Noun Phrases in mixed Martinican Creole and French: Evidence for an Underspecified Language Model

Pascal Vaillant

Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé, LIMICS, INSERM, Sorbonne Université, F-93000, Bobigny, France

vaillant@univ-paris13.fr

Abstract

Contact between French and Martinican Creole (MC) takes place in a society where bilingualism is the standard, in a situation of constant language mixing. French and MC, although related, show significant typological divergences on some specific features, e.g. the order between noun and definite determiner in the noun phrase, or the use of a linker to mark a possessive embedded noun phrase. In this paper, I explore the possible combination of the different values of these features in mixed noun phrases occurring in corpora. I inquire about the possible parameters which may influence the outcome and explain the relative frequencies of these different combinations. It appears that there is a partially common pool of elementary structures. Many utterances fall into the category termed by Muysken (2000) 'congruent lexicalization'. I also observe that apparent complex double embeddings have an internal logic, as they result from adjunction of multi-word modifiers. Finally I propose a model which accounts for the observed occurrences by postulating a level in the speech generation process where language itself is underspecified, and where it is in a position to be specified on the fly by contextual factors, coming either from the lexicon or from the constructional frame.

Keywords

Language Mixing, Noun Phrase, Martinique, Creole, French, Mixed Language Model, Constraints on Code-switching

1. Introduction and context: French and Creole in the Martinique

The island of Martinique, in the West Indies, is a place where a majority of the population commonly speak two languages: Martinican Creole and French.

1.1. French and Martinican Creole (MC)

French is an Indo-European language of the Romance family, in all estimates one of the four languages with the largest community of speakers in the world today. French mainland dialects resulted from the continuous evolution of Vulgar Latin in the former Gallic Provinces of the Roman Empire, with the influence of contacts with various other languages, from the 4th century up to the present day (Brunot & Bruneau 1961; Wartburg 1946/1969).

French — or rather, a mixture of French regional varieties, and this is a distinction which deserves to be stressed (Chaudenson 2001) — has been exported to other continents by French settlers between the 16th and the 19th century. Its presence in Martinique dates back to 1635. Since that time, French has gone through a process of official standardization in the 17th century, and is taught to all children at school in French regions (including overseas regions like Martinique) since the end of the 19th century. Vernacular French spoken in Martinique also has evolved since that time from the original mix, in the context of the language contact situation (see below, Section 1.2), and exhibits some differences with

standard official French; but for the purpose of tagging linguistic features to compare them, we will use the latter as the standard.

Martinican Creole (MC), like other French based creoles in the same area, emerged during the 18th century, at a time when French colonists, settled in the Caribbean islands since the half of the 17th century, started to massively import slave workers from different regions of West Africa (Valdman 1978; Chaudenson 2001). MC went through a period of chaotic emergence in the 18th century, as the population of the island underwent rapid change as a function of the slave trade: the population of black slaves of African descent, roughly equal in number to that of European settlers or indentured labourers in the middle of the 17th century, before the rise of the sugar economy (Sainton et al. 2004), was multiplied by 8 between 1682 and 1804, whilst during the same period the white population was multiplied by 2, with a strict social segregation established between the two populations (Elisabeth 2003). The continuous flow of new speakers during this period led to a restructuring of French into a new linguistic system, a process which is very often taken to be one definition of a creole language ('a language that has come into existence at a point in time', Muysken and Smith 1995a).

The two languages share a great part of their lexical stock, and some typological features (e.g. both are mostly head initial with SVO clause structure and prepositions), but MC is devoid of any flexional morphology, and there are also many differences in other syntactic features. There has been some debate as to whether these many differences should lead one to interpret the restructuring process of creolization as a kind of fast evolution (emphasizing the role of French) or as relexification (emphasizing the role of substrate African languages), but this discussion is beyond the scope of the present article.

1.2. Language contact situation

Martinican Creole has a history of continuous contact with French, from its beginnings to the present day. There is no evidence whatsoever of any historical moment in which MC speakers might have had no contact with French. However, the sociolinguistic circumstances of the contact situation have evolved over the past century. In rural Martinique in the mid-19th century, it is reasonable to suppose that French was restricted to a small subset of social interactions (among educated, urban people, in some specific communication contexts). But from the end of the 19th century on, French has gained ground in most layers of the population. In 1910 it was estimated that 60% of the adult male population, and 70% of the adult female population, was illiterate in French (Fauquet 1912). A century later the rates were respectively 16% and 13% (Coupin & Forgeot 2008). Before the 1960s, French elementary school (compulsory since 1882), and the military conscription (drafting young adult males for many months in other regions of France), may have played a major role in the familiarity with French of a growing proportion of the Martinican population, without much affecting the use of MC in everyday life.

Since then, other variables have shifted the language balance situation. The decline of the sugar industry (only partly replaced by the culture of banana) has led to a decline in the proportion of people living in rural areas: cane sugar production, for three centuries the major production of the island's economy, declined by 71.5% between 1960 and 1970 (Nicolas 1998). In parallel, since 1963, the French government has promoted an active policy of emigration of adult workers from French overseas territories to mainland France (BUMIDOM); this planned emigration has concerned a total of 42,000 people from Martinique between 1962 and 1981 (Milia-Marie-Luce 2007) (from a net population of 292,000 in 1961), and has led many families to found offshoots in major urban centres in France, with second generation cousins speaking mostly French (Anselin 1990). The growing influence of the mass media, with French as their main language, has also played a role in the pervasive presence of French. Finally, MC has suffered, like many regional languages in mainland France, from a strong feeling of inferiority that was first brought into the families via the school system, and then internalized by the people themselves: the prejudice of inferiority has remained some decades after any official discourse

has ceased to convey it. In a survey, March (1996) shows that a majority of Martinican mothers chose to speak only French to their children, even if they were competent in MC, and even if they would not explicitly state any negative valuation about MC, because they felt it was a better choice for their children's school performance[1]. As a consequence, more and more children (and mostly male children) start growing their competence in MC at middle school age, when it becomes a marker of manhood (Murray 1997); and as a corollary consequence, in the youngest generations of adults, MC has been increasingly marked as a language reserved for the male. In the neighbouring island of Guadeloupe (where Creole is purportedly more widespread than in Martinique), Pustka (2006) has described how in two generations, in the second half of the 20th century, French has become the first language of most children, even in families where grandparents were monolingual Creole speakers.

During the same period however (from 1960 to present), MC has gained ground in other fields. Utterly clandestine in the public media (radio mostly) in the first two decades of their existence in Martinique, it began to assert itself as a language of political expression in the workers' unions of the 1970s, at a time of social claims and political struggles linked to the crisis of a declining traditional rural economy, and some years later it even earned a legitimate place in broadcasts on public radio (Pulvar 2005). In the general context of the emergence of post-colonial studies, and of renewed interest in regional languages and cultures in many regions of France and other European countries, numerous attempts have been made to promote MC as a language for uses that were traditionally reserved to French: literature, popular science, political discourse.

The resulting situation is one of generalized bilingualism, where the sociolinguistic and pragmatic borders of the respective niches of French and MC are less clear-cut, and in any case different (in the age or gender specializations), from what they used to be when the reference descriptions were given, three decades ago.

According to Reutner (2005), MC monolinguals are now hard to find, but the sociolinguistic situation does not look like a language attrition situation, since MC also has extended its perimeters of use, and the connotations traditionally assumed in diglossic situations (one language associated with social distance and formality, one language associated with familiarity and vulgarity) do not apply. Pustka (2006) considers that the language contact situation in Guadeloupe and Martinique is typical of the early stage of a process of language death. More recently, Managan, who assessed the language situation in Guadeloupe, found no evidence of strict functional compartmentalization of the two languages: 'both French and Krèyol may be heard in nearly **all** contexts' (2016: 270). She thinks it is too soon to tell whether language shift is underway, but notes that Creole is undergoing internal change: 'It is acquiring syntactic features and lexical items from French, which may indicate that decreolization is taking place' (ibid.: 278). Her account describes the situation in Martinique too, to a great extent.

1.3. Diglossia or Creole Continuum: what is in the language mix?

Two distinct salient models have been proposed to describe language contact situations between a creole and its lexifier language. One is centred on the concept of **diglossia**, used for decades to describe the linguistic situation of Greece or Arabic-speaking countries, and which was applied by Ferguson (1959) to contact between French and Creole in Haiti, and more generally by Valdman (1978) to contacts between French and French-based creoles. The term describes a situation where two related languages are used in a country, in very specific — and non-overlapping — sets of situations of use (usually characterized as an opposition high/formal *vs.* low/informal).

1 This probably evidences not only a groundless despise against Creole, but also groundless prejudices about child bilingualism.

Another model for creole language contact situations is centred on the concept of **continuum**, such as described by DeCamp (1971) and extended by Bickerton (1973), originally about the set of speech varieties used in (British) Guyana. It captures the virtually continuous variation along an axis (thought of as unidimendional) lying between two extremes, an **acrolect** (English) on one end, and a **basilect** (radical Creole) on the other end. Speakers have command over a given segment of the whole continuum, and in the available span, choose the way they express themselves depending on pragmatic and social circumstances. The notion of a (post-) creole continuum is often associated with the idea that there is a slow diachronic evolution towards the acrolect, the basilect tending to disappear, and that the observed variation is like a geological stratum, displaying the temporary coexistence of different stages along the path leading to a merger of the creole into the lexifier language (Bickerton 1973; de Rooij 1995).

Traditionally, the notion of continuum has not been considered appropriate for describing situations in French creole areas like Haiti or Martinique, the two languages involved being seen as too far away from each other, and the contexts of use too clearly distinct (DeCamp 1971). Lefebvre (1974) concluded from a field study that there was no such thing as a continuum in Martinique, because the speakers seemed to be perfectly secure in basilectal Creole, but even in 1974, her conclusions may have been influenced by the test material chosen and the instructions of the experiment (telling local folk tales). A few years later, based on the observation of informal conversations, Prudent (1981) challenged this claim and postulated an interlectal zone between French and Creole.

At any rate, the notion of continuum has a flaw. It assumes that linguistic variation occurs along a single axis, where the lexifier language and the (basilectal) creole language are two idealised endpoints. It also assumes that the lexifier language is a frozen homogeneous language system, and that any variation simply amounts to making the linguistic variety ‘more’ or ‘less’ creole. In fact, though, in Martinique, there is variation both in MC and in French.

A first undebatable observation is that some linguistic features of French are open to variation, and that not all of their possible alternative values mirror the one described as the French standard. This holds for Martinique as for any country or region where French is used as a vernacular language, and in this meaning, variation cannot be represented on a single axis. This also does not imply that variation in these linguistic features is necessarily specific to Martinique. For instance, Roberts (2014) has chosen to focus on three variable grammatical features (the doubling of subject NPs, the omission of the negative particle *ne*, and the frequent expression of future temporal reference by an analytic future tense using the auxiliary verb *aller*) and has studied their use in Martinique French. Unsurprisingly for anyone who knows vernacular French, there is no significant differences, in these three features, between French spoken in Martinique and French spoken in mainland France.

There also are features that are characteristic of regional varieties of French in the French West Indies (Guadeloupe and Martinique). Several studies have given descriptions of these features, including Hazaël-Massieux & Hazaël-Massieux (1996), Telchid (1997), Sobotta (2006), Pustka (2007). They mostly insist on phonological and lexical features, but there are also regional features in morphology and syntax. A typical example is the structure of the causative construction (Thibault 2018): *je fais les enfants manger* instead of *je fais manger les enfants* in standard French (‘I make the children eat’). A few examples also concern the internal structure of the NP, and so they must be considered specifically in the context of this study (see below, Section 2).

There is also internal variation within MC, on many features. Many of them parallel the variations observed in regional spoken French: for instance, there is great variability on the use of relative pronouns to introduce subordinate clauses (Lengrai et al. 2006), and of course also in the NP (see Section 2).

In sum, an accurate understanding of the language situation in Martinique cannot be provided by the models of diglossia or continuum. There is heavy code mixing (amply shown in the corpus), but also code copying, between two language systems. Both have internal variation and are in a situation of mutual influence reinforced by the high proportion of shared lexical roots. So in most contexts, MC shows an extensive influence of French, much more so than in the neighbouring country of St-Lucia, for example (St-Hilaire 2013: 15); while at the same time, regional French shows specificities which are easy to interpret as the result of an influence of Creole (Sobotta 2006: 8-13).

Bernabé (1983) proposed a mixed explanation ('continuum-discontinuum') that amounts to acknowledging the fact that there is variation in both MC **and** French, while still recognizing the existence of two distinct language systems. This conservative model is a good starting point for analysing data in the Martinican language contact situation.

1.4. Data used for the present study

In the present study, I mainly rely on data recorded on radio broadcasts in Martinique in 2005-2006. The corpus has been recorded and transcribed by two masters students, and annotated by the author. Some texts are transcriptions of a discourse by a single speaker, and some of conversations involving multiple speakers. Some are discourses of popular science (that would be expected to involve a heavier influence from French), and some are related to more common daily-life topics (e.g. speaking habits, children's performance at school).

For the quantitative analyses, I have extracted a sample of 743 occurrences (tokens) of noun phrases representing 500 distinct noun phrases (types). When not otherwise explicitly stated, the figures given are percentages of the total number of types (500), not of the number of occurrences. The sample has been drawn from 5 distinct transcriptions (Table 1), where the speakers all are fluent Martinican Creole speakers. Roughly the same number of distinct noun phrases has been extracted from each of the 5 texts (a little more than 100, because some common noun phrases occur in more than one text). They are the first ~100 NPs in the annotated transcriptions, which occur, depending on the text, in the first 3 to the first 8 minutes of the recording.

Text	Title	Genre	# Speakers	# NP occurrences	# NP types
1	*Les jambes lourdes*	Popular science	1	139	103
2	*Les reins et la vessie*	Popular science	1	133	103
3	*Le fossoyeur Kokiyol*	Life story interview	2	145	102
4	*Le créole dans les médias*	Debate about social issue	6	183	105
5	*Annou kozé divini yich nou*	Debate about social issue	5	143	101
Total			**14**	**743**	**500**

Table 1. *Sample of noun phrases extracted from the annotated corpora.*

2. Possession and determination in the Noun Phrase

In this article, I shall concentrate on the expression of possessive (genitive) relations, of definite or indefinite determination, and of the combinations between the two of them within the noun phrases. I use common-ground concepts of linguistic typology, in terms of Croft (2002a).

2.1. Possession

The relation of possession (or, to be more general, genitive) between two nouns can be expressed in different ways across languages of the world, including simple juxtaposition ('N_2 N_1', 'N_1 N_2')[2], the use of

2 N_1 is conventionally used for the possessum, N_2 for the possessor.

an independent morpheme (English 'N_1 *of* N_2', Chinese 'N_2 *de* N_1'), or the use of an indexical morpheme attached to one of the nouns (genitive in Indo-European languages, construct state in Hebrew).

In standard contemporary French, possessive is mainly expressed by the preposition *de*. Between two nouns, it is the norm: 'N_1 *de* N_2'. 'N_1 *à* N_2' is also fairly common in spoken French. It probably has something to do with the genesis of the possessive joiner 'N_1 *a* N_2' in Guadeloupean Creole. However, it does not exist in MC, and is not strikingly more common in regional French than in any variety of spoken French. It is absent from our corpus data. Additionally, some 'N_1 N_2' constructions may exist in specific contexts in spoken French (*une pause café* 'a coffee break'); they express noun specification rather than possession. The frequency of 'N N' constructions has been reported as typical of regional French in the French West Indies (Hazaël-Massieux & Hazaël-Massieux 1996), but the examples given often are set expressions that may be analysed as Creole calques (*punch coco*, *pomme cythère*, *abricot pays*). There are 4 such constructions on our 500 sample NPs, and they all appear in a fairly unambiguously MC context. So, globally, it is safe to assume that in the context of our corpus, the expression of possession in French is the one that is referenced as the French standard.

When the determiner (possessor) is a personal pronoun, French makes use of a preposed possessive determiner indexing the person of the possessor, and the number and gender of the possessum: Prsn$_2$ N_1.

In MC, according to descriptions of the basilectal variety (Bernabé 2003), there is no adposition to express genitive. Possession is expressed by juxtaposition. The order used is 'N_1 N_2', in all cases. It is the same when the possessor is a noun and when it is a personal pronoun ('N_1 Prsn$_2$'). In the latter case, unlike in French, there is no specific possessive form: the personal pronoun in a possessor's role has the same form as when it is used alone as a predicate argument (*kay Manman* [house mother] 'mother's house'; *kay mwen* [house PRN.1SG] 'my house').

2.2. Determination

In contrast to the relation of possession, which has to be expressed some way or another in all languages, noun determination, and in particular definiteness, is not a semantic universal. In many European languages it is encoded in a specific part of speech known as an article. But as noted by Heine & Kuteva:

> The majority of the world's languages do not have definite and/or indefinite articles. According to a survey carried out by Dryer (1989), about two-thirds of the languages of the world do not employ articles, and less than 8% of the languages of his sample (…) have both definite and indefinite article (2006: 98).

Heine & Kuteva give the following tentative definition: 'the use of articles is anchored in general human conceptualization capacities such as foregrounding (or individuating) of an entity against the background of a developing "textual" world or a shared situational world: both the definite and the indefinite articles individuate an entity out of a group of entities. The function of the definite articles is to individuate qualitatively' (2006: 97).

A traditional account of indefinite and definite determination, in the language of logical semantics, has been proposed by Russell (1905): it posits that indefiniteness presupposes existence, while definiteness expresses both existence and uniqueness.

More recent works in formal semantics looked for ways to model the anaphorical value of definiteness: in **Discourse Representation Theory**, for example, indefinite is defined as the introduction of a new referent in the discourse world, while definite is defined as referring to an already-existing referent (Kamp et al. 2011). Another way to look at the problem of definiteness made use of the notion of

hierarchy of salience: ' "the F" denotes x if and only if x is the most salient F in the domain of discourse, according to some contextually determined salience ranking' (Lewis 1979: 348).

Of course 'foregrounding' a particular referent in the discourse is something that can be done in any language, but this does not imply that there must be a particular grammaticalized way of doing so, 'other means being in particular word order, sentence stress, case oppositions, verbal agreement suffixes, or adjectival suffixes' (Heine & Kuteva 2006: 97).

Now, in as much as there is a grammatical category for determination, and a specialized morpheme used for overt expression of definite or indefinite determination on nouns, and leaving aside the (less frequent) case where determination is carried by another unit than the head noun, there leaves the possibility for the specialized morpheme (typically, the article) to be before or after the noun, and to be either free or bound (D N, N D, DN or ND).

In French, following the classical grammar tradition (and school grammar books), the noun may have four degrees of determination: zero (*elle travaille avec soin* 'she works carefully'), indefinite [INDF] (*un chien* 'a dog'), definite [DEF] (*le chien* 'the dog [you and I know about]'), and partitive [PART] (du pain 'some [quantity of] bread') (Brunot & Bruneau 1961 : 214, 227).

In contemporary French, the set of constructions where a noun may appear with no article is limited to some specific functions: within a prepositional phrase used as an adverbial clause modifier or noun modifier, as an qualifying attribute with a copula, and in a few other frozen V N contexts. When a noun is the head of a complete noun phrase that has the role of a predicate argument (e.g., subject, or direct object), an overt article is basically compulsory. However, this was not the case in earlier stages of the language (until the period of classical French, viz. approximately until the end of the 17th century), which explains the presence of bare nouns in set expressions or proverbs.

In French, both the indefinite and definite article come before the noun (D N). Both also index the category of number, and the category of gender — only in the singular. The article occupies the leftmost position in a NP, which means that if some prenominal modifiers are adjoined (such as preposed adjectives), they are inserted between the article and the noun (*un gros chat* 'a big cat').

In addition to the basic, 'Russellian', values of the indefinite and definite article, French makes use of both articles with a generic value (*Le cheval est un mammifère* 'horses are mammals'; *Un cheval retrouve toujours son chemin* 'a horse always finds its way back home'). The nuance between the two has been explained by Guillaume (1919) by saying that the definite article expresses a 'movement of thought from the singular to the general', and the indefinite a 'movement of thought from the general to the singular' (e.g. within an implicit syllogism). In most contexts, the former would also not be translated by a definite article in English, but rather by a generic bare noun or a plural. This nuance is described as the difference between **kind-referring** genericity, as opposed to **characterizing sentences** (Krifka et al. 1995)[3]. As many authors have noted (e.g. Corblin 2011), the first type of genericity is a typical value of the definite article in French, called 'intensional' definite.

Finally, let us include in this description the demonstrative determiner (traditionally classified, in French grammar textbooks, as an 'adjectif démonstratif'), which works much the same way as articles do. It indexes the categories of number and gender. It carries a deictic value which may range from strong (defined in the situation context) to weak (defined in the discourse context). It is a broadly acknowledged fact that the border between weak deictic value and definiteness is tenuous, which also explains why in

3 In terms of denotational semantics, if a sortal noun refers to an entity type <e>, the 'kind-referring' type of genericity (*The potato was first cultivated in South America*) is described by λP (P(e)), while the type encountered in characterizing sentences (*A potato contains vitamin C and amino-acids*) is described by λP (∀*x* (IS_A(*x*,e) ⇒ P(*x*)).

languages which mark definite determination, the morpheme often has evolved from a deictic marker (Heine & Kuteva 2002: 109-111).

The demonstrative is often reinforced by post-nominal invariable locative adverbs specifying either proximal (*ci*) or distal (*là*) deixis (*ce verre-ci est à moi, ce verre-là est à toi* 'this glass is mine, that one is yours'), where the locative value has been bleached out)[4].

Table 2 summarizes the grammaticalized system of determination in French.

determiner		zero	indefinite	definite	partitive	demonstrative
singular	masculine	Ø	*un*	*le*	*du*	*ce*
	feminine	Ø	*une*	*la*	*de la*	*cette*
plural		Ø	*des*	*les*	*des*	*ces*
value(s)		(seldom used)	indefinite generic	definite generic	partitive	deictic

Table 2. *Expression of determination in Standard contemporary French*

In MC, the category of determination is not parallel to the one of Standard French.

The structure of the Noun Phrase in French-lexifier Creoles (FLC) has been extensively studied in earlier works. In a cross-Creole perspective, Déprez (2001) made a convincing case for a common architecture in FLC, involving an underlying hierarchy with an article (definite or indefinite) at the top, an intermediate demonstrative, and, at the bottom (closer to the noun), an optional plural marker (DP > DemP > NumP > NP). She accounts for the surface variation in the order of the markers, in a minimalist frame, by arguing that some of them (most notably, the definite article) are more heads than specifiers (the noun phrase hence being a DP rather than an NP), a fact which triggers NP movement to the front of the determiner. In (Déprez 2008), she connects this property to consequences of grammaticalization and semantic bleaching. In a less theoretical frame, Vaillant (2008) proposes a unification-based model to factor common properties of the NP architecture across French-lexifier Creoles of the Atlantic area.

Reference grammars (e.g. Bernabé 2003) identify two types of articles: an indefinite and a definite article. Vaillant (2008) identifies four levels in a closed class: zero, indefinite, definite and demonstrative.

The indefinite article in MC is an invariable morpheme, *an* (pronounced [ã]), preposed to the noun (DN). Like in French, it has to be the first unit in the NP, which means that adjoined preposed modifiers are inserted between the indefinite article and the noun (*an vié rèv* 'an unpleasant dream'). In terms of X-bar theory, the article is positioned at the left of the N', not merely of the N. Unlike in French, but like in English, the indefinite article does not have the same form as the numeral 'one' (*yon*). It is used only in the singular.

According to Bernabé (2003), there is no plural indefinite article: plural indefinite is expressed by bare nouns (see below). However, in an earlier work, Bernabé (1983: 633, 719) mentions a plural indefinite morpheme *dé*, used 'in cases where it is necessary to avoid an ambiguity between singular and plural', since nouns don't bear a mark for plural (*Mi dé boug kouyon* 'What silly blokes!') (ibid.: 719).

4 The *-ci* and *-là* adverbs filled a gap after the disappearance of the proximal and distal demonstrative determiners of Old French *cist* and *cil* (Brunot and Bruneau 1961: 246).

As will become evident in the corpora I analysed, this morpheme *dé* is far from becoming extinct[5]. Unlike what Bernabé suggests, it appears in many contexts, not restricted to presentative constructions. But, unlike in French, it is not compulsory in noun phrases used as predicate arguments.

The definite article is a morpheme postposed to the noun (N D). Its basic form is *la*. It is a highly likely hypothesis that it was derived from the French deictic adverb *là* (there)[6]. In the case of Haitian Creole, Fattier (2000) has given a very convincing explanation of how West African substrate influence, in the first generations of slaves, could lead to reanalysis of the post-nominal French reinforcement deictic *là*, ultimately giving it the properties it has in Creole (post-nominal definite article and post-phrasal NP specifier— see footnote 6). Fattier's analysis of the origin of this morpheme in Haitian Creole are very plausibly applicable to Martinican Creole, since the language genetic and sociolinguistic parameters were similar during the Creole genesis period, before 1791 (Singler 1995).

This definite article bears no indexical mark of number or gender. Actually, it is questionable to refer to categories such as gender or number for nouns in MC, and it is quite possible that the grammatical descriptions that do so, do it because their authors unconsciously follow a model acquired by the prior study of the grammar of French. The lexical units in MC are invariable and bear no morphological mark whatsoever. When gender has to be explicitly specified (for example to give a precision on the gender of an animate being), it has to be expressed by an independent lexical morpheme meaning 'female' or 'male'. When plural has to be explicitly specified — which is anything but compulsory (see below about the use of bare nouns) — it is expressed by the use of an independent, invariable grammatical morpheme. Those independent morphemes need appear only once, as there is no phenomenon such as agreement.

However, the definite article has four different possible forms: *a*, *la*, *an* [ã] or *lan* [lã]. The choice among those forms is guided by a principle of harmony (or 'sandhi', Bernabé 1983: 645) between two phonological features: the forms beginning with [l] are used when the preceding word ends with a consonant, and the forms ending with [ã] are used when the preceding word ends with a nasal syllable. This alternation is not an agreement with an inner category of the noun, rather it is a phonological harmony with the last unit in the N': when this unit is not the noun itself but a modifier such as a postposed adjective or a possessor, it is often with this last unit that harmony occurs (*loto a* 'the car', *loto nèf la* 'the new car', *loto mwen an* 'my car').

In many European languages (for instance French or German), articles are polyfunctional units which carry not only determination, but also number, gender, and case. In MC (and more generally in

5 The fact that Bernabé mentions it in his work of 1983 and not in his work from 2003 is not to be interpreted as meaning that this plural indefinite is in the process of disappearing; it is linked to the difference between the descriptive nature of (Bernabé 1983), and the prescriptive nature of (Bernabé 2003). The latter aims at setting a language standard for MC, consciously seeking to keep a safe distance from French, to avoid 'decreolization'.

6 The locative deictic adverb *là* is widely known to be used as a post-nominal specifier (with locative and/or anaphorical value) in many varieties of French, including standard contemporary French. As a deictic specifier, its locative content is bleached out in most contexts, although it keeps the value 'distal' (see footnote 4) as opposed to *ci* ('proximal') (Heine & Kuteva 2007: 84). In contemporary French, as a reviewer of this article pointed out, it has become a **simple phrasal clitic**: it **encliticises** (rather than post-cliticises) to the last word of its NP-support. It is expected to appear only in combination with the prenominal demonstrative determiner *ce/celle*, although in spoken language it is also very frequently combined with the definite article. In many regional varieties of French, including most notably Picard and Québec French, but also popular Parisian French two centuries ago, there has been a long history of weakening of the marking of the categories of gender and number on the definite article, which has led to a rise of the use of demonstrative determiners (pre-nominal *ce*, *che* in Picard) and demonstrative adverbs (post-nominal *là*) in addition to the article (Wittmann 1995: 299; Barbaud 1998: 111-112; see also Conwell & Juilland 1963: 177-181 for examples in Louisiana French), which may have paved the way for a loss of the French preposed definite article in favour of the exclusive use of post-nominal specifier *là* (Wittmann 1996: 134) in Creole.

Caribbean French creoles[7]), two different morphemes are used to express number and definiteness, and they are used for that purpose only[8]. The morpheme used to express plural in MC, *sé*, is invariable and preposed to the noun. Like the indefinite article, it has to occupy the leftmost position in the NP. It is compatible with the definite article (*sé timoun lan* 'the kids'), and with the demonstrative (*sé timoun tala* 'those kids'), but not with the indefinite article.

In MC, the definite article is dropped when a noun is followed by another noun in a genitive complement, in cases where the second noun itself is followed by a definite: *mèt la* 'the holder'; *kay la* 'the house'; *mèt kay la* 'the householder' (not *mèt la kay la*) (Bernabé 1983: 749). This behaviour is propagated backward to every noun in a chain of genitive dependencies: *tianmay mèt kay la* 'the householder's children', *jwè tianmay mèt kay la* 'the householder's children's toy'.

However, the definite article has to be repeated at the rightmost position of the NP in the cases where a long adjoint like a relative clause is put after the noun (Bernabé 1983: 924). This means that in such cases, the definite article appears twice: firstly just after the noun and any postposed articles, and lastly after the relative clause. Bernabé (1983) calls this phenomenon a 'demarcating function' of the relative article.

Finally, a very common degree of determination in MC is the degree zero: bare nouns. Unlike in French and other 'SAE'[9] languages, bare nouns have a very broad range of use, and are able to constitute NP by themselves. They are the standard expression of the 'generic' value, as can be seen in their pervasive use in proverbs (*ravèt pa ka ni rèzon douvan poul* 'a roach does not get the better of a hen').

Krifka (2003) showed that in English, bare nouns denote intensional properties that can be interpreted either as kind-referring or as characterizing, depending on the context. The same could be said about MC, with an additional note: as Zribi-Hertz and Jean-Louis convincingly demonstrated (2014), MC has also developed a specific morpheme to express a particular type of 'intensional' genericity, historically based on the French preposed definite article (*l-, la-* or *lé*), but having lost its status of determiner ('lè difé pri, fòk kriyé *lé* ponpyé' = when a fire has broken out, one must call **the Fire Brigade** – not just some particular firemen). So, unlike in French, the 'intensional' type of genericity is not expressed in the same way as definiteness.

Bare nouns are the most common way of expressing the indefinite value (introducing new discourse entities) for plural referents: *sa sé bagay ki ka rivé* [DEM COP thing REL.SBJ IMPF happen] 'this is the type of things that happen'. They also are used in contexts where French would make use of a partitive article, that is, when mentioning a new discourse entity belonging to a 'mass noun' category (*ba mwen dlo* 'give me [some] water').

The demonstrative determiner in MC is an invariable postposed unit with the same syntactic properties as the postposed definite article; its surface form is always *tala*. It can be further analysed as a demonstrative morpheme and a definite morpheme, as a comparison with other French-based creoles endorses.

Table 3 summarizes the grammaticalized system of determination in MC.

7 A comparative sketch of the core structures of NP and VP in four different French Caribbean Creoles (Haiti, Guadeloupe, Martinique and French Guiana) is given in (Vaillant 2008).

8 This example of uniform mapping of one morpheme to one function is a case of **semantic transparency**, a notion which has been posited by some authors (Seuren & Wekker 1986) as being maximized in Creole languages genesis.

9 SAE: Standard Average European, borrowed from Whorf, used in Heine & Kuteva (2006).

determiner	zero	indefinite	definite	demonstrative
singular	Ø	*an N*	*N la*	*N tala* (or *ta'a*)
plural		*dé N*	*sé N la*	*sé N tala* (or *ta'a*)
value(s)	generic indefinite plural partitive	indefinite singular indefinite plural	definite ('plural' is expressed by another morpheme)	deictic ('plural' is expressed by another morpheme)

Table 3. *Expression of determination in Martinican Creole (MC)*

3. Noun Phrase configurations in language mixing

3.1. Congruent and non-congruent structures

As may be seen from the preceding section, despite the relatedness of the two languages, the NP systems for expressing genitive and determination in French and in MC are congruent only to a limited extent; the expression of indefinite determination is the same in the singular, and similar in the plural in some contexts only. In all other aspects, the two languages differ, in what is supposed to be their 'core' system. Table 4 sums up the comparison.

value	French	MC
N/GNR (generic noun)	INDF.SG N | DEF N	N (in some special cases: *lé* N)
N/PART (partitive)	*de* DEF N (or contracted form)	N
N/INDF/SG (indefinite SG)	INDF.SG N	INDF.SG N
N/INDF/PL (indefinite PL)	INDF.PL N	N | INDF.PL N
N/DEF/SG (definite SG)	DEF.SG N	N DEF
N/DEF/PL (definite PL)	DEF.PL N	PL N DEF
N/DEM/SG (deictic SG)	DEM.SG N	N DEM
N/DEM/PL (deictic PL)	DEM.PL N	PL N DEM
NP1/NP2 (genitive between full NPs, NP2 = possessor)	NP1 *de* NP2	NP1 NP2
NP/PP*n* (genitive between a full NP and a person index *n*)	POSS.*n* NP (possessive determiner)	NP PRN.*n* (full personal pronoun)

Table 4. *Comparison of French and MC in some typical cases of simple or nested constructions*

In this table, I try to compare similar basic semantic values — to the extent that they may be considered the same — and the constructions allowing their expression in French and MC. In the column 'value', I note a dependency relation by a slash (e.g. N/DEM), irrespective of the actual linear position of the different elements. When more than one expression is possible for a given value, the least marked is first.

In Martinique, French and MC are extensively mixed in everyday use. It is therefore interesting to observe what the speakers actually produce in mixed language contexts. In monolingual settings, the lexical items are in the same language that imposes the construction. If we suppose for instance that we have two different languages α and β, that α imposes the order DN and β the order ND, then the expected monolingual productions should be $[\ D_\alpha\ N_\alpha\ ]_\alpha$ and $[\ N_\beta\ D_\beta\ ]_\beta$.

In mixed language settings, a lexical item of one language may appear in a construction of another language. So, if all free combinations were allowed, we could find $[\ D_\alpha\ N_\beta\ ]_\alpha$ (structurally 'α-like' NP where the lexical matter of the noun is from language β) or $[\ D_\beta\ N_\alpha\ ]_\alpha$ (where the lexical matter of the determiner is from language β). Symmetrically, we could find $[\ N_\alpha\ D_\beta\ ]_\beta$ or $[\ N_\beta\ D_\alpha\ ]_\beta$. Additionally, if the

structure is permitted to be from one language and the lexical matter from another, it should also be possible to observe structures like [D_β N_β] $_\alpha$ or [N_α D_α] $_\beta$.

As Chan (2009: 193, 197) observes, in the case of diverging order in the two languages, two other possibilities are also theoretically allowable, even if more rarely observed: that the two possible orders are activated at the same time (doubling, which violates the economy constraint: [D_α N D_β]), or that neither of them is (deletion, which violates the isomorphism constraint: [N]). I will consider possible cases of doubling in Section 5.1.

3.2. Purported constraints on the possible combinations

Several hypotheses on the internal constraints inherent to language mixing predict that some of the combinations are very unlikely.

Sankoff and Poplack (1981), for instance, posit that bilingual speakers are subject to an implicit **equivalence constraint**, which requires that switches only be permitted at points where the grammars of the two languages are equivalent with regards to the surface syntax of the current constituent ('the order of the sentence constituents immediately adjacent to and on both sides of the switch point must be grammatical with respect to both languages involved simultaneously', (1981: 5)). Under this constraint, for example, occurrences of the patterns [D_α N_β], [D_β N_α], [N_α D_β], and [N_β D_α], in the situation described above (non-congruent order of NP constituents in α and β) should not be observed.

Di Sciullo, Muysken and Singh (1986) also think there are constraints on possible combinations, although they prefer to express these in terms of deep structure rather than in linear terms. Their proposition is that, apart from possible constraints specific to some language contact situations, there is one universal constraint controlling code-mixing occurrences, the **government constraint**: major syntactic categories impose their language on the syntactic nodes they govern ('if X has language index *q* and if it governs Y, Y must have language index *q* also', (1986: 5)). Since the authors' understanding of 'government' in this context is defined as minimal (not maximal) c-command, this constraint would not prevent switches between a D and a N in a NP, but it would require, for example, relative clauses to be in the same language as their governing noun.

Belazi, Rubin and Toribio (1994), observing that in subordinate clauses, the functional head (relative pronoun or complementizer) is more commonly in the same language as the subsequent clause, than in the same language as its governing noun or verb (counter to the prediction of Di Sciullo et al.'s government constraint), suggest that a universal constraint applying to intrasentential code switching is the **functional head constraint**: there can be no switch between a **functional** head and the following phrase, although switches may occur freely after a **lexical** head ('The language feature of the complement f-selected by a functional head, like all other relevant features, must match the corresponding features of that functional head', (1994: 228)). This constraint applies successfully to many observed utterances containing a subordinate clause, but it also predicts that no switches should be observed between a D and a N, even when the NP constituent orders are congruent in the two languages; however, many such switches are found in code switching corpora. Belazi et al. dismiss such counter-examples by explaining that in every one of these cases, what is observed is not an instance of code switching but one of borrowing. In the same article, the authors also assume a deeper universal constraint, which they call the **word-grammar integrity corollary**. It is a statement of the intuition arising from the very general observation that words with a complement structure have a local grammar which is drawn along from the same language as the word itself ('A word of language X, with grammar Gx, must obey grammar Gx', ibid.: 232). However, as Mahootian and Santorini noted, since this statement has not been expressed in terms of complement structure but in more general terms

('obey grammar Gx'), it 'reduces in effect to the well-known Equivalence Constraint' (1996: 469), and actually has the same set of counter-examples.

In a somewhat different approach, inspired by psycholinguistic insights, Myers-Scotton (1993) proposes a model where it is always possible to define, for a given fragment of discourse, a **matrix language** (ML) playing a predominant role in the mix of languages. Under this assumption, she contends that two principles apply in code switching productions: the **morpheme order principle** and the **system morpheme principle**. The former states that even when there are some insertions from an **embedded language** (EL), the overall morpheme order of a sentence is a valid order in the matrix language, except within embedded language 'islands', whose domain of locality are limited ('In ML+EL constituents consisting of singly-occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML', (1993: 83)). The latter states that system morphemes that reach at the sentence level are always imposed by the matrix language ('In ML+EL constituents, all system morphemes which have grammatical relations external to their head constituent (i.e. which participate in the sentence's thematic role grid) will come from the ML.', (ibid.: 83)).

The two principles posited by Myers-Scotton apply in some extent to the structure of NPs (1993: 85). The morpheme order (MO) principle predicts that 'in ML+EL constituents, morpheme order is that of the ML'. This applies only to mixed NPs, not to EL islands. The system morpheme (SM) principle makes predictions only for morphemes which have grammatical relations (or a domain of co-indexicality) outside the NPs —this is not frequently applicable to the internal structure of NPs in languages which do not use case-marking morphemes, but in the case studied here, it could apply to possessive determiners (which index the possessor). In this (1993) model, a NP may either be a full EL island (in which case its 'internal' system morphemes, namely the determiner, would have no reason to be ML), or a ML NP with an EL island appearing at a deeper level (N' or N), in which case the determiner should be ML. In short, (a) the combinations [D_α N_α] and [N_β D_β] are always possible; (b) the combination [D_α N_β] is only possible when α is the ML, and the combination [N_α D_β] is only possible when β is the ML; (c) the combinations [D_β N_α] and [N_β D_α] are never possible. In a later work on the same theoretical assumptions, Jake, Myers-Scotton and Gross add a quantitative prediction:

> The **Bilingual NP hypothesis**: the system morphemes in mixed NPs come from only one language, called the ML. An asymmetry between mixed NPs and full NPs from the EL obtains: full EL NPs are dispreferred because their system morphemes (and their uninterpretable features) do not match other system morphemes and their uninterpretable features elsewhere in the bilingual CP (Jake et al. 2002: 78-79).

If this contention is correct, the combinations [D_α N_α] and [N_β D_β] should seldom be observed when the ML is β or α, respectively (i.e. there should be few 'EL islands' NPs).

As to genitive constructions, the MLF (in its 1993 version) predicts that 'system morphemes' should be ML; but a finer expression of the constraints (Myers-Scotton & Jake 2000a), the '4-M model', allows the authors to make finer predictions. They distinguish four types of morphemes: two of them are 'conceptually activated': the **content morphemes** (morphemes with full lexical content, e.g. nouns, verbs), and the **early system morphemes**, that belong to the lemmas which come along with the content morphemes (e.g. adpositions that introduce the complements of some verbs), but unlike the content morphemes, cannot be assigned a thematic role. The two other types of morphemes are 'late' system morphemes. Here Myers-Scotton & Jake distinguish two types of late system morphemes. **Outsider system morphemes** have links outside their constituent (e.g. agreement markers), when **bridge system morphemes** are simply compelled by the internal constructional rules of the language of the constituent (and here a typical example is the joiner morpheme *of*).

Among the two types of late system morphemes, only the outsider system morphemes are expected to be ML, whilst the bridge system morphemes may as well be EL if the constituent is an EL island. This theoretical standpoint predicts that the French preposition *de*, being of the latter type, is allowed (and indeed expected) to be present if the NP with an internal genitive construction is a French NP (even when French is not the ML). The only morphemes in the French (or mixed) NPs which would count as 'outsider system morphemes' are possessive determiners: as noted above, these are expected to be in the Matrix Language.

The Matrix Language approach brings a limitation along: it requires the identification of the matrix language of any utterance — before being able to make predictions about its structure —, and to provide criteria for doing so (if it didn't, as Myers-Scotton herself acknowledges (1993: 66), it would be based on a circular definition). In (Myers-Scotton 1993: 68), the criterion proposed to identify the ML is based on morpheme frequency: the language with the most morphemes in a discourse span surrounding a given utterance should be considered the ML of that utterance. Since that criterion alone leaves room for ambiguity in some cases (the dominant language may evolve in the course of a conversation), an additional criterion is proposed in more recent works (Myers-Scotton & Jake 1995): the ML can be identified by a structural criterion at the level of every finite verbal clause, namely, as the language providing the grammatical frame of the CP (Complementizer Phrase[10]). This criterion, however, might well remain ambiguous in the case when the two languages involved are closely related, and congruent with respect to the structure of the verbal clause.

Authors who have proposed and discussed such grammatical constraints on code switching have offered theoretical arguments and empirical support. The matrix language frame model for instance has been tested independently by other researchers on different corpora, and has shown robustness in explaining a large amount of language mixing data (Myers-Scotton & Jake 2000b: 1). It has, specifically, been used to explain how the structure used to express possession in the noun phrase is determined by the language of the lexical item for the possessum, in the case of Ewe-English code-switching (Amuzu 2014).

However, it is still debated whether there actually is a need to hypothesize universal constraints on code-mixing, and researchers from the field of formal grammar have proposed that simply unifying the grammatical constraints of the languages involved in the mix can account for the observed data (and for the impossibility of non-occurring data) in every single specific language mixing situation. As MacSwan states, 'the history of CS reveals a common intuition among researchers that theories about CS should be free of grammatical mechanisms and constraints specific to it', (2014a: 18). Mahootian (1993) for instance expresses this intuition in terms of lexicalized Tree-Adjoining Grammars, proposing a 'null theory of codeswitching'. MacSwan (1999) expresses it in terms of Chomsky's Minimalist Program, with a view to proposing an operational model allowing the description of code-switched sentences and the explanation of why some switches are not possible, without appealing to constraints specific to code switching situations (Stabler and MacSwan 2014).

This family of explanations, that posits that it is possible to merge grammatical structures from two (or more) distinct languages with no overhanging constraint, is based on two axioms, which authors like Mahootian or MacSwan explicitly acknowledge. The first one is that there is no need to specify 'language' as a feature in any deep layer of generation (this clearly contradicts models such as Belazi, Rubin & Toribio's, or Myers-Scotton's, which make use of this language feature); in fact, the models of grammar used by Mahootian (L-TAGs) or MacSwan (Minimalism) are based on the idea that grammatical structures are anchored in the lexicon, and may be merged by some simple operations.

10 In formal grammar, a CP is a sentence with an optional complementizer (that may be Ø). This definition includes complete, finite sentences, as well as subordinate clauses which actually are sentences with a complementizer that allows them to fill an argument slot in verbs which require clausal complements (like 'that I know nothing').

The second axiom is that different languages may contain similar grammatical categories (in a maximalist view, universal categories; in a less demanding one, categories, in a given pair of languages, with enough structural and cognitive overlap to allow merging). Although there is debate about the validity of the claim that there should be universal part-of-speech categories (Vaillant 2014), the notion that some categories could be common to two languages or more ('congruence') does seem realistic, especially when languages are related. And even when they are not related, situations of prolonged contact and borrowing may cause varieties used in close contact situation to grow structures (grammatical categories, constructions, etc.) which display a high degree of congruence. Well-known examples include the case of the Arnhem Land (Australia) languages studied by Heath (1978), and of the Kupwar (India) languages studied by Gumperz and Wilson (1971).

Even more so than with grammatical categories, one may expect to find that some languages share common lexical items. As a matter of fact, in many language contact situations, it is possible for some constituents to be 'undecidable': i.e. it may happen that one does not know whether they should be considered as being elements of language α or β. This is typically the case when the languages are close to one another, like in a situation of contact between related languages (like Hindi-Urdu and Punjabi (Gumperz & Wilson 1971), English and Dutch or German (Clyne 1987), Czech and Slovak (Nábělková 2014), or Belarusian and Russian (Hentschel 2014)); between related dialects (like High German and Low German (Höder 2012)), or between a non-standard dialect and a standardized dialect (like Swiss German and Swiss Standard High German (Werlen 1988)); between a creole and its lexifier language (like Creole and French in the Réunion (Ledegen 2012)); or between different creoles (like the Eastern Maroon Creoles of Suriname and Sranan Tongo (Migge & Léglise 2013: 123-140)). Even when languages are not genetically related, it is possible that a long history of borrowings and convergence has given rise to a situation in which some words become indistinguishable within the different languages involved in the contact situation (this is the case with the Australian languages studied by Heath (1981), among which there was a diffusion not only of structural features, but also of many lexical items).

Additionally, the case of prolonged contact between a creole and its lexifier language is known to be somewhat exceptional in that it gives rise to what has been called a creole continuum (see Section 1.3 above). In such situations, there may never have been any break of the living ties between the two distinct languages; the position of any given variety along the continuum is a matter of sociolinguistic parameters linked to the speaker and to the interaction type. The task of drawing a border is then an intractable one, as Bickerton explains:

> To speak of 'dialects' or even perhaps 'languages' may be misleading; these terms merely seek to freeze at an arbitrary moment, and to coalesce into an arbitrary whole, phenomena which in nature are ongoing and heterogeneous. (1973: 643).

This does not mean that there are no constraints on what can occur (there actually is an 'implicational scale'), but that languages α and β are not located in a discrete space, but in a gradual one. It has also been noticed that in such situations, virtually any word of the lexifier language may be incorporated in a creole utterance without it being perceived as a borrowing:

> Dans les aires créolophones (…) la considérable osmoticité entre français et créoles fait que presque tout lexème français peut être "créolisé" et que, dans l'autre sens, la quasi-totalité des termes créoles peut apparaître en français. (Chaudenson 1993: 391)[11].

When not only isolated words, but whole phrases, can be ambiguously analysed as manifesting α or β, Ledegen (2012) talks about 'floating' segments. My data from Martinique displays a lot of such cases,

11 In Creole-speaking areas (…), as a consequence of the considerable osmoticity between French and Creole, any French lexical unit may be 'creolized' and, in the other way, virtually all Creole terms may appear in French.

as in the examples (1a-d), where floating segments are transcribed over two tiers, in Ledegen style. Only segments which appear on only one of the two tiers can unequivocally be said to be MC (upper tier) or French (lower tier).

This situation makes it difficult to make strong predictions about which combinations should be possible and which impossible, without leaving the door open to requiring an explanation for every single apparent exception (in terms of lexical similarity, language permeability, or 'nonce borrowings'). Actually it is debatable to speak of situations of contact between a creole and its lexifier language in terms of alternation or insertion of constituents (see below, Section 5.2).

(1) a. kon tout lé zétud ka montré (MC)
Et de fait toutes les études montré (French)
'and as a matter of fact, like all studies are showing'

b. mem dé konstatasion ke dé nonm ka fè ka montré (MC)
même des constatations que des montré (French)
'and even the observations that some people do are showing'

c. ke ni dé relasion ant condision travay yo (MC)
que des relations ent(re) conditions de travail (French)
'that there are links between their work conditions'

d. épi dégradasion sirkulasion ka fèt vèn (MC)
dégradation de circulation au niveau veines (French)
'and the deterioration of blood flow that is happening in the veins'

4. Observations in actual corpora

4.1. As evidenced in mixed language corpora

Table 5 sums up the combinations which are predicted unlikely to occur according to the various purported code switching constraints mentioned above.

An important comment is necessary when interpreting the results of Table 5. Some lexical items of MC are similar in their phonetic form to lexical items of French, with syntagmatic combination properties that are also partly identical, but a semantic value that is not. For instance, assuming we have a noun with a similar phonetic form in MC and French, e.g. *tas* (MC) / *tasse* (French) 'cup', 'mug'. An NP realized /setasla/ could as well correspond to (2a) as to (2b):

(2) a. sé tas la (MC)
PL cup DEF
'the cups'

b. ces tasses là (French)
DEM.PL cup DEIC
'those cups'

This implies that in cases such as lines 12 to 19, or line 22, of Table 5, identification not only of lexical units, but also of system morphemes, is subject to doubt.

In the process of annotating the corpus, I have kept the information about 'floating' segments wherever there was such an ambiguity. However, in order to decide whether a word should be dominantly classified as French or as MC (when counting occurrences of patterns in Table 5), I had to make choices, in order to be able to measure the phenomena.

The criteria I used to make choices were the following, in order of priority:

1. when a word is unambiguously MC (e.g. *yich* 'son'), it is counted as MC

2. when a word sounds approximately similar in French and MC, but its phonetic realization is congruent with one of the two phonological systems only, it is counted accordingly (e.g. /faktɛ/ 'factor' is counted MC, /faktœχ/ is counted French; /linivɛʁsite/ 'the university' MC, /lyniverχsite/ French)

3. in all other cases, when no clear lexical or phonological clue allows to choose, the word is alloted to the matrix language of the surrounding span, determined following Myers-Scotton's criteria (see below).

As MC is the matrix language of most of the corpus (unsurprisingly, since it was recorded in MC-speaking radio broadcasts), the third rule amounts to adopting a conscious bias in favour of counting words as MC even when they occur at the margin of French-like structures (a very frequent pattern), whenever a reasonable doubt exists. This bias favours the risk of overestimating switches over the risk of missing actual contact phenomena by classifying all ambiguous words as French.

Another comment is that the predictions of the MLF Model (Myers-Scotton 1993; Jake et al. 2002) are barely testable in all possible configurations, since in the major part of my sample, MC is the dominant matrix language (as identified following Myers-Scotton's criteria: (a) the language providing the majority of the morphemes in a surrounding span of the corpus (Myers-Scotton 1993), (b) the language which is least marked for the interaction type (Myers-Scotton 1995: 237), (c) the language which the speaker himself identifies as the language of the interaction (ibid.), and (d) the language providing the syntactic frame of the main clause (CP) (Myers-Scotton & Jake 2000b); the four criteria give the same results). Of the 743 NP occurrences counted here, only 17 belong to a corpus span where the ML is French (and there is only one such span). This means that in lines (9, 10, 19, 31, 32) of Table 5, the figures given are not very significant.

Despite these cautionary remarks, some phenomena appear clearly:

1. The high counts in lines 11, 1, and 18 of Table 5 can all be subsumed under a generic explanation expressed by the maxim:

   (M1) Any French N' can potentially be used as a MC N' in a MC NP[12].

   This maxim is verified in a number of cases behind a MC indefinite article ('an *aspect commun*', 'an *station assise prolongée*', an '*système de canalisations*', 'an *place officielle*'), where it seems to violate the Functional Head constraint on code-switching posited by Belazi, Rubin & Toribio (1994) (line 1 in Table 5).

   It is verified in an even greater number of cases in a definite NP where the MC definite morpheme follows the noun ('*chasse veineuse* a', '*écoulement urine* a', '*institution scolaire* la', '*premier type de robinets* a'…) (line 11 in Table 5). In these cases, it also violates the Equivalence constraint (posited by Sankoff & Poplack 1981), since the configurations for marking definiteness are not congruent in MC and in French (N DEF *vs.* DEF N).

   It is also verified in a significant number of cases where the MC plural morpheme (*sé*) is used before a noun ('sé *clapets* a', 'sé *types de profession* a'…) (line 18 in Table 5).

12 N': N-bar: a noun, or a noun-headed constituent with a noun and some of its possible modifiers (adjectives, relative clauses, possessive complements...), but without a determiner. Some authors (Muysken 2000: 61) use NP for what I here call N', and DP (Determiner Phrase) for what I call an NP.

Table 5. Purportedly impossible or dispreferred NP types, according to some theories of code switching

	combination:	*predicted:*	*according to:*	*specific feature:*	*# in sample*	*%*	*example (from the corpus unless otherwise specified (when #=0)):*
1	INDF.SG*c* N*f*	impossible	FH		39	7.8	**an** service public
2	INDF.SG*f* N*c*	impossible	FH		0	0.0	une **tifiy** *(made up example)*
3	INDF.PL*f (de)* N*c*	impossible	FH		2	0.4	des **moun**
4	[DEF.SG*f* N*c*] *c*	impossible	MO		3	0.6	(**ou ka bat ba**) la **kilti péyi a**
5	[DEF.SG*f* N*f*] *c*	dispreferred	BNP		15	3.0	(**ni difikilté** à **lité** contre) l'effet la pesanteur
6	DEF.SG*f* N*c*	impossible	E; FH		3	0.6	la **kilti péyi a**
7	DEF.SG*f (lə/la)* N DEF*c (a)*	impossible	E; FH	doubling	3	0.6	l'air **la**
8	N*c* DEF.SG*f*	impossible	E		0	0.0	**moun** le *(made up example)*
9	[N*f* DEF*c*] *f*	impossible	MO		2	0.4	(des écoulements principalement de) urine **a**
10	[N*c* DEF*c*] *f*	dispreferred	BNP		0	0.0	(les ouvriers veulent) **lajan an** *(made up example)*
11	N*f* DEF*c*	impossible	E; FH		62	12.4	institution scolaire **la**
12	[DEF.PL*f (le)* N*c*] *c*	impossible	MO		8	1.6	les **jenn manmay** (**ka chwazi**)
13	[DEF.PL*f (le)* N*f*] *c*	dispreferred	BNP		20	4.0	(**ki ka sityé ko y** entre) les parties génitales
14	DEF.PL*f (le)* N*c*	impossible	E; FH		8	1.6	les **jenn manmay**
15	DEF.PL*f (le)* N DEF*c (a)*	impossible	E; FH; MO	doubling	1	0.2	les **moun lan**
16	PL*c (se)* DEF.PL*f (le)* N	impossible	E; MO	doubling	0	0.0	**sé** les chiens *(made up example)*
17	DEF.PL*f (le)* PL*c (se)* N	impossible	MO	doubling	0	0.0	les **sé** chiens *(made up example)*
18	PL*c* N*f*	impossible	E		10	2.0	**sé** muscles vessie **a**
19	[PL*c* N*c* DEF*c*] *f*	dispreferred	BNP		1	0.2	(à partir du moment où) **sé tiyo a**
20	DEM.SG*f* N*c*	impossible	E; FH		0	0.0	cette **kay** *(made up example)*

	combination:	*predicted:*	*according to:*	*specific feature:*	*# in sample*	*%*	*example (from the corpus unless otherwise specified (when #=0)):*
21	N*c* DEM.SG*f*	impossible	E		0	0.0	**timanmay** ce *(made up example)*
22	DEM.PL*f* *(se)* N*c*	impossible	E; FH		0	0.0	ces **chimin** *(made up example)*
23	PL*c* *(se)* DEM.PL*f* *(se)* N	impossible	E	doubling	0	0.0	**sé** ces **moun** *(made up example)*
24	NP1*c* *də* NP2*f*	impossible	G		1	0.2	**an sòt** de filtration
25	NP1*c* *də* NP2*c*	impossible	FH		1	0.2	**an serten nomb** de **faktè**
26	NP1*f* *də* NP2*c*	impossible	FH		17	3.4	évacuation vers l'avant de **pisa a**
27	[POSS.n*f* NP*c*] *c*	impossible	MO		0	0.0	son **yich** (**sòti kay-la**) *(made up example)*
28	[POSS.n*f* NP*f*] *c*	dispreferred	BNP		0	0.0	sa mère (**sòti kay-la**) *(made up example)*
29	POSS.n*f* NP*c*	impossible	E		0	0.0	son **zanmi** *(made up example)*
30	NP*c* POSS.n*f*	impossible	E; MO if ML=*f*; SM if ML=*c*		0	0.0	**tifiy** ma *(made up example)*
31	[NP*f* PRN.n*c*] *f*	impossible	MO		0	0.0	(je voudrais revoir) maison **mwen** *(made up example)*
32	[NP*c* PRN.n*c*] *f*	dispreferred	BNP		0	0.0	(je voudrais revoir) **kay mwen** *(made up example)*
33	NP*f* PRN.n*c*	impossible	E; G		14	2.8	propre emploi **mwen**
34	PRN.n*c* NP*f*	impossible	E; G; MO if ML=*c*; SM if ML=*f*		0	0	**zot** enfants *(made up example)*

Table 5. *Constraints on possible combinations of MC and French NP structures, as predicted by theories implying generic constraints on (intrasentential) language switching. In the second column, indications on the actual phonological form of some function words are in IPA. Index letters indicate the language of an element:* ***c*** *for Martinican Creole and* ***f*** *for French (keeping in mind reservations expressed above regarding the limits to unequivocal identification of the language); when a matrix language is involved in the prediction, it is noted as an index letter spanning the whole constituent (in square brackets). In the fourth column, the following abbreviations are used to refer to hypothesized constraints on code switching: E = Equivalence Constraint (Sankoff & Poplack 1981); G = Government Constraint (Di Sciullo, Muysken & Singh 1986); FH = Functional Head Constraint (Belazi, Rubin & Toribio 1994); MO = Morpheme Order Principle (Myers-Scotton 1993), SM = System Morpheme Principle (Myers-Scotton 1993); BNP = Bilingual NP hypothesis (Jake et al. 2002). In the fifth column, I note when some specific issue appears, e.g. suspected cases of 'doubling' (when it may appear that the same grammatical or semantic value occurs twice in a constituent); this issue raises another question, namely whether one can be sure to correctly identify the two morphemes involved in the suspected cases. Figures in the sixth and seventh columns are numbers of types, not tokens. In the eighth column, I give an example. Contexts within parentheses give an idea of the matrix language. When there is at least one instance of the structure in the corpus, the example is drawn from the corpus. If none is found, the example is made up (just to show how the structure would look like, should it exist). Grey backgrounds signal the most frequent patterns.*

2. The significant counts in lines 13, 5, and 33 of Table 5 suggest a generalization of maxim M1, namely:

   (M2) Any French NP can potentially be used as a MC NP in a MC frame.

   In fact, there seems to be little restriction on the possibility of using a complete French NP in a MC clause, be it plural ('san an ki an didan vèn moun la ka lité *contre les effets de la pesanteur*', 'yo ka di w *les apprentissages fondamentaux*'…) (line 13 of Table 5), or singular ('manniè ou ka viv sa andidan *la télé*', 'sé manmay tala ki dan *le supérieur*'…) (line 5 of Table 5). This type of 'EL island' (in Myers-Scotton's terms) is only predicted **dispreferred**, and not **blocked**, by Jake, Myers-Scotton and Gross's 2002 adaptation of the Matrix Language Frame model (2002: 78-79) — a prediction that my data does not disprove, in any case.

   Complete French NPs also appear embedded in MC possessive constructions ('*conditions d'exercice de profession* yo', '*propre emploi* mwen'…) (line 33 of Table 5), a significant fact which violates the predictions of both Sankoff & Poplack's Equivalence constraint, and Di Sciullo et al.'s (1986) Governement constraint, but is completely in line with the predictions of the MLF model.

3. There is a (not high but) significant number of occurrences of /le/ (which I have summed up in the counts of lines 14-15 of Table 5 as if they were occurrences of the French plural definite article *les*), in front of MC nouns. Those constructions, if they are what they seem to be, violate several constraints observed in earlier works on code-switching, and they are particularly counter-intuitive in contexts where MC is the dominant language. They even include cases where this morpheme /le/ is present simultaneously with the MC postponed definite determiner (1 occurrence in the sample of 500 NP types — line 15 of Table 5; but there are too many other examples in the remainder of my corpus for it to be possible to explain them as transcription errors, or to throw them in the 'performance bucket'). This raises the question of the status of the morpheme *lé* in MC, a question to which I shall return below (Section 5.1).

4. Finally, a noteworthy observation is the relatively high number of possessive constructions of the type NP_1 *f də* NP_2 *c* (line 26 in Table 5 — also to be found as lines 2, 3 and 5 of Table 6, below). This, I think, demonstrates that some N-headed constructions are stored in the lexicon as ready-made frames with multiple lexical anchors (not only the noun, but also the relational morphemes — here, prepositions —, that connect its modifiers). I return to this question in Section 5.3.

4.2. Embedded structures

The most interesting observations in the corpus concern NPs containing embedded constituents (relative clauses, genitive complement NPs), since they provide a testbed for possible structural explanations of code-mixing. I will concentrate here on genitive NP/NP constructions.

As shown in the two last lines of Table 4, MC and French have non-congruent ways of expressing genitive. If we note NP_1 the head constituent (the 'possessum' or 'determined' one) and NP_2 the dependent constituent (the 'possessor' or 'determining' one), MC has one single structure, 'NP_1 NP_2', used both when the two constituents are actual 'full' NPs (*kay manman* 'mother's house') and when NP_2 is a person pronoun (*kay mwen* 'my house')[13]. French has a specific possessive determiner that replaces the use of a full pronoun when the 'possessor' is a person index (*ma maison* 'my house'); with

13 The expression of possession is a point of divergence between MC and other closely related French-based Creoles. The Creole of Cayenne (used in the central part of French Guiana) uses a possessive determiner, as in French: *mo kaz* 'my house'. The Creole of Guadeloupe uses a joiner word (*a/an*), as in French *à*: *kaz a Piyè* 'Pierre's house'.

two full NPs, it uses the same head-first order as MC, but with a joiner preposition (*maison de Maman* 'mother's house').

When the depth-level of embedded constituents is equal to—or greater than—two, the question is: which structure will be adopted by mixed NPs? Will they tend to use the order of the embedding constituent's head, or the order of the embedded one? Will this choice be the same if the embedding constituent is in the same language as the 'matrix language' (the whole sentence's dominant frame, provided there is one)? Is the structure of the constituent imposed by the language of its lexical head? The data give no absolute answer to these questions, but there are tendencies.

The first obvious fact is that in my corpus, genitive NP/NP constructions are sensitive to language mixing. Of 108 such NP types in the sample of 500, only 38 look like monolingual NPs (26 look like 'pure' MC and 12 'pure' French, both on lexical and structural criteria).

Examining the remaining 70 NPs to look for lexical or structural patterns, we get the figures shown in Table 6.

The first observation to be made is that practically all possible combinations are actually present, making it hard to see clear evidence of any particular constraint on code switching (such as those mentioned in Section 3.2). There is a broad ambiguous zone, due to the fact that Creole and French constructions are indistinguishable for some frequent values, most notably the indefinite (INDF N) and the bare noun (N). For example, most instances of cases listed in lines 4, 5 and 11,12 of Table 6 (ambiguous structure of bottom NP node) contain an embedded bare noun, which is intrinsically unable to display structural features of either language.

no	top NP node		bottom NP node		example	# in sample	%
	structure	*head*	*structure*	*head*			
1	French	French	French	French	goutte-à-goutte de [production de [**Ø** urine par rein **a**]]	3	4.3
2	French	French	Creole	French	des contractions de [**sé** muscle **a**]	15	21.4
3	French	French	Creole	Creole	évacuation vers l'avant de [**pisa a**]	1	1.4
4	French	French	F/C	French	**an** certain nombre de [professions]	11	15.7
5	French	French	F/C	Creole	**an** certain nombre de [**faktè**]	1	1.4
6	French	Creole	French	French	**an sòt** de [filtration de [sang **ta'a**]]	1	1.4
7	French	Creole	F/C	F/C	witè (huit heures) de [tan (temps)]	1	1.4
8	Creole	French	French	French	les différentes parties **Ø** [l'organisme]	7	10.0
9	Creole	French	Creole	French	paroi musculaire **Ø** [vessie **a**]	11	15.7
10	Creole	French	Creole	Creole	les parties génitales **Ø** [lé **misié épi** lé **madanm**]	8	11.4
11	Creole	French	F/C	Creole	veines **Ø** [**moun**]	6	8.6
12	Creole	F/C	F/C	F/C	problem (problème) **Ø** [batri (batterie)]	3	4.3
13	Creole	Creole	Creole	French	**bout Ø** [**tout** système]	2	2.8

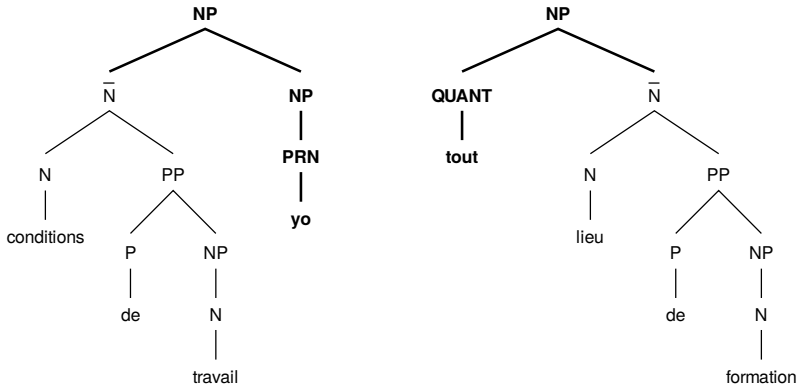
Table 6. *Different configurations of NP-NP embeddings in mixed genitive constructions. There are 70 mixed genitive NPs (from a total of 108 genitive NPs). A genitive construction involves a dependency between two NP nodes, a top node and a bottom node. I noted the values of two parameters for each of these two NP nodes: the language that governs its structure—determined by unambiguous morphosyntactic features, e.g. the use of the joiner morpheme* de *(French) or postponed definite article* (l)a *(Creole)—; and the language of its lexical head. The languages are tagged French, Creole, or F/C in ambiguous cases. To make the examples easier to read, (1) it is explicitly indicated (Ø) where there is no joiner morpheme, (2) the embedded (bottom) NP node is in square brackets, (3) the two lexical heads are underlined, (4) morphemes that are unambiguously Creole are in bold face. Mixed NPs are expected to display heterogeneous values for the 4 expected parameters, with the exception of the instances of line #1, where the two top levels are French both in vocabulary and in structure; they are however counted as mixed NPs because there are deeper levels of embedding which clearly are Creole. Grey backgrounds signal the most frequent patterns.*

The most important parameter giving a clue to a specific language, on the level of the top NP node, is the use of an overt joiner morpheme (*de* vs. Ø). Here the number of top NPs using the Creole structure (Ø) for expressing possession (lines 8-13) is only slightly above the number of those using the French structure (*de*), despite the fact that the matrix language, where there is one, is mainly Creole in my corpus[14]. This does not strictly make the Matrix Language Frame model invalid, since under the premises of its recent version (Myers-Scotton & Jake 2000a), a joiner morpheme is just a 'bridge' system morpheme that does not have to be bound to the matrix language; this simply falls into the zone where the MLF model has nothing to predict.

On the whole, the examples in this corpus fail to test any purported universal constraint on code switching. They have to be interpreted either as contradicting them (see Table 5 above, Section 4.1), or as falling into the cases that would be treated as exceptions or invalid test cases for the code switching theories (e.g. N' insertion). In any case, hypotheses of universal constraints do not seem to shed much light on this data.

We are left with an open question: how is it possible to describe the structure of mixed embedded NPs? In this corpus we are faced with some cases of 'wild' internal mixing, with an overall embedding structure seemingly in language A, a head noun in language B, a determiner in A, an embedded NP in B, along with all permutations of these elements (like in Figure 1).

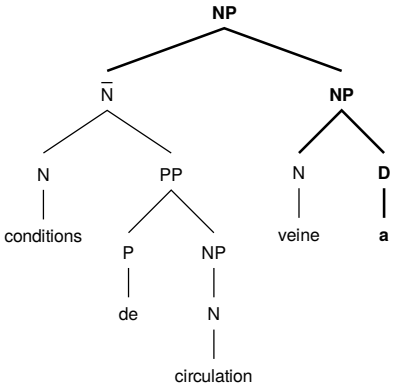
Such multiple-level embeddings seem to challenge any claim to general rules governing the insertion of constituents in alternate languages—since two successive insertions would seem an odd sort of phenomenon. It is actually easier to understand them in the TAG (Tree-Adjoining Grammars) framework, using the complementary linguistic operations of substitution and adjunction, as I will propose below, in Section 5.3.



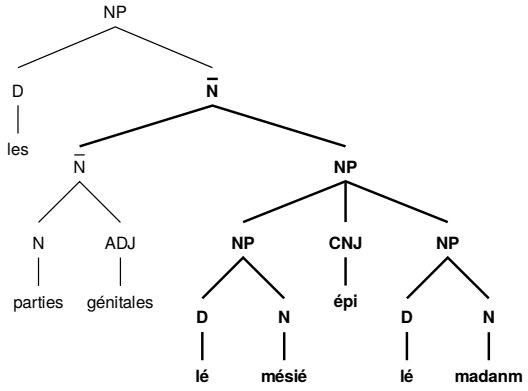
1.a. conditions de travail **yo**

1.b. **tout** lieu de formation

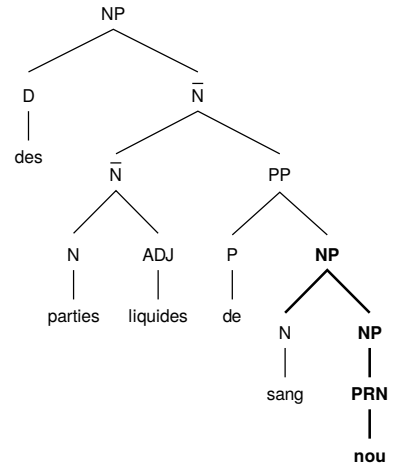
14 On the 108 genitive NPs, 4 have been tagged as having French as their matrix language (based on Myers-Scotton's (1993) morpheme frequency criterion).



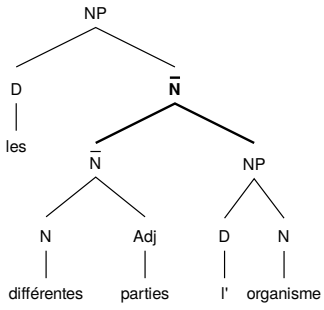
1.c. conditions de circulation veine **a**



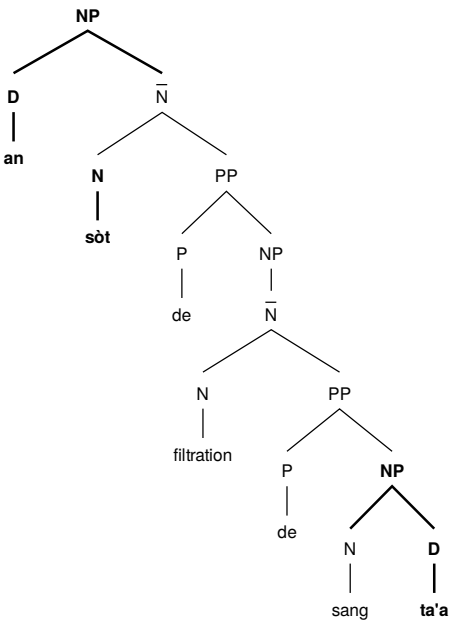
1.d. les parties génitales **lé mésié épi lé madam**



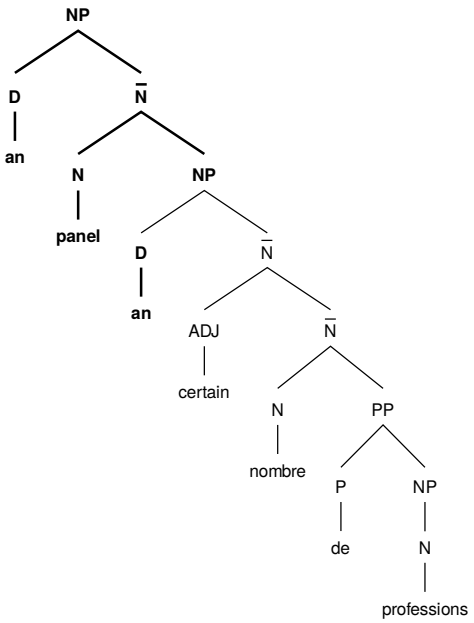
1.e. des parties liquides de sang **nou**



1.f. les différentes parties l'organisme



1.g. **an sòt** de filtration de sang **ta'a**



1.h. **an panel an** certain nombre de professions

Figure 1. *Typical examples of mixed NPs displaying language mixing at both lexical and structural levels. In these examples, unambiguously Creole morphemes or structures are displayed in bold. Note: in 1.b* ***tout*** *is pronounced* [tut]*, which is what makes it Creole; in 1.f all the morphemes are French but the structure of the top N' node is Creole (no joiner morpheme* de*).*

5. Discussion

5.1. Suspected cases of doubling

In the sample that has been studied here, there are 44 NP occurrences (representing 26 distinct NP types, i.e. 5.2% of the total) that display the pre-nominal morpheme /le/ in the context of a plural. In French, this phonetic form corresponds to the gender-neutral plural definite article (written *les*).

The morpheme *lé* (/le/) is often considered to be either a borrowing of the plural definite article from French, or to be nothing more than a regional variant of the plural morpheme *sé*, used in the North-Western Martinican district of Sainte-Marie (Pinalie & Bernabé 1999: 21; Bernabé 2003: 109). In the first hypothesis, it would be expected to appear in French NPs of the 'EL-island' type (and to be correlated with French nouns, most of the time); in the second, it should always be accompanied by a post-nominal definite (*la*) or demonstrative (*tala*) marker, and should be less frequent than the more widespread *sé* form. The data presented here does not show the expected distribution. Counter to the first hypothesis (or at least a monofactorial understanding of it): a significant number of occurrences of *lé* (19 on 44) appear with a typically Creole noun (like *lé moun*, *lé jenn manmay*). Counter to the second hypothesis: first, only in a minority of cases (3 on 44 in this sample) do these occurrences display a post-nominal determiner; second, there are more NP occurrences with *lé* (44) than with *sé* (29)[15].

There is an alternative explanation to *lé*. Zribi-Hertz & Jean-Louis (2014) have shown that *lé*, with plural nouns, carries a 'kind-referring' value (as opposed to the 'set-referring' value carried by *sé*[16]). Their analysis explains many of the examples found in my sample, such as (3a,b).

(3) a. lé **lé** misié épi **lé** madanm (MC)
les **les** **les** (French)
DEF.PL parts genital DEF.PL man with DEF.PL woman
'the genital organs of men and women' (i.e. the male and female genitals)

b. ki plas kreyol la pé ni andidan **lé** média (MC)
place créole **les** média (French)
which place Creole DEF can have within DEF.PL media
'which place can Creole have on the (mass) media' (as a place of expression)

However, the ordinary, set-denoting, value of the plural definite article seems to be retained in some occurrences, even when the noun in the NP is rather typically MC than French (4).

(4) **lé** jenn manmay ka chwazi an branch (MC)
les choisi(r) branche (French)
DEF.PL young kid IPFV choose INDF branch
'the young kids choose a branch (of study)'

Another fact that fails to find an explanation within any of the proposed models is that some occurrences simultaneously display the preposed *lé* and the postposed definite marker *la*, like in (5).

(5) **lé** jenn **lan** pa lé travay pou anyen
DEF.PL young DEF NEG want work for nothing
'the young [people] don't want to work for nothing'

Zribi-Hertz & Jean-Louis assert that '*Lé* NP is incompatible in its DP with the specific determiner -LA' (2014: 303). In fact, the analysis that they propose for its value excludes its use in contexts such as (5):

15 I do not have biographical data for all of the speakers in the corpus from which my sample is drawn; however, it is unlikely that a majority of them come from the district of Sainte-Marie, which accounts for 4.84% of the population of Martinique (INSEE 2009 census).

16 See above: Section 2.

the ‘intensional’ definiteness marker is expected to denote the concept itself, and not a set of instances of the concept; hence, it is ‘anti-specific’, and therefore ‘crucially’ (ibid.: 285) incompatible with the postposed definiteness marker of MC, which bears a /+specific/ value.

Examples such as (5), however, are not exceedingly rare. They represent only 4 NP types in the 500 sample that has been studied in Table 5 (lines 7 and 15), (on 44 which display *lé*), but I have observed more of them in the rest of the corpus. They are puzzling cases for existing models: it is impossible to explain them in Zribi-Hertz & Jean-Louis’ analysis; it is also highly unlikely that they are all instances of the regional variant mentioned above (see footnote 15). If they were to be interpreted as mere ‘borrowings’ of the French definite article, their presence with phonetically typical MC nouns[17] would be surprising.

A possible interpretation of the simultaneous presence of preposed *lé* and postposed *la* could be that they are instances of morphological doubling, induced by an attempt to simultaneously activate diverging word orders in the two languages in contact, a phenomenon which has been described by Hicks: ‘When two source languages have different constituent orders, most code switches resulting from these languages will adhere to the constituent order of one source language or the other. (...) A rare type of switch, however, occurs when **both** constituent orders are (at least partially) adopted. When this happens, the resulting code switch may contain at least one morphosyntactic element (a phrase, word, or morpheme) which is produced twice: the first occurrence of the element appears in the unmarked position for that element in one source language, while the second occurrence appears in the unmarked position in the other source language’ (2012: 45).

I suspect that there is more in this phenomenon than mere doubling of a definiteness marker. In some extracts of the corpus, two or three of the four possible variants ‘N’, ‘*lé* N’, ‘*lé* N *la*’ and ‘*sé* N *la*’ appear at a short distance from one another; in those cases, there appears to be a significant contrast between those different uses, like the examples (6a-d) show.

(6) a. montré **lé** manmay ki manniè lang kreyol ka woulé (MC)
montrer les (Fr)
show DEF.PL kid which way language creole IPFV roll
‘[Our goal was to] show the kids the way the Creole language functions’

b. montré **lé** moun **lan** ki manniè lang lan ka woulé (MC)
montrer les (Fr)
show DEF.PL people DEF which way language DEF IPFV roll
‘[there are volunteers who] show the people the way the language functions.’

c. yonn-dé moun té lé sav ki jan pou yo matjé (MC)
one-two people PST want know which way for 3PL write
‘A couple of people wanted to know how they ought to write’

d. pas **sé** moun **lan** ni lanvi (MC)
because PL people DEF have desire
‘because those people had the desire to do so.’[18]

The contrast between the values of the different plural definite constructions is typically illustrated by (6a), (6b) and (6d), that consecutively appear in a span of text #4. In the discourse context, it seems that the first construction in (6a), *lé manmay* is a generic definite description of the category of target population for the action that is the topic of the current talking (teaching Creole to Creole-illiterate

17 In the audio track of the recording, the lexical item ‘jenn’ is unambiguously pronounced the way it is expected to be in the MC phonological system (i.e. /ʒɛn/, with an unrounded vowel, cmp. /ʒœn/ in French).

18 In the examples (6a-d), the following glosses are used: DEF = definite; PL = plural; IPFV = imperfective; PST = past tense.

adults)[19]; the second in (6b), *lé moun lan*, is another generic definite description of the same (above-mentioned) target group; the third in (6d), *sé moun lan*, is a definite description that more specifically denotes the 'couple of people' that were introduced in (6c). Very few extracts display the complete set of combinations of values that may be seen in (6a-d), but some other instances displaying just two of the four possible combinations confirm the present interpretation. Two things appear clearly: (1) *lé* is not merely a regional variant of *sé*, since both forms appear in specific contexts in utterances from the same speaker; (2) *lé* is not 'crucially' incompatible with postposed *la*, since they sometimes appear simultaneously.

(6a)	*lé N*	kind-referring	– deictic/anaphora	N as a category
(6b)	*lé N la*	kind-referring	+ deictic/anaphora	N as a category, definite in the context
(6d)	*sé N la*	set-referring	+ deictic/anaphora	A definite set of instances of category N

Intuition leads me to think that in this type of contexts, *lé* actually has the kind-referring value that was identified by Zribi-Hertz and Jean-Louis, in contrast with *sé* which has a set-referring value. So finally, what is the value of the MC postposed morpheme *la*? It seems to me that it may be loaded with the same anaphorical value that has been identified in the French deictic adverb *là* (see above, footnote 6; ex. (2a,b)). If we remember that we are in a language-contact situation where a majority of speakers are balanced bilinguals, and where many morphemes sound like their reflex in the other language, it seems perfectly plausible that the semantic value of a morpheme in MC should in certain settings be influenced by the semantic value of the reflex morpheme in French. This explains cases like (6b), where *lé* is simultaneously present with *la*: the discourse anaphorical value is not incompatible with the intensional value.

5.2. Limits of the 'constraints on code-switching' approach: congruent lexicalization

As some researchers have pointed out before (e.g. Muysken 2000; MacSwan 2014a), theories that have tried to explain the structural constraints on code-switching by positing universal constraints have all described a part of the truth, but also have been proven wrong on some sets of counter-examples.

In the opinion of Muysken (2000), a problem of global theories on code-mixing is that they have described different phenomena under a single generic term ('code-switching'). This also explains why they sometimes come up with contradictory predictions. For instance, what Sankoff, Poplack and co-workers have studied under the terms of 'flagged switching', or 'code-switching under equivalence', mainly fall into the category of of phenomena that Muysken calls 'alternation': the possibility to switch languages between constituents that are not embedded in a rigid syntactic frame (in cases like parataxis, or freely positioned adjuncts). Models like Di Sciullo, Muysken and Singh's, or Myers-Scotton and Jake's, on their side, have focused on 'insertion': which constraints do apply when embedding a constituent of one language in a syntactic frame of another one (Sankoff & Poplack deal with such cases under other terms, like 'nonce borrowings').

Muysken tried to make a distinction between what he considered to be three types of code-mixing phenomena: **insertion**, **alternation** and **congruent lexicalization**. The three distinct types of code-mixing are represented in a schematic way by the formulas: 'A [$_A$ B]' for insertion, 'AB' for alternation, '$A_1 \ldots A_n \sim B_1 \ldots B_n$' for congruent lexicalization (2000: 31)[20].

19 Here, like is often the case in MC, *manmay* (/mɑ̃maj/) 'kids' does not refer to actual children, but is a familiar description for (friendly) adult people.

20 In (Muysken, 2013) he added a fourth strategy he calls **backflagging**, that can be seen in situations of language attrition, when speakers just 'flag' their linguistic identity by inserting some elements of the heritage language in the dominant

Insertion involves the possibility to insert a constituent of language B into a frame of language A 'under categorial equivalence': it is possible 'when the switched element has the same status in the two languages, is morphologically encapsulated, shielded off by a functional element from the matrix language [e.g. a determiner or preposition], or could belong to either language'.

Alternation includes 'extraposition, the suspension of syntax, fronting, adverbial constructions, pauses, flagging, fillers, and, morphologically, agglutination'.

Congruent lexicalization, he argues, is what has mostly been studied by researchers like Clyne, who have worked on situations involving related language pairs: 'In a third set of cases, it appears that there is a largely (but not necessarily completely) shared structure, lexicalized by elements from either language, **congruent lexicalization**.' (ibid.: 5-6).

As will be obvious to people familiar with contact between a creole and its lexifier language, much of the MC/French data may fall into this last category (see above, end of Section 3.2, esp. ex. (1)). It is therefore not surprising that in Table 5 one can find many counter-examples to constraints that simply do not apply to the situation (especially line 1).

Under the light of this concept, in a situation of congruent lexicalization, maxims M1 and M2 above (Section 4.1), which expressed recurring observations in my corpora (Any French N' can be used as a MC N' in a MC NP; any French NP can be used as a MC NP in a MC frame), could be reformulated in more general terms:

(M3) In the context of MC/French code-mixing discourse, in MC (resp. French) phrases, any pre-terminal syntactic node of category X may be filled by a French (resp. MC) lexical form, provided that form matches a congruent category X, and the elementary trees[21] of the phrases anchored in X have the same structure in the two languages.

This obviously is all the more likely to be observed as the lexical forms in the two languages are themselves very similar; however, it is important to stress that surface similarity is a facilitating factor **under the provision** of structure similarity, not alone.

Given the many syntactic similarities between MC and French, the generalization expressed by (M3) predicts that many possible structural combinations will be observed, which actually is the case. For example, expressions of NP-NP genitive determination may follow the French structure (involve a joiner morpheme *də*—even with Creole NPs), or the Creole structure (with no joiner morpheme—even with French NPs). This does not, however, imply that such phenomena happen randomly, and that we have no conceptual tool to understand them.

5.3. The difference between substitution and adjunction; derivation trees

Mahootian (1993) has proposed a 'null-theory of code-switching', claiming that 'phrase structure is projected from the lexicon, with each lexical item projecting its own language-specific syntactic requirements' (1993: 186), and that 'these structures combine with each other at maximal projection nodes to form clausal structures. Codeswitched utterances are generated by substituting the lexical structures of one language, L1, into empty complement nodes of trees anchored by L2 heads' (ibid.: 187). Her model, to my knowledge, is the most robust to account for possible combinations, across different language-contact situations involving diverging syntagmatic structures. It does not postulate that word orders have to be the same in two languages for a switch to occur, simply that there has to be

language.

21 Elementary trees are a way to represent the local syntactic structure projected by a lexical item, e.g. whether a complement should appear left or right of the head. They are a key concept in the Tree-Adjoining Grammars (TAG) model, introduced in the next section.

some categorial equivalence between saturated ('maximal projection') nodes across the languages involved, so that e.g. a NP in L1 may be (roughly) equivalent to a NP in L2. The substitution operation may fill an object NP slot in an English sentence with a well-formed Farsi NP like '**you'll buy** xune-ye jaedid' [house-PART new] (example (106) from Mahootian 1993: 152).

Substitution of obligatory complements is not the only way linguistic structures can be combined. Another way is adjunction, an operation which allows an optional constituent to plug into an already complete tree. These two operations have been formalized as the two basic mechanisms of a syntactic model (TAG: Tree-Adjoining Grammars) (Joshi & Schabes 1997), along with rules that control the output of the combinations and account for the complement/adjunct distinction. Substitution is an operation that is compulsory on some specific nodes (without it, the elementary tree is not complete); it models the obligatory complements of a syntactic head, like the subject and object in the example given in Figure 2.a. Adjunction is an operation that allows a special tree (its head node cannot be the root of a phrase, it can only be adjoined) to be inserted in the structure of an another tree, by splitting one of its internal nodes in two (an upper node and a lower node), like the adjective in the example of Figure 2.b. It allows for the integration of optional modifiers in a sentence.

My corpora provide plenty of evidence supporting Mahootian's model of mixed constituents by unification of elementary trees in a tree-adjoing grammars (actually, I have not been able to find a counter-example). An important notion I now turn to is the operation of adjunction.

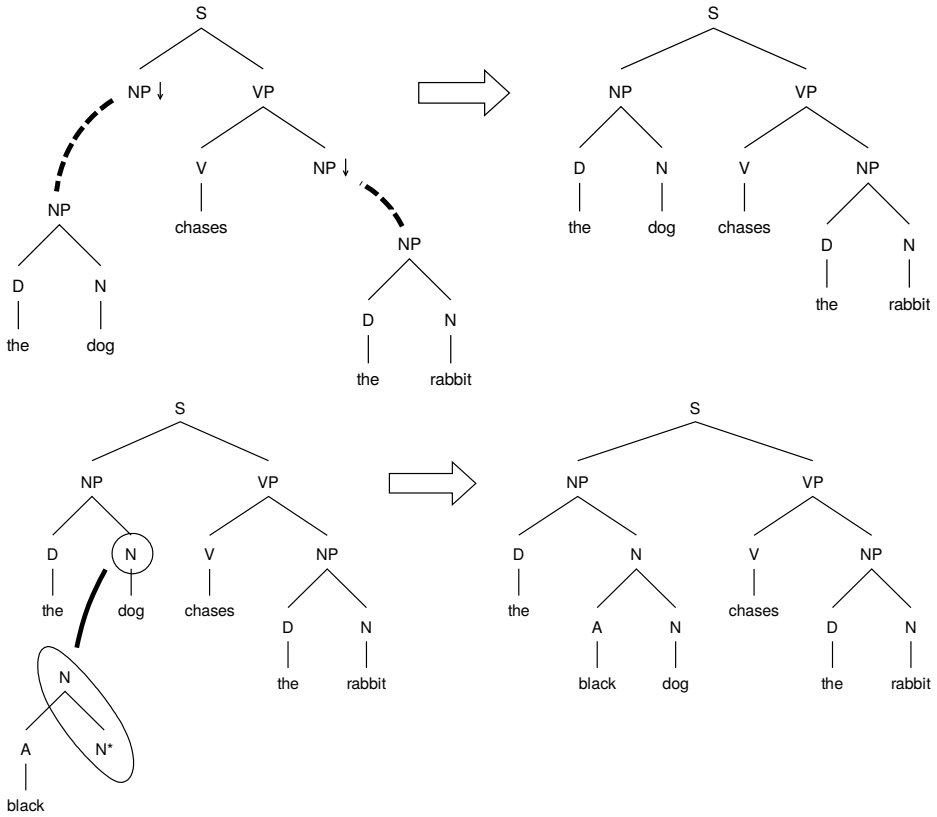


Figure 2. *a. (above) Substitution of complements. b. (below) Adjunction of adjuncts. The fact that elementary structures are trees allows the model to naturally encode elementary constructions, like the english SVO order in 2.a, but also, if needed, set multi-word expressions (like in 'John kicked the bucket', where 'kick the bucket' has to be processed as a single syntactic and semantic unit). It hence offers a natural account of the widely-acknowledged fact that in CS, lexical heads seem to project the grammatical structure of the language they have been chosen from*[22].

22 This intuition has been captured by many authors, however different their models of code-switching might be. Bentahila & Davies say that 'the impossibility of [some switches] can be explained in terms of the two languages' subcategorisation

Mahootian and Santorini (1996) have explored the distinction between complements and adjuncts, mainly in an attempt to account for data that seemed to contradict the tree unification model, in cases of code switching between languages where the position of adnominal adjectives were not congruent. Their main point was that adjunction does not impose such a strict constituent ordering than substitution. This is not the point that I would like to discuss here.

What seems important here, to explain mixed data in the MC/French corpora, is the role that adjunction may have in the production of NPs containing complex (multi-word) determiners or quantifiers. If we look close enough to the [NP1*f də* NP2] examples (line 26 in Table 5), we observe that many of them actually fall into that category, e.g. '*un (an) certain nombre de* N', '*une (an) sorte (sòt) de* N', '*un (an) ensemble (ansamb) de* N', '*un (an) espèce de* N' (examples from my corpora: 7a-d).

(7) a. an serten nomb
un certain nombre de professions
'a certain number of professions'

b. an sòt filtrasion san ta'a
sorte de filtration de sang
'a sort of filtration of that blood'

c. an ansamb
un ensemble de particules
'a set of particles'

d. an espès triang
une espèce de triangle
'a kind of triangle'

Here I am using a distinction between nouns that project subcategorisation frames (nouns that expect actual complements with thematic roles) and nouns that are simply used in multi-word compounds that function as quasi-determiners, quantifiers of classifiers. The elementary trees for the first category have substitution nodes; those for the second category are auxiliary trees that may optionally be adjoined. An example for this distinction is given by Abeillé (1993): in French, the NPs 'verre à vin' (wine glass) and 'verre de vin' (glass of wine), despite their apparent structural similarity, have very different semantic and distributional properties. The first one denotes a type of glass, the second one a given quantity of wine. This example is represented in Figure 3, trees (α5) and (β1).

The semantic difference between 'verre à vin' and 'verre de vin' is not visible anymore in the final derived tree of utterances like 'Jean casse un verre à vin' or 'Jean boit un verre de vin' (Figure 4.a-b). However, the TAG model has a way to 'memorize' the derivation history of a sentence: the derivation tree. For instance, the difference in the generation history of the two contrasting example sentences in 4.a/b show up in their derivation trees in Figure 4.a'/b'.

My hypothesis is that some examples that seem to imply multiple switches (and hence to violate principles of economy) in fact simply result from adjunction of an auxiliary tree drawn from the lexicon-grammar of another language. Let's illustrate with the example sentence 5 in Table 6 (8).

(8) ni an serten nomb faktè (MC)
un certain nombre de (French)
'there are a certain number of factors'

rules' (1983: 321). Belazi, Rubin & Toribio say that 'a word of language X, with grammar Gx, must obey grammar Gx' (1994: 232). Mahootian says that 'syntactic structure is projected from the lexicon and bilingual speakers have access to the lexicon (and therefore the syntactic structures) of both languages' (1993: 139).

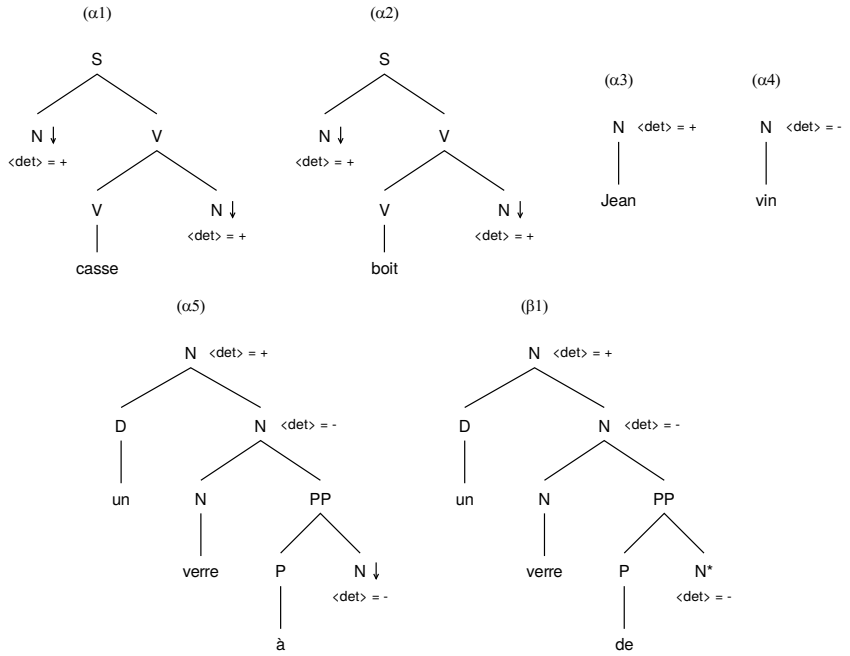


Figure 3. *A set of elementary trees (or partly derived trees) for some phrases, among which (α5)* un verre à [vin] *'a wine glass', and (β1)* un verre de [vin] *'a glass of wine'. The former has the same distributional properties as* verre *(it can be broken); the latter as* wine *(it can be drunk). Inspired from Abeillé (1993: 209).*

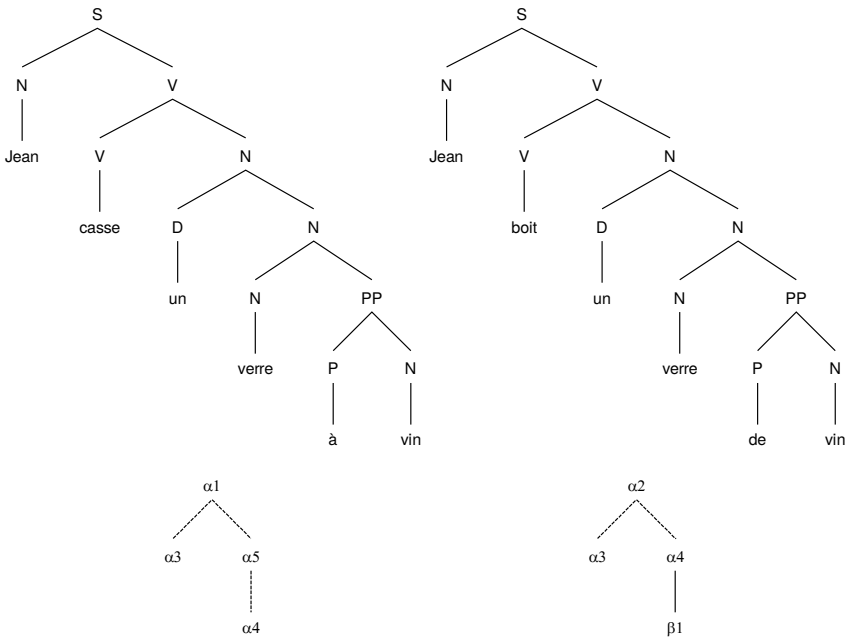


Figure 4. *Derived trees for sentences (a)* Jean casse un verre à vin *'Jean breaks a wine glass' (up left), and (b)* Jean boit un verre à vin *'Jean drinks a glass of wine' (up right); the trees have a very similar static structure, but their difference is best explained by their dynamic derivation history, represented in the derivation trees (a') (bottom left) or (b') (bottom right): (a) results from the insertion of 'wine' in a pattern 'Jean breaks a glass', while (b) results from the adjunction of 'a glass of' in a pattern 'Jean drinks wine'.*

The basic sentence pattern for '*There are* N' in MC is '*Ni* N' (no plural indefinite needed). The sentence observed in (8), which seems to imply a first switch from MC to French between the predicate verb and the NP, then a second switch from French to MC between the PP and the embedded NP, results in my

analysis from the adjunction of a French multi-word set expression used as a quasi-quantifier (‘*un certain nombre de*’)—very common in argumentative speech—in a MC sentence frame (Figure 5)[23].

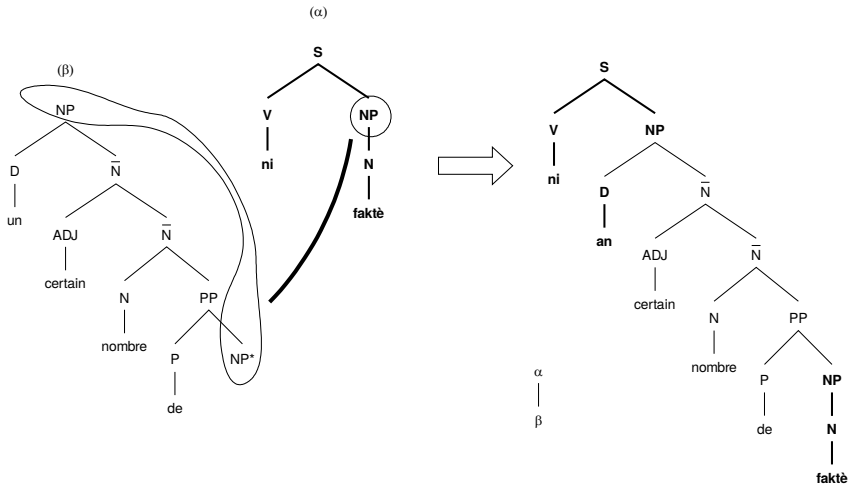


Figure 5. *Adjunction of a French quasi-quantifier (β)* un certain nombre de *into an MC sentence (α)* ni faktè. *The derived tree (rightmost) appears to contain two switches, when in fact the derivation tree (bottom) is very simple.*

6. Underspecified Language Model

The model that best explains the data under scrutiny does not have to include strong hypotheses about a matrix language governing the general pattern of sentences or the use of system morphemes. Actually, the ‘null-theory’ proposed by Mahootian (1993), that postulates nothing more than the ability, for bilingual speakers, to unify elementary structures where unification is possible, does a very good job in accounting for the observations in my corpus. It is a natural frame to understand the observations formulated as maxims M1 and M2 (Section 4.1, above), that allows substitution of a N' node in a language α by a N' in another language β.

I only need to depart from Mahootian’s description frame, or to refine it, in some respects.

1. Not all elementary structures in MC grammar may be anchored in the lexicon. The genitive NP-NP subordination relation that is expressed by joiner morphemes like *de* in French or *-e-* in Farsi (that led Mahootian’s developments on the ‘ezafe phrase’), is expressed by juxtaposition in MC. So, the model should also be able to include non-lexicalized elementary structures, in order to explain how phrases like the example in Figure 1.f (*les différentes parties l’organisme*) may exist. The example 1.f results from the substitution of two French NPs in a Creole NP-NP pattern.

   Another argument for allowing non-lexicalized constructions is to be found in data from other researchers. It has been shown that in cases of contact between languages with different word orders, some constructions did not follow the order prescribed by the lexical head. This is the case in examples like ‘I have to ***ttakē* my hands**’ ‘I have to wash my hands’ (Korean [OV] verb in an English VO verb phrase; from Choi, 1991, quoted by Chan 2009: 190) or ‘want ou Tex laat ons **daai *group join***’ ‘Because old Tex made us join that group’ (English [VO] verb in an Tsotsitaal OV verb phrase; from Slabbert and Myers-Scotto, 1997, quoted by Chan 2009). Although less frequent, such examples, as Chan put it, ‘resist a **lexicalist** account of verb-object order in CS, where verb-object order is specified in the head verb’ (2009: 191). They are

23 As one of the reviewers of this article pointed out, *un certain nombre de* would be analysed in Myers-Scotton and Jake’s terms (2000a: 1064) as an Embedded Language Island from French with a French bridge morpheme *de* ‘of’ (see above, Section 3.2).

a strong argument for the MLF model. So, if a 'null' theory of code-switching has an ambition to be generalized to different types of contact situations, it cannot be 100% lexicalized.

2. When seeking to understand the structure underlying utterances with internal language mixing, derivation history should not be forgotten. Some facts that appear contradictory, if one looks only at the static structure of language production, may be understood in the light of different derivation histories.

   The TAG formalism (used by Mahootian) offers an interesting concept for displaying that history: the derivation tree (Figure 4.a'-b'). The fact that about half of the NP-NP genitive constructions that can be observed (line 1-7 in Table 6) follow the French *N de N* structure, even in what could be described MC matrix language contexts, is best explained by the adjunction of French-like quasi-quantifiers or quasi-classifiers of the *a kind of* sort (see Section 5.3).

3. In a context where two languages in contact have a high number of items in common ('items' being used here for both surface lexical forms and elementary structures), language production often falls in the zone called by Muysken 'congruent lexicalization', that allows for what Ledegen calls 'floating' segments. In this zone, (1) congruent elementary structures are freely mixed, and (2) pre-terminal nodes may be freely filled by surface forms from languages α or β, all the more so that those forms are similar (see Section 5.2 above). Sometimes it is not easy, and perhaps even not relevant, to determine whether a surface form is α or β. As Clyne explained about his Dutch-English or German-English data from immigrant communities in Australia, 'some verbs are 'compromise forms' or lexical transfers promoted by partial phonological correspondence and are therefore common to both systems: '***Dit kan*** *be anywhere, You don't see* ***dat in Australië***' (1987: 760). The same could definitely be said about nouns in MC/French corpora. When surface forms are not outright identical in the received descriptions of both languages (like *espès/espèce* [espes]), some surface forms like *an* [ã], *nomb* [nõb] or *sòt* [sɔt] are very similar in Creole and in some variants of spoken French. In the grey zone of congruent lexicalization, my observation, formulated as maxim M3 in Section 5.2 above, is that the phonological realization of any given lexical item is not strongly constrained, and may be influenced by linear phonological constraints or recency effects.

In line with a tendency to trim down unification-based models of language from unnecessary categories (Vijay-Shanker & Joshi 1988; Abeillé 1993), which can also be found in recent evolutions of the minimalist school (Chomsky 1995), I did not try to describe specified NPs as DP (determiner phrases), QP (quantifier phrases), or KP (case-marked phrases), like Mahootian did. First, these categories are not always transferable across languages with different typological profiles (Chan 2009: 194). Moreover, many languages are flexible as to the conditions for merging constituents (MC certainly is), and increasing the specificity of categories can only unnecessarily tighten the syntactic limitations, when incorporating feature structures in syntactic models allows to give satisfactory accounts of all the desired phenomena. Abeillé (1993) eliminated all the categories except for the major ones, and represented (compulsory or optional) specifications as valued features, including the bar-levels (projection scope) of constituents. This can be applied here, and has been, for example, in Figures 3 and 4. I have kept the distinction (N, N', NP) in Figure 1 for the sake of notation brevity. Some have even argued (Vaillant 2014) that syntactic categories could be totally dispensed of, as long as feature structures allow to model the necessary constraints on unification (which are fewer than is generally thought). This point has not been explored here, but it certainly is stronger in code switching contexts — or even in informal spoken language — than in constrained written language with strict grammar checking.

To go further, I think it is possible to embrace the points 1, 2 and 3 above in a single model of bilingual syntax that will not need to postulate universal rules or to force the identification of a matrix language

even in dubious cases. That model is particularly adapted to cases of language contact between related languages (in the broader sense: genetically close languages, dialects of a same dialect family, creole **continua**). Vaillant (2008) has proposed, on grounds of economy of modelling, to factor the common structures of groups of related dialects, including a 'language' feature to allow for practical restriction, among a common set of elementary structures. However, as MacSwan convincingly argues (contra Belazi et al., 1994), the idea of a 'language' feature is problematic in that it posits as a primitive of grammar a mere name that is given to an arbitrarily defined social phenomenon, that is at most the description of a loose collection of features common to some thousands of individual I-languages (2000: 41). So, if using something like a 'language' feature might be useful in an NLP applicative context, to filter out single standardized languages from a repository of partially pooled language structures, it very probably isn't in the context of real-life language contact among multilingual speakers.

To sum up, the model of bilingual speech that adequately describes the MC-French corpora that I have observed and described in this paper has the following properties:

1. It is based on unification of tree-like elementary structures (either anchored in the lexicon, or reflecting frequent grammatical constructions that assign features to syntactic positions).
2. Unification may happen within two types of basic operations: substitution (of complements) or adjunction (of adjuncts).
3. At substitution nodes, feature checking determines which elementary structures may be unified; at adjunction nodes, since the node is split in an upper and a lower node in the operation, it is possible that 'top' features have different expectations than 'bottom' features.

Note: These first three points are fulfilled by the FS-TAG model (Vijay-Shanker & Joshi 1988).

4. No unnecessarily specific categories are needed (perhaps even none at all), as long as feature checking blocks uninterpretable matches.
5. No unnecessary language constraints are needed, as long as feature checking blocks the unification of incompatible structures (like DN and ND).
6. At surface level, when syntactic structures are compatible down to pre-terminal nodes, a degree of sloppiness is tolerated in the phonological realization of a lexical item that has similar forms in languages α or β. This can lead to the α form, to the β form, or to an interpolated (αβ) form.

In other words, the model I propose is very much like Mahootian's, but it doesn't postulate that elementary structures have to clearly belong to one language or another. Language is, in some way, underspecified. (E-)Language is not a concept that is in all cases relevant to describe the (I-)language of a bilingual speaker in contexts involving related (E-)languages.

Conclusion

The language contact situation in Martinique involves two languages, Martinican Creole (MC) and French, that have many lexical and structural similarities, but that also show diverging structural features, especially in Noun Phrases (NPs). Most MC noun modifiers are post-nominal, except for the indefinite article, which is used only in the singular, and a specialised plural morpheme. French, on the contrary, has pre-nominal articles, pre-nominal possessive personal determiners, and an obligatory pre-nominal plural indefinite article. MC and French also differ in the way they express the genitive relation between two full NPs: while MC simply puts the determiner ('possessor') NP after the determined NP, the same way it does with personal pronouns, French uses a joiner morpheme */də/*.

Bilingualism in MC and French is widespread in Martinique, and 'pure' monolingual speech is rare in natural communication contexts. So, it is interesting to see which structures are preferred in mixed discourse. In this article, I have observed, in particular, the structure of NPs in a corpus of MC-French language contact.

Remotivation through language contact

A first observation is the frequent use of what looks like French pre-nominal article in MC NPs. The French style pre-nominal plural indefinite /*de*/ and plural definite /*le*/ articles frequently appear along with MC nouns. The analysis proposed here is that these articles perform quantifier functions that are not in the 'base' system of (basilectal) MC.

In the case of /*de*/, it is a pure plural indefinite, that selects a set of countable elements, and can not be confused with a generic or partitive value (Table 3).

An analysis has been proposed to explain the role of the morpheme /*le*/ in MC: it actually is a MC morpheme in its own right, that carries an intensional (kind-referring) form of genericity (see Section 5.1 above). This analysis is confirmed by the data, but it fails to explain why /*le*/ also appears with the post-nominal definite article /*la*/, a phenomenon that was predicted impossible by Zribi-Hertz & Jean-Louis (2014). My suggestion is that the deictic/anaphoric value of the French adverb *là*, which is at the origin of the MC definite article /*la*/, has not completely disappeared in the latter, and may be reactivated in situations of language mixing.

Congruent lexicalization

A second very important observation in the MC/French corpora is the number of segments that are 'floating' between the two languages, where neither phonological, nor lexical, nor grammatical indices allow a clear-cut assignment to one language or another. This is a situation that is known to linguists who have studied contact between related languages (Section 5.2). I show that the notion of 'congruent lexicalization', proposed by Muysken (2000) as one of the three major types of code switching (along with the 'alternational' and 'insertional' types), is a good conceptual tool to understand what is happening in these floating zones. Everything is happening as if elementary lexical choices, where the surface forms of lexical items are similar and the local syntactic structures are compatible in the two languages, were underspecified up to the last moment, and could be determined by local rules of phonological harmony, or recency effects.

Importance of the derivation history to understand multiple switches

A last point that I make here is the importance of the derivation history of sentences to explain the structure of some mixed phrases. In fact, it appears that although in the bulk of my data MC plays the role of matrix language (according to several definitions of it), a great diversity of possible mixed outcomes are observed. For example, French-like genitive constructions (N *de* N) are nearly as frequent as MC-like ones (N N). It is also fairly frequent to observe multiple switches (or multiple levels of foreign constituent embedding) in NPs. Such observations could seem to evidence unpredictability in the way switches occur, when in fact they don't: it is possible to analyse them in terms of adjunction of multi-word modifiers within simpler constituents. For example, I analyse a sentence like 'ni an *certain nombre de* faktè' as an MC sentence 'ni faktè' ('there are factors') where a French NP adjunct, playing the role of a quasi-quantifier ('*un certain nombre de ...*': 'a certain number of') has been adjoined. It is therefore to be analysed not as two switches, but as a single elementary operation of syntactic merging (adjunction) drawing an elementary structure in an embedded language. In terms of the TAG formalism, that I have used to illustrate the process, the derivation tree (displaying the derivation history) brings more light on

the underlying structure of the observed utterance than the (static) phrase-structure tree, where the single adjunction operation appears as two embeddings.

An underspecified language model

To conclude, the data collected in MC/French mixing situations seems to be best explained with a model where speakers have access to a pool of elementary structures (most of them—but not all—anchored in the lexicon) that they are free to combine together as long as unification is not blocked by clashing feature structures.

The model I propose is very much inspired by the one used by Mahootian (1993) and termed 'null-theory of code switching', except that instead of having 'two separate systems produce acceptable mixed utterances' (Mahootian 1993: 138-139), I suggest that here we have a common, partly pooled system of elementary structures. In this system where language is underspecified, only some structures exist in two different alternative versions (like the post-nominal definite article of MC *vs.* the pre-nominal definite article of French). Unification of feature structures is the only mechanism necessary to account for the fact that these alternative versions do not clash.

Acknowledgements

I am indebted to Christelle Lengrai and Juliette Moustin, at that time master students in linguistics at the Université des Antilles et de la Guyane at Schoelcher (Martinique), who recorded the corpus and worked on the first stage of the transcriptions. Most of the work of further data annotation and analysis was done in the frame of the CLAPOTY research project, led by Isabelle Léglise, where I had the opportunity to share ideas and concepts with a team of talented linguists (Evangelia Adamou, Sophie Alby, Claudine Chamoreau, Gudrun Ledegen, Isabelle Léglise, Bettina Migge, Claire Saillard, Duna Troiani). In particular I would like to thank Claudine Chamoreau, who read an early draft of this article and made very useful comments. I also thank the three reviewers who helped me clarify many points in this article, and pointed me to new insights and research literature. I owe many thanks to Huw Sanderson, who took much time and did his best to reformulate the most awkward pieces of non-native English into regular English.

References

Abeillé, Anne. 1993. *Les Nouvelles Syntaxes: Grammaires d'unification et analyse du français*. Paris: Armand Colin.

Amuzu, Evershed K. 2014. Producing composite codeswitching: The role of the modularity of language production. *International Journal of Bilingualism* 18(4). 384-407.

Anselin, Alain. 1990. *L'émigration antillaise en France : la Troisième Île*. Paris: Karthala.

Barbaud, Philippe. 1998. Dissidence du français québécois et évolution dialectale. *Revue Québécoise de Linguistique* 26(2). 107-128.

Belazi, Hedi M., Edward J. Rubin & Almeida Jacqueline Toribio. 1994. Code-switching and X-Bar theory: The Functional Head Constraint. *Linguistic Inquiry* 25(2). 221-237.

Bentahila, Abdelali & Eirlys E. Davies. 1983. The syntax of Arabic-French code-switching. *Lingua* 59. 301-330.

Bernabé, Jean. 1983. *Fondal-natal : grammaire basilectale approchée des créoles guadeloupéen et martiniquais : approche sociolittéraire, sociolinguistique et syntaxique* (3 volumes). Paris: L'Harmattan.

Bernabé, Jean. 2003. *Précis de syntaxe créole*. Matoury (Guyane Fr.): Ibis Rouge.

Besters-Dilger, Juliane, Cynthia Demarkar, Stefan Pfänder & Achim Rabus (eds.). 2014. *Congruence in Contact-Induced Language Change: Language families, typological resemblance, and perceived similarity*. Linguae & Litterae 27. Berlin: De Gruyter.

Bickerton, Derek. 1973. The Nature of a Creole continuum. *Language* 49(3). 640-669.

Brunot, Ferdinand & Charles Bruneau. 1961. *Grammaire historique de la langue française*, 5th edn. Paris: Masson (1st edn. 1933).

Chan, Brian Hok-Shing. 2009. Code-switching between typologically distinct languages. In Bullock, Barbara E. & Almeida Jacqueline Toribio (eds.), *The Cambridge Handbook of Linguistic Code-switching*, 182-198. Cambridge: Cambridge University Press.

Chaudenson, Robert. 1993. Francophonie, « français zéro » et français régional. In Robillard, Didier de & Michel Beniamino (eds.), *Le Français dans l'espace francophone*, 385-404. Paris: Champion.

Chaudenson, Robert. 2001. *Creolization of language and culture*. London: Routledge.

Chomsky, Noam. 1995. Bare Phrase Structures. In Campos, Héctor and Paula Kempchinsky (eds.), *Evolution and Revolution in Linguistic Theory*. 51-109. Washington: Georgetown University Press.

Clyne, Michael. 1987. Constraints on code switching: how universal are they? *Linguistics* 25(4). 739-764.

Conwell, Marilyn J. & Alphonse Juilland. 1963. *Louisiana French Grammar*. The Hague: Mouton.

Corblin, Francis. 2011. Des définis para-intensionnels : être à l'hôpital, aller à l'école. *Langue Française* (171). 55-75.

Coupin, Sandrine & Gérard Forgeot. 2008. Lire, écrire, compter : la maîtrise des compétences-clé en Martinique. *Enquête IVQ Martinique 2006-2007.* Paris: INSEE. <http://www.insee.fr/fr/statistiques/1290005> (8 January 2020).

Croft, William. 2002. *Typology and Universals*, 2nd edn. Cambridge: Cambridge University Press (1st ed. 1990).

Croft, William. 2002a. Typological classification. In Croft (2002), 31-48.

DeCamp, David. 1971. Toward a generative analysis of a post-creole speech continuum. In Hymes, Dell (ed.), *Pidginization and Creolization of Languages*, 349-370. Cambridge: Cambridge University Press.

De Rooij, Vincent. 1995. Variation. In Muysken and Smith (eds.), 53-64.

Déprez, Viviane. 2001. Determiner Architecture and Phrasal Movement in French Lexifier Creoles. In Quer, Josep, Jan Schroten, Mauro Scorretti, Petra Sleeman & Els Verheugd-Daatzelaar (eds.), *Romance Languages and Linguistic Theory 2001: Selected papers from 'Going Romance', Amsterdam, 6–8 December 2001*, Current Issues in Linguistic Theory (245), 49-74. Amsterdam: John Benjamins.

Déprez, Viviane. 2008. Nominal constituents in French lexifier creoles: Probing the structuring role of grammaticalization. *Journal of Pidgin and Creole Languages*, vol. 22, no. 2, 263-307.

Di Sciullo, Anne-Marie, Pieter Muysken & Rajendra Singh. 1986. Government and code-mixing. *Journal of Linguistics* 22. 1-24.

Elisabeth, Léo. 2003. *La société martiniquaise aux XVII^e et XVIII^e siècles (1664-1789)*. Paris: Karthala.

Ferguson, Charles A. 1959. Diglossia. *Word* 15. 325-340.

Fattier, Dominique. 2000. Genèse de la détermination nominale en haïtien : l'empreinte africaine. *L'information grammaticale* 85. 29-46.

Fauquet, Georges. 1912. Note sur la population de la Martinique (Éléments ethniques et catégories sociales). *Bulletins et mémoires de la Société d'anthropologie de Paris* VI/3(3-4). 154-161.

Guillaume, Gustave. 1919. *Le problème de l'article et sa solution dans la langue française*. Paris: Hachette.

Gumperz, John J. & Robert Wilson. 1971. Convergence and creolization: A Case from the Indo-Aryan/Dravidian border in India. In Hymes, Dell (ed.), *Pidginization and Creolization of Languages*, 151-167. Cambridge: Cambridge University Press.

Hazaël-Massieux, Guy & Marie-Christine Hazaël-Massieux. 1996. Quel français parle-t-on aux Antilles ? In Robillard, Didier de, Michel Beniamino & Claudine Bavoux (eds.), *Le français dans l'espace francophone : Description linguistique et sociolinguistique de la francophonie*, 665-687. Paris: Honoré Champion.

Hazaël-Massieux, Marie-Christine. 2008. *Textes anciens en créole français de la Caraïbe*. Paris: Publibook.

Heath, Jeffrey. 1978. *Linguistic diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies. Quoted in Heine & Kuteva (2005), 166-167.

Heath, Jeffrey. 1981. A case of intensive lexical diffusion. *Language* 57(2). 335-367.

Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Heine, Bernd & Tania Kuteva. 2006. *The changing languages of Europe*. Oxford: Oxford University Press.

Heine, Bernd & Tania Kuteva. 2007. *The genesis of grammar*. Oxford: Oxford University Press.

Hentschel, Gerd. 2014. Belarusian and Russian in the Mixed Speech of Belarus. In Besters-Dilger et al. (eds.), 93-121.

Hicks, Caleb. 2012. A dual-structure analysis of morphosyntactic doubling in code switching. *Studies in the Linguistic Sciences: Illinois Working Papers* 37. 44-57.

Höder, Steffen. 2012. Multilingual constructions: a diasystematic approach to common structures. In Braunmüller, Kurt & Christoph Gabriel (eds.), *Multilingual Individuals and Multilingual Societies*, Hamburg Studies on Multilingualism 13. 241-258. Amsterdam: John Benjamins.

Jake, Janice, Carol Myers-Scotton & Steven Gross. 2002. Making a minimalist approach to codeswitching work: Adding the Matrix Language. *Bilingualism: Language and Cognition* 5(1). 69-91.

Joshi, Aravind K. & Yves Schabes. 1997. Tree-adjoining grammars. In Rozenberg, G. and Salomaa, A. (eds.), *Handbook of formal languages (3): Beyond Words*, 69-123. New-York: Springer.

Kamp, Hans, Josef Van Genabith & Uwe Reyle. 2011. Discourse Representation Theory. In Gabbay, Dov M. & Franz Guenthner (eds.), *Handbook of philosophical logic* (15), 125-394. Amsterdam: Springer.

Krifka, Manfred, Francis J. Pelletier, Gregory N. Carlson, Alice ter Meulen, Gennaro Chierchia & Godehard Link. 1995. Genericity: an introduction. In Carlson, Gregory N. & Francis J. Pelletier (eds.), *The Generic Book*, 1-124. Chicago: The University of Chicago Press.

Krifka, Manfred. 2003. Bare NPs: Kind-referring, indefinites, both, or neither? *Semantics And Linguistic Theory SALT* 13, Washington, 9th-11th May 2003. 180-203.

Ledegen, Gudrun. 2012. Prédicats « flottants » entre le créole acrolectal et le français à la Réunion : Explorations d'une zone ambiguë. In Chamoreau, Claudine & Laurence Goury (eds.), *Changements linguistiques et langues en contact: Approches plurielles du domaine prédicatif*, 251-270. Paris: CNRS Éditions.

Lefebvre, Claire. 1974. Discreteness and the linguistic continuum in Martinique. *Anthropological Linguistics* 16(2). 47-78.

Lengrai, Christelle, Juliette Moustin & Pascal Vaillant. 2006. Interférences syntaxiques et lexicales du français dans le créole martiniquais des émissions radiophoniques. In *3. Freiburger Arbeitstagung zur Romanistischen Korpuslinguistik: Korpora und Pragmatik*. Freiburg in Breisgau (Germany).

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1). 339-359.

MacSwan, Jeff (ed.). 1999. *A minimalist approach to intrasentential code switching*. New York: Garland.

MacSwan, Jeff. 2000. The architecture of the bilingual language faculty: evidence from intrasentential code switching. *Bilingualism: Language and Cognition* 3(1). 37-54.

MacSwan, Jeff (ed.). 2014. *Grammatical Theory and Bilingual Codeswitching*, Cambridge (Massachussets): MIT press.

MacSwan, Jeff. 2014a. Programs and proposals in codeswitching research: Unconstraining theories of bilingual language mixing. In MacSwan (ed.) (2014), 1-33.

Mahootian, Shahrzad. 1993. *A null theory of codeswitching*. Evanston, IL: Northwestern University Ph.D. Dissertation.

Mahootian, Shahrzad & Beatrice Santorini. 1996. Code switching and the complement/adjunct distinction. *Linguistic Inquiry* 27(3). 464-479.

Managan, Kathe. 2016. The sociolinguistic situation in Guadeloupe: Diglossia reconsidered. *Journal of Pidgin and Creole Languages* 31(2). 253-287.

March, Christian. 1996. *Le discours des mères martiniquaises*. Paris: L'Harmattan.

Migge, Bettina & Isabelle Léglise. 2013. *Exploring Language in a Multilingual Context: Variation, Interaction and Ideology in Language Documentation*. Cambridge: Cambridge University Press.

Milia-Marie-Luce, Monique. 2007. La grande migration des Antillais en France ou les années BUMIDOM. In Calmont, André & Cédric Audebert (eds.), *Dynamiques migratoires de la Caraïbe*, 93-104. Paris: Karthala.

Milroy, Lesley & Pieter Muysken (eds.). 1995. *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. Cambridge: Cambridge University Press.

Murray, David A. B. 1997. The cultural citizen: negations of race and language in the making of Martiniquais. *Anthropological Quarterly* 70(2). 79-90.

Muysken, Pieter & Norval Smith (eds.). 1995. *Pidgin and creoles: an introduction*, Creole Language Library (15). Amsterdam: John Benjamins.

Muysken, Pieter & Norval Smith. 1995a. The study of pidgin and creole languages. In Muysken and Smith (eds.), 3-14.

Muysken, Pieter. 2000. *Bilingual Speech. A Typology of Code-Mixing*. Cambridge: Cambridge University Press.

Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition 16(4)*. 709-730.

Myers-Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Oxford University Press.

Myers-Scotton, Carol. 1995. A lexically based model of code-switching. In Milroy and Muysken (eds.), 233-255.

Myers-Scotton, Carol & Janice Jake. 1995. Matching lemmas in a bilingual language competence and production model: evidence from intrasentential code-switching. *Linguistics* 33(5). 981-1024.

Myers-Scotton, Carol & Janice Jake. 2000a. Four types of morpheme: evidence from aphasia, code switching, and second-language acquisition. *Linguistics* 38(6). 1053-1100.

Myers-Scotton, Carol & Janice Jake. 2000b. Testing the 4-M model: An introduction. *International Journal of Bilingualism* 4(1). 1-8.

Nábělková, Mira. 2014. The case of Czech-Slovak language contact and contact-induced phenomena. In Besters-Dilger, J. et al. (eds.), 61-92.

Nicolas, Armand. 1998. *Histoire de la Martinique : de 1939 à 1971*. Paris: L'Harmattan.

Pinalie, Pierre & Jean Bernabé. 1999. *Grammaire du créole martiniquais*. Paris: L'Harmattan.

Prudent, Lambert-Félix. 1981. Diglossie et interlecte. *Langages* 15(61). 13-38.

Pulvar, Olivier. 2005. Créole et scène publique aux Antilles-Guyane françaises. *Études Caribéennes* 2. 75-83.

Pustka, Elissa. 2006. Le mythe du créole L1– et la naissance du français régional aux Antilles. *Romanistisches Jahrbuch* 57(1). Berlin: Walter de Gruyter.

Pustka, Elissa. 2007. Le français régional émergent en Guadeloupe. *Bulletin PFC (Phonologie du Français Contemporain)* (7). 261-271. Nanterre: MoDyCo.

Reutner, Ursula. 2005. *Sprache und Identität einer postkolonialen Gesellschaft im Zeitalter der Globalisierung. Eine Studie zu den französischen Antillen Guadeloupe und Martinique*. Kreolische Bibliothek (20). Hamburg: Buske.

Roberts, Nicholas S. 2014. *A Sociolinguistic Study of Grammatical Variation in Martinique French*. Ph.D. thesis, University of Newcastle (UK).

Russell, Bertrand. 1905. On Denoting. *Mind* 14(56). 479-493.

Sainton, Jean-Pierre, Richard Château-Degat & Lydie Ho-Fong-Choy Choucoutou. 2004. Vers le deuxième âge colonial des Petites Antilles : le tournant sucrier et ses incidences (1640-1685). In Sainton, Jean-Pierre (ed.), *Histoire et civilisation de la Caraïbe (Guadeloupe, Martinique, Petites Antilles), Tome I : Le temps des Genèses (des origines à 1685)*, 255-314. Paris: Maisonneuve & Larose.

Sankoff, David & Shana Poplack. 1981. A formal grammar for code-switching. *Papers in linguistics: International Journal of Human Communication* 14(1). 3-45.

Seuren, Pieter A. M. & Herman Wekker. 1986. Semantic transparency as a factor in Creole genesis. In Muysken, Pieter & Norval Smith (eds.), *Substrata versus Universals in Creole Genesis*, 57-70. Amsterdam: John Benjamins.

Singler, John Victor. 1995. The Demographics of Creole Genesis in the Caribbean: A Comparison of Martinique and Haiti. In Arends, Jacques (ed.), *The Early Stages of Creolization*, 203-232. Amsterdam: John Benjamins.

Sobotta, Elissa (= Pustka). 2006. Französisch und Kreolisch in Guadeloupe. In König, Torsten, Christoph Oliver Mayer, Laura Ramírez Sáinz & Nadine Wetzel (eds.), *Rand-Betrachtungen: Beiträge zum 21. Forum Junge Romanistik*, 99-114. Bonn: Romanistischer Verlag.

St-Hilaire, Aonghas. 2013. Creolizing globalization: pan-cultural identities and language in Saint-Lucia. *Caribbean Studies* 41(1). 3-25.

Stabler, Edward & Jeff MacSwan. 2014. A Minimalist parsing model for codeswitching. In MacSwan (ed.), 257-282.

Telchid, Sylviane. 1997. *Dictionnaire du français régional des Antilles – Guadeloupe, Martinique*. Paris : Bonneton.

Thibault, André. 2018. La syntaxe du factitif en francophonie et ses corrélats en créole. In Thibault, André (ed.), *Le causatif : perspectives croisées*, 185-223. Strasbourg: Eliphi.

Vaillant, Pascal. 2008. A layered grammar model: using Tree Adjoing Grammars to build a common syntactic kernel for related dialects. In *Proceedings of TAG+9 [International Workshop on Tree-Adjoining Grammars and Related Formalisms]*, Tübingen, 7-8 June 2008.

Vaillant, Pascal. 2014. La syntaxe, c'est de la sémantique. In Ablali, Driss, Sémir Badir & Dominique Ducard (eds.), *Documents, Textes, Œuvres : Perspectives sémiotiques*, 297-320. Presses Universitaires de Rennes, 2014.

Valdman, Albert. 1978. *Le créole. Structure, statut et origine*. Paris: Klincksieck.

Vijay-Shanker, K. & Aravind K. Joshi. 1988. Feature Structures Based Tree Adjoining Grammars. *Proceedings of COLING 1988 [12th International Conference on Computational Linguistics]*, Budapest, 22-27 August 1988, 714-719.

Wartburg, Walther von. 1969. *Évolution et structure de la langue française*, Bibliotheca Romanica, series prima (I). Berne: A. Francke (1st edn. 1946).

Werlen, Iwar. 1988. Swiss German Dialects and Swiss Standard High German. In Auer, Peter & Aldo Di Luzio (eds.), *Variation and convergence: Studies in social dialectology*, 94-124. Berlin: De Gruyter.

Wittmann, Henri. 1995. Grammaire comparée des variétés coloniales du français populaire de Paris du 17e siècle et origines du français québécois. *Revue Québécoise de Linguistique Théorique et Appliquée* 12. 281-334.

Wittmann, Henri. 1996. L'ouest français dans le français des Amériques: le jeu des isoglosses morphologiques et la genèse du dialecte acadien. In Cesbron, Georges (ed.), *L'Ouest français et la francophonie nord-américaine*, 127-136. Angers: Presses de l'Université d'Angers.

Zribi-Hertz, Anne & Loïc Jean-Louis. 2014. From Noun to Name: on Definiteness marking in modern Martinikè. In Cabredo Hofherr, Patricia & Anne Zribi-Hertz (eds.), *Crosslinguistic studies on Noun Phrase Structure and Reference*, 268-315. Leiden/Boston: Brill.