ are more fronted in the urban group, whereas /ɨ/ is more fronted in the rural group.

As for data reported in the scatterplot in Figure 7, /u/ and /ɨ/ appear more fronted in the rural group, whereas urban /i/ has a more fronted position. The scatterplot reveals a trend which went unobserved in the boxplots: urban /e/ appears to have a lower F1. Given the inverse correlation of F1 with vowel height, this might indicate that urban /e/ is higher than rural /e/. However a closer look to the F1 boxplot for /e/ reveals that an outlier in the rural group is the cause of the higher mean F1.

With regard to nasal vowels, an analysis of mean vowels mostly confirms the initial observations made in the boxplot distribution. Indeed, the plot shows that urban /ĩ/, /ũ/ seem to be more fronted, given their higher F2 values.

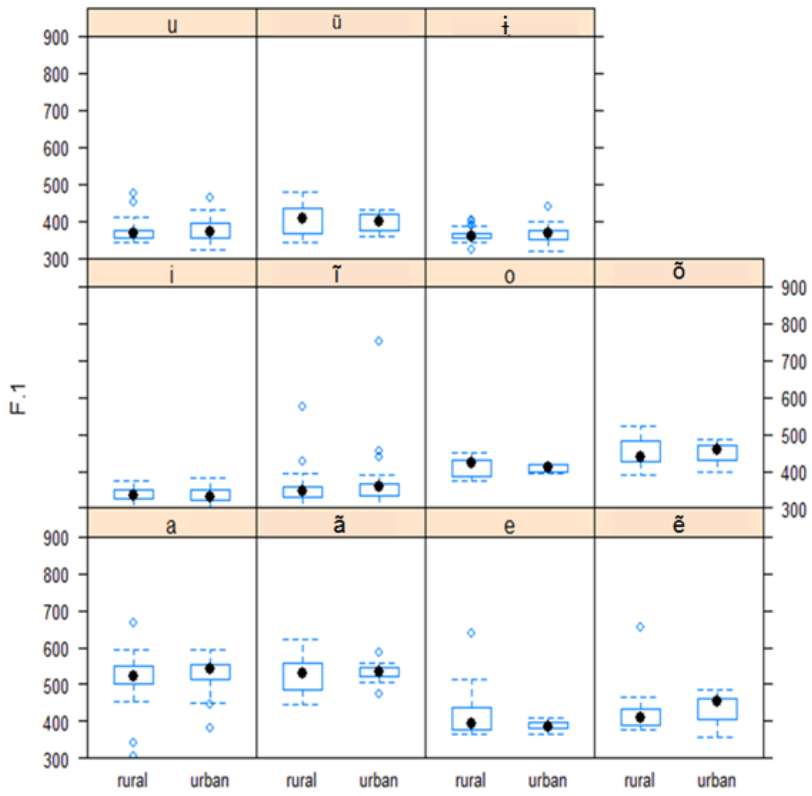

Figure 5. Boxplots for F1 by area (urban vs. rural) grouped by vowel

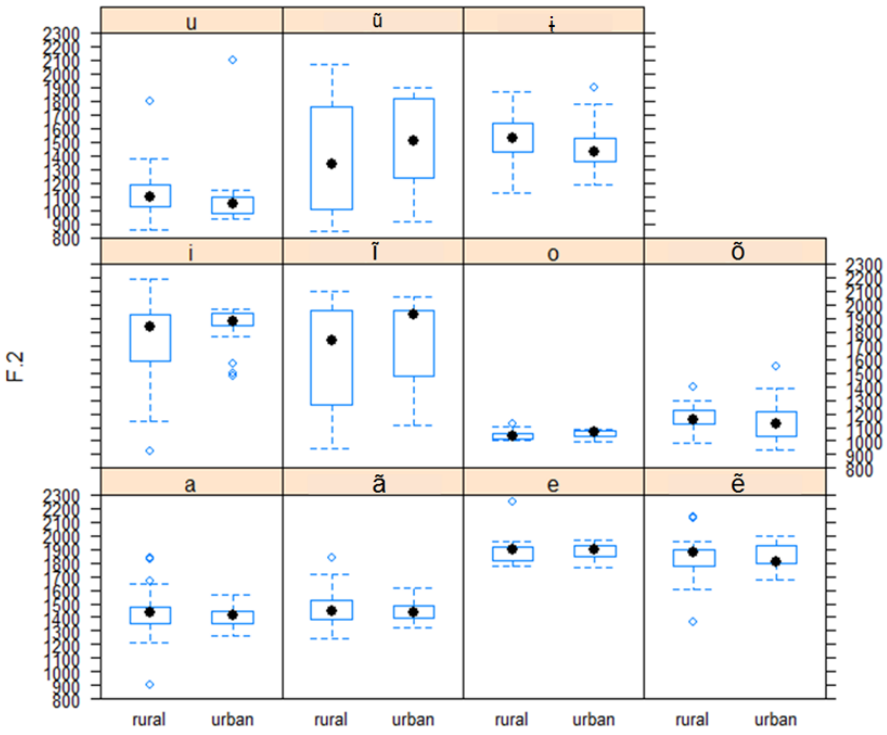

Figure 6. Boxplots for F2 by area (urban vs. rural) grouped by vowel

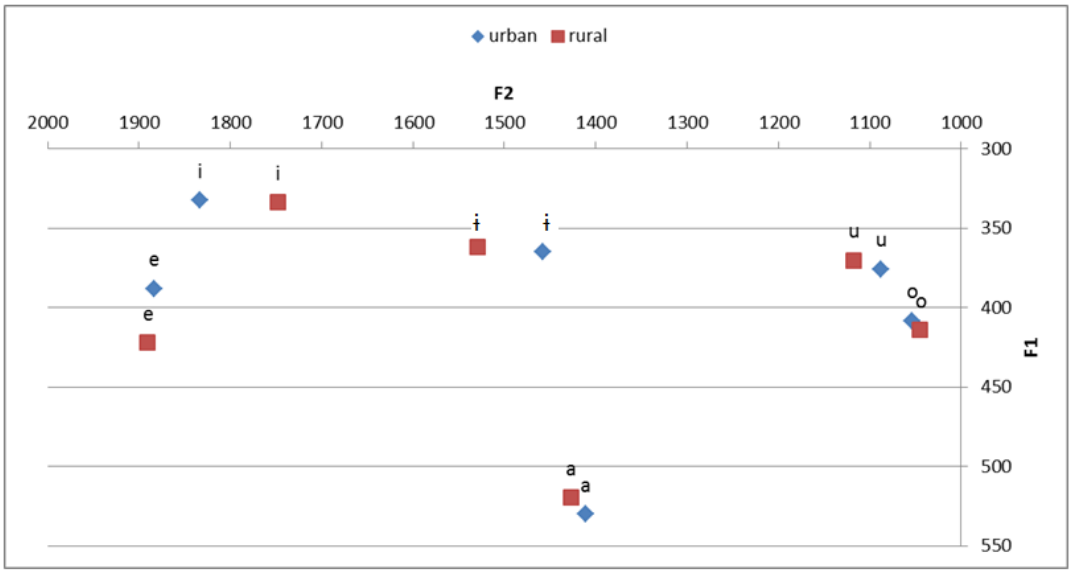

Figure 7. Mean F1 and F2 values in urban and rural group, oral vowels

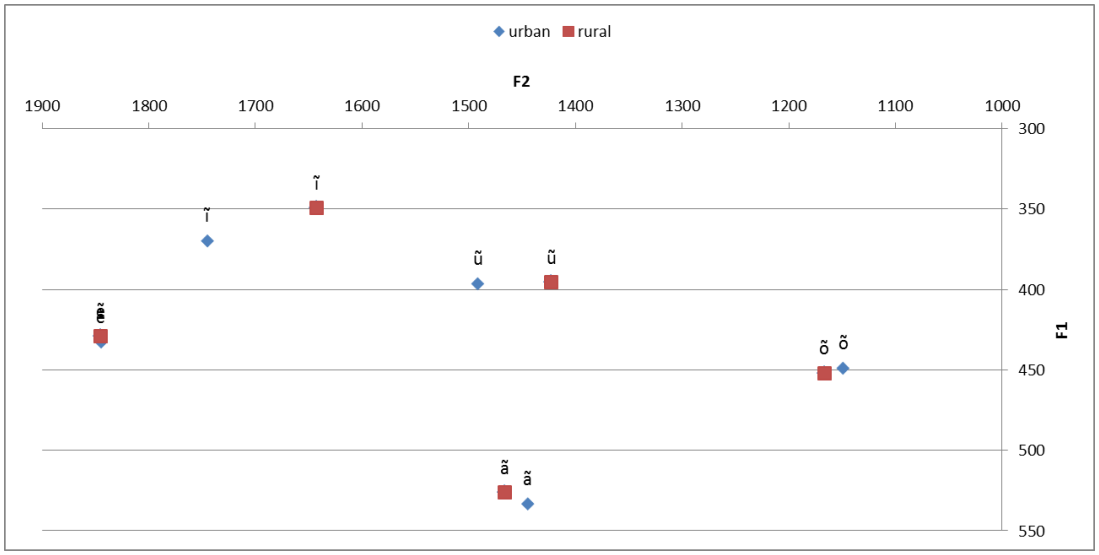

Figure 8. Mean F1 and F2 values in urban and rural group, nasal vowels

3.3.4.2. Linear mixed effect model and variables

The exploratory analysis presented in the previous sections revealed some differences between the urban and the rural group. In order to test whether these differences can be extended to the whole population this section fits the data to linear mixed effect models. For this, I used R, a freely available software for statistical analysis, and more concretely the R lme4 package, which is used for fitting linear mixed effects models.

Linear mixed effect models include two types of independent variables or, more precisely, predictors: fixed effects and random effects. Fixed effects are the variables that are expected

to impact the dependent variable and to be generalizable beyond the specific data sample. In this case, fixed effects are the urban vs. rural condition and the phonological environment. Phonological environment was further divided in two factors, namely, consonant preceding the target vowel and vowel in the previous syllable.

Separate linear mixed effect models were used to test whether F1 and F2 of each vowel was significantly different in the urban vs. the rural group.

The dependent variable is thus F1 and F2 of each vowel. A separate model was built for each vowel pair and F1 or F2. Random effects include variation that is specific to the data sample and that therefore, one does not expect to be generalizable beyond the experiment at hand. In this case, random effects are speaker and word. In other words, there might something specific about the speakers chosen for the experiment (e.g. the size of the vocal folds or of the oral cavity) or about the specific words (e.g. their average frequency of use in the language) that might have an effect on the F1 and the F2 of the target vowels. Random effects capture this idiosyncratic variation and prevent its generalization to the whole population. More technically, if random effects were not included, the independence assumption of linear models would be violated. Indeed, these models assume that all the observations of the experiment are independent from each other (Winter 2013: 2ff.). However, in our data this is not the case. For instance, all the utterances coming from the same speaker will be related since each speaker has a specific baseline pitch that is going to condition all her utterances. Since subject is modelled as a random factor, unlike in the previous section, F1 and F2 are not normalized. Raw F1 and F1 values in Hz are used instead.

The two fixed effects (‘area’ ‘preceding consonant’) are treated as categorical factors with a set of possible values, also known as levels in R terminology. The area factor has the levels ‘urban’ and ‘rural’. The ‘preceding consonant’ factor has the following nine levels:

Preceding consonant		
Levels	**Abbreviation**	**Phoneme**
velar stop	vs	k
alveolar stop	as	t
labial stop	ls	p
prenasalized stop	ps	^{n}d, ^{m}b
alveolar fricative	af	s
bilabial fricative	bf	β
palato-alveolar fricative	pf	ʃ
alveolar tap	at	ɾ
nasal	n	n

Table 7. Possible values of the 'preceding consonant' factor

The formula representing the model in R syntax is as follows:

F1 ~ area+ preceding_consonant +(1|subject)+(1|word)+ ε

F2 ~ area+ preceding_consonant +(1|subject)+(1|word)+ ε,

where $_i$ stands for each of the vowels measured in the experiment, and ε is the error term representing deviations from the initial predictions due to factors that cannot be controlled experimentally.˜ should be read as 'is a function of', so F1 and F2 are, respectively, a function of the fixed effects 'area' and 'preceding consonant', and the random effects 'subject' and 'word', including an error term.

3.3.4.3. Results: differences in F1 and F2 between urban and rural vowels

As already predicted by the descriptive statistics analysis presented in the previous section, significant differences between the urban and the rural group are found in F2, whereas F1 values are quite similar between the two groups. This implies that the main differences between both groups have to do with the degree of vowel fronting (correlated with F2) rather than with vowel height (correlated with F1).

In the following tables, statistically significant coefficients are highlighted in bold. $p<0.05$ was taken as the cutoff point for statistical significance, which means that the probability of a false positive is below 5%. P values, which are used for determining the significance of a statistical hypothesis, are not provided anymore by the current version of lme4[18]. They were obtained using the alternative method suggested by Barr *et al.* (2013: 266) to obtain specific p values for each parameter with lme4. These authors suggest obtaining p values by using the standard normal approximation as a reference, thus treating the t-statistic as if it were a z-statistic[19].

Among oral vowels, statistically significant differences were only confirmed for /ɨ/ and /i/. The linear mixed effect models used here provided an estimate of the coefficients needed to predict the values of F2 in the urban sample, starting from values corresponding to the rural population, which is taken as the baseline[20]. As shown in Table 8 and Table 10, area alone did not yield statistically significant results. However, statistically significant coefficients were observed for the interaction of area and preceding consonant. For /ɨ/, a significant interaction was found for area and velar stop /k/. For /i/ a significant interaction was found for area and prenasalized stop /^{m}b/.

In both cases, thus, the differences in F2 observed in the scatterplot of section 3.3.1.2 cannot be attributed generally to differences in the vowel's F2, but to an interaction with the preceding consonant segment. In the case of /ɨ/ the urban group presented a significantly lower F2 ($p= 0.009$), pointing to a more posterior place of articulation than in the rural group.

[18] Previous versions of lme4 provided a mcmcsamp function which generated a Markov chain Monte Carlo sample from the posterior distributions of the parameters and provided p values. This function has been withdrawn from lme4 versions posterior to 1.0.0, since it was difficult to guarantee the reliability of the output of the function.

[19] T-statistic and z-statistic are both statistical tests to determine the significance of a statistical hypothesis. The main difference between both tests lies on the fact that the z-statistic follows a normal distribution whereas the t-test follows Student's t-distribution. In a normal distribution most values of a data set cluster in the middle of the range whereas the rest are symmetrically distributed in the two extremes. Conversely, the Student's t-distribution uses a number of variables (called degrees of freedom) which change the shape of the normal distribution.

[20] Rural values were taken as the baseline by R only because R takes the level that comes first in alphabetical order as the baseline.

	estimate	standard error	t value	p value
intercept	1761.28	70.75	24.894	0
area urban	-66.29	94.49	-0.702	0.4829604
preceding as	71.56	54.3	1.318	0.1875279
preceding ls	-193.59	62.49	-3.098	**0.001947569**
preceding pf	323.67	71.87	4.503	**0.000006686334**
preceding vs	58.36	63.36	0.921	0.3570001
urban*preceding as	-104.98	91.04	-1.153	0.2488638
urban*preceding ls	-27.85	100.36	-0.278	0.78139
urban*preceding pf	28.38	112.16	0.253	0.8002806
urban*preceding vs	-265.87	102.53	-2.593	**0.009508888**

Table 8. Estimates in the linear mixed effects model predicting F2 of /ɨ/ from the variables 'area' and 'preceding consonant' and the interaction between the last two

However, the factor 'preceding consonant' is actually highly correlated with preceding vowel in the case of the vowel /ɨ/. Thus it is difficult to tell whether the statistical difference should be attributed to the preceding consonant or to the preceding vowel. In other words, since there is only one word in the dataset where /k/ precedes /ɨ/, namely *juky*, and this is also the only word in which the vowel in the preceding syllable is /u/ (in the other words containing /ɨ/ the vowel of the preceding syllable is /a/ or /o/), one might wonder whether the relevant interaction is with /u/ rather than /k/. In order to rule out this possibility a test was done by comparing the F2 of /ɨ/ in *aky* with that in *juky*. The word *aky* 'ripe' was actually not elicited from all the participants of the urban group and therefore was excluded from the dataset aimed at comparing the urban and the rural group. However, it was possible to elicit it in the whole rural group, and so data of F2 of /ɨ/ in *aky* and in *juky* were fit to a linear mixed effects model with preceding vowel and speaker as random effect. Results show that F2 of /ɨ/ is not significantly different in *juky* ($p>0.05$). This seems to indicate that, at least in the rural group, the preceding vowel is not having an effect on the F2 of / ɨ/ in *juky*.

Nevertheless, it remains to be seen whether in the production of the urban group F2 of /ɨ/ is lower due to the presence of the back vowel /u/ in the preceding syllable. The word *aky* was only elicited from one urban speaker. A comparison of the mean F2 in *aky* and *juky* for this speaker reveals a lower value in *juky*: mean F2 for *aky* is 1600 Hz, and for *juky* 1585 Hz[21]. There is thus a possibility that the low F2 value of /ɨ/ in urban *juky* is motivated by the back vowel in the preceding syllable, rather than by the velar environment itself. In the absence of sufficient data, this remains only a tentative hypothesis.

	estimate	standard error	t value	p value
intercept	2099.5029	99.70579	21.05698	0
preceding vowel u	-199.7337	108.02546	-1.84895	0.06446505

Table 9. Estimates in the linear mixed effects model predicting F2 of /ɨ/ from the variable 'preceding vowel' in the rural group

In the case of /i/ the relevant coarticulation involves the prenasalized stop /^{m}b/. Indeed, in the urban group, F2 of /i/ is in this context significantly higher than in the rural group ($p=0.03$),

[21] Since data come from the same speaker in this case F2 values were not normalized.

indicating a more anterior place of articulation than in the rural group. However, as in the preceding case, there is only one lexical item with this consonant-vowel combination, namely, *panambi*. It is thus impossible to test for word-specific effects. However, unlike the preceding case, in this case another lexical item was included in the sample that allows to exclude the effect of the vowel in the preceding syllable as an explanation for the higher F2 in /i/. Indeed, /i/ in *avati* did not present any differences between the urban and the rural speakers, so the differences observed in *panambi* should be attributed either to word-specific effects, or to the consonant environment, namely /^{m}b/.

	estimate	standard error	t value	p value
intercept	2248.62	183.38	12.262	0
area urban	-139.15	220.64	-0.631	0.52824445
preceding n	-107.84	143.57	-0.751	0.45257302
urban*preceding ps	477.81	219.45	2.177	**0.02945489**

Table 10. Estimates in the linear mixed effects model predicting F2 of /i/ from the variables 'repetition', 'area', 'preceding consonant' and the interaction between the last two

These results seem to indicate that, contrary to the initial hypothesis, there is no significant difference between the vowel segments of urban and rural Guarani. Differences might rather lie in coarticulatory patterns. This conclusion is of course only tentative, since in the present dataset there is only one word for each consonantal context. Only with a larger dataset will it be possible to confirm this initial finding.

3.3.5. Discussion: coarticulatory patterns and bilingualism

The fact that L1 Guarani speakers and L2 Guarani speakers present different coarticulatory patterns is coherent with previous literature claiming that coarticulation is language specific (e.g. Hardcastle 1982; Boyce 1990; Kondo 2000). According to this literature, whereas some coarticulatory phenomena might be motivated by low-level constraints and thus might be universal, other coarticulatory patterns might be conditioned by L1 phonology. However, as noted earlier, at this stage it is difficult to identify the phonological constraints that might be motivating the observed differences. Here I will only describe the possible explanatory hypotheses and discuss their plausibility.

3.3.5.1. Differences in F2 of /ɨ/ in juky

In the urban group F2 of /ɨ/ is significantly lower after the velar consonant /k/ and, conversely, it is significantly higher in the rural group. The effect observed in the rural group is consistent with the findings of Hillenbrand and Clark (2001), who reported an upward shift in F2 for back and central vowels in velar environments.

However, the difference between the urban and the rural group might also be interpretable in the opposite sense: it might be seen as a decrease in F2 in the urban group due to the effect of the presence of /u/ in the preceding syllable. As noted earlier, with the available data it is impossible to tell if the downward shift in F2 in /ɨ/ in the urban group is due to a coarticulation with the /u/ in the preceding syllable. It is however plausible that the articulation of /u/ followed by /k/ causes a /ɨ/ target undershoot, making it more to the back (lower F2) than it regularly is. This would point to a failure of the tongue body to readjust to the more fronted articulatory target of /ɨ/ after being positioned for the back target of /u/ in the urban group.

The rural group might have a smoother articulatory transition between back and central vowels. However, if this is the case, it is surprising that the same effect is not observed when the preceding vowel is /o/ (as for example in *yvoty* and *koty*).

3.3.5.2. Differences in F2 of /i/ in panambi

As shown in Table 10, urban speakers produce a significantly higher F2 of /i/ in this particular lexical item. This suggests a more anterior place of articulation as compared to rural speakers. However, due to the limitations of the available data it is not possible to draw a general conclusion from this result, since only one word contained /i/ in the environment of a prenasalized stop /^{m}b/.

4. Conclusions

Several conclusions can be drawn from the preceding sections. Most of them are directions for further research rather than definite conclusions. In this sense, the thesis can be seen as a first exploratory study setting the ground for further fieldwork on Paraguayan Guarani. Firstly, it has discussed extensively the challenges to define Paraguayan Guarani and thus to undertake its description; secondly the findings of the acoustic study point to the hypothesis of coarticulation differences to be examined in the future; and thirdly, it has identified specific methodological points to be taken into account in future work.

In this sense, sections 1 and 2 have led to a better definition of the object of study itself. Several variables have been identified that should be considered in order to describe the language, such as the register, the communicative context, the semantic domain or the bilingual profile of the speakers. Indeed, it has been shown that certain semantic domains (e.g. professional activities, academic subjects) are widely dominated by Spanish, and in the same way Guarani and Spanish stand in different power relations depending on the pragmatic context (e.g. speakers feel the urge to show they are able to communicate in Spanish in a formal context, such as while interacting with the public administration, but conversely they need to show they are equally good speakers of Guarani in a more intimate situation, such as in a gathering of friends). These contextual variables should be carefully taken into account when collecting linguistic data, since they may have an effect on the type of mixed structures used by the speakers and potentially also on the articulation of phonetic categories. At the same time, as shown in section 2, the existence of a great variety of bilingual profiles, partly correlated with geographical location and with socioeconomic status, makes it clear that a proper description of Paraguayan Guarani depends on the accurate identification of those profiles. In the current state of affairs, where a shared linguistic norm is still under development, the safest option may be to adopt a comprehensive approach and to aim at a description of the varieties of Paraguayan Guarani spoken by speakers with different bilingual profiles and in different urban and rural settings of the country. A particularly interesting methodological demarche would be precisely to compile corpora recorded in different contexts, targeting different semantic domains and involving speakers of different ages. By comparing features in each type of corpus, such as types of mixed structures (e.g. which language takes the role of the matrix language), it would be possible to establish a clearer picture of the extent of language mixing and its stability. This kind of empirical analysis, providing measures of stability across different contexts of communication and bilingual profiles, could throw light on the unsettled debate about the mixed status of Paraguayan Guarani.

This proposal seems to run counter to the widely held belief that Paraguayan Guarani is extremely homogeneous. And the findings of section 3 indeed provide evidence of this so far unexplored potential variation. The phonetic study presented in section 3 reveals that there are significant differences between the Paraguayan Guarani spoken in urban and rural areas, although these differences may involve more complex phonological components, such as coarticulation patterns, rather than individual phonemes. Indeed, it was found that /ɨ/ had a significantly more posterior place of articulation in the urban group when it is preceded by the velar /k/. It was also observed that /i/ had a significantly fronter articulation in the urban group

when following the prenasalized stop /^{m}b/. This opens up an interesting line of research for the future in the domain of Paraguayan Guarani phonetics. Instead of targeting individual segments as potential elements of variation, studies should focus on specific coarticulatory patterns that may be influenced by the speakers' L1.

As for the methodological implication of the study, at least three facts can be highlighted. The first one is that the methodology for eliciting test items was adequate. Asking speakers to translate sentences from Spanish into Guarani and to name pictures yielded good results and it allowed circumventing the fact that most speakers of the rural communities were only incipient literates.

The second one has to do with the design of the elicitation questionnaire itself. The questionnaire designed for this study aimed at eliciting lexical items that would allow studying individual phonemes. No significant differences were found in this domain between the different groups of speakers, but there is evidence to believe that an in-depth analysis of coarticulatory patterns may reveal significant differences. Thus, in the future the questionnaire should be specifically tailored to the study of coarticulatory patterns, and thus should include several items for each combination of vowel and target consonant. Only by including several items for each vowel-consonant combination it will be possible to test for word-specific effects. It was also observed that some semantic domains are not equally known in urban and rural areas, so in the future this should be taken into account if the same questionnaire is to be used in urban and rural areas.

The third fact concerns the interaction with the speech community. This is the terrain where a considerable improvement will have to be made in the future. Through this study, a first contact with the speech community in rural areas was established, and I could get a more precise knowledge of the type of bilingual speaker in those areas (mostly passive knowledge of Spanish in higher age ranges, with increasing active knowledge of Spanish in younger generations). However, due to the constraints inherent to the fieldwork carried out for this study (I accompanied the monitoring team of the Ministry of Culture), I could work with the speakers only for a very limited amount of time. In most cases I interacted with each speaker only during the interview, which did not last more than 30 minutes. Moreover, the participants had not been previously informed about the project and sometimes this made the interaction awkward since they had never participated in a linguistic research project before and I only had a couple of minutes to tell them about the background and goal of the project. Not having completely grasped the actual purpose of the study some speakers did not feel confident about their answers and sometimes tried to give me alternative sentences that would be more useful to "learn the language". Now that the initial contact has been made, future fieldwork in the area will have to foresee a relatively extended period of time to develop a trust relationship with the speech community. This will ensure that a much higher percentage of recorded data will be usable for analysis. Indeed, for this study around 40 speakers were interviewed in different locations, but only a small number of those interviews were usable for the study.

References

Auer, P. (1999). From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism*, *3*(4), 309–332.

Backus, A. (2003). Can a mixed language be conventionalised alternational codeswitching? In *The Mixed Language Debate: theoretical and empirical advances* (pp. 237–270). Berlin: Mouton de Gruyter.

Bakker, P. (1997). *A language of our own : the genesis of Michif, the mixed Cree-French language of the Canadian Métis*. New York: Oxford University Press.

Bakker, P. (2000). Convergence intertwining: An alternative way towards the genesis of mixed languages. In D. G. Gilbers, J. Nerbonne, & J. Schaeken (Eds.), *Languages in Contact* (pp. 29–35). Amsterdam, Atlanta: Rodopi.

Bakker, P. (2003). Mixed languages as autonomous systems. In P. Bakker & Y. Matras (Eds.), *The Mixed Language Debate: theoretical and empirical advances* (pp. 107–150). Berlin: Mouton de Gruyter.

Bakker, P., & Mous, M. (Eds.). (1994). *Mixed languages : 15 case studies in language intertwining*. Amsterdam: IFOTT .

Bakker, D., Gómez Rendón, J., & Hekking, E. (2008). Spanish Meets Guaraní, Otomí, and Quechua: a Multilingual Confrontation. In T. Stolz, D. Bakker, & S. Salas Palomo (Eds.), *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes. Empirical Approaches to Language Typology 35* (pp. 165–238). Berlin: Mouton de Gruyter.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Aunthenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studies in Second Language Acquisition*, *19*(4), 447–465.

Caballero, R. V. (1911). Contribution à la connaissance de la phonétique du guarani. *Revue de Phonétique. Tome Premier*, 138–162.

Caballero, V. C. (2008). *When policy becomes practice: teacher's perspective on official bilingualism and the teaching of Guaraní as a second language*. Arizona State University.

Centurión Cervin, C., & Davalos Arce, J. (2009). *Gramática Guaraní*. Perú: Milibri Ediciones.

Cerno, L. (2011). *Descripción fonológica y morfosintáctica de una variedad de la lengua guaraní hablada en la provincia de Corrientes (Argentina)*. Universidad Nacional de Rosario.

Choi, J.K. (2004). La planificación lingüística y la revaloración del guaraní en el Paraguay: Comparación, evaluación e implicación. *Language Problems and Language planning*, 28(3), 241-259.

Comisión nacional de rescate y difusión de la Historia Campesina(1992). *Kokuegara Rembiasa. Experiencias campesinas. Ligas Agrarias Cristianas 1960-1980. Tomo III. Misiones y Paraguarí*. Asuncion: CEPAG.

Correa, D., Taxler, G., & Hite, D. (2007). Poverty, language, and participation in non-farm labor markets in rural Paraguay.*American Agricultural Economics Association, 2007 Annual Meeting, July 29-August 1, Portland, Oregon*.

Dietrich, W. (2010). Lexical evidence for a redefinition of Paraguayan Jopara. *STUF - Language Typology and Universals Sprachtypologie Und Universalienforschung*, *63*(1), 39–51.

Estigarribia, B. (2015). Guaraní-Spanish Jopara mixing in a Paraguayan novel. Does it reflect a third language, a language variety, or true code-switching? *Journal of Language Contact*, *8*, 183–222.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169–186.

Flege, J. E., Schirru, C., & MacKay, R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 467–491.

Flege, J., & Mackay, I. (2011). What accounts for "age" effects on overall degree of foreign accent? In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010, Vol. 2* (pp. 65–82). Bern, Switzerland: Peter Lang.

Gabbert, W. (2004). *Becoming Maya. Ethnicity and social inequality in Yucatán since 1500*. Tucson: The University of Arizona Press.

Galeano Olivera, D. (2011). UH tiene razón: Guarani es la asignatura más odiada (En realidad lo que odian es el Jopara). *Ateneo de Lengua Y Cultura Guaraní*. http://ateneoguarani.edu.py/index.php/charlas-y-conferencias/80-uh-tiene-razon-guarani-es-la-asignatura-mas-odiada-en-realidad-lo-que-odian-es-el-jopara

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, *32*(2), 141–174.

Gregores, E., & Suárez, J. (1967). *A Description of Colloquial Guaraní*. The Hague: Mouton de Gruyter.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*, 3–15.

Grosjean, F. (1999). Studying bilinguals: Methodological and conceptual issues. *Language and Cognition*, *1*, 117–130.

Guasch, A. (1948). *El idioma Guarani. Gramática y antología de prosa y verso* (Author's e.). Buenos Aires, Argentina.

Hamel, R. E. (1994). Linguistic rights for Amerindian peoples in Latin America. In T. Skutnabb-Kangas, R. Phillipson, & M. Rannut (Eds.), *Linguistic human rights: overcoming linguistic discrimination* (pp. 289–303). Berlin and New York: Mouton de Gruyter.

Harcastle, W. (1982). Constraints on coarticulatory processes. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 33–49). London: Edward Arnold.

Boyce. (1990). Coarticulatory organization for lip-rounding in Turkish and English. *Journal of the Acoustical Society of America*, *88*, 2584–2595.

Kondo, Y. (2000). Production of schwa by Japanese speakers of English: an acoustic study of shifts in coarticulatory strategies from L1 to L2. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V. Acquisition and the lexicon* (pp. 29–39). Cambridge: Cambridge University Press.

Hauck, J. D. (2014). La construcción del lenguaje en Paraguay: fonologías, ortografías e ideologías en un país multilingüe. *Boletin de Filología*, *Tomo XLIX*(2), 113–137.

Heath, S. B. (1972). *Telling tongues: Language policy in Mexico, colony to nation*. New York and London: Teachers College Press.

Herbert, R. K. (1986). *Language universals, markedness theory and natural phonetics*. Amsterdam: Mouton de Gruyter.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, *109*, 748–763.

Jiquilín-Ramírez, D., & Albano, E. (2014). The effect of adjacent vowels on the spectrum of back fricatives in Guarani and Spanish. In *1st International DINAFON Meeting, Campinas. Book of abstracts*.

Johnston, J. (2007). *A phonetic description of Jopará: Analysis of the phonetic influence of Spanish and Guarani on Jopará*. University of Kansas.

Kaiser, E. (2008). Nasal spreading in Paraguayan Guarani: Introducing long-distance continuous spreading. In *La structure des langues amazoniennes, Phonologie et Grammaire I, a. Alignements morphosyntaxiques, b.Numéro Spécial de Amerindia 32, Revue d'ethnolinguistique amérindienne. Harmonie Nasale.* (pp. 283–300). Paris: CELIA/CNRS.

Kallfell, G. (2010). *Grammatik des Jopara. Gesprochenes Guaraní und Spanisch in Paraguay*. Westfälischen Wilhelms-Universität.

Kluck, P. (1990). The society and its environment. In D. M. Hanratty & S. W. Meditz (Eds.), *Paraguay: a country study*. Washington, DC: Library of Congress, Federal Research Division.

Krivoshein de Canese, N., & Corvalán, N. (1987). *El español del Paraguay. En contacto con el guaraní*. Asunción: CEPAG.

Krivoshein de Canese, N., & Acosta Alcaraz, F. (2007). *Gramática Guaraní*. Asunción, Paraguay: Servilibro.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley and Sons.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, *49*, 606–608.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251–285.

Lustig, W. (1996). Mba'éichapa oiko la Guarani? Guarani y Jopara en el Paraguay. *Papia*, *4*(2), 19–43.

Matras, Y., & Bakker, P. (2003). The study of mixed languages. In Y. Matras & P. Bakker (Eds.), *The Mixed Language Debate: theoretical and empirical advances* (pp. 1–20). Berlin: Mouton de Gruyter.

Meakins, F. (2013). Mixed languages. In P. Bakker & Y. Matras (Eds.), *Contact Languages* (pp. 159–228). Boston, Berlin: De Gruyter Mouton.

Melià, B. (1975). Hacia una "tercera lengua" en el Paraguay. In J. Pla & B. Melià (Eds.), *Bilingüismo y tercera lengua en el Paraguay*. Asuncion: Universidad Católica.

Melià, B. (1986). *El Guaraní conquistado y reducido. Ensayos de etnohistoria*. Asunción, Paraguay: Centro de Estudios Antropológicos.

Melià, B. (2003). Elogio del monolingüismo guaraní. In *Discriminaciones y Medidas Antidiscriminatorias. Debate Teórico Paraguayo y Legislación Comparada. Derechos Humanos; Documento de Trabajo 4* (UNFPA., pp. 37–46). Asunción, Paraguay.

Melià, B. (2010). *Pasado, presente y futuro de la lengua guaraní*. Asunción, Paraguay: CEADUC.

Melià, B. (2013). *La tercera lengua del Paraguay y otros ensayos*. Asunción, Paraguay: Servilibro.

Ministry of Education and Culture (2006). La Educación Bilingüe en la Reforma Educativa Paraguaya. Asunción: Ministerio de Educación y Cultura.

Mortimer, K. (2006). Guaraní académico or Jopará? Educator perspectives and ideological debate in Paraguayan bilingual. *Working Papers in Educational Linguistics*, *21*(2), 45–71.

Munro, M. J., Flege, J. E., & MacKay, R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Pshycholinguistics*, *17*, 313–334.

Muysken, P. (1979). La mezcla de Quechua y Castellano. *Lexis*, *3*(1), 41–56.

Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. In A. Highfield & A. Valdman (Eds.), *Historicity and variation in Creole studies* (pp. 52–78). Ann Arbor, MI: Karoma.

Muysken, P. (1995). Code-switching and grammatical theory. In L. Milroy & P. Muysken (Eds.), *One speaker, two languages. Cross-disciplinary perspectives on code-switching* (pp. 177–198). Cambridge: Cambridge University Press.

Muysken, P. (1997). Media Lengua. In S. G. Thomason (Ed.), *Contact languages: a wider perspective* (pp. 365–426). Amsterdam: John Benjamins.

Muysken, P. (2000). *Bilingual speech: a typology of code-mixing*. Cambridge: Cambridge University Press.

Muysken, P. (2004). *Bilingual speech. A typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). *Duelling languages. Grammatical structure in codeswitching*. New York: Oxford University Press.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, *11*(1), 73–89.

Pearce, A. J., & Heggarty, P. (2011). Introduction. History, linguistics and the Andean past: A much-needed conversation. In P. Heggarty & A. J. Pearce (Eds.), *History and language in the Andes* (pp. 1–18). New York: Palgrave Macmillan.

Penner, H. (2007). Se habla. Es Guaraní. No es Guaraní. Es Castellano. No es Castellano. Es Guaraní y Castellano. No es ni Guaraní ni Castellano. ¿Qué es? *Signos Lingüísticos*, *5*, 45–95.

Pfeiler, B. (2014). Maya and Spanish in Yucatán: an example of continuity and change. In S. S. Mufwene (Ed.), *Iberian imperialism and language evolution in Latin America* (pp. 205–225). Chicago and London: The University of Chicago Press.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2. *Journal of Phonetics*, *29*, 191–215.

Pottier, B. (1970). La Situation Linguistique Du Paraguay. *Cahiers Du Monde Hispanique et Luso-Brésilien*, *14*, 43–50.

PREAL. (2013). *Informe de progreso educativo. Paraguay*. Instituto Desarrollo.

Ross, M. D. (1999). Exploring metatypy: How does contact-induced typological change come about? *Keynote Talk, Australian Linguistic Society's Annual Meeting, Perth*.

Rubin, J. (1985). The special relation of Guaraní and Spanish in Paraguay. In N. Wolfson & J. Manes (Eds.), *Language and Inequality* (pp. 111–120). The Hague: Mouton de Gruyter.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Service, E. R. (1971). *Spanish-Guarani relations in early colonial Paraguay*. Westport, Conn.: Greenwood Press.

Socolow, S. M. (2000). *The women of colonial Latin America*. Cambridge: Cambridge University Press.

Price, S., & Stewart, J. (2013). Nasal harmony fading in Guarani. In *Canadian Linguistic Association, 2013 Conference. University of Victoria*.

Telesca, I. (2009). *Tras los expulsos. Cambios demográficos y territoriales en el Paraguay después de la expulsion de los jesuitas*. Asunción, Paraguay: CEADUC.

Shain, Cory and Tonhauser, J. (2010). The synchrony and diachrony of differential object marking in Paraguayan Guaraní. *Language Variation and Change*, *22*(3), 321–346.

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite.

Thomason, S. G. (1995). Language mixture: Ordinary processes, extraordinary results. In C. Silva-Corvalán (Ed.), *Spanish in four continents: Studies in language contact and bilingualism* (pp. 15–33). Washington, DC: Georgetown University Press.

Thun, H. (2005). "Code switching", "code mixing", "reproduction traditionnelle" et phénomènes apparentés dans le guarani paraguayen et dans le castillan du Paraguay. *Rivista Di Linguistica*, *17*(2), 311–346.

Thun, H., & Pereira Jacquet, M. G. (Eds.). (2002). *Atlas lingüístico Guaraní-Románico: sociología (ALGR-S). T. II: Mapas.* Kiel: Westensee.

Tonhauser, J. (2006). *The temporal semantics of noun phrases: evidence from Guarani*. PhD Thesis. Stanford University.

Tonhauser, J. (2011). Temporal reference in Paraguayan Guaraní, a tenseless language. *Linguistics and Philosophy*, *34*, 257–303.

UN. (2015). *World urbanization prospects. The 2014 revision*. New York.

Walker, R. (1999). Guaraní Voiceless Stops in Oral versus Nasal Contexts: An Acoustical Study. *Journal of the International Phonetic Association*, *29*(1), 63–94.

Walker, R. (2013). *Nasalization, Neutral Segments, and Opacity Effects*. New York: Routledge.

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications.Tutorial 2.*

Yoshioka, H. (2010). Indigenous language usage and maintenance patterns among indigenous people in the era of neoliberal multiculturalism in Mexico and Guatemala. *Latin American Research Review*, *45*(3), 5–34.

Zajícová, L. (2009). *El bilingüismo paraguayo. Usos y actitudes hacia el guaraní y el castellano*. Madrid: Iberoamericana.