HAL
open science

Extended Projections in a Guadeloupean TAG Grammar

Emmanuel Schang

To cite this version:

Emmanuel Schang. Extended Projections in a Guadeloupean TAG Grammar. Workshop on High-level Methodologies for Grammar Engineering@ ESSLLI 2013, 2013, Düsseldorf, Germany. hal-01690399

HAL Id: hal-01690399

https://hal.science/hal-01690399v1

Submitted on 23 Jan 2018

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.

Extended Projections in a Guadeloupean TAG Grammar

Emmanuel Schang

Univ. Orléans, LLL UMR 7270, F-45065 Orléans, France

emmanuel.schang@univ-orleans.fr

1 Introduction

This paper presents a TAG (meta-)grammar of Guadeloupean Creole (GC). Guadeloupean is a French-based creole language spoken on the island of Guadeloupe (French West Indies). While sharing most of its lexicon with French, GC differs from French in its grammar (see Bernabé (1983) for a comprehensive description, and Damoiseau (2012) for a comparative approach). In particular, GC has preverbal Tense-Aspect Markers while French has inflectional suffixes for Tense. I will show that these TMAs can be correctly described as extended projections of verbs. They are generated as inflected forms of a verbal lexeme (i.e. in morphology rather than in syntax).

First, I present succinctly the key concepts of a TAG grammar to the reader. Next, section 2 describes the Tense-Aspect markers (TMAs) in GC and section 3 presents the way they are incorporated in the GC metagrammar [1]. Section 4 briefly presents how the same approach can be applied to the nominal domain.

Before addressing the question of Tense and Aspect markers, let me briefly present the key notions of TAG.

Tree-Adjoining Grammar (TAG) allows two operations (for a complete description, see Joshi and Schabes (1997), from which I take the following definitions):

- " *Substitution* takes only place on non-terminal nodes of the frontier of a tree. [...] By convention, the nodes on which substitution is allowed are marked by a down arrow (↓). When substitution occurs on a node n, the node is replaced by the tree to be substituted. When a node is marked for substitution, only trees derived from initial trees can be substituted for it." (Joshi and Schabes, 1997, p.4)
- " Adjoining[2] builds a new tree from an auxiliary tree β and a tree α (α is any tree, initial, auxiliary or derived). Let α be a tree containing a non-substitution node n labeled by X and let β be an auxiliary tree whose root node is also labeled by X. The resulting tree γ, obtained by adjoining β to α at node n is built as follow:

[1] I would like to thank Simon Petitjean, Yannick Parmentier and Denys Duchier for their precious help. I would also like to express my gratitude to the anonymous reviewers for their precious comments.

[2] I will prefer the term *adjunction* in the remainder of the paper.

- the sub-tree of α dominated by n, call it t, is excised, leaving a copy of n behind.
- the auxiliary tree β is attached at the copy of n and its root node is identified with the copy of n.
- the sub-tree t is attached to the foot node of β and the root node of t (i.e. n) is identified with the foot node of β. ”

(Joshi and Schabes, 1997, p.4)

As proposed in Frank (2002), Substitution and Adjunction are supposed to be universal operations. The differences between languages can only reside in the shape of the elementary trees. The way elementary trees are built is thus a crucial matter, and yet the locus of divergences between several TAG grammars. It is thus necessary to explain the principles that govern the building of the elementary trees.

The major reference for a TAG French grammar is Abeillé (2002)[3] which presents the following (linguistic) principles of elementary trees well-formedness:

Lexical Anchoring: An elementary tree must have (at least) one non-empty lexical head.

Predicate-Argument Co-occurrence: A predicate elementary tree must have a node for each of its arguments.

Semantic Anchoring: A syntactic elementary tree must correspond to a (non-empty) semantic element.

Compositionality Principle: An elementary tree corresponds to one and only one semantic unit.

Furthermore, I adopt the Conditions on Elementary Tree Minimality (CETM) (Frank, 2002, 54) :

CETM: The syntactic heads in an elementary tree and their projections must form an extended projection of a single lexical head.

This leads me to adapt Grimshaw (2000)'s definition of head and projection to the TAG framework and say that an elementary tree is a maximal projection of one (or several) lexical head(s) within which the categorial features are shared. This opens discussions about what is functional and what is lexical. In the case of prepositions, this is a difficult matter for which the debate remains open (see Cinque (2010)).

2 Tense-Aspect Markers

2.1 A Brief Description of Tense-Aspect Markers in GC

Creole languages are known to make use of independent[4] markers to express Tense and Aspect (see Winford (2012) for a synthesis) usually gathered under the label *TMA markers* (Tense, Mood and Aspect markers) or TAM (Tense and Aspect Markers). Guadeloupean behaves like other creoles w.r.t. TMAs. The inflectional morphology found in

[3] For English, see the XTAG project (XTAG Research Group (2001))

[4] In a sense that will be clarified later.

French verbs is replaced in GC by independent morphemes. For comprehensive analyses, the reader may refer to McCrindle (1999) and Pfänder (2000). In Table 1, I will use the description of Vaillant (2008a), taking as a paradigmatic example the verb *dansé* 'to dance'[5] [6].

Table 1. Tense and Aspect markers.

VALUE	FORM
Accomplished/Aoristic	dansé
Unaccomplished / Present	ka dansé
Frequentative	ka dansé
Progressive	ka dansé
Future	ké dansé
Unaccomplished Future (seldom)	ké ka dansé
Accomplished past (pluperfect)	té dansé
Unaccomplished past	té ka dansé
Irrealis (Past)	té ké dansé
Irrealis unaccomplished (extremely rare)	té ké ka dansé
Conditional / Optative	té dansé

Vaillant (2008b) also notes that *ké ka* and *té ké ka* are attested, although rare[7]. The precise semantic value and uses of these TMAs are beyond the scope of this paper and the reader may refer to Pfänder (2000) for a complete description. The following lines are just intended to present the useful key points for a non-specialist reader.

Bare Verbs: As in many languages, bare verbs in GC are used to express the (past) perfective (or preterite) with dynamic processes (as in (1-a)) and express the present tense with stative verbs[8], as in (1-b).

(1) a. Jan rivé.
Jean come
'Jean came.'
b. Jan enmé Sofi.
Jean love Sophie
'Jean loves Sophie.' (and not: * Jean loved Sophie)

Tense: The anterior marker of GC is *té*. When combined with non-stative verbs, *té* (ANT) provides a perfective interpretation:

[5] There is another marker *kay*, which is sometimes analysed as part of this system, however, as I will show later, it can't be included among the TMAs.

[6] This table shows the main uses of the TMA markers to give the reader a quick overview of the TMAs, but, as it will be claimed below, the interpretation of a TMA sequence is highly dependent on the context.

[7] I leave aside the marker *laka* which can plausibly be analysed as *la* (locative) and *ka*.

[8] And more generally with non-stative predicates, such as adjectival predicates.

(2) Sofi té palé ba Jan
Sophie ANT speak to Jean
'Sophie had spoken to Jean.'

And a past imperfective reading with stative verbs:

(3) Jan té enmé Sofi.
Jean ANT love Sophie
Litt.:'(At this time,) Jean was loving Sophie'

Aspect: GC Aspect markers are *ka* and *ké* for Imperfective and Prospective respectively. When *ka* is combined with stative verbs, the reading must be Iterative, as in:

(4) I ka tini onlo lajan a fen a chak mwa.
3sg ASP have a-lot-of money at end of each month
'He has a lot of money at the end of each month.' (example from (Delumeau, 2006, 117))

With non-stative predicates, the favored interpretation is Imperfective:

(5) Jan ka manjé.
Jean ASP eat
'Jean is eating'

ké triggers a future reading:

(6) Jan ké manjé.
Jean ASP eat
'Jean will eat'

2.2 Projecting TMAs

TMAs in GC have already been described in the TAG framework by Vaillant (2008a). This analysis differs from the one I present here on several aspects: First, it does not rely on a metagrammar, a point which matters since the implementation I present here relies heavily on the concept of metagrammar. Second, it calls upon adjunction to describe the TMA markers, an approach I will question here.

In this section, I will provide arguments in support of integrating TMA markers into the elementary trees as extended projection of a verbal head[9], instead of using adjunction as a way to integrate TMAs (as auxiliary trees) into the structure. Naturally, it is well known that every grammar using substitution can be rewritten using adjunction only. Formal considerations are therefore of little help in this discussion as far as the derived tree is concerned, but if derivation trees as intended to reflect the meaning of the sentence, the way the trees combine is a crucial matter. I base my argumentation on linguistic arguments and propose below a series of tests which show that TMAs behave differently from verbs.

[9] This can be extended to predicates in general, since there are non-verbal predicates in GC.

Several tests have been proposed for Romance languages (in Abeillé and Godard (2003) that cannot be used for GC. These are based on clitic-climbing and infinitival constructions which do not exist as such in GC and are therefore of little help. Instead, I will base my arguments on coordination and cleft structures.

Coordination: While coordination can apply to lexical items almost without restriction (see (Bernabé, 1983, 1396ff.) for a review of coordination in GC), TMA markers cannot be coordinated on the same position:

(7) *Jan ka é ké manjé.
Jean IMPERF and PROSP eat
'Jean is and will be eating'

But lexical items can be coordinated:

(8) sèvolan la ka monté, monté, (é) monté !
kite DEF IMPERF go-up go-up (and) go-up
'The kite goes up, up, up!' (from (Bernabé, 1983, 545))

(9) shows that the repetition of the predicate is the norm, since coordination of TMA is blocked, and (10) shows that verb coordination can appear below TMAs.

(9) an fouté y, an ka fouté y, an ké fouté y !
1sg win 3sg 1sg IMPERF win 3sg 1sg PROSP win 3sg
'I won (against him in the past), I won (this time) and I will win!'

(10) Jan ka dansé é chanté
Jean IMPERF dance and sing
'Jean is dancing and singing'

Predicate Cleft: Predicate cleft is a frequent construction in GC (while impossible in French). In (11), the lexical verb is clefted without the TMA marker. The same sentence with clefted TMA is ungrammatical, as in (12).

(11) sé monté nou ka monté pou nou rivé la nou ka alé.
it-is climb 1pl IMPERF climb for 1pl arrive there 1pl IMPERF go
'We are CLIMBING to arrive where we are going' (intensive meaning)

(12) a. *sé **ka** monté nou ka monté... [Aspectual marker prohibited]
b. *sé **té** monté nou té monté...[Tense marker prohibited]

This test highlights the difference between (still) real periphrastic elements and TMAs. For instance, it is usually accepted that the form *kay (ka+ay)* 'Imperfective+go' (example (13)) is one of the TMA (see Vaillant (2008a) and (Bernabé, 1983, 1035)).

(13) Jan kay vini
Jean IMPERF+go come
'Jean is going to come'

But *ay* 'go' differs from the TMAs w.r.t this test, as shown in (14):

(14) sé ay Jan kay vini
it-is go Jean IMPERF+go come
'Jean is GOING to come.' (stress on the movement)

Then, the verb *ay* 'to go' will be inserted using adjunction, as in Abeillé (2002) for aspectual verbs and auxiliaries in French. It anchors its own tree and is stored in the Lexicon, contrarily to the TMAs.

The negative marker is also excluded from the clefted position (15), lending weight to the hypothesis that the clefted element is below any verbal functional projections [$_{VP}$ V (NP)].

(15) Sé (*pa) monté, nou pa monté
it-is NEG climb 1pl IMPERF climb
'We DIDN'T climb up.'

A noteworthy exception seems to be (16):

(16) Sé vlé pa, i vlé pa !
It-is will NEG 3SG will NEG
'He really doesn't want !'

But the behavior of *vlé*, here, is atypical, since the negative marker follows the verb[10].

From these tests, I conclude that TMA markers are functional elements, just like inflectional affixes in French are. The only – but significant– difference is that these markers are not expressed via synthesis but via a particular form of periphrasis (a situation where two or more words express the meaning expected for a single word). Syntactic intermediate projections are available between Tense and Aspect and allow some (rare) adverbs to be inserted between the TMA markers, as in (17):

(17) Pyè té ja ka vin.
Pierre ANTERIOR already IMPERF come
'Pierre was already coming.' (from (Bernabé, 1983, 1060))

Thus, I treat TMA markers as functional elements that form an extended projection of the lexical head, in conformity with the CETM.

This work is in accordance with many recent works attempting to reconsider the traditional division between morphology and syntax. For instance, Brown et al. (2012) try to avoid the binary division of labor between morphology and syntax and claim that periphrasis are both. I will follow here some key assumptions presented in the Paradigm Function Morphology (PFM) framework (Stump (2001); Bonami and Stump (prep)) which are highly compatible with a lexicalist approach of syntax. This separates clearly the TAG grammar I propose here from the standard GB analyses (in the wake of Chomsky (1981)) which are non-lexicalist. In particular, I adopt the idea that "some syntactic constructions express the pairing of a lexeme with a morphosyntactic set, and should thus be incorporated into inflectional paradigms as inflectional periphrases" (Bonami and Stump, prep, p.20).

[10] It might illustrate a case of lexical integration of *pa* into the verbal stem.

In the cases of the GC's TMAs, it is reasonable to claim that the TMAs and the lexical verb realize different forms of the same lexeme. Bonami and Webelhuth (2013) note that most of the morphological descriptions of functional periphrasis are rudimentary in their description of the periphrasis syntactic structures and that their syntactic behavior differs from one language to another (see also Abeillé and Godard (2003). I suggest that TMAs are in part similar to proclitics, an analysis which has also been proposed in the HPSG framework by Henri and Kihm (2013), but still allow syntactic nodes for adjunction. The present study – as well as Henri and Kihm (2013) and Schang et al. (2012) for other creole languages – suggest a finer-grained characterization of the TMAs from a typological point of view.

3 TMAs in the TAG Metagrammar

3.1 eXtensible Meta-Grammar

XMG[11] is a declarative language for specifying tree-based grammars at a meta-level (Crabbé and Duchier (2004); Crabbé et al. (2013)). XMG allows the linguist to capture generalizations on his grammar by defining tree fragments (*Classes*) that can combine via feature unification (conjunctive / disjunctive combinations of fragments). Once defined, Classes can be reused in distinct contexts, allowing elegant generalizations. A *core* grammar is described by fragments of elementary trees and these fragments combine to form the *expanded* grammar which is made of elementary trees[12]. I will use these concepts to describe the TMA part of the *core* grammar of GC (Crabbé and Duchier, 2004, p.1).

3.2 Fragments of Functional Projections

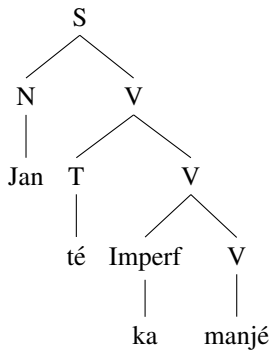
From the preceding section, I take for granted that TMAs are not stored in the Lexicon (they don't anchor any tree properly) but are co-anchors of the elementary tree associated with verbs[13].

For instance, (19) illustrates the structure of (18).

(18) Jan té ka manjé
Jean ANT IMPERF *eat*
'Jean was eating.'

[11] I am using XMG-2 version.
[12] See Crabbé (2005) for a large metagrammar of French using XMG.
[13] For a similar approach on Santomense's Tense and Aspect markers, see Schang et al. (2012) .

(19)



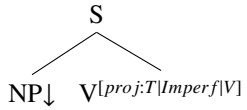
In (19), S is a projection of V, the maximal functional stretching of the verb.

In XMG's framework, the structure (19) is broken down into four pieces (*i.e.* classes) each containing minimal information. These Classes are listed below.

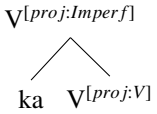
- *CanSubject*: for the External Argument of the verb. It is described in (20-a) .
- *Imperf*: as a projection of the Imperfective (Aspect) marker. It is described in (20-b).
- *Tensed*: as a projection of Tense.
- *Intransitive* verb: the minimal projection of V. It is described in (20-d).

(20)

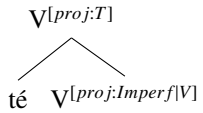
a.



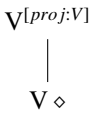
b.



c.



d.



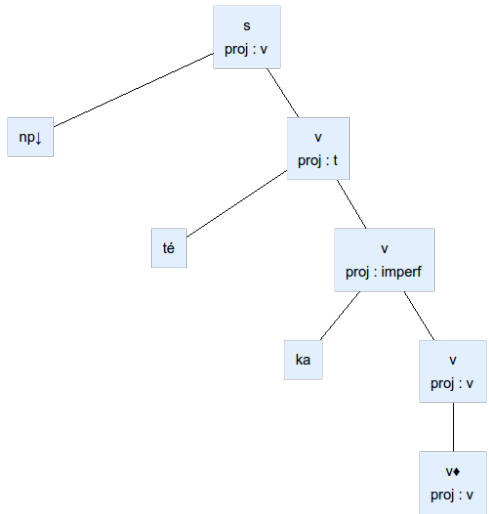
Thus, (19) is built up from the following conjunction of Classes:

$$CanSubject \wedge Intransitive \wedge Imperf \wedge Tensed$$

As in Schang et al. (2012) for Santomense, the feature *proj(ection)* is used to rule out invalid combinations in the output elementary tree.[14]

[14] In the values associated with feature proj, "|" refers to disjunction.

From the conjunction of classes given above, the result of the metagrammar compilation are elementary trees for intransitive verbs, inflected with Tense and Aspect, as shown in (21)[15].



(21)

The same mechanism is used to describe verb families. For instance, the class of Intransitive verbs inflected for TMAs can be described as a class *IntransV* gathering the classes of the inflected forms for intransitive verbs[16]:

$$BareV \mid ImperfV \mid ProspImperfV \mid TensProspImperfV$$

$$\mid TensedV \mid ProspV \mid TensImperfV \mid TensProspV$$

As expected from the morphological approach I defend here, the TMAs do not appear as adjuncts but are co-anchors of the verb. (22) shows the derivation tree[17] for the sentence *Jan té ka manjé* 'Jean was eating'.

(22) $\alpha 1$-manjé[Ant; $Imperf$]

$\alpha 2$-Jan

The benefit we have here with regards to adjunction is that the semantic interpretation of the sequence *N0 té ka V* is directly derivable from the features on the verb (see Table

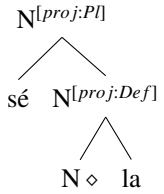
[15] Which is a picture of the elementary tree as displayed by XMG.
[16] The abbreviations are: Imperf(ective), Prosp(ective), Tens(ed).
[17] Where α is the conventional label for an elementary tree.

1). There is no need to postulate ambiguous TMA markers triggering various interpretations or different syntactic positions for the same marker, as is commonly proposed in creole languages descriptions (see Winford (2012)), i.e. postulating several distinct *ka* or Ø markers to derive the correct interpretation. On the contrary, it is compatible with the PFM idea that words (or multiword expressions) realize the morphosyntactic property set of the lexeme they are member of. Then, *té ka manjé* is the realization of the morphosyntactic set $\{Anterior, imperfective\}$ for the lexeme MANJÉ.

4 Extended Projections in the Nominal Domain

The same methodology developed for the TMAs is adopted in the Nominal domain, incorporating Definite, Demonstrative and Plural markers as co-anchors of a lexical Noun, as in (23), where the Plural marker *sé* precedes the N and the Definite marker follows the N, as illustrated in (24). Note that in GC, the Plural requires the Definite, **sé timoun* being agrammatical.

(23)



(24) [sé timoun la] vin
PL child DEF come
'The children came.'

(23) is the conjunction of the classes:

$$Plural \mid Definite \mid Noun$$

where *Plural* can only appear if *Definite* is present.

The genitive preposition *a*[18] 'of' in (25), which only has a functional role akin to a Genitive case marker is also treated as a co-anchor.

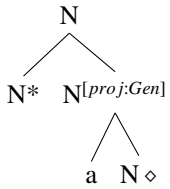
In (25), the NP [a Lelette] is adjoined to the head noun *kaz*.

(25) [kaz [a Lelette]]
house GENITIVE Lelette
'Lelette's house'

The genitival form of the noun (as head of the genitive NP) is an auxiliary tree (for adjunction) containing the co-achor *a*, as shown in (26).

[18] And its allomorph *an* when preceding a nasal.

(26)



The comparison of GC with a Matiniké Creole (spoken on the island of Martinique, closely related to GC and diachronically linked to GC) lends weight to this approach, since in Matiniké, the same NP as in (25) would be *kaz Lelette*, without preposition. The difference between these creoles relies only in the fact that Matiniké has a covert (silent) preposition for the same structure as (26).

To summarize, the inflected forms of a noun can be represented as a conjunction of classes:

$$Demonstrative - N \mid Plural - N \mid Definite - N \mid Bare - N \mid Genitive - N$$

Note that while Plural, Demonstrative, Definite and Genitive appear as syntactic projections in GC, the same functional elements are concatenated at the word level in other languages.

5 Conclusion

I have shown that Tense and Aspect markers of Guadeloupean Creole are functional projections that can be described as co-anchors of a lexical head (extended projections). Their response to the tests of coordination and cleft structures shows that TMAs are not as free as other periphrastic elements (such as the verb *ay* 'to go'). Their place is somewhere between concatenative morphology at the word level and 'free' periphrastic elements, which I take to be adjoined elements. In the nominal domain, I have suggested that the definite, demonstrative, plural and genitive markers are also functional elements of the same sort. These functional elements are integrated in the elementary trees that form the GC grammar as co-anchors of the lexical item. I have demonstrated how these elements form fragments of (elementary) trees and how they combine to form the expanded grammar. To do this, I have used the XMG formalism for metagrammar. Since the combination of the fragments constitute inflected forms of a lexeme (as the compound tenses are still members of the verbal paradigm, see Ackerman and Stump (2004)), the building of Elementary trees is as much a morphological as a syntactic operation. It thus casts a new light on Creole languages which are commonly thought to have little (or even no) morphology.

Bibliography

Abeillé, A. (2002). *Une Grammaire électronique du Français*. CNRS Editions, Paris.

Abeillé, A. and Godard, D. (2003). Les prédicats complexes dans les langues romanes. *Les langues romanes. Paris: CNRS Editions*, pages 125–184.

Ackerman, F. and Stump, G. (2004). Paradigms and periphrastic expression: a study in realization-based lexicalism. *Projecting Morphology. CSLI.*, pages 111–58.

Bernabé, J. (1983). *Fondal-natal*. l'Harmattan Paris.

Bonami, O. and Stump, G. (in prep). Paradigm Function Morphology. In Spencer, A., editor, *The Handbook of Morphology, 2nd ed.* Wiley-Blackwell.

Bonami, O. and Webelhuth, G. (2013). The phrase-structural diversity of periphrasis: a lexicalist account. In Chumakina, M. C. G. G., editor, *Periphrasis: The role of syntax and morphology in paradigms*. Oxford University Press.

Brown, D., Chumakina, M., Corbett, G., Popova, G., and Spencer, A. (2012). Defining 'periphrasis': key notions. *Morphology*, 22(2):233–275.

Chomsky, N. (1981). Lectures on the Theory of Government and Binding. *Dordrecht: Foris*.

Cinque, G. (2010). The syntax of adjectives: a comparative study.

Crabbé, B. (2005). *Représentation informatique de grammaires d'arbres fortement lexicalisées : le cas de la grammaire d'arbres adjoints*. PhD thesis, Université Nancy 2.

Crabbé, B. and Duchier, D. (2004). Metagrammar Redux. In *International Workshop on Constraint Solving and Language Processing*, Copenhagen.

Crabbé, B., Duchier, D., Gardent, C., Le Roux, J., and Parmentier, Y. (2013). XMG : eXtensible MetaGrammar. *Computational Linguistics*, 39(3):1–66.

Damoiseau, R. (2012). *Syntaxe créole comparée*. Karthala et cndp-crdp edition.

Delumeau, F. (2006). *Une description linguistique du créole guadeloupéen dans la perspective de la génération automatique d'énoncés*. PhD thesis, Université Paris X.

Frank, R. (2002). *Phrase Structure Composition and Syntactic Dependencies*. MIT Press, Cambridge, Mass.

Grimshaw, J. (2000). Locality and extended projection. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, pages 115–134.

Henri, F. and Kihm, A. (2013). The Morphology of TMA in Mauritian and Creole.

Joshi, A. K. and Schabes, Y. (1997). Tree-Adjoining Grammars. In Rozenberg, G. and Salomaa, A., editors, *Handbook of Formal Languages*, volume 3, pages 69–124. Springer, Berlin, New York.

McCrindle, K. L. (1999). Temps, mode et aspect, les creoles des Caraibes a base lexicale française.

Pfänder, S. (2000). *Aspekt und Tempus im Frankokreol*. G. Narr.

Schang, E., Duchier, D., Ekoukou, B. M., Parmentier, Y., and Petitjean, S. (2012). Describing São Tomense Using a Tree-Adjoining Meta-Grammar. In *11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+ 11)*.

Stump, G. (2001). *Inflectional morphology: a theory of paradigm structure*, volume 93. Cambridge Univ Pr.

Vaillant, P. (2008a). A layered grammar model: Using tree-adjoining grammars to build a common syntactic kernel for related dialects. *arXiv preprint arXiv:0810.1207*.

Vaillant, P. (2008b). Une grammaire formelle du créole martiniquais pour la génération automatique. *arXiv preprint arXiv:0810.1199*.

Winford, D. (2012). Creole Languages. *The Oxford Handbook of Tense and Aspect*.

XTAG Research Group (2001). A Lexicalized Tree Adjoining Grammar for English. Technical report, Institute for Research in Cognitive Science, Philadelphia. Available from `ftp://ftp.cis.upenn.edu/pub/xtag/release-2.24.2001/tech-report.pdf`.