Analyzing the structure of code-switched written texts

The case of Guarani-Spanish Jopara in the novel Ramona Quebranto

Bruno Estigarribia and Zachary Wilkins
University of North Carolina at Chapel Hill

As more written language data become available, the interest in written language mixing / codeswitching (LM/CS) is increasing (Sebba, Mahootian & Jonsson 2012; Sebba 2013). LM/CS in non-naturalistic (e.g., literary) texts raises issues related to gauging (1) the authenticity and representativity of a textual corpus, and deciding (2) whether categories/mechanisms of spoken LM/CS apply to written LM/CS.¹ We focus on Guarani-Spanish LM/CS (Jopara) as represented in the Paraguayan novel Ramona Quebranto (RQ). We apply the framework of Muysken (1997; 2000; 2013), developed as a taxonomy of spoken LM/CS. Our contribution extends its applicability to written LM/CS. We show that Jopara has a mix of insertional and backflagging strategies, with infrequent alternations.

Keywords: language mixing, codeswitching, literary corpora, insertion, alternation, backflagging, authenticity, naturalistic language, matrix language, composite ML

1. Guarani and Jopara in Paraguay

Paraguay’s bilingualism is unique because it is the only country in the Americas in which the majority non-indigenous population speaks an indigenous language (Guarani) (Gynan 2001, among others). Preliminary 2012 census figures put Guarani speakers at 77% (of which 8% do not speak Spanish) and Spanish speakers

¹. We do not focus on representativity here. Authenticity is a necessary but not sufficient condition for representativity, which is related to sampling of data to exemplify a maximum of the language. A one-speaker corpus can be authentic but only represents one idiolect.
at 73% (of which 5% do not speak Guarani). Guarani-Spanish bilinguals total 69%, and speakers of either Guarani or Spanish total 82%.\(^2\)

Guarani is a Tupí-Guarani language, genetically unrelated to Spanish. Classical Guarani is its old form. Guaraníete or Academic Guarani is a recent standardized form that minimizes Spanish loanwords and grammatical influences. Guarani is agglutinative (1), mildly polysynthetic (2), and postpositional (3), hence typologically rather different from Spanish (Gregores & Suárez 1967; Krivoshein de Canese & Acosta Alcaraz 2007; Tonhauser 2006; Velázquez-Castillo 2004). Example (3b) additionally shows that some postpositions have an allomorph for bases with a nasal segment (nasal harmony: \(\text{pl -} \text{kuéra} > \text{nguéra} \); \(\text{loc -} \text{pe} > \text{-me}\)).\(^3\)

(1) \textit{nderejehechaukavéi.}  
\textit{nde-re-je-hecha-uka-vé-i}  
\textit{NEG-2SG-PASS-SEE-CAUS-CMPR-NEG}  
\textit{‘You don’t make yourself seen anymore.’}  
\textit{(Velázquez-Castillo 2004:1427)}

(2) \textit{ajepohéi.}  
\textit{a-je-po-(j)o(h)éi}  
\textit{1SG-REFL-hand-wash}  
\textit{‘I wash my hands.’}  
\textit{(Velázquez-Castillo 2004:1428)}

(3) a. \textit{Ahákuri cherógaguive nderógape kavajuári.}  
\textit{A-há-kuri che-róga-guive nde-róga-pe kavaju-ári}  
\textit{1SG-GO-PST 1SG-HOUSE-ABL 2SG-HOUSE-LOC horse-on}  
\textit{‘I went from my house to your house on horse.’}  

b. \textit{Ohechapa neirũnguéra cheirũme.}  
\textit{O-hecha-pa ne-irũ-\textit{nguéra} che-irũ-me}  
\textit{3-SEE-Q 2SG-FRIEND-PL 1SG-FRIEND-LOC}  
\textit{‘Did your friends see my friend?’}

Paraguayans commonly speak a variable mixing of Guarani and Spanish named \textit{Jopara} (Bakker, Gómez Rendón & Hekking 2008; Dietrich 2010; Gómez Rendón)

---

2. Additionally, Brazilian Portuguese is spoken by almost 7% of the population. These speakers are also largely fluent in Guarani and Spanish. (Source “Censo 2012: hay un absoluto predominio del uso del guaraní y el castellano,” http://ea.com.py/v2/censo-2012-hay-un-absoluto-predominio-del-uso-del-guaraní-y-el-castellano/, accessed 10 November 2014).

3. We minimize technical glosses for Spanish morphemes since their role is less important in the paper.

4. A reviewer notes there is apheresis of the first syllable of the root in this example.

5. Guarani spelling rules attach only monosyllabic postpositions to their stem. Application of this rule is inconsistent in RQ. We attach all postpositions following Velázquez-Castillo (2004) and Tonhauser (2011).

The labels Castení and Guarañol imply stable varieties, perhaps even new mixed languages (Matras & Bakker 2003). We reject this implication. As Estigarribia (2015) discusses, Jopara lacks the degree of conventionalization and stability of mixed languages. Here, we use Castení and Guarañol solely as convenient shortcuts.

2. The corpus: The novel Ramona Quebranto

Our corpus is Ramona Quebranto (RQ; Ayala de Michelagnoli 1989), the first novel written using Jopara. Margot Ayala de Michelagnoli was born to Paraguayan parents in Paris, but at the age of three emigrated to Paraguay, where she came into frequent contact with Jopara (Díaz de Espada de Ramírez Boettner 1989; Méndez-Faith 1996). RQ’s storyline follows the life of Ramona, a Paraguayan living in an impoverished neighborhood of Asunción called La Chacarita. The novel features both passages by an omniscient narrator (exclusively in Spanish), and very frequent first-person narration passages and dialogues with extensive Guarani-Spanish mixing. The author’s intended audience is primarily Paraguayans with varying degrees of exposure to Guarani (and thus Jopara), as the novel is impossible to understand for monolingual Spanish speakers (Estigarribia 2015).

2.1 The authenticity of our corpus

RQ, being written as a fictional text, and moreover, as a mixed language text, may raise a question as to its language authenticity. However, note that authenticity has many facets.

6. Gómez-Rendón (2008) considers RQ to be written in Castení. We do not agree, since many passages are mostly Guarani-based (see Estigarribia 2015).
7. Note that these terms are purely technical and are not of common use among Paraguayans.
 Perhaps the most obvious sense of authenticity is “nativeness” or “vernacularness”. Montes-Alcalá (2012:85) considers CS literary productions authentic if literary characters are “members of a speech community where CS is the norm” and if the author is “a member of that speech community.” The first condition is unequivocally met in RQ. As for the second, we note that authors of fiction routinely portray characters whose identities are quite different from their own. Ayala de Michelagnoli is not a member of the Chacarita community she depicts. Yet, prior to writing the novel, the author spent considerable time living there, recording its inhabitants’ speech (Peiró Barco 2002). The author carefully constructed the novel’s Jopara to mirror real use. Many Paraguayan critics have lauded her for her astonishing precision in accomplishing this (Lustig 1996; Peiró Barco 2002; Peiró & Rodríguez Alcalá 1999; Rubí 2011).

A related sense of authenticity is “naturalness.” A text is authentic in this sense if it is produced by a speaker without reflection or conscious monitoring (Eckert 2003), and occurs in natural and representative conditions (Coupland 2003; Joy 2011). Written sources, especially literary, are more carefully planned and edited, are under greater prescriptivist scrutiny, and are removed from the addressee(s). Yet these properties do not automatically make such productions unavailable for linguistic analysis, historical linguistics being an obvious example where researchers need to assume that written prose, poetry, and legal documents authentically reflect naturalistic practices. Moreover, Renouf’s (2007:46) claim that “the issue of how to classify conversation within novels [i.e., as authentic or inauthentic] has remained largely unresolved” is particularly relevant, since RQ is mostly composed of conversational passages.

A third sense of authenticity is related to the notions of “linguistic isolationism” and “essentialism” (Bucholtz 2003), that is, of a “pure” language that somehow reveals and instantiates the original essence of the speakers. This sense is obviously flouted by all contact varieties. This can help explain why the authenticity of a text seems to become particularly more suspect in the case of mixing. As Callahan (2004:1) observes: “Fiction writers who use codeswitching in their characters’ dialogue were assumed to be merely imitating, not always in a very accurate manner, what might be uttered in actual conversation.” In fact, written LM/CS inherits the worst of both worlds: not only is it written, it is also victim

8. In reality, RQ is not carefully edited. Typos abound, spelling is erratic, and orthographic recommendations are not always followed.

9. Historically, scholarship on language was conducted using written sources. In the 20th century speech became for linguists the “gold standard”. Perhaps for this reason, written, composed language is seen often nowadays as “inauthentic” or non-representative of “real” language (Trujillo 1979; Valdés-Fallis 1977).
to the “monolingual bias in industrialized societies” (Sebba 2012:1). But a fourth and more culturally sensitive meaning for \textit{authenticity} is determined by a text’s \textit{systemic coherence} (originality in some important social or cultural matrix), \textit{consensus} (high degree of acceptance within a constituency), and \textit{value} (cultural relevance) (Coupland 2003). Little, Devitt & Singleton (1988) consider a text authentic if its creation fulfills a social purpose in the community. These fundamental properties are noted by Peiró & Rodríguez Alcalá (1999:153): “[La] novela Ramona Quebranto (1989) merece una atención particular por ser la primera narración que reproduce el lenguaje \textit{real} paraguayo […]; el jopará, mezcla de español y guaraní.”\footnote{The novel \textit{Ramona Quebranto} (1989) deserves special attention because it is the first narrative work that reproduces the \textit{real} Paraguayan language […]: Jopara, a mixture of Spanish and Guarani.” (Our translation and emphasis.)}

Our conclusion is that authenticity is not a monolithic concept and cannot be decided a priori. Importantly, calling a literary text “authentic” does not mean it has to be the same as spoken language.

2.2 Applying spoken CS models to written CS

We believe that analytical frameworks for spoken LM/CS can be applied to RQ. Both the novel’s conversation-like passages and first-person narrations are written in a conversational style. Moreover, Guarani and Spanish are treated, for visual purposes, as a single language. Sebba (2013) calls such texts \textit{written CS} and observes their similarity to spoken CS, noting that tools for the latter are likely to be applicable to the former.

Moreover, the analysis of LM/CS often has recourse to written sources. Mahootian (2005) studies CS in a lifestyles magazine, and Graedler (1999) compares CS in magazines to CS in letters. Hinrichs (2006) and Negrón Goldbarg (2009) examine CS in emails. Literary sources have been recently examined by

\footnote{“[In RQ] there is the implicit intention of, first and foremost, denouncing the suffering in this social reality, and of contributing to elevate Jopara to literary status; RQ is at the same time a humble contribution to understanding Paraguay’s bilingualism.” (Our translation.)}
Thomas & Sayahi (2012). Montes-Alcalá (2012: 85) notes that “the sociopragmatic functions that have been traditionally ascribed to oral discourse could [i.e., were] found in a bilingual literary corpus.” Myers-Scotton (1998) argues that her Matrix Language Frame Model (MLF), based on spoken CS, is applicable to the analysis of literary works. Following these suggestions, Callahan (2004) analyzes a corpus consisting of fictional texts and finds no significant difference between written and spoken corpora in the applicability of the MLF.12

Given these considerations, we set out to test the applicability to RQ of a different structural CS framework, developed by Muysken (1997; 2000; 2013). We hypothesized that this analysis would cast light on many structural aspects of the Jopara in RQ, and show that models designed for spontaneous speech apply to contemporary literary texts.

3. Muysken’s typology of CS

Muysken presents a typology of spoken CS corpora. We will introduce the three strategies of Insertion, Alternation, and Backflagging.13

Insertion is a strategy where well-defined chunks of a language B (LB) appear in a clause whose general structure is in another language A (LA). LB insertions are generally content words or short constituents, sometimes with morphological integration into LA (e.g., affixes from LA attached to LB bases), usually completely nested between two fragments in LA. Consider in this regard Example (4):

\[(4) \quad \text{vile vitu zake zi-me-spoil-iw-a.} \quad \text{[Swahili-English]}
\]

\[\text{those things hers they-

\text{PRF-spoil-PASS-IND}
\text{‘Those things of her were spoiled for her.’}
\]


Deuchar, Muysken & Wang (2007) propose using Myers-Scotton’s concepts of a matrix language (ML) and an embedded language (EL) (Myers-Scotton 1997; Myers-Scotton & Jake 2000). Crucially, the ML of a clause is defined as the language that contributes outsider system morphemes, that is, non-content morphemes required by material outside of their own maximal projection, whose

12. We will discuss this model in some detail in Section 3, although it is not central to our own analysis.

13. We do not introduce congruent lexicalization because this strategy is observed mainly between languages that are typologically very similar. Given the typological distance between Guarani and Spanish, this strategy is predicted not to be viable. Consistent with this prediction, we have identified only a few cases of possible, but unclear, congruent lexicalization in RQ.
occurrence builds the structure of the clause as a whole. The ML is, then, not necessarily the language that contributes the most morphemes or words. The clearest examples of outsiders are SV agreement, clitics/affixes, and grammatical function marking (Myers-Scotton 2002:75–76). Morphemes like plurals (which are required by the semantics of their own constituent) or possessive de in Spanish (which is required by the internal syntax of its immediate maximal projection) are not outsiders. Importantly, insertion is most often an intra-clausal strategy, since MLs can be determined only in reference to a clause, and can change inter-clausally.

*Alternation* is a strategy where fragments in LA alternate with others in LB, in sentences not clearly in either LA or LB. Alternation usually involves switches that are longer, complex constituents, or involve several constituents, switches that are not nested, and those that occur at major clause boundaries (5a).

(5) a. *But it’s sort of like ‘n bietjie van dit en ‘n bietjie van dat.*

   a bit of this and a bit of that

   ‘But it’s sort of like a bit of this and a bit of that.’

   [English-Afrikaans]

   (Van Dulm 2007:7; Cited by Muysken 2015:243)

Hence, while insertions are intra-clausal, alternations are often (but not necessarily) inter-clausal. That the switch is an adverb, conjunction, discourse marker, or a tag suggests an alternation strategy, rather than insertion. Alternation is also likely in cases where some elements appear in both languages (“doubling”), or where the switch is “flagged” by an LB marker, as in (5b).


   thus cart-ins-1pl water far for that

   ‘Thus we carted the water from afar in order to drink it.’


Finally, *backflagging* involves mainly the insertion of heritage language discourse markers (simple and frequent, clause-peripheral single items) in L2 discourse. This definition presupposes a community where some speakers at least have switched to L2 dominance, but continue to mark their discourse with L1 elements to signal their ethnic identity, as in (6).

(6) a. *No mang vrouw-tje, vind juist z’n accent hinderlijk.* [Sranan-Dutch]

   no man woman-dim, find exactly his accent irritating

   ‘No woman, I really find his accent irritating.’

   (Muysken 2015:245, from the chat site www.mamjo.nl)
b. *Ik ben een dokter wella ik ben een ingenieur.* [Dutch-Moroccan Arabic]
   I am a doctor or I am an engineer
   ‘I will become a doctor or an engineer.’

(Nortier 1990: 142; cited by Muysken 2013: 713)

We will show data that suggest this definition may be too restricted, and we will suggest ways in which other kinds of morphemes can participate in backflagging. Notably, we will pursue a backflagging analysis for clauses with a composite ML, that is, clauses with outsider morphemes from both participating languages where an insertional analysis is not viable.

Muysken (2000; 2013) notes that insertion is prevalent in post-colonial settings or recent migrant communities, when languages have unequal power, and when speakers have asymmetric proficiency. Alternation occurs when there is a tradition of stable bilingualism with strict language separation, with both languages vying equally for political power. Congruent lexicalization is mainly found in situations of long contact between languages that are typologically and lexically very similar, with equal prestige, no tradition of language separation, and high bilingual proficiency. Finally, backflagging is favored in cases of second or third generation language shift. Each community where CS is found tends to use predominantly one of these strategies, but often a secondary strategy is found.

Paraguay’s situation, at least in the urban, non-indigenous context depicted in the novel, is that of long contact between typologically dissimilar languages that have unequal prestige, with a strong tradition of separation between the two. In the next few sections we will show that RQ’s primary profile is a mix of insertional and backflagging strategies, with infrequent cases of alternation. Hence, even though RQ is a fictional text, its representation of Jopara seems to coincide with Muysken’s predictions.

A caveat regarding the status of CS involving Guarani affixes is in order. The agglutinative morphemes of Guarani do not exhibit a high degree of selection of their bases (unlike many Spanish fusional affixes). They easily attach to Spanish lexemes, often resulting in morphologically integrated insertions. This often happens with outsider morphemes in Guarani used in a clause where other outsiders are from Spanish, giving rise to a composite ML, or when the clause containing the Guarani morpheme is clearly in Spanish. Yet this case is not contemplated in Muysken’s typology. One option is to regard these cases as insertions of morphology. We present them here instead as a kind of backflagging by bound morphology. It shares with backflagging by discourse markers the status as single items, simple and frequent, with a clear ethnic connotation (insofar as they denote Paraguayan-ness). This analysis seems adequate especially in the many cases in which the category or slot the affix occupies does not exist in the ML. A potential problem is
that these affixes may not be clause-peripheral (as in prototypical backflagging) but instead well-integrated in the syntax of the clause.\textsuperscript{14}

4. Application of Muysken’s model to RQ

First, we present insertional (Section 4.1), then backflagging switches (4.2), and lastly alternational switches (4.3). To give a general idea of the language distribution in RQ, the first 272 sentences (excluding sentences from the omniscient narrator and a character that never speaks Guarani) contain 522 clauses. Mixed sentences (inter- and intra-clausally) number 186 (68\%), containing 504 switch points.\textsuperscript{15,16} Table 1 gives the distribution of word and morpheme tokens in this corpus.

<table>
<thead>
<tr>
<th>Word tokens</th>
<th>Spanish</th>
<th>Guarani</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1146</td>
<td>798</td>
<td>261</td>
<td>87</td>
</tr>
<tr>
<td>69.6%</td>
<td>22.8%</td>
<td>7.6%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Morpheme tokens</th>
<th>Spanish</th>
<th>Guarani</th>
</tr>
</thead>
<tbody>
<tr>
<td>1860</td>
<td>1273</td>
<td>587</td>
</tr>
<tr>
<td>68%</td>
<td>32%</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 suggests that, as the primary lexical contributor, Spanish is the base language of the novel, and that the novel is in Castení (consistent with Gómez Rendón 2008). Yet Estigarribia (2015) demonstrates that understanding RQ requires good competence in both Guarani and Spanish. Entire passages in the novel are totally incomprehensible without at least some knowledge of Guarani. However, as our later analysis will suggest, the true picture is much more complicated than that.

The total number of switches that fall in each of the discussed strategies is 119 for backflagging, 114 for insertion, and 57 for alternation. In the sections that follow, of all the CS examples we found in the novel, we chose here the most representative ones for each strategy.

\textsuperscript{14} Sociolinguistic criteria can also be problematic, since for backflagging the item needs to be in a former L1, in a community where speakers have switched to a new L1 that was formerly the L2. L1/L2 status may already be difficult to determine on an individual basis in Paraguay, but it is certainly much harder at the community level.

\textsuperscript{15} An additional advantage in studying RQ is the sheer density of intra-sentential switches.

\textsuperscript{16} Note that the count of switch points is provided solely as a rough measure of how much the text goes “back and forth” between the languages. For that reason, it overestimates the actual switch count roughly by a factor of 2.
4.1 Insertional strategy

As we mentioned above, interpreting a switch in LA as a result of insertion depends on determining that LB is the ML of the clause. (Different clauses in the same sentence can and often have different MLs.) We will consider as insertions single content word switches (noun, verb, or adjective) or single-constituent switches, especially if they are nested (followed by content in the other language, in the same clause). Other single words such as adverbs, conjunctions or discourse markers generally do not qualify because they are more peripheral to the clause. We first discuss switches with morphological integration and then without. Importantly, the fact that Guarani is the ML of these clauses with Spanish insertions has been captured by the label Guarañol, introduced in Section 1.

4.1.1 With morphological integration

Several Guarani affixes appear commonly on Spanish bases in RQ. This is expected because Guarani is agglutinative, creating many opportunities for expressing grammatical relations via affixation of Guarani morphemes onto Spanish insertions. Moreover, Guarani affixes and particles show little to no selection of their bases (barring allomorphs dictated by nasal harmony with the base), further favoring their affixation to Spanish words.

4.1.1.1 Guarani SV agreement on inserted Spanish verbs

In (7) the sentence contains two clauses, the first entirely in Spanish, the second with ML Guarani and a Spanish verb inserted. In (8) on the other hand, the Spanish verb with Guarani morphology is in a clause showcasing a Composite ML, since (a) the Spanish preposition *de* is required by the subcategorizing predicate, and is therefore an outsider morpheme, and (b) SV agreement is in Guarani (also an outsider morpheme). In (9), we see insertion of a Spanish adjective fulfilling a predicate function in the Guarani clause.

(7)  
\[\text{Todo tu(s) prima(s) e(s)tá(n) todo(s) en ciudá(d), o-trabaja-porã...}\]
\[\text{all your cousins.fem they.are all in city 3-work-well}\]
\[\text{Guarañol-Spanish}\]
\[\text{‘All your (female) cousins, they are all in the city, they have good jobs…’} \]
\[\text{Ayala de Michelagnoli 1989: 33}\]

(8)  
\[\text{A-de(s)confiá luego de ese pitogüé…}\]
\[\text{1sg-mistrust after of that kiskadee}\]
\[\text{Guarañol-Spanish}\]
\[\text{‘I am mistrustful of that bird of ill omen…’} \]
\[\text{Ayala de Michelagnoli 1989: 18}\]
We must mention that many of these cases are actually integrated borrowings in modern Paraguayan Guarani that have almost completely replaced their original Guarani counterparts. That is the case for deconfiá, or vale and gutó below. However, we do not consider borrowings and CS to be necessarily different. Muysken (2000: 274) explicitly states that lexical borrowing and relexification “rely primarily on insertion,” and defends a unified grammatical explanation for borrowing and CS. Myers-Scotton (2002) also defends a position in which borrowing is not antithetical to CS, highlighting CS as a process yielding borrowing as a result. It is quite possible that early insertional CS of Spanish words by Guarani-dominant speakers led to replacement of the original Guarani expression (where it existed) in some cases. This would also explain why these Spanish words appear with Guarani morphology, since Guarani speakers’ competence with Spanish morphology was probably not very high in the earlier stages of contact.

4.1.1.2 Guarani negation on inserted Spanish verbs

These cases of Spanish insertions in a Guarani ML clause (in (10) and (11)) involve the Guarani negation circumfix n(d)+ … +(r)i. These clauses have ML Guarani because the negation and the SV agreement prefixes are Guarani.

(10) Nd-ai- gu(s)to -i-ete .
    NEG-1sg-please-NEG-very
    ‘I don’t like that at all.’ (Ayala de Michelagnoli 1989: 38)

(11) n-a- enterde -i e(s)e capricho.
    NEG-1sg-understand-NEG this whim
    ‘I don’t understand this whim’ (Ayala de Michelagnoli 1989: 36)

4.1.1.3 Modification by the suffix -ete

The suffix -ete is an intensifier that can mean “true” or “very.” Modification by this suffix is mostly observed in negative constructions in RQ, although there are a few instances with positive polarity. Instances in connection with the insertion of a Spanish word or constituent in a Guarani clause are shown in (12) and (13).

17. (Ni)ko is one of Guarani’s emphatic markers.
18. A similar position is defended by Mahootian (2005).
(12) *Sapy’a che-ofendé grati-ete*.

suddenly 1sg-offend for.free-very

‘All of a sudden, he offends me totally for nothing.’

(Ayala de Michelagnoli 1989: 63)

(13) *Tarde-ete a-menda*.

late-very 1sg-marry

‘I married very late.’

(Ayala de Michelagnoli 1989: 69)

There are examples, however, where the affix alone is inserted in a Spanish clause, as in (14) (note that here the TAM markers all come from Spanish), which may be more appropriately seen as cases of backflagging, since *-ete* is not a content word morphologically integrated in the clause and equivalent to a Spanish lexeme.

(14) *Quiero-ete que sea(s) felí(z)…*

I.want-very that you.be happy

‘I really want you to be happy…’

(Ayala de Michelagnoli 1989: 25)

4.1.1.4 Guaraní -ma suffix

The suffix *-ma* is used when an event has been completely finalized, or with the meaning “already.” The examples we give involve insertions of Spanish nouns or verbs in Guaraní clauses (since SV agreement and negation are from Guaraní).

(15) *n-o-aguanta-vé-i-ma tejuruguái.*

neg-3-put.up.with-CMPR-NEG-already whip

‘They don’t put up with the whip (of oppression) anymore.’

(Ayala de Michelagnoli 1989: 81)

(16) *Heta mitá porã re-reclutá-ma*.

many children good 2sg-recruit-already

‘You have already recruited many pretty girls.’

(Ayala de Michelagnoli 1989: 82)

4.1.1.5 Future marker -ta

The suffix *-ta* is obligatory to mark future tense in Guaraní. (Unmarked roots can be present or past.)

(17) *Ni radio ni diario o-solucioná-ta la ñande situación.*

nor radio nor newspaper 3.solve-FUT the our situation

‘Neither the radio nor the newspapers will solve our situation.’

(Ayala de Michelagnoli 1989: 28)

To summarize, person prefixes, negation, and aspectual and temporal markers are all found affixed to Spanish bases. It is noteworthy that these do not appear “doubled” by the Spanish counterparts, a criterion that would have otherwise signaled their status as alternations instead of insertions (according to Muysken's system).

4.1.2 Insertions without morphological integration

These are almost always Guarani insertions in Spanish clauses (but see 4.1.2.5), which contrary to Spanish insertions are never morphologically integrated. Examples like these, together with those discussed as backflagging in Section 4.4, have prompted the use of the label Castení introduced in Section 1.

4.1.2.1 Insertion of Guarani N/NP

The examples in (18) through (20) show insertion of Guarani nouns in Spanish clauses. Example (21) shows that a nominal constituent bigger than a noun can be inserted (here an NP).

(18) ¿Vo(s) piko no sabé(s) que kuimba'e e(s) como gallo? [Spanish-Guarani] 
you no you know that man is like rooster 
‘Don’t you know that man is like a rooster?’ (Ayala de Michelagnoli 1989: 16)

(19) ¡Dio(s) mío, me hace todo pý’a-jere! [Spanish-Guarani] 
God mine me makes all entrail-turn 
‘God, it makes my stomach turn!’ (Ayala de Michelagnoli 1989: 18)

(20) porque no saca su(s) ojo(s) de mi retyma [Spanish-Guarani] 
because no he takes his eyes of my leg 
‘…because he does not take his eyes off my legs’ 
(Ayala de Michelagnoli 1989: 18)

(21) I-mena nomá(s) e(s) demasiado jodido! [Spanish-Guarani] 
3-husband just is too screwed 
‘Only that her husband is too difficult!’ (Ayala de Michelagnoli 1989: 17)

4.1.2.3 Guarani V/VP insertions

Our single example of a V/VP insertion involves the Guarani verb oiko ‘is’. We take it to be an insertion because it is a single content word, clearly nested in between Spanish constituents.20

20. However, these cases highlight an issue with Muysken's use of Myers-Scotton's ML concept. One could argue that since these examples have Guarani SV agreement, they have Guarani ML in which case oiko cannot qualify as a switch.
(22) Su cabeza oiko en otra parte.  
her head is in other part  
‘Her head is somewhere else.’ 
(Ayala de Michelagnoli 1989: 21)

4.1.2.4 Adjective insertions
Adjectives are the only Spanish insertions without Guarani affixal morphology (23). This is expected, since Guarani does not mark gender on nouns or adjectives, or plural on adjectives.

(23) O-heka otro kuña avei!
3-seek other woman too  
‘He looks for other women too.’ 
(Ayala de Michelagnoli 1989: 18)

4.2 Backflagging strategy
The markers discussed in this section are not insertions. Insertions are usually content items (often selected by a predicate) that are used in “replacement” of an equivalent content item in the other language; however, for most of the morphemes discussed in this section, Spanish clearly has no equivalent. When a categorically equivalent item exists, as in the case of mante ‘only’, discussed below, often linear equivalence is absent (i.e., the equivalent words do not appear in the same position in the sentence). Backflagging in most cases involves the insertion of a discourse marker or other marker with sentential scope. The cases we discuss here correspond well to Muysken’s original formulation of this strategy. Nevertheless, we will show that an extension to other kinds of morphemes is needed.

4.2.1 Interrogative markers
Interrogative markers are obligatory in all questions in Guarani (which are not marked by prosody). Interestingly, almost all questions in RQ – whether in Guarani or Spanish – bear one of those markers.

(24) ¿Sabé(s) pa qué dice mi comagre?  
you.know Q what says my bosom.friend  
‘You know what my neighbor says?’ 
(Ayala de Michelagnoli 1989: 18)

(25) ¡Y qué piko le vamos(s) (a) hacé(r)!
and Q to.it we.go to do  
‘And what are we gonna do!’ 
(Ayala de Michelagnoli 1989: 17)

(26) ¿Por qué pa quiere mi des(s)gracia?  
why Q wants my misfortune  
‘Why does he wish me bad luck?’ 
(Ayala de Michelagnoli 1989: 17)
4.2.2 Emphatic and evidential markers

Guarani makes extensive use of emphatic and evidential markers in cases where Spanish (like English) simply would not use any means of conveying speaker perspective. These markers also surface often in Spanish sentences. The examples in (27) and (28) illustrate insertions of emphatic (ni)ko.

(27) A mí qué si anda por ella, no me hace niko falta. [Spanish-Guarani] to me what if goes by her no me makes EMPH lack ‘So what if he also pursues her, I don’t need him.’ (Ayala de Michelagnoli 1989: 16)

(28) Así ko nomá(s) son la(s) mujere(s) malagradecida(s). [Spanish-Guarani] thus EMPH just are the women ungrateful ‘Women are just like that, ungrateful.’ (Ayala de Michelagnoli 1989: 17)

Another word that can have an emphatic use is mante ‘only’, as in (29):

(29) Tiene que aguantá(r) mante. [Spanish-Guarani] she.has that stand only ‘She has to just take it without complaining.’ (Ayala de Michelagnoli 1989: 25)

As for evidentials, consider the Guarani reportative ndaje in (30):

(30) Dice luego que le bu(s)camo(s) a su hijo ndaje. [Spanish-Guarani] says after that him we.seek to his son it-is-said ‘She says that we are after his son.’ (Ayala de Michelagnoli 1989: 25)

Another evidential marker used is ra’e (spelled rae in our examples). Velázquez-Castillo (2013) classifies it as a proximal mediative/cogitative marker, that is, marking that some piece of information has been arrived at by reasoning, and it has associated interrogative meaning.

(31) ¿Por qué pa no le hice caso a mi comagre rae? why Q no her paid attention to my bosom.friend before ‘Why hadn’t I paid attention to my bosom friend?’ (Ayala de Michelagnoli 1989: 18)

4.2.3 Tag and interjection insertions

Guarani tags (as in (32)) and interjections (as in (33) and (34)) can also occur in Spanish sentences. These examples are very similar to the original examples given by Musyken cited in (6).
(32) **O-pensá que cualquié(r) sortado le viola mba’e.** [Spanish-Guarani] 3-think that any soldier her rape thing ‘She thinks that some soldier will rape her or something.’ (Ayala de Michelagnoli 1989:29)

(33) …*si él e(s) fayuto por mí. ¡Cherejápe!* [Spanish-Guarani] if he is false for me leave.me ‘…if he cheats on me. Get out of here!’ (Ayala de Michelagnoli 1989:70)

(34) ¡A la pinta! ¡Ndee! [Spanish-Guarani] to the pint you ‘Doggone it! Wow!’ (Ayala de Michelagnoli 1989:56)

Interrogative, emphatic, and evidential markers, as well as tags and interjections, can straightforwardly be considered backflagging because they are simple, frequent, and clause-peripheral. In the next subsections we show that backflagging is also an attractive analysis for morphemes that do not fit the “discourse marker” label as conceived by Muysken, mainly because they are rather more integrated in the clause. We classify the following uses of Guarani morphemes as backflagging because of their perceived functions as signals of Paraguayan identity.

### 4.2.4 Guarani gua’u constructions

Guarani *gua’u* means ‘false’, ‘in jest’, or ‘product of fantasy’. It is often used as a predicate in conjunction with a complementizer, with a meaning similar to ‘pretend’ (35) or ‘don’t tell me that’. It also appears as a modifier, with a meaning similar to ‘so-called’ (36). In this case, it does not have clausal scope (as discourse markers do).

(35) **Gua’u que e(s)tá haciendo ge(s)tione(s) para saca(r)no(s) de Chacarita.** false that is doing paperwork for get.out-us of Chacarita [Spanish-Guarani] ‘She says that she is doing some paperwork to get us out of Chacarita.’ (Ayala de Michelagnoli 1989:19)

(36) **Me va mandá(r) gua’u exilio Clorinda-pe.** me goes send false exile Clorinda-loc [Spanish-Guarani] ‘He will send me, he says, to exile in Clorinda.’ (Ayala de Michelagnoli 1989:50)

### 4.2.5 Imperatives and requests

The Guarani suffix *-na* is often added to mitigate imperatives, occasionally with the optional *-mi.*
(37) Que caminá-na así. [Spanish-Guarani] 
that walk-IMP(polite) thus
‘Walk this way.’ (Ayala de Michelagnoli 1989: 30)

(38) Quitale-na esa e(s)tupidé(z) de su cabeza. [Spanish-Guarani] 
remove.it-IMP(polite) that stupid.thing from his head
‘Get that stupid idea out of his head.’ (Ayala de Michelagnoli 1989: 76)

4.2.6 Plural -kuéra
Many examples of Guarani -kuéra pluralization on Spanish lexemes are found in the RQ corpus. In (39), the sentence also shows CS by emphatic niko, strengthening the claim that this is a case of backflagging.

(39) Siempre niko e(s)toy sola para da(r) de comé(r) a mi(s) hijo-kuéra. 
always EMPH.I.am alone for give of to.eat to my child-pl
‘I am always alone to feed my children.’ [Spanish-Guarani] 
(Ayala de Michelagnoli 1989: 73)

However, (40) is more clearly a case of alternation, since the sentence continues in Guarani, triggered perhaps by the use of the plural affix.

(40) No ve(n) que todo(s) mi(s) hermano-kuéra omano akue. 
no they.see that all my sibling-pl 3-die before
‘Don't they see that all my siblings died.’ [Spanish-Guarani] 
(Ayala de Michelagnoli 1989: 81)

4.2.7 Nominal temporal marker -kue
The past marker -kue (and its future counterpart -rã, not found in RQ) attaches to nominal bases to give a meaning usually equivalent to English ‘former’, as in (41).21

In RQ there is also an example of nominal modification with the variant akue (42).

(41) E(s) rancho-kue de mi papá. [Spanish-Guarani] 
is house-past of my dad
‘It was my dad's house.’ (Ayala de Michelagnoli 1989: 20)

(42) E(s) mi hijo akue de otro hombre. [Spanish-Guarani] 
is my son past of other man
‘He is the son I had before with another man.’ (Ayala de Michelagnoli 1989: 17)

21. See Tonhauser (2007) and Nordlinger and Sadler (2008) for extensive discussion of these markers.
4.2.8 Postpositional locative -pe

Finally, in RQ, locative -pe appears very frequently on Spanish bases, with directional and stative meanings. In (43) and (44), it heads an adjunct phrase:

(43) que murió guerra-pe
    that died  war-loc
    ‘that died in the war’
    [Spanish-Guarani][Ayala de Michelagnoli 1989: 20]

(44) Portón-pe e(s)tá ese tu comagre.
    gate-loc is  that your bosom.friend
    ‘That friend of yours is at the gate.’
    [Spanish-Guarani][Ayala de Michelagnoli 1989: 21]

In (45) and (46), the postpositional -pe phrase is subcategorized for by the predicate, giving rise to a composite ML clause in Myers-Scotton’s framework.

(45) Vamo(s) sin falta infierno-pe.
     we.go without fail  hell-loc
     ‘We will certainly go to hell.’
     [Spanish-Guarani][Ayala de Michelagnoli 1989: 18]

(46) De(s)de ese día que entra la cárcel-pe
     since that day that enters the jail-loc
     ‘from that day that he went to jail’
     [Spanish-Guarani][Ayala de Michelagnoli 1989: 22]

A possible alternative to the analysis of these clauses as Spanish (with backflagging switches to Guarani) is to take them to be composed of two insertions, as follows: the -pe headed Guarani phrase is inserted in the Spanish sentence, and at the same time a Spanish noun or NP is inserted in the Guarani postposition. This alternative analysis has the advantage of explaining why the Spanish prepositions are omitted. Its disadvantage is that it posits two switches where only one is evident.

4.3 Alternational strategy

If syntactic subordination or embedding is key in insertions, lack thereof suggests alternation. Alternations also tend to occur at clause boundaries. (If not, a congruent lexicalization analysis is sometimes possible.) In some cases (as in (48) below), the alternation seems to be triggered by the occurrence of a word in the other language (juky). This strategy occurs in RQ, but only sporadically.

(47) ¡Marã piko che-mombe’u si no podemo remedíá!
    why  1sg-tell if no we.can remedy
    ‘Why are you telling me if we can't fix it!’
    [Guarani-Spanish][Ayala de Michelagnoli 1989: 17]
5. Conclusion

In this paper we analyzed Guarani-Spanish mixing, or Jopara, as portrayed in the first novel ever written in this variety, *Ramona Quebranto* (RQ). We applied Muysken’s (1997; 2000; 2013) typology and showed that Jopara is the result of insertional and backflagging strategies, with recourse to secondary alternation strategies. We have also clarified the import of the theoretical dichotomy Guarañol and Castení, recently proposed in the literature to explain the gamut of possible realizations of Jopara. We find that the use of insertional strategies in a Guarani ML clause leads to productions resembling a “Guarani with Hispanisms” (Fernández Guizzetti 1966), or Guarañol. However, Castení is not the exact mirror image of Guarañol, since it does not derive from Guarani insertions, but rather from the extensive use of backflagging by Guarani discourse markers or affixes in Spanish clauses.

22. This integrated borrowing of *la* from Spanish is Guarani’s singular determiner. The plural is *lo*.
Additionally, and most importantly for the goals of this volume, we believe our work strongly argues in favor of using corpus research with literary works that contain CS alongside with vernacular speech. By successfully applying a framework designed for spoken LM/CS to a literary text, we have provided evidence that these tools can be extended to literary written LM/CS under certain conditions that guarantee authenticity. Moreover, as our analysis of backflagging cases shows, textual LM/CS can suggest additions or corrections to already existing theories of language mixing. Finally, engaging more forcefully and fully with literary LM/CS sources would enable researchers to conduct extremely fruitful comparisons between spontaneously produced LM/CS and other linguistic productions with varying degrees of planning, monitoring, and editing.

References


Appendix.  Abbreviations not in the standard Leipzig list

CPMR  Comparative (¬ve)

Address for correspondence

Bruno Estigarribia
University of North Carolina at Chapel Hill
Department of Romance Studies
Dey Hall, Room 332
Chapel Hill, NC 27599
USA
estigarr@email.unc.edu

Co-author information

Zachary Wilkins
Artificial Solutions
zacharyawilkins@gmail.com