



HAL
open science

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

Pascal Vaillant

► **To cite this version:**

Pascal Vaillant. Noun Phrases in mixed Martinican Creole and French: Evidence for an Underspecified Language Model. 2016. hal-01451392

HAL Id: hal-01451392

<https://hal.science/hal-01451392>

Preprint submitted on 1 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Pascal Vaillant (E-mail: vaillant@univ-paris13.fr)¹

Abstract

The contact between French and Martinican Creole — a French-lexified Creole language spoken in the Lesser Antilles island of Martinique — takes place in a society where bilingualism is the standard, among people who are in majority fluent in both languages, and this situation leads to constant language mixing. French and Martinican Creole are historically related (which is acknowledgeable from the vast common lexical stock), and this might be a factor promoting language mixing, while possibly also blurring the distinction between the languages at the lexical level. Interesting questions arise from the fact that despite this relatedness, the two languages show significant typological divergences on some specific features. Some of those differences lie in the order between noun and definite determiner in the noun phrase (DN in French, ND in Creole), and in the use of a preposition to mark a possessive embedded noun phrase (compulsory in French, reputedly absent in Martinican). In the present paper, we are exploring the possible combination of the different values of these different features in mixed noun phrases occurring in corpora that exhibit a degree of language mixing. We inquire about the possible parameters which may influence the outcome and explain the relative frequencies of the different combinations. It appears that the prolonged contact with French has created a situation where both systems are present at the same time in the bilingual speakers minds, leading to a partially common pool of elementary structures. Many utterances fall into the category termed by Muysken (2000) “congruent lexicalization”. We also observe that apparent complex double embeddings have an internal logic, as they result from adjunction of multi-word modifiers. Finally we 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.

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

The island of Martinique, in the West Indies, is a place where two languages are commonly spoken by a majority of the population: 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, 1933; Wartburg, 1946/1969). French in its standard contemporary form, assumed to be the norm adopted by French speakers in Martinique, is generally considered to be the result of a process of *Ausbausprache* building among the cultured elites, mostly in Paris, after the 13th century (Cerquiglioni, 1991/2013), and among written varieties (Greub, 2007), although some claims have also been made to its being the result of a popular oral *koineization* in Paris (Lodge, 2004). In either hypotheses it involves dialect levelling among Northern French varieties (langue d’oil). French — or 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. In parallel, 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.

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 workforce from different regions of West Africa (Valdman 1978, Chaudenson 2001). MC has gone through a period of chaotic emergence in the 18th century, as the population of the island was changing very fast within the process of slave trade: the population of black slaves of African descent, roughly in the same numbers as 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, when at the same period the white population was multiplied by 2, and precisely when strict social segregation was established between the two populations (Elisabeth, 2003). The continuous flow of new speakers during the period led to a restructuring of French into what formed a new linguistic system, a process which is very often taken to be a definition of what is a creole language (“a language that has come

¹ LIMICS (UMRS INSERM 1142) — Université Paris 13, 74 rue Marcel Cachin, 93000 Bobigny, France.

into existence at a point in time”, Muysken and Smith, 1995a). This period is linguistically ill-documented, as many observers were not aware of witnessing the emergence of a new language, and mentioned linguistic productions as corrupted pronunciations of French (Hazaël-Massieux, 2008). After the abolition of slave trade (1815), then of slavery (1848), from the mid-19th century onwards, the population of the island, and the proportion of its components, has stabilized (Fauquet, 1912; Benoist, 1963), allowing for a stabilization of the linguistic system of MC. The structure of MC, and a sketch of its two-tier dialectal system (with diatopic and diastratic variation) was described by Bernabé (1983).

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 those 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 out of topic in the present article.

1.2. *Language contact situation*

Martinican Creole has a continuous history of contact with French, from its beginnings to the present day. No evidence whatsoever shows that there could have been any given moment in history where some MC speakers would not have been in contact with French. However, the sociolinguistic circumstances of the contact situation have evolved in 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)². 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% (INSEE, 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 of a growing proportion of the Martinican population with French, 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, was divided by 3.5 between 1960 and 1970 (Nicolas, 1998). In parallel, since 1963, the French government has promoted an active policy of emigration of adult workers from the French overseas department 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) (on 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 despise 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) has shown 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 performance³. As a consequence, more and more children start growing competences in MC at the age of middle school, 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 (personal observation of the author as a university teacher in Martinique between 2002 and 2007). During the same period though (from 1960 to present), MC has also gained ground on other fields. Utterly clandestine in the public media (radio mostly) in the first two decade of their existence in Martinique, it has begun to assert itself as a language of political expression of the workers' unions in the 1970s, at a time of social claims and political struggles linked to the crisis of traditional rural economy, and some years later it even earned a legitimate place in broadcasts on the public radio station (Pulvar, 2005). In a general context of emergence of post-colonial studies, and of renewed interest into regional languages and cultures in many regions of France and other European countries, numerous attempts have been made to promote MC as a language in uses that were traditionally reserved to French: literature, scientific popularization, 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 (see the age or gender

2 A catholic priest, l'Abbé Goux, witnesses in 1842 that the black people in plantations do not understand French (this is a rationale for writing a catechism in their language, not yet termed "Creole"; see Hazaël-Massieux, 2008).

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

specializations), from what they used to be when the reference descriptions were given, three decades ago.

1.3. Diglossia or Creole Continuum

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, 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 related⁴, but in practice mutually unintelligible, languages are used in a country, in very specific — and non-overlapping — sets of situations of use (usually simplified as a formal/informal opposition). Another 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 a (unidimensional) axis 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 pragmatical and social circumstances. The notion of a (post-) Creole continuum is often associated with the idea that there is a 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 fit 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) has challenged this claim and postulated an interlectal zone between French and Creole. Bernabé (1983) has 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, until proven wrong, is a good starting point for analyzing data in the Martinican language contact situation, although it is not desirable in all cases to insist on labelling every speech production, *a priori*, by pinpointing it to a specific position on the double continuum model.

1.4. Data used for the present study

In the present study, we mainly rely on data recorded on radio broadcasts in Martinique in 2005-2006. The corpus has been recorded and transcribed by two master students of the University of Antilles-Guyane in Schoelcher (Christelle Lengrai and Juliette Moustin), and annotated by the author. Some texts are transcriptions of a discourse by one single speaker, and some of conversations involving multiple speakers. Some are discourses of science popularization (that are expected to involve heavier influence from French), and some are related to more common-life topics (speaking habits, children performance at school...)

For our quantitative analyses, we have extracted a sample of 743 occurrences of noun phrases representing 500 distinct types. When not explicitly stated otherwise, 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 tapped from each of the 5 texts (a little more than 100, because some common noun phrases occur in more than one text).

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

Table 1. Sample of noun phrases tapped from our corpora.

4 In the classical Fergusonian view, diglossia implies relatedness of the involved languages, which display relationships such as ancient state vs. modern state of a language (Greek); standard supranational norm vs. regional norm (Switzerland); a mixture of both parameters (Arabic); or Creole vs. lexifier language (Haiti).

2. Possession and determination in the Noun Phrase

In this article, we 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.

2.1. Possible encodings of genitive

What follows is a very general overview of fairly common-ground categories used in linguistic typology, in the terms given e.g. by Croft (1990/2002a).

The expression of a grammatical relation like possession (or, more broadly speaking, *genitive*) between two nouns (or between a noun and a personal pronoun) may be encoded following different strategies in various languages. Let us suppose in the following that N stands for the possessed noun, and G for the possessor (noun or personal pronoun). The different possible encodings may involve either simple *juxtaposition* — which may as well occur between free lexical items (N G or G N, depending on the preferred order), as involve morphological binding (NG or GN) —, or the use of an *additional morpheme*. The additional morpheme, if present, may be indexical (if it merely denotes one of the units by re-encoding one of its closed-class grammatical categories and attaching it to the other unit), or relational (if it denotes the category of the grammatical relation itself, e.g. in this case genitive). In either cases, the additional morpheme (m) may be attached to one or the other units (N(m) G, G N(m), N G(m) or G(m) N), or be an independent unit (N m G, G m N, m N G, m G N, N G m, or G N m). Typical examples of bound relational morphemes are case affixes, whereas typical examples of free relational morphemes are adpositions (prepositions or postpositions, depending on the language's preferred order)⁵. When the case affix is attached to G, it is the typical case of what is called *genitive* (following the classical tradition used for the description of Greek and Latin); in the languages where it is attached to N, it is traditionally known under another name (like the *construct state* in Hebrew). In some cases ("fusion") one of the nouns or pronouns may have a specific form denoting its function as possessor or possessed, like in the case of the English or French possessive determiner (G' N)⁶.

2.2. Possible encodings of determination

In contrast to the relation of possession, which has to be expressed some way or another in all languages⁷, 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 article⁸. As noted by Heine and 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). In comparison to "the worldwide figure of 8%, over 39% of European languages have both definite and indefinite articles, and an additional 15% have definite but no indefinite articles" (2006: 99).

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." (Heine & Kuteva, 2006: 97). Dryer defines the indefinite article as a morpheme which "accompanies the noun and signals that the noun phrase is pragmatically indefinite in the sense that it denotes something not known to the hearer" (2005/2013a), and the definite article as a morpheme which "accompanies nouns and which codes definiteness or specificity" (2005/2013b). A more general definition of the article's denotation has been proposed by Guillaume: an article is a word which, given a set extension of the concept at the level of the noun type (e.g. *dog* as opposed to the more general *animal* or the more specific *terrier*), allows a finer, dynamic tuning of the extension at the level of the occurrence (1963: 260).

A well-known 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. By insisting on uniqueness, Russell seeks to capture the intuition

5 Free relational morphemes may also be "relational" nouns or verbs, which are used to this purpose, while remaining at the same time part of the open-class lexicon. It is the case for instance with Mandarin Chinese 給 (*gěi*), literally the verb "to give", used as a dative preposition. MC makes use of the same principle with the verb *ba(y)* ("to give", < Middle French "bailler"), also used as a dative preposition in competition with *pou* (< French preposition "pour") (Vaillant, 2003, p. 5); but MC has no such equivalent for the expression of genitive. Such relational nouns or verbs may get involved in a process of semantic bleaching, whereby they lose their original lexical meaning, and ultimately of grammaticalization, and become no more than an adposition (Heine & Kuteva, 2007, p. 62, 71).

6 This is also the case with "fused" case marking, e.g. when an archaic genitive mark has been fused by *umlaut* into a noun radical, and is not identifiable as a separate morpheme anymore.

7 Although the boundaries of its contents, around the core meaning of "possession", may of course be highly variable.

8 A notion rooted in the Greek category of *arthron*, coined by Denys Thrax' *Techné Grammatiké* in the 2nd century B.C.

that in a sentence like “The father of Bertrand Russell was English”, it is the reliable belief (shared by speaker and hearer) that “the father of Bertrand Russell” has a unique referent that makes the sentence interpretable. In a more recent formulaic representation, it can be stated by the following equivalences (here N stands for the predicate denoted by the noun in the NP, and P for the clause predicate):

$$[a \text{ N P}] = \lambda N \lambda P \exists x (N(x) \wedge P(x))$$

(there exists a x such that x is N and x does P)

$$[the \text{ N P}] = \lambda N \lambda P \exists x \forall y ((N(y) \Leftrightarrow y = x) \wedge P(x))$$

(there exists a x such that x is N, and that if anything is N, then it must be x, and x does P)

The universal quantifier used by Russell is of course dependent on the universe of the discourse, a fact that remains hidden as long as examples such as “the central mass of the solar system” or “the king of England” are used, but that must be accounted for with examples such as “the shop around the corner”. Actually, as noted by Löbner, “The definite article associates the referent with previous experience. [...] It is thus not uniqueness, but non-ambiguity, which is essential for definiteness” (1985: 291). More recent work in formal semantics has 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 has also 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). A discussion on those complementary views can be found in (Heusinger, 1997).

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 (a N, N a, aN or Na).

2.3. Encoding of genitive in French and in MC

In French, possessive is mainly expressed by the preposition ‘de’. In Old French (before the 14th century), there still was a form of case marking on nouns: a two-cases system (*cas sujet* and *cas régime*), relic of the fully-fledged nominal morphology of Latin (Banniard, 1997). This case system allowed some genitive constructions between two nouns to be expressed by juxtaposition, in constructions of the type ‘N G’, where G would be in the *cas régime* form (*le fils Hugon*: Hugues’ son). That construction was already coexisting with prepositional constructions like *le fils d’Hugon* or *le fils à Hugon*, and “[avait] peut-être, dès l’ancien français, un caractère archaïque” (Brunot & Bruneau, 1933). It is now found only in fossilized form like in some place names (*Bois-le-Roi*: the King’s Wood). In standard modern French the ‘N de G’⁹ form is considered the universal way of expressing possession between two nouns (1), even if some elliptical ‘N N’ constructions may exist in specific contexts (*une pause café*: a coffee break)¹⁰.

- (1) Le fils de Hugues
 DEF.M son of Hugues
 ‘Hugues’ son’

When possession is to be expressed between a noun and a personal pronoun, French makes use of a preposed possessive determiner indexing the person of the possessor, and the number and gender of the possessed noun (2a-2f). Note that the gender of the possessor (opposite to English) is not indexed (2d-2f may equally be used if the possessor is male or female). So in this case the preferred strategy for encoding possession is ‘G’ N’.

- (2) a. Mon bateau
 POSS.1SG>SG.M boat(M)
 ‘My boat’
 b. Ma maison
 POSS.1SG>SG.F house(F)
 ‘My house’

⁹ ‘N à G’ being also very common in spoken French.

¹⁰ Their interpretation as an expression of “genitive” would be subject to debate.

- c. Mes bateaux
POSS.1SG>PL boat(M).PL
'My boats'
- d. Son bateau
POSS.3SG>SG.M boat(M)
'His/her boat'
- e. Sa maison
POSS.3SG>SG.F house(F)
'His/her house'
- f. Ses maisons
POSS.3SG>PL house(F).PL
'His/her houses'

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 G', in all cases. It is the same when the possessor is a noun (3) and when it is a personal pronoun (4). 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.

- (3) Man pa enmen liv Kanmi
1SG NEG like book Kanmi
'I don't like Camille's books'
- (4) Man wè papa ou
1SG see father 2SG
'I have seen your father'

2.4. Encoding of determination in French and in MC

In French, following the classical grammar tradition (and school grammars), the noun may have four degrees of determination: zero (5a,5b), definite [DEF] (5c), indefinite [INDF] (5d), partitive [PART] (5e) (the glosses in 5c-5e are inspired from Brunot & Bruneau, 1933).

- (5) a. Céline travaille avec soin
Céline work.PRS.3SG with care
'Céline works with care'
- b. Je cherche une cage pour chats
1SG look_for.PRS.1SG INDF.SG.F cage(F) for cat.PL
'I am looking for a cat cage'
- c. Je viens prendre le pain
1SG come.PRS.1SG take.INF DEF.SG.M bread(M)
'I come to fetch the bread [the type and quantity I use to buy every day]'
- d. Je viens prendre un pain
1SG come.PRS.1SG take.INF INDF.SG.M bread(M)
'I come to fetch a [loaf of] bread'
- e. Je viens prendre du pain
1SG come.PRS.1SG take.INF PART.SG.M bread(M)
'I come to fetch some bread [e.g. half a pound of bread from a 2lb loaf]'

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 (5a) or noun modifier (5b), as an qualifying attribute with a copula (*Je suis docteur*), 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 some set expressions or proverbs.

Gary-Prieur describes the way the French articles, possessive pronouns, and demonstratives (and other adjectival determiners such as *quelques*, *plusieurs*, *certains*, *différents*, *divers*, *aucun*, *zéro*, *chaque*, *tout*, *quel*, *tel*) implement the universal discourse functions mentioned above. She synthesizes the role of the definite, indefinite and partitive articles by "*un N*: pose l'existence, dans la situation du discours, d'un objet particulier correspondant au sens de N [+dénombrable]; *du N*: pose l'existence, dans la situation du discours, d'un objet particulier correspondant au sens de N [-dénombrable]; *le N*: pp: il existe un objet de

catégorie N dans la situation du discours, p: donne l'instruction d'identifier cet objet et d'en envisager la totalité/unicité¹¹ (2011: 121). Gross (2012: 175-196) offers a detailed analysis of the category of noun determination in contemporary French, including the four degrees mentioned above (5a-e), but also including demonstrative, negative, interrogative, and quantifying determiners, as well as a wealth of other frozen constructions such as "a S of N", where S is a specifier or measure word (somehow similar in function to Chinese measure words, although not compulsory). In this section, however, we will stick to the basic, received description.

It is worth noting that the surface form of what is known as the partitive article is exactly similar, in all cases, to the addition of the genitive preposition "de" and of the definite article, both in contracted forms ("du" [GEN;SG;M] < "de" [GEN] + "le" [DEF;SG;M]; "des" [GEN;PL] < "de" [GEN] + "les" [DEF;PL]) and not contracted forms ("de la" [GEN SG.F])¹². It is also important to say that there is no distinction between partitive and indefinite in the plural (the unique form is "des"). This has been interpreted by some authors as signifying that there actually is no partitive in the plural; by others that any plural, even of a count noun, is automatically treated as a mass noun, which induces the use of the partitive and disables the possibility of using the indefinite¹³.

In French, both the indefinite and definite article are before the noun (a 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* ; *la gentille petite fille*). The indefinite article has evolved from a former cardinal number (Latin *unum, una*: one), a grammaticalization trend which is quite common across languages (Heine & Kuteva, 2002: 220). The definite article has evolved from a former demonstrative (Latin *ille, illa*: this), also a very common grammaticalization trend (Heine & Kuteva, 2002: 109-111).

In addition to the basic, "Russellian", values of the indefinite and definite article, French makes use of both articles with a generic value, in contexts like (6a-b).

- (6) a. Le cheval est un mammifère
 DEF.SG.M horse(M) be.PRS.3SG INDF.SG.M mammal(M)
 'Horses are mammals'
- b. Un cheval re-trouve toujours son chemin
 INDF.SG.M horse(M) back-find.PRS.3SG always POSS. 3SG>SG.M way(M)
 'A horse always finds its way back'

There is a nuance between (6a) and (6b), which Guillaume explained 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) (Guillaume, 1919). In most contexts, (6a) would 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)¹⁴. 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 the 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), as in (7a) and (7b) respectively. It is a broadly acknowledged fact that the border between weak deictic value and definiteness is tenuous, which also explains why in languages which mark definite determination, the morpheme often has evolved from a deictic marker (Heine and Kuteva, 2002, mentioned above).

- (7) a. Ouvrez cette porte
 Open.IMP DEM.SG.F door(F)
 'Open this door' (the door in front of me)
- b. Je vivais à douze ans dans les codes de ce monde
 1SG live.PST.IMP at 12 year.PL in DEF.PL code.PL of DEM.SG.M world(M)

11 In Gary-Prieur (2011), following Ducrot, "pp" stands for "présupposé" and "p" for "posé".

12 Compare with Russian, where the notion of partitive is expressed by using the genitive flexion of the noun (*хочу хлеба* [want.1SG bread.GEN.SG] = I want some bread).

13 There actually was a genuine plural indefinite article in Old French (*uns chevaliers* = some knights), that disappeared in Middle French to be replaced by the plural partitive (*des chevaliers*).

14 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 $\lambda P (P(e))$, while the type encountered in characterizing sentences (*A potato contains vitamin C and amino-acids*) is described by $\lambda P (\forall x (is_A(x,e) \Rightarrow P(x)))$.

'At the age of 12, I lived within the conventions of this world' (of the *milieu* described earlier)

The demonstrative is often reinforced by post-nominal invariable locative adverbs specifying either proximal (*ci*) or distal (*là*) deixis, as in (8) (there the locative value has been bleached out)¹⁵.

- (8) Ce verre-ci est à moi, ce verre-là est à toi
 DEM.SG.M glass(M)-PROX be.PRS at 1SG DEM.SG.M glass(M)-DIST be.PRS at 2SG
 'This glass is mine, that glass is yours'

Table 2 summarizes the grammaticalized system of determination in French.

determiner		zero	indefinite	definite	partitive	demonstrative
singular	masculine	∅	<i>un</i>	<i>le</i>	<i>du</i>	<i>ce</i>
	feminine	∅	<i>une</i>	<i>la</i>	<i>de la</i>	<i>cette</i>
plural		∅	<i>des</i>	<i>les</i>	<i>des</i>	<i>ces</i>
value(s)		(seldom used)	indefinite (5d) generic (6b)	definite (5c) generic (6a)	partitive (5e)	deictic (7a,7b)

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

In MC, the category of determination is not parallel to the one of Standard French. Reference grammars (e.g. Bernabé, 2003) identify two types of articles: an indefinite and a definite article. In a former description of the category of determination in MC (Vaillant, 2008), we have proposed to identify 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 (a N). 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 rêv* = an unpleasant dream). In terms of X-bar theory, the article is positioned at the left of the N', not merely of the N¹⁶. 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 (9).

- (9) Man ni an loto nèf
 1SG have INDF car new
 'I have a new car'

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" (10) (*ibid.*: 719).

- (10) Mi dé boug kouyon!
 PSTV INDF.PL man stupid
 'What silly blokes!'¹⁷

As will become evident in the corpora we analyzed, this morpheme *dé* is far from becoming extinct¹⁸. Unlike what Bernabé suggests, it appears in many contexts, not restricted to presentative constructions (11). But, unlike in French, it is not compulsory in noun phrases used as predicate arguments.

- (11) Ès ou pé kompwann dé bagay konsa?
 INT 2SG be_able understand INDF.PL thing like_that(ADV)
 'Can you understand things of that sort?'¹⁹

15 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, 1933, p. 246).

16 According to the classical description frame known as the X-bar theory (Jackendoff, 1977), it is useful to describe phrases as projections of the core part-of-speech categories, and it is relevant to distinguish three levels of projection: (1) the lexical unit itself, e.g. the noun: N⁰ (*boy*); (2) the lexical unit which may have been adjoined modifiers, but which still does not constitute a full phrase by itself: N' (*big boy*); (3) the complete phrase, which is the maximal projection of the category: N", or NP (*the big boy*).

17 The abbreviation PSTV is used for "presentative". The morpheme *mi* is a lexicalized presentative in MC. Other possible presentative constructions are cleft constructions that make use the verbs *sé* (BE) or *ni* (HAVE), like in many other languages (Lambrecht, 2001).

18 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 way 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, which consciously seeks to remain at a safe distance from French, in order to avoid "decreolization".

19 The abbreviation INT is used for "interrogative". The morpheme *ès* is a polar question marker, somehow like sentence final 嗎 in Mandarin Chinese, or sentence initial *est-ce que* in spoken French (of which it is a cognate).

The definite article is a morpheme postposed to the noun (N a). Its basic form is *la*. It is a highly likely hypothesis that it was derived from the French deictic adverb *là* (there)²⁰. This definite article bears no indexical mark of number or gender. Actually, it is highly objectionable to refer to categories such as gender or number 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: *la*, *a*, *lan* [lã] or *an* [ã]. The choice among those forms is guided by a principle of harmony (or “sandhi”, Bernabé 1983: 645) on two phonological features: the forms beginning with [l] are used when the preceding word ends with a consonant (10a, 10c), and the forms ending with [ã] are used when the preceding word ends with a nasal syllable (10c, 10d). As examples such as (10e) and (10f) display, this alternation is really not an agreement with an inner category of the noun, but 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 does occur.

- (12) a. l pa pé wè chat la
 3SG NEG can see cat DEF
 ‘He could not see the cat’
- b. l pa pé wè loto a
 3SG NEG can see car DEF
 ‘He could not see the car’
- c. l pa pé wè fanm lan
 3SG NEG can see woman DEF
 ‘He could not see the woman’
- d. l pa pé wè chien an
 3SG NEG can see dog DEF
 ‘He could not see the dog’
- e. l pa pé wè loto nèf la
 3SG NEG can see car new DEF
 ‘He could not see the new car’
- f. l pa pé wè chat mwen an
 3SG NEG can see cat 1SG DEF
 ‘He could not see my cat’

In many European languages (like in French, or German), articles are polyfunctional units which carry not only determination, but also number, gender, and case. In MC (and more generally in Caribbean French Creoles²¹), two different morphemes are used to express number and definiteness, and they are used only to that purpose²². 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 (13a), and with the demonstrative (13b), but not with the indefinite article.

20 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 contexts like *cet homme-là* (that man), although it keeps the value ‘distal’ as opposed to *ci* (‘proximal’) (Heine & Kuteva 2007, p. 84). In Standard French, 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, p. 299; Barbaud, 1998, p. 111-112; see also Conwell & Juilland, 1963, p. 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, p. 134) in Creole.

21 For a sketch of the compared core structures of NP and VP in four different French Caribbean Creoles (Haiti, Guadeloupe, Martinique and French Guiana), see Vaillant (2008).

22 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.

- (13) a. Sé timoun lan pa sa fè anyen
 PL child DEF NEG know do nothing
 'The children can't do anything'
- b. Sé timoun tala pa sa fè anyen
 PL child DEM NEG know do nothing
 'These children can't do anything'

In MC, the definite article is dropped out when a noun is followed by another noun in a genitive complement, in cases when 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, e.g.: *tianmay mèt kay la* = the householder's child, *jwè tianmay mèt kay la* = the householder's child's toy.

However, the definite article has to be repeated at the rightmost position of the NP in the cases where a long adjunct like a relative clause is put after the noun (Bernabé, 1983: 924). This means that in such cases, the definite article appears twice: first just after the noun and possible postposed articles, and last after the relative clause (14).

- (14) Boug la man ja palé ba ou la
 bloke DEF 1SG already talk to 2SG DEF
 'The bloke I already told you about'

Bernabé calls this phenomenon a "demarcating function" of the relative article²³.

Finally, a very common degree of determination in MC is the degree zero: bare nouns. Unlike in French and other "SAE"²⁴ languages, bare nouns have a very broad range of use, and can perfectly constitute NP by themselves. They are the standard expression of the "generic" value, as witnesses their pervasive use in proverbs (15).

- (15) Ravèt pa ka ni rèzon douvan poul
 roach NEG IMPF have right in_front_of hen
 '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 exposed (2014), MC also has 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 (16). So, unlike in French, the "intensional" type of genericity is not expressed in the same way as definiteness.

- (16) Lè difé pri, fòk kriyé lé ponpyé
 when fire break_out must call lé fireman
 'When a fire has broken out, one must call the Fire Brigade' (not: some specific firemen)

Bare nouns are the most common way of expressing the indefinite value (introducing new discourse entities) for plural referents (17).

- (17) Sa sé bagay ki ka rivé
 DEM COP thing REL;SBJ IMPF happen
 'These are things that happen'²⁵

Bare nouns are also 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 (18).

- (18) Ba mwèn dlo
 give 1SG water
 'Give me (some) water'²⁶

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* (19). It can be further analysed as a demonstrative morpheme and a definite morpheme, as a comparison with other French-based Creoles endorses.

23 I take it that it suggests that at the end of verbs or verb phrases, *la* plays the role of a nominalizer, somehow like *one* in Singapore English, that may appear in relative clauses like *the fruit they grow one very sweet* = the fruit that they grow is very sweet (Alsagoff & Lick, 1998, p. 134), in some respects on the model of Chinese 的.

24 SAE: Standard Average European — a term borrowed to Whorf and much used in Heine & Kuteva (2006).

25 Here the copula verb typically is used to introduce a cleft construction carrying predicate focus (Lambrecht, 2001).

26 Note that in this example, the French partitive article (which frequently precedes mass nouns in French: *de l'eau*) has been reanalyzed and incorporated in the Creole noun (*dlo*).

- (19) Man té sav ou té kay pozé mwen kèsion ta-la
 1SG PST know 2SG PST PROSP set 1SG question DEM-DEF
 'I knew you would ask me this question'²⁷

Table 3 summarizes the grammaticalized system of determination in MC.

determiner	zero	indefinite	definite	demonstrative
singular	∅	<i>an N</i>	<i>N la</i>	<i>N tala</i>
plural		<i>dé N</i>	<i>sé N la</i>	<i>sé N tala</i>
value(s)	generic (15) indefinite plural (17) partitive (18)	indefinite singular (9) indefinite plural (11)	definite (12a,13a) (‘plural’ is carried by another morpheme)	deictic (19) (‘plural’ is carried 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 restricted 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.

In this table, we 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”, we 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 are possible for a given value, we give the least marked first.

value	French	MC
N/GNR (generic noun)	INDF.SG N DEF N	N (in some special cases: <i>lé N</i>)
N/PART (partitive)	<i>de</i> 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 <i>de</i> NP2	NP1 NP2
NP/PP _n (genitive between a full NP and a person index <i>n</i>)	POSS. <i>n</i> NP (possessive determiner)	NP PRN. <i>n</i> (full personal pronoun)

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

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 DET N and β the order N DET, then the expected monolingual productions should be $[\text{DET}_\alpha \text{N}_\alpha]_\alpha$ and $[\text{N}_\beta \text{DET}_\beta]_\beta$.

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 $[\text{DET}_\alpha \text{N}_\beta]_\alpha$ (structurally “ α -like” NP where the noun is borrowed from language β) or $[\text{DET}_\beta \text{N}_\alpha]_\alpha$ (the determiner being borrowed from language β). Symmetrically, we could find $[\text{N}_\alpha \text{DET}_\beta]_\beta$ or $[\text{N}_\beta \text{DET}_\alpha]_\beta$. Additionally, if the structure may be

²⁷ The combination of the past (PST) and prospective (PROSP) preverbal particles, in sentences like (19), yields a conditional value.

in a language and the lexical material in another, it will also be possible to observe structures like [DET_β N_β]_α or [N_α DET_α]_β.

As Chan (2009: 193) observes, in the case of diverging order in the two languages, two other possibilities are also theoretically possible, even if they occur more scarcely: the two possible orders are activated at the same time (doubling, which violates the economy constraint), or none of them do (deletion, which violates the isomorphism constraint). We will consider possible cases of doubling in Section 5.1 below.

3.2. Purported constraints on the possible combinations

Some hypotheses on the internal constraints on 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 imposes that switches are only possible at points where the grammars of the two languages are equivalent as 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 [DET_α N_β], [DET_β N_α], [N_α DET_β], and [N_β DET_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 the possible combinations, although they prefer to express them in terms of deep structure rather than in linear terms. Their proposition is that beside possible constraints specific to some language contact situations, there is one universal constraint on 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’ conception of “government” in this context is defined as minimal (not maximal) c-command, this constraint would not prevent switches between a DET 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 often in the same language as the following clause, than in the same language as its governing noun or verb (opposite to what Di Sciullo *et al.*’s government constraint predicts), 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 adapts to many observed utterances containing a subordinate clause, but it also predicts that no switches should be observed between a DET and a N, even when the NP constituent orders are congruent in the two languages; however, many such switches are attested 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 G_x, must obey grammar G_x”, *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 G_x”), 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 hints, Myers-Scotton (1993) proposes a model where there always is a possibility 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 inserts 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 to some extent to the structure of NPs (1993: 85). The morpheme order principle (MO) 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 has not many reasons to apply to the internal structure of NPs in languages which do not use case-marking morphemes, but in our case 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, 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 [DET_α N_α] and [N_β DET_β] are always possible; (b) the combinations [DET_α N_β] is only possible when α is the ML, and the combination [N_α DET_β] is only possible when β is the ML; (c) the combinations [DET_β N_α] or [N_β DET_α] are never possible. In a later work on the same theoretical assumptions, Jake, Myers-Scotton and Gross (2002) 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” (1993: 78-79). If this contention is correct, the combinations [DET_α N_α] and [N_β DET_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”, distinguishes “outsider system morphemes” (which have links outside their constituent) and “bridge system morphemes” (which are simply compelled by the internal constructional rules of the language of the constituent). The former are expected to be ML, where the latter may as well be EL if the constituent is an EL island. This theoretical standpoint predicts that the French preposition *de*, being a “bridge” system morpheme, 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 the possessive determiners: as noted above, they are expected to be in the Matrix Language.

The Matrix Language approach brings a constraint along: it compels to identify the matrix language of any utterance — before being able to make predictions about its structure —, and to provide criteria to do 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 to 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 complemented verbal clause, namely, as the language providing the grammatical frame of the CP (Complementer Phrase²⁸). This criterion, however, might 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 exhibited 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 a certain robustness to explain many language mixing data (Myers-Scotton & Jake, 2000b: 1). 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 it, “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 propose 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 like such as Belazi, Rubin and Toribio’s, or Myers-Scotton’s, who make use of this language feature); actually, 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. The second one is that there may

28 In formal grammar, a CP is a sentence with an optional complementer (that may be \emptyset). This definition includes complete, finite sentences, as well as subordinate clauses which actually are sentences with a complementer that allows them to fill an argument slot in verbs which require sentence-like complements (like “that I know nothing”).

even makes it quite debatable whether it is appropriate to speak of situations of contact between Creole and their lexifier language in terms of alternation or insertion of constituents (see below, Section 5.2).

4. Observations in actual corpora

4.1. Attested in mixed language corpora

However we sum up, in Table 5, the combinations which are predicted to be unlikely to occur according to the different purported constraints on code switching.

	<i>combination:</i>	<i>predicted:</i>	<i>according to:</i>	<i>specific feature:</i>	<i># in sample</i>	<i>%</i>
1	INDF.SG _c N _f	impossible	FH		39	7.8
2	INDF.SG _f N _c	impossible	FH		0	0.0
3	INDF.PL _f (<i>de</i>) N _c	impossible	FH		2	0.4
4	[DEF.SG _f N _c] _c	impossible	MO		3	0.6
5	[DEF.SG _f N _f] _c	dispreferred	BNP		15	3.0
6	DEF.SG _f N _c	impossible	E; FH		3	0.6
7	DEF.SG _f (<i>la/la</i>) N DEF _c (<i>a</i>)	impossible	E; FH	doubling	3	0.6
8	N _c DEF.SG _f	impossible	E		0	0.0
9	[N _f DEF _c] _f	impossible	MO		2	0.4
10	[N _c DEF _c] _f	dispreferred	BNP		0	0.0
11	N _f DEF _c	impossible	E; FH		62	12.4
12	[DEF.PL _f (<i>le</i>) N _c] _c	impossible	MO		8	1.6
13	[DEF.PL _f (<i>le</i>) N _f] _c	dispreferred	BNP		20	4.0
14	DEF.PL _f (<i>le</i>) N _c	impossible	E; FH		8	1.6
15	DEF.PL _f (<i>le</i>) N DEF _c (<i>a</i>)	impossible	E; FH; MO	doubling	1	0.2
16	PL _c (<i>se</i>) DEF.PL _f (<i>le</i>) N	impossible	E; MO	doubling	0	0.0
17	DEF.PL _f (<i>le</i>) PL _c (<i>se</i>) N	impossible	MO	doubling	0	0.0
18	PL _c N _f	impossible	E		10	2.0
19	[PL _c N _c DEF _c] _f	dispreferred	BNP		1	0.2
20	DEM.SG _f N _c	impossible	E; FH		0	0.0
21	N _c DEM.SG _f	impossible	E		0	0.0
22	DEM.PL _f (<i>se</i>) N _c	impossible	E; FH		0	0.0
23	PL _c (<i>se</i>) DEM.PL _f (<i>se</i>) N	impossible	E	doubling	0	0.0
24	NP1 _c <i>də</i> NP2 _f	impossible	G		1	0.2
25	NP1 _c <i>də</i> NP2 _c	impossible	FH		1	0.2
26	NP1 _f <i>də</i> NP2 _c	impossible	FH		17	3.4
27	[POSS. <i>nf</i> NP _c] _c	impossible	MO		0	0.0
28	[POSS. <i>nf</i> NP _f] _c	dispreferred	BNP		0	0.0
29	POSS. <i>nf</i> NP _c	impossible	E		0	0.0
30	NP _c POSS. <i>nf</i>	impossible	E; MO if ML= <i>f</i> ; SM if ML= <i>c</i>		0	0.0
31	[NP _f PRN. <i>nc</i>] _f	impossible	MO		0	0.0
32	[NP _c PRN. <i>nc</i>] _f	dispreferred	BNP		0	0.0
33	NP _f PRN. <i>nc</i>	impossible	E; G		14	2.8
34	PRN. <i>nc</i> NP _f	impossible	E; G; MO if ML= <i>c</i> ; SM if ML= <i>f</i>		0	0

Table 5. Constraints on possible combinations between 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 the reservations expressed above about the limited possibility to unequivocally identify the language); when a matrix language is involved in the prediction, we note the matrix language as an index letter spanning the whole constituent (in square brackets). In the third 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 fourth column, we note when some specific issue appears, like suspected cases of “doubling” (when it could seem that the same grammatical or semantic value is exposed twice in the constituent); this issue raises another question, namely whether we are sure to correctly identify the two morphemes involved in the suspected cases. Figures in the fifth and sixth columns are numbers of types, not tokens.

An important comment is necessary when interpreting the results of Table 5. Some lexical items of MC are similar in their phonetic form with some lexical items of French, and their syntagmatic combination properties also are partly identical, although their semantic value is not. For instance, assuming we have a noun with a similar phonetic form in MC and in French, e.g. *tas* (MC) / *tasse* (French) (cup, mug). An NP realized /setasla/ could as well correspond to (21a) as to (21b):

- (21) a. sé tas la (MC)
 PL cup DEF
 ‘the cups’
- b. ces tasses là (French)
 DEM.PL cup DEIC²⁹
 ‘those cups’

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

In many cases, in order to decide whether a word should be dominantly classified as French or as MC (when counting occurrences of some patterns in Table 5), we had to resort to dictionaries (which reflect a partial, and biased, view of the lexicon), or to our own judgement in deciding if it was more “typically” Creole or more “typically” French. We have adopted a tolerant judgement, favouring the risk of overestimating over the risk of missing actual phenomena.

Another comment is that the predictions of the MLF Model (Myers-Scotton, 1993; Jake *et al.*, 2002) are hardly testable in all possible configurations, since in the major part of our sample, MC is the dominant matrix language (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 the more unmarked 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). On our 743 NP occurrences, 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 much 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 which can be expressed by the maxim:

(M1) Any French N' can potentially be used as a MC N' in a MC NP³⁰.

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).

2. The significant counts in lines 13, 5, and 33 of Table 5 allow to suspect 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 restrictions on the possibility of using a complete French NP in a MC

²⁹ The French locative (distal) adverb *là* often post-cliticizes to the noun and takes up a deictic or anaphorical value (see footnote 19).

³⁰ N': N-bar: a noun, or a noun-headed constituent with a noun and some of its possible complements (adjectives, relative clauses, possessive complements...), but without a determiner (see footnote 15). Some authors (Muysken, 2000, p. 61) use NP for what we here call N', and DP (Determiner Phrase) for what we call an NP.

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 our data does not prove wrong 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 Sciullio *et al.*’s (1986) Gouvernement constraint, but are 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 at the same time with the MC postponed definite determiner (1 occurrence in the present 500 NP types sample — line 15 of Table 5 —; but there are too many other examples in the remainder of our 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 a morpheme *lé* in MC, to which we will go back further down (Section 5.1).
4. Finally, a noteworthy observation is the relatively high number of possessive constructions of the type NP1_f d_ə NP2_c (line 26 in Table 5 — also to be found as lines 2, 3 and 5 of Table 6, below). This, we think, witnesses for the fact 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 complements). We return to that question in Section 5.3.

4.2. Embedded structures

As a matter of fact, the most interesting observations in the corpus concern NPs containing embedded constituents (relative clauses, genitive complement NPs), since they are a testbed for possible structural explanations of code-mixing. We 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 NP1 the head constituent (the “possessee” or “determined” one) and NP2 the dependent constituent (the “possessor” or “determining” one), MC has one single structure, NP1 NP2, used both when the two constituents are actual “full” NPs (*kay manman*: mother’s house) as when NP2 is a person pronoun (*kay mwen*: my house)³¹. 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 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? Considering the data, there is no absolute answer to these questions, but there are statistical tendencies.

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

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

no	top NP node		bottom NP node		example	# in sample	%
	structure	head	structure	head			
1	French	French	French	French	<i>goutte-à-goutte de [production de [Ø urine par rein a]]</i>	3	4.3
2	French	French	Creole	French	<i>des contractions de [sé muscle a]</i>	15	21.4
3	French	French	Creole	Creole	<i>évacuation vers l’avant de [pisa a]</i>	1	1.4
4	French	French	F/C	French	<i>an certain nombre de [professions]</i>	11	15.7

31 The expression of possession is a point of divergence between MC and closely related French-based Creoles. The Creole of Cayenne (used in the central part of French Guiana) has a possessive determiner, like French: *mo kaz*: my house. The Creole of Guadeloupe uses a joiner word, like French (*a/an*): *kay an mwen*: my house.

5	French	French	F/C	Creole	an certain <u>nombre</u> de [faktè]	1	1.4
6	French	Creole	French	French	an sòt de [<u>filtration</u> de [sang ta'a]]	1	1.4
7	French	Creole	F/C	F/C	<u>witè</u> (huit heures) de [tan (temps)]	1	1.4
8	Creole	French	French	French	les différentes <u>parties</u> \emptyset [<u>l'organisme</u>]	7	10.0
9	Creole	French	Creole	French	<u>paroi</u> musculaire \emptyset [<u>vessie</u> a]	11	15.7
10	Creole	French	Creole	Creole	les <u>parties</u> génitales \emptyset [<u>lé misié épi</u> lé madanm]	8	11.4
11	Creole	French	F/C	Creole	<u>veines</u> \emptyset [moun]	6	8.6
12	Creole	F/C	F/C	F/C	<u>problem</u> (problème) \emptyset [batri (batterie)]	3	4.3
13	Creole	Creole	Creole	French	bout \emptyset [tout système]	2	2.8

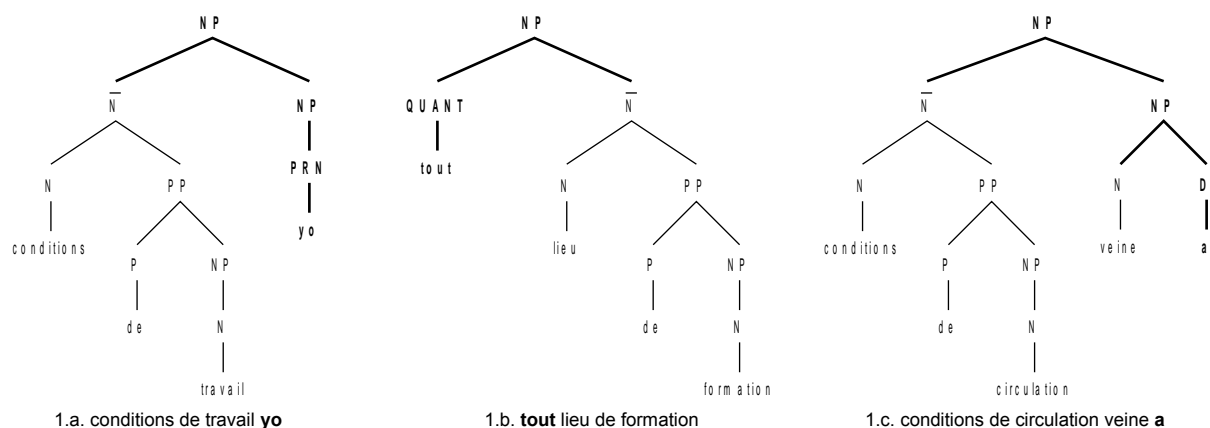
Table 6. *Different configurations of NP-NP embeddings in mixed genitive constructions. There are 70 mixed genitive NPs (on a total of 108 genitive NPs). A genitive construction involves a dependency between two NP nodes, a top node and a bottom node. We have observed the values of two parameters for each of these two NP nodes: the language that governs its structure — determined by unambiguous morphosyntactic features, e.g. 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 clearer, (1) it is explicitly indicated (\emptyset) when 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.*

The first observation to be made is that practically all of the possible combinations are actually present, which makes 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 undistinguishable for some frequent values, most notably the indefinite (INDF N) and 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 obviously displays no structural features of either language.

The most important parameter that gives a clue to a specific language, on the level of the top NP node, is the use of an overt joiner morpheme (*de* vs. \emptyset). Here the number of top NPs using the Creole structure (\emptyset) for expressing possession (lines 8-13) is only slightly over the number of those using the French structure (*de*), despite the fact that the matrix language, provided there is one, is mainly Creole in our corpus³². 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 but 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 our 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 our data.

We are left with an open question: how is it possible to describe the structure of mixed embedded NPs? In our 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, and any type of mixture of the sort (examples are shown in Figure 1).



32 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).

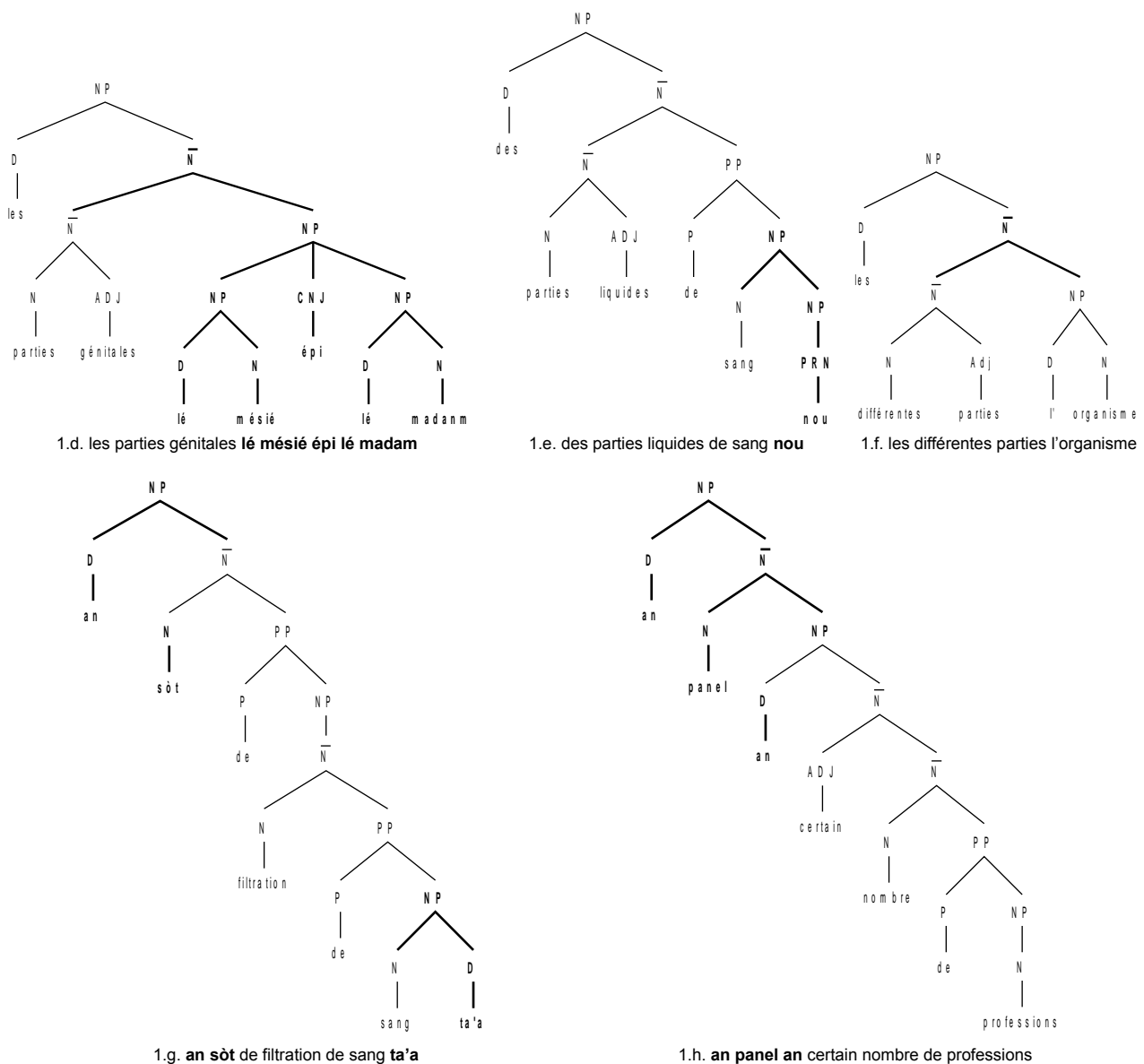


Figure 1. Typical examples of mixed NPs displaying language mixing at both lexical and structural levels. In these examples, unambiguously Creole morphemes or structures are displayed in bold. Note: in 1.b **tout** is pronounced [tut], which is what makes it Creole; in 1.f all the morphemes are French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

Such multiple-level embeddings seem to challenge any claim to general rules governing the insertion of constituents in alternate languages — since two successive insertions 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.

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 our 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 should 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 come along with a post-nominal definite (*la*) or demonstrative (*tala*) marker, and should be less frequent than the more widespread *sé* form. Our data does not show the expected distribution. Against 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*...). Against 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)³³.

There is an alternative explanation to *lé*. Zribi-Hertz & Jean-Louis (2014) have shown that *lé*, with plural nouns, carries an “kind-referring” value (as opposed to the “set-referring” value carried by *sé*³⁴). Their analysis explains many of the examples found in our sample, like in (22).

- (22) a. *lé les parties génitales lé misié épi les madanm* (MC)
 DEF;PL parts genital DEF;PL man with DEF;PL woman (French)
 ‘the genital organs of men and women’ (i.e. the masculine and feminine genital organs)
- b. *ki plas kreyol la pé ni andidan lé média* (MC)
 which place créole DEF can have within DEF;PL média (French)
 ‘which place can Creole have on the (mass) media’ (i.e. media in general, 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 (23).

- (23) *lé jenn manmay ka chwazi an branch* (MC)
les 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 the fact that some occurrences simultaneously display the preposed *lé* and the postposed definite marker *la*, like in (24).

- (24) *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 (24): 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 (24), 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 we 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 31). If they were to be interpreted as mere “borrowings” of the French definite article, their presence with phonetically typical MC nouns 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 at simultaneously respecting 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 example (25) shows.

- (25) a. *lidéal sété di montré lé manmay ki manniè lang kreyol ka woulé* (MC)

33 We do not have biographical data for all of the speakers in the corpus from which our 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).

34 See above: Section 2.4, especially example (16).

- l'idéal c'était de montrer les (French)
 ideal was to show DEF;PL kid which way language creole IPFV roll
 'Our goal, ideally, was to show the kids the way the Creole language functions'
- b. [...] ki ka montré lé moun lan ki manniè lang lan ka woulé (MC)
 qui montrer les (French)
 who IPFV 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 willing
 'because those people had the willing to do so.'

The contrast between the values of the different plural definite constructions is typically illustrated by (25a), (25b) and (25d), that consecutively appear in a span of text #4. In the discourse context, it seems that the first construction in (25a), *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 adults)³⁵; the second in (25b), *lé moun lan*, is another generic definite description of the same (above-mentioned) target group; the third in (25d), *sé moun lan*, is a definite description that more specifically denotes the "couple of people" that were introduced in (25c). Very few extracts display the complete set of combinations of values that may be seen in (25), 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.

(25a)	<i>lé N</i>	kind-referring	– deictic/anaphora	N as a category
(25b)	<i>lé N la</i>	kind-referring	+ deictic/anaphora	N as a category, definite in the context
(25d)	<i>sé N la</i>	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 19, example 21, footnote 28). 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 (25b), where *lé* is simultaneously present with *la*: the 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), the 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 all 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 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 has tried to make a distinction between what he considers 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, "A...B" for alternation, "A₁ ... A_n ~ B₁ ... B_n" for congruent lexicalization (2000: 31).

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

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**. Consider the following example:

Weet jij [**whaar**] Jenny is?

‘Do you know where Jenny is?’ (Dutch: waar Jenny is)

(English/Dutch; Crama and van Gelderen 1984)

The sequence *where Jenny is* could as easily be English in structure as Dutch. Furthermore *where* is close to Dutch *waar* (particularly when pronounced by bilinguals), *Jenny* is a name in both languages, and *is* is homophonous.

(...) The term congruent lexicalization refers to a situation where the two languages share a grammatical structure which can be filled lexically with elements from either language.” (*ibid.*: 5-6).

As will be obvious to people familiar with contact between a Creole and its lexifier language, many of our data may fall into this last category (see above, end of Section 3.2, esp. example (20)). It is therefore not surprising that in Table 5 we find many counter-examples to constraints that simply do not apply to the situation (especially line 1 in Table 5).

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 our 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) syntagms, 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 of the syntagms 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, perhaps identical; 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 *de* — 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 our knowledge, has proven 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 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 in example (26) (example 106 from Mahootian, 1993: 152).

(26) **you'll buy** xune-ye jaedid (English/Farsi)
house-PART. new
‘you'll buy a new house’

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 optional operation that allows a special tree (its head node cannot be the root of a syntagm, 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.

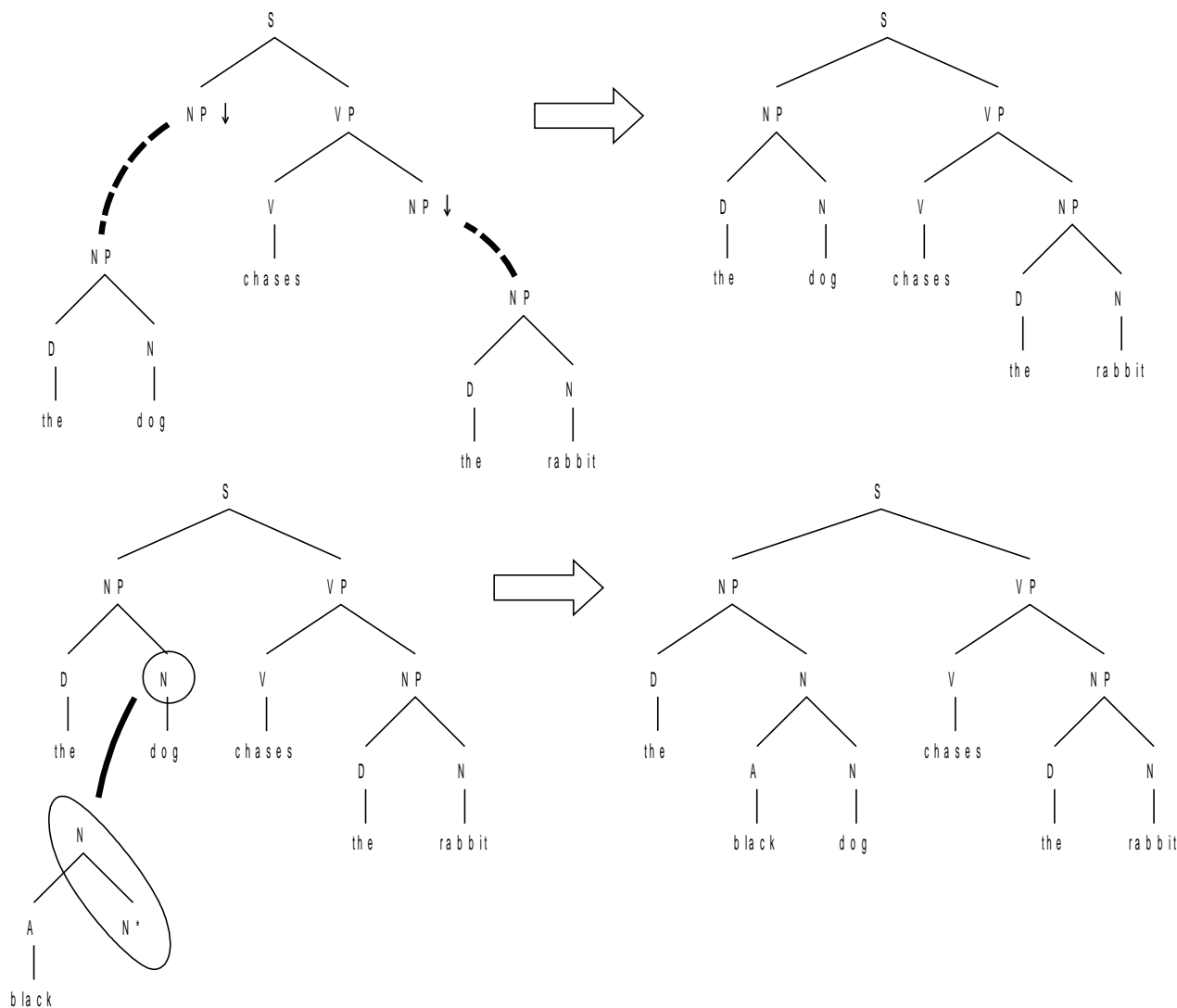


Figure 2. a. (above) Substitution of complements. b. (below) Adjunction of adjuncts. The fact that elementary structures are trees allows the model to naturally encode elementary constructions, like the english SVO order in 2.a, but also, if needed, set multi-word expressions (like in "John kicked the 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 code-switching, lexical heads seem to project the grammatical structure of the language they have been chosen from³⁶.

Our corpora provide plenty of evidence supporting Mahootian's model of mixed constituents by unification of elementary trees in a tree-adjoining grammars (actually, I have not been able to find a counter-example). A notion I now would like to insist on is the interest of the operation of adjunction.

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

36 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 rules" (1983: 321). Belazi, Rubin & Toribio say that "a word of language X, with grammar G_x, must obey grammar G_x" (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).

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 d_ə 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 (sôt) de N*, *un (an) ensemble (ansamb) de N*, *un (an) espèce de N* (examples from our corpora: 27.a-d).

- (27) a. an serten nomb
 un certain nombre de professions
 'a certain number of professions'
- b. an sôt 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 we are 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 or 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).

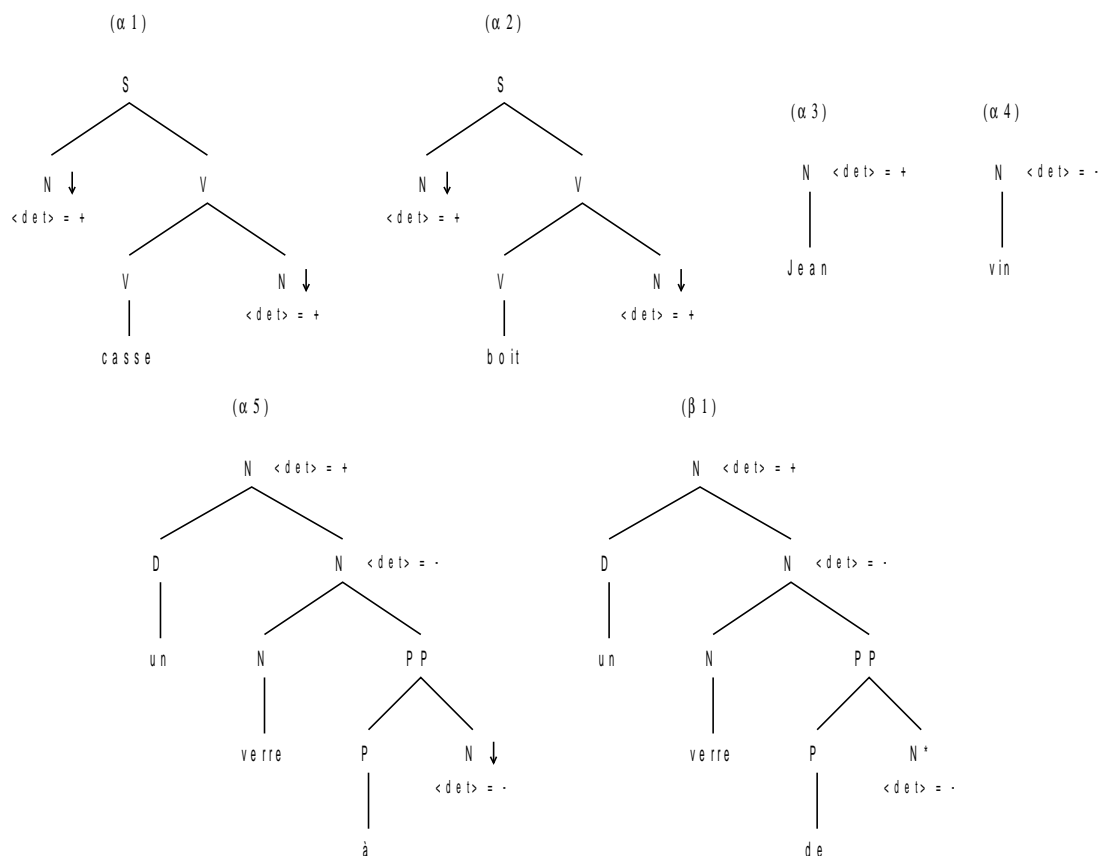


Figure 3. A set of elementary trees (or partly derived trees) for some phrases, among which (α5) "un verre à [vin]" (wine glass), and (β1) "un verre de [vin]" (glass of wine). The first one has the same distributional properties as "verre" (it may be used in sentences like "Jean casse un verre à vin" [Jean breaks a wine glass]); the second one as "wine" (it may be used in sentences like "Jean boit un verre de vin" [Jean drinks a glass of wine]). Inspired from Abeillé (1993: 209).

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. So for instance, the difference in the generation history of sentences involving the two contrasting example of 4.a and 4.b would show up in their derivation trees in Figure 4.a’ and 4.b’.

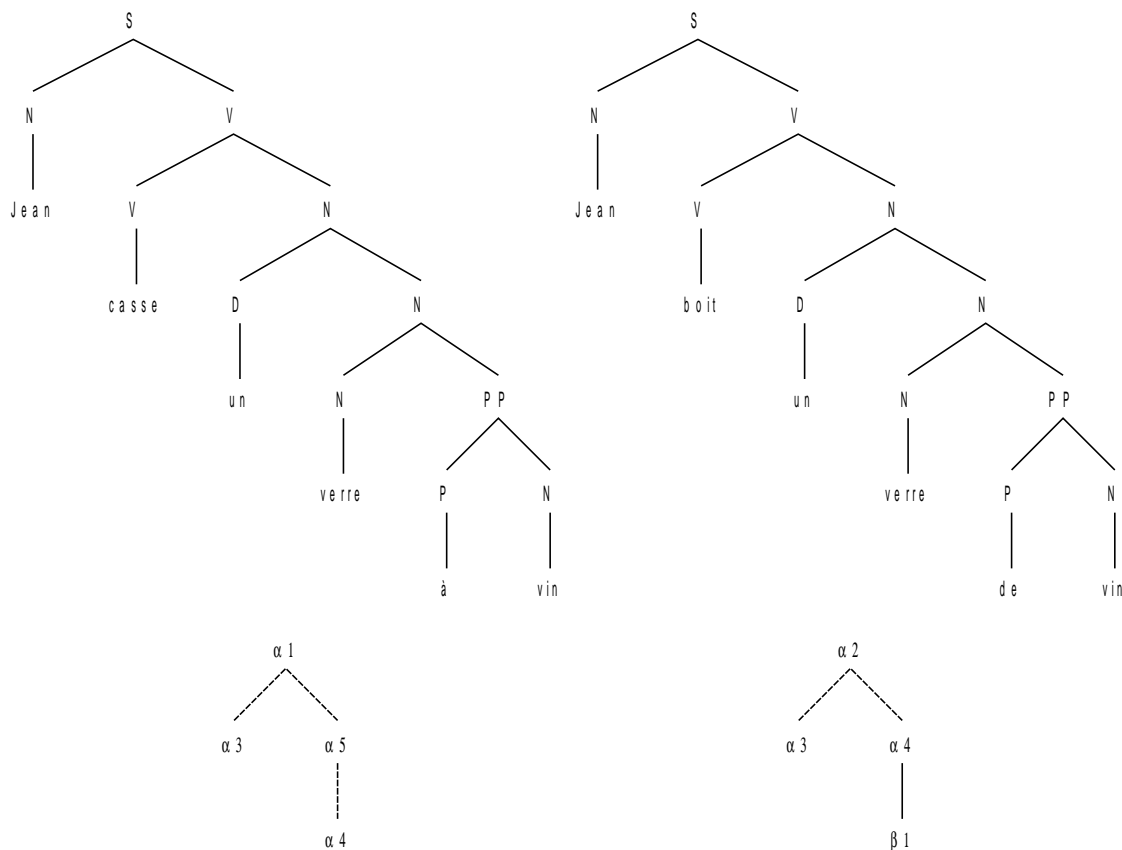


Figure 4. Derived trees for sentences (a) “Jean casse un verre à vin” [Jean breaks a wine glass] (up left), and (b) “Jean boit un verre de vin” [Jean drinks a glass of wine] (up right); the trees have a very similar static structure, but their difference is best explained by their dynamic derivation 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”.

Our 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 sentence that provides example 5 in Table 6: “Ni an certain nombre de faktè [...]” (28).

- (28) ni an serten nomb faktè (MC)
 un certain nombre de (French)
 ‘there are a certain number of factors’

The basic sentence pattern for “There are N” in MC is “Ni N” (no plural indefinite needed). The sentence observed in (28), 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 our analysis from the adjunction of a French multi-word set expression used as a quasi-quantifier (“un certain nombre de”) — very common in scientific or argumentative speech — in a MC sentence frame (Figure 5).

6. Underspecified Language Model

The model that best explains our data does not have to include strong hypotheses about there being 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 our corpus. It is a natural frame to understand our 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 .

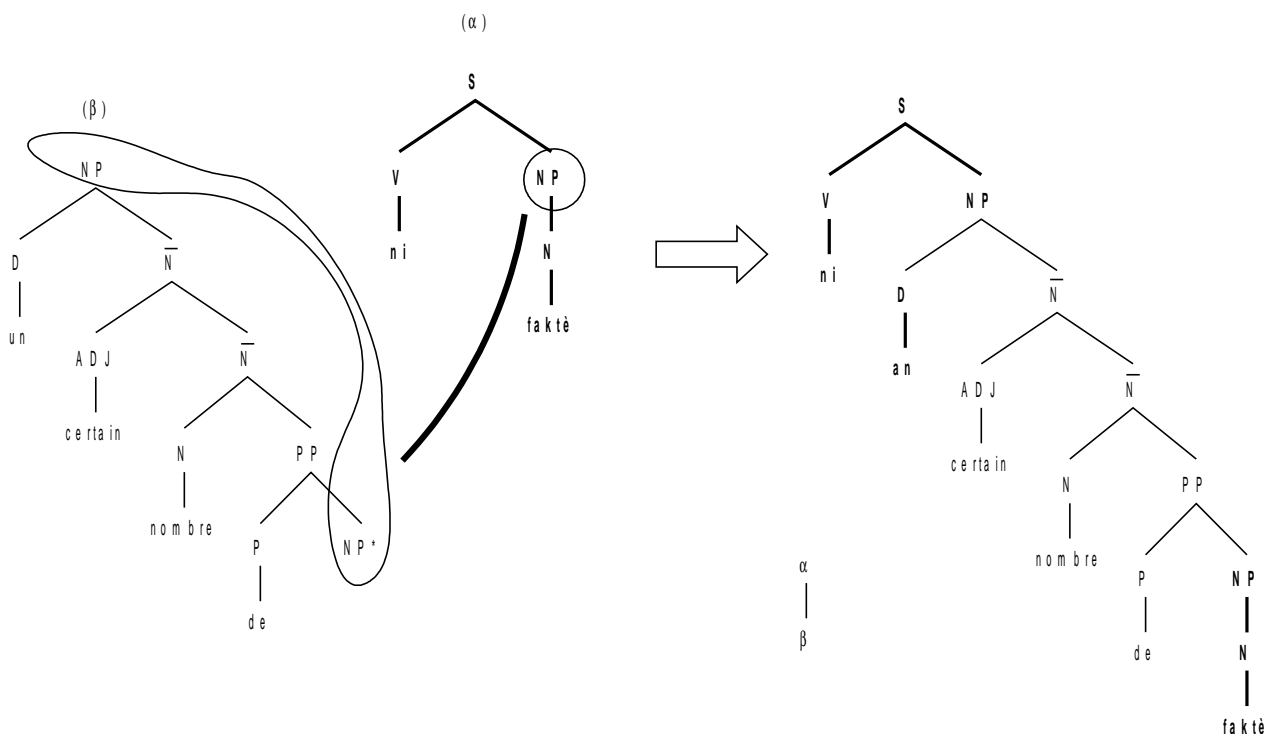


Figure 5. Adjunction of a French quasi-quantifier (β) “un certain nombre de ...” to an MC sentence (α) “ni faktè”. The derived tree (rightmost) appears to contain two switches, when in fact the derivation tree (bottom) is very simple.

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, it is necessary to include in our model the possibility for some 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 laa 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-Scotton, 1997, quoted by Chan, *ibid.*). 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 a strong argument for the MLF model. So, if a “null”-like 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 we observe (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-quantifier or quasi-classifiers of the “*a kind of ...*” sort (see Section 5.3 above).

3. In a context where two languages in contact have a high number of items in common (“items” being taken here to mean 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, or 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 our MC/French corpora. When surface forms are not outright identical in the received descriptions of both languages (like *espèce* (*espès*)), some surface forms like *an* [ã], *nomb* [nõb] or *sòt* [sõt] are similar in Creole and in some variants of spoken French. In the grey zone of congruent lexicalization, our 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 in some formal schools, in the last decades, to trim down syntactic models from unnecessary categories (Vijay-Shanker & Joshi, 1988; Abeillé, 1993), and opposite some evolutions of the generativist school, 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 very portable 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 corset, when incorporating feature structures in syntactic models allows to give satisfactory accounts of all the desired phenomena. Abeillé (1993) got rid of 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. I have argued elsewhere (Vaillant, 2014), that syntactic categories could even 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 developed here, but it certainly is stronger in code switching contexts — or even in “sloppy language” contexts — 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*). I have proposed earlier, on grounds of economy of modelling, to factor the common structures of groups of related dialects (Vaillant, 2008). At that time, I proposed to include a “language” feature to allow for practical restriction, among a common set of elementary structures, of the lexicon and grammar of a single, “pure”, language. 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 we have observed and described in this paper has the following properties:

1. It is based on unification of tree-like elementary structures, anchored either in the lexicon, or in a repository of pregnant constructions.
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 ($\alpha\beta$) form.

In other terms, the model we 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. It is not a concept that is in all cases relevant to describe the language of a bilingual speaker in contexts involving related 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. 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 a compulsory 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 very rare in natural communication contexts. So, it is interesting to see which structures are preferred in mixed discourse. In this article, we 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 [də] and plural definite [lə] articles frequently appear along with MC nouns. The analysis proposed here is that those articles perform quantifier functions that are not in the “base” system of (basilectal) MC.

In the case of [də], 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 [lə] 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 [lə] 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). We 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 is made 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 our 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 appear to witness 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, we 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 with the others as long as unification is not blocked by clashing feature structures.

The model we 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), we here 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 those alternative versions do not clash.

References

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

Alsagoff, L., Lick, H. C. (1998). The relative clause in colloquial Singapore English. *World Englishes*, vol. 17, no. 2, p. 127-138.

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

Banniard, M. (1997). *Du latin aux langues romanes*. Paris: Nathan.

Barbaud, Ph. (1998). Dissidence du français québécois et évolution dialectale. *Revue québécoise de linguistique*, vol. 26, n°2, p. 107-128.

Belazi, H. M., Rubin, E. J., Toribio, A. J. (1994). Code-switching and X-Bar theory: The Functional Head Constraint. *Linguistic Inquiry*, vol. 25, no. 2, p. 221-237.

Benoist, J. (1963). Les Martiniquais : Anthropologie d'une population métissée. *Bulletins et mémoires de la Société d'anthropologie de Paris*, XI^o Série, tome 4, fasc. 2, p. 241-432.

Bentahila, A., Davies, E.E. (1983). The syntax of Arabic-French code-switching. *Lingua*, vol. 59, p. 301-330.

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

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

Besters-Dilger, J., Demarkar, C., Pfänder, S., Rabus, A. (eds.) (2014). *Congruence in Contact-Induced Language Change: Language Families, Typological Resemblance, and Perceived Similarity*. *Linguae & Litterae* (27). Berlin: De Gruyter.

Bickerton, D. (1973). The Nature of a Creole continuum. *Language*, vol. 49, no. 3, p. 640-669.

Brunot, F., Bruneau, Ch. (1933). *Grammaire historique de la langue française*. Paris: Masson.

Cerquiglini, B. (2013). *La Naissance du français*, Que sais-je ? (2576). Paris: PUF (1st ed. 1991).

Chan, B. H.-S. (2009). Code-switching between typologically distinct languages. In Bullock, B. E., Toribio, A. J. (eds.), *The Cambridge Handbook of Linguistic Code-switching*, p. 182-198. Cambridge: Cambridge University Press.

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

Chaudenson, R. (2001). *Creolization of language and culture*. London: Routledge.

Clyne, M. (1987). Constraints on code switching: how universal are they? *Linguistics*, vol. 25, no. 4, p. 739-764.

Croft, W. (2002). *Typology and Universals*. Cambridge: Cambridge University Press (1st ed. 1990).

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

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

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

- Di Sciullo, A.-M., Muysken, P., Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, vol. 22, p. 1-24.
- Elisabeth, L. (2003). *La société martiniquaise aux XVII^e et XVIII^e siècles (1664-1789)*. Paris: Karthala.
- Ferguson, C. A. (1959). Diglossia. *Word*, vol. 15, p. 325-340.
- Fauquet, G. (1912). Note sur la population de la Martinique (Éléments ethniques et catégories sociales). *Bulletins et mémoires de la Société d'anthropologie de Paris*, VI^e Série, tome 3, fasc. 3-4, p. 154-161.
- Gary-Prieur, M.-N. (2011). *Les déterminants du français*. Paris: Ophrys.
- Greub, Y. (2007). Sur un mécanisme de la préstandardisation de la langue d'oïl. *Bulletin de la Société de Linguistique de Paris*, vol. 102, n^o 1, p. 429-434.
- Gross, G. (2012). *Manuel d'analyse linguistique*. Lille: Presses Universitaires du Septentrion.
- Guillaume, G. (1919). *Le problème de l'article et sa solution dans la langue française*. Paris: Hachette.
- Guillaume, G. (1973). *Principes de linguistique théorique. Recueil de textes inédits*. Valin, R. (ed.). Paris: Klincksieck.
- Gumperz, J., Wilson, R. (1971). Convergence and creolization: A Case from the Indo-Aryan/Dravidian border in India. In Dell Hymes (ed.), *Pidginization and Creolization of Languages*, p. 151-167. Cambridge: Cambridge University Press.
- Hazaël-Massieux, M.-Ch. (2008). *Textes anciens en créole français de la Caraïbe*. Paris: Publibook.
- Heath, J. (1978). *Linguistic diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies. Quoted in Heine & Kuteva (2005), p. 166-167.
- Heath, J. (1981). A case of intensive lexical diffusion. *Language*, vol. 57, no. 2, p. 335-367.
- Heine, B., Kuteva, T. (2002). *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Heine, B., Kuteva, T. (2005). *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
- Heine, B., Kuteva, T. (2006). *The changing languages of Europe*. Oxford: Oxford University Press.
- Heine, B., Kuteva, T. (2007). *The genesis of grammar*. Oxford: Oxford University Press.
- Hentschel, G. (2014). Belarusian and Russian in the Mixed Speech of Belarus. In Besters-Dilger, J. et al. (eds) (2014), p. 93-121.
- Hicks, C. (2012). A dual-structure analysis of morphosyntactic doubling in code switching. *Studies in the Linguistic Sciences: Illinois Working Papers*, vol. 37, p. 44-57.
- Höder, S. (2012). Multilingual constructions: a diasystematic approach to common structures. In Braunmüller, K., Gabriel, C. (eds.), *Multilingual Individuals and Multilingual Societies*, Hamburg Studies on Multilingualism (13), p. 241-258. Amsterdam: John Benjamins.
- von Heusinger, K. (1997). Saliency and Definiteness. *The Prague Bulletin of Mathematical Linguistics*, no. 67, p. 5-23.
- INSEE (2008). Enquête IVQ Martinique 2006-2007.
<URL:http://www.insee.fr/fr/insee_regions/martinique/themes/etudes_detaillees/IVQ/IVQ_2008_04_ma.pdf>
- Jackendoff, R. (1977). *X[̄] syntax: a study of phrase structure*. Cambridge, Massachusetts: MIT Press.
- Jake, J., Myers-Scotton, C., Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the Matrix Language. *Bilingualism: Language and Cognition*, vol. 5, no. 1, p. 69-91.
- Joshi, A.K., Schabes, Y. (1997). Tree-adjointing grammars. In Rozenberg, G. and Salomaa, A. (eds), *Handbook of formal languages, vol. 3: Beyond Words*, p. 69-123. New-York: Springer.
- Kamp, H., Van Genabith, J., Reyle, U. (2011). Discourse Representation Theory. In Gabbay, D. and Guenther, F., *Handbook of philosophical logic* (15), p. 125-394. Amsterdam: Springer.
- Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Link, G., Chierchia, G. (1995). Genericity: an introduction. In Carlson, G. N. and Pelletier, F. J. (eds), *The Generic Book*, p. 1-124. Chicago: The University of Chicago Press.
- Krifka, M. (2003). Bare NPs: Kind-referring, indefinites, both, or neither? In *Proceedings of SALT 13*, Washington, 9th-11th May 2003, p. 180-203.

- Lambrecht, K. (2001). A framework for the analysis of cleft constructions. *Linguistics*, vol. 39, no. 3, p. 463-516.
- Ledegen, G. (2012). Prédicats « flottants » entre le créole acrolectal et le français à la Réunion : Explorations d'une zone ambiguë. In Chamoreau, C., Goury, L. (eds.), *Changements linguistiques et langues en contact: Approches plurielles du domaine prédictif*, p. 251-270. Paris: CNRS Éditions.
- Lefebvre, C. (1974). Discreteness and the linguistic continuum in Martinique. *Anthropological Linguistics*, vol. 16, no. 2, p. 47-78.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, vol. 8, no. 1, p. 339-359.
- Löbner, S. (1985). Definites. *Journal of Semantics*, vol. 31, no. 4, p. 279-326.
- Lodge, A. R. (2004). *A Sociolinguistic history of Parisian French*. Cambridge: Cambridge University Press.
- MacSwan, J. (ed.) (1999). *A minimalist approach to intrasentential code switching*. New York: Garland.
- MacSwan, J. (2000). The architecture of the bilingual language faculty: evidence from intrasentential code switching. *Bilingualism: Language and Cognition*, vol. 3, no. 1, p. 37-54.
- MacSwan, J. (ed.) (2014). *Grammatical Theory and Bilingual Codeswitching*, Cambridge (Massachusetts): MIT press.
- MacSwan, J. (2014a). Programs and proposals in codeswitching research: Unconstraining theories of bilingual language mixing. In MacSwan (ed.) (2014), p. 1-33.
- Mahootian, S. (1993). *A null theory of codeswitching*. Ph.D. Dissertation. Evanston (Illinois): Northwestern University.
- Mahootian, S., Santorini, B. (1996). Code switching and the complement/adjunct distinction. *Linguistic Inquiry*, vol. 27, no. 3, p. 464-479.
- March, Ch. (1996). *Le discours des mères martiniquaises*. Paris: L'Harmattan.
- Migge, B., Légise, I. (2013). *Exploring Language in a Multilingual Context: Variation, Interaction and Ideology in Language Documentation*. Cambridge: Cambridge University Press.
- Milia-Marie-Luce, M. (2007). La grande migration des Antillais en France ou les années BUMIDOM. In Calmont, A., Audebert, C. (eds.), *Dynamiques migratoires de la Caraïbe*, p. 93-104. Paris: Karthala.
- Milroy, L., Muysken, P. (1995). *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. Cambridge: Cambridge University Press.
- Murray, D. A. B. (1997). The cultural citizen: negations of race and language in the making of Martiniquais. *Anthropological Quarterly*, vol. 70, no. 2, p. 79-90.
- Muysken, P., Smith, N. (eds.) (1995). *Pidgin and creoles: an introduction*, Creole Language Library (15). Amsterdam: John Benjamins.
- Muysken, P., Smith, N. (1995a). The study of pidgin and creole languages. In Muysken and Smith (eds.) (1995), p. 3-14.
- Muysken, P. (2000). *Bilingual Speech. A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, C. (1993). *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Oxford University Press.
- Myers-Scotton, C. (1995). A lexically based model of code-switching. In Milroy and Muysken (eds.) (1995), p. 233-255.
- Myers-Scotton, C., Jake, J. (1995). Matching lemmas in a bilingual language competence and production model: evidence from intrasentential code-switching. *Linguistics*, vol. 33, no. 5, p. 981-1024.
- Myers-Scotton, C., Jake, J. (2000a). Four types of morpheme: evidence from aphasia, code switching, and second-language acquisition. *Linguistics*, vol. 38, no. 6, p. 1053-1100.
- Myers-Scotton, C., Jake, J. (2000b). Testing the 4-M model: An introduction. *International Journal of Bilingualism*, vol. 4, no. 1, p. 1-8.
- Nábělková, M. (2014). The case of Czech-Slovak language contact and contact-induced phenomena. In Besters-Dilger, J. et al. (eds) (2014), p. 61-92.
- Nicolas, A. (1998). *Histoire de la Martinique : de 1939 à 1971*. Paris : L'Harmattan.
- Pinalie, P., Bernabé, J. (1999). *Grammaire du créole martiniquais*. Paris : L'Harmattan.

- Prudent, L.-F. (1981). Diglossie et interlecte. *Langages*, 15^{ème} année, n° 61, p. 13-38.
- Pulvar, O. (2005). Créole et scène publique aux Antilles-Guyane françaises. *Études Caribéennes* 2, p. 75-83.
- Russell, B. (1905). On Denoting. *Mind*, vol. 14, no. 56, p. 479-493.
- Sainton, J.-P., Château-Degat, R., Ho-Fong-Choy Choucoute L. (2004). Vers le deuxième âge colonial des Petites Antilles : le tournant sucrier et ses incidences (1640-1685). In Sainton, J.-P. (ed.), *Histoire et civilisation de la Caraïbe (Guadeloupe, Martinique, Petites Antilles), Tome I : Le temps des Genèses (des origines à 1685)*, p. 255-314. Paris: Maisonneuve & Larose.
- Sankoff, D., Poplack, S. (1981). A formal grammar for code-switching. *Papers in linguistics: International Journal of Human Communication*, vol. 14, no. 1, p. 3-45.
- Seuren, P. A. M., Wekker, H. (1986). Semantic transparency as a factor in Creole genesis. In Muysken, P., Smith, N. (eds.), *Substrata versus Universals in Creole Genesis*, p. 57-70. Amsterdam: John Benjamins.
- Stabler, E., MacSwan, J. (2014). A Minimalist parsing model for codeswitching. In MacSwan (ed.) (2014), p. 257-282.
- Vaillant, P. (2003). Une grammaire formelle du créole martiniquais pour la génération automatique. In *Actes de TALN 2003*, Batz-sur-mer, 10-14 juin 2003.
- Vaillant, P. (2008). A layered grammar model: using Tree Adjoining Grammars to build a common syntactic kernel for related dialects. In *Proceedings of TAG+9*, Tübingen, 7th-8th June 2008.
- Vaillant, P. (2014). La syntaxe, c'est de la sémantique. In Ablali, D., Badir, S., Ducard, D. (eds.), *Documents, Textes, Œuvres : Perspectives sémiotiques*, p. 297-320. Presses Universitaires de Rennes, 2014.
- Valdman, A. (1978). *Le créole. Structure, statut et origine*. Paris: Klincksieck.
- Vijay-Shanker, K., Joshi, A.K. (1988). Feature Structures Based Tree Adjoining Grammars. *Proceedings of COLING 1988* (12th International Conference on Computational Linguistics), Budapest, 22-27 August 1988, p. 714-719.
- von Wartburg, W. (1969). *Évolution et structure de la langue française*, Bibliotheca Romanica, series prima (I). Berne: A. Francke (1st ed. 1946).
- Werlen, I. (1988). Swiss German Dialects and Swiss Standard High German. In Auer, P., Di Luzio, A. (eds.), *Variation and Convergence: Studies in Social Dialectology*, p. 94-124. Berlin: De Gruyter.
- Wittmann, H. (1995). Grammaire comparée des variétés coloniales du français populaire de Paris du 17^e siècle et origines du français québécois. *Revue québécoise de linguistique théorique et appliquée*, vol. 12, p. 281-334.
- Wittmann, H. (1996). L'ouest français dans le français des Amériques: le jeu des isoglosses morphologiques et la genèse du dialecte acadien. In Cesbron, G. (ed.), *L'Ouest français et la francophonie nord-américaine*, Angers: Presses de l'Université d'Angers, p. 127-136.
- Zribi-Hertz, A., Jean-Louis, L. (2014). From Noun to Name: on Definiteness marking in modern Martinikè. In Cabredo Hofherr, P. and Zribi-Hertz, A. (eds.), *Crosslinguistic studies on Noun Phrase Structure and Reference*, Leiden/Boston: Brill, p. 268-315.