A Metagrammatical Approach to Periphrasis in Gwadloupéyen
Emmanuel Schang

To cite this version:

HAL Id: halshs-01876042
https://halshs.archives-ouvertes.fr/halshs-01876042
Submitted on 18 Sep 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.
A Metagrammatical Approach to Periphrasis in Gwadloupéyen

Emmanuel Schang
LLL UMR 7270, Université d’Orléans & CNRS
(<emmanuel.schang@univ-orleans.fr>)

Abstract:
In this paper, I show that verbal and nominal functional elements of Gwadloupéyen can be described in the Tree-Adjoining Grammar as pertaining to morphological periphrasis. This challenges the claim that Creoles have fully analytical morphology.

Keywords: Guadeloupean Creole, periphrasis, Tree-Adjoining Grammar

1. Introduction

Creole languages have been claimed to be simple languages on morphological arguments (see Seuren and Wekker 1986; McWhorter 2001, a.o.). The fact that they present analytic constructions (instead of synthetic morphology) is taken as an argument for their simplicity.

Analytic constructions are supposed to be transparent and syntactically driven. In this paper, I will show that the Tense and Aspect markers (TMAs) in Gwadloupéyen, but also some elements of the nominal domain can be described in the Tree-Adjoining Grammar as pertaining to morphological periphrasis.

In Section 1.1, I give a brief overview of Gwadloupéyen. Section 2 describes the framework I will use to organize the data. The concept of metagrammar will be explained in Section 2.2. The source code of the metagrammar developed in this paper is available on GitHub and can be freely uploaded and tested. In Section 3, I present an analysis of the TMA markers as periphrasis. Section 3 extends this analysis to the nominal domain.
1.1 A quick presentation of Gwadloupéyen

Guadeloupean Creole (or Gwadloupéyen) is spoken by approximately 850,000 speakers both on the Island of Guadeloupe (and its dependencies) and in ‘mainland’ France. The level of proficiency in Creole varies highly between speakers. The degree of exposure to French (the official language) differs according to the individual (Jno-Baptiste 2015). As explained in Jeannot-Fourcaud and Jno-Baptiste (2008), the first language of many Guadeloupean children is not exclusively Creole, and before any schooling, they are educated in French and Creole in variable proportions.

This leads to difficulties to define what is Creole and what is not. In this work, I will use as reference grammaticality judgments from Creole speakers and examples taken from a spoken corpus of Guadeloupean (Glaude 2013) available online. My informants are students in Linguistics and persons of various ages met during fieldworks. They all are native speakers of Gwadloupéyen.

2. Building a TAG Grammar of Gwadloupéyen

2.1 Tree-Adjoining Grammar

Tree-Adjoining Grammar is a grammar formalism developed in the mid-70s (Joshi and Schabes 1997; Joshi 2012). As its name clearly indicates, it is a formal tree rewriting system, with a domain of locality and a tree depth different from Context-Free Grammars. As an example, the sentence S “John loves peanuts” combines three Elementary Trees (α1 John, α2 likes and α3 peanuts) together to form a Derived Tree γ in Fig. 2 (proving that S can be generated by the grammar). The operation that combines the Elementary Trees at ↓ nodes in Fig. 1 is called substitution.

![Figure 1. Substitution in TAG](image-url)
Note that each tree in Fig.1 has a **lexical anchor** (lexical item). A TAG grammar in which trees are obligatorily anchored by (at least) one lexical item is a **Lexicalized Tree-Adjoining Grammar** (LTAG). In this paper, I will refer interchangeably to LTAG and TAG.

The second operation available in TAG is **adjoining**, which involves inserting a tree into another (Fig.3). An **auxiliary tree** $\beta$ has a special node (a foot node marked *).

**Figure 2. Derived Tree in TAG**

**Figure 3. Adjoining in TAG**

**Figure 4. Derived Tree after Adjoining**
Abeillé (2002) proposes several linguistic principles to build a correct LTAG grammar (tree well-formedness). These are:

- 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.

I adopt these principles here and, as explained in Schang (2013), in accordance with the Compositionality Principles functional items are considered as co-head of a lexical item.

2.2 MetaGrammar with XMG-2

The concept of metagrammar has been implemented initially in Candito (1999) to describe a TAG grammar of verbs in Italian and French. This description was based on a three-dimensional view of language which combines a) the subcategorization frames of verbs, b) the transformations (functional rearrangements between the initial frames and the morphologically derived forms, e.g. active/passive transformation) and c) the syntactic surface realizations (included word-order variation).1

Later, Crabbé (2005) proposed a more flexible implementation of the metagrammar for French (named XMG) and Petitjean (2014) developed XMG-2, a modular metagrammar compiler which allows for the description of various linguistic phenomenon (see Duchier et al. 2017, 2014, for instance). XMG2 proposes a set of languages of description which includes:

- a language of description for feature structures,
- a language of description of syntactic trees,
- a language for flat semantics, see Bos (1996),
- a language for frame semantics, see Lichte and Petitjean (2015).

This development of a modular metagrammar for morphology opened the door to investigations in computational morphology and syntax (Duchier et al. 2012a; Schang et al. 2012; Duchier et al. 2017) which rely on XMG2 to model some grammatical phenomena in different ‘little-studied’ languages, such as Santomense and Ikota.

1 See Abeillé (2002: chap. 7).
2.2.1 A Metagrammar of Trees

For French, a TAG grammar must have to express the link between two constructions of the verb *manger* ‘to eat’:
- the sentence *Jean mange* ‘Jean eats’,
- the NP *L’homme qui mange* ‘the man who eats’.

That is, it has to make an explicit link between *manger* with its canonical subject and *manger* with a relative subject. Both are part of the syntactic combinations allowed with *manger*.

This can be expressed in the TAG framework as two elementary trees, as in Fig. 5 and 6.

![Figure 5. Elementary Tree of 'NP mange'](image)

![Figure 6. Elementary Tree of 'NP qui mange'](image)
The assumption behind XMG2 is that these trees and their relation can be described as a set of block (called classes) that combine using a disjunctive or conjunctive composition.

That is Fig. 5 is obtained via the composition of the CanonicalSubject class and the Intransitive class (conjunction at the node VP). The boxed node represents the node where the fragments are glued together.

This conjunction can be expressed as:

\{ \text{CanonicalSubject} \land \text{Intransitive} \}

In contrast, the RelSubject class expresses the part of the tree describing a relativized subject argument.
As a result, one can define a class Subject that combines in a disjunction the tree fragments (classes) RelSubject and CanonicalSubject:  
\[
\text{Subject} = \{ \text{CanonicalSubject} \mid \text{RelSubject} \}
\]

2.2.2 A Metagrammar for Morphology

Petitjean, Samih and Lichte (2015) have used XMG2 for their morphological description of verbs in Arabic, (Magnana Ekoukou 2015; Duchier et al. 2012b) presented an analysis of Ikota’s verbs as a set of position classes and Duchier et al. (2014) described nominal morphology of Somali.

As for Ikota, the verbal morphology was described in Duchier et al. (2012b) as a conjunction of classes, as formulated in (1). A verb is composed of six classes which are linearly ordered around the Verbal Root (VR).

\[(1) \text{Verb} \land \text{Subj} \land \text{Tense} \land \text{VR} \land \text{Aspect} \land \text{Active} \land \text{Proximal} \]

The composition of verbs in Ikota is similar to the composition of trees in French (example above) in the fact that it is a composition of fragments. What differs is the level (the domain) of application, i.e. word vs. elementary trees.

2.2.3 Periphrasis: A Challenge for Lexicalist Grammars

As already said above, morphology and syntax form two distinct levels in Lexicalist Grammars. 3 This question is still a matter of debate among linguists (Borer 1998; Sproat 1998).

In a lexicalist framework such as TAG where the lexicon is inserted at a particular leaf node (called the anchor) in Elementary Trees, this question is clearly set. But this may appear as a downside when it comes to investigate the properties of TMAs in Creole: if syntax cannot interfere with the properties of words (syntax can only read features provided by words), is it possible to account for periphrastic elements such as the TMAs?

3. TMAs as Periphrastic Expressions

3.1 TMAs and periphrasis

Gwadloupéyen’s verbal inflection is, at least at first look, strongly different from French, its lexifier (superstrate language). Whereas French has a synthetic morphology (2a), Creole (2b) has Tense and Aspect preverbal markers.

---

2 There is no room in this paper to present the description language (code) of XMG2 in detail; I let the reader look at <http://dokufarm.phil.hhu.de/xmg/doku.php?id=start> for more details.

3 In XMG2’s terminology, these constitute distinct dimensions.
At first look, this difference seems dramatic as it sets the two languages in two different typological groups, Creole grammars being crucially different from the grammar of their lexifiers.

However, Chaudenson (2004) has shown that, at the time of the creolization period, French also had periphrastic constructions which were in competition with synthetic forms. And it is still the case. Indeed, many, if not all, French speakers will use the periphrastic future \textit{il va mourir} ‘he will die’, \textit{je vais coudre} ‘I will sew’ instead of the rarely used (if even known) synthetic future of \textit{mourir} and \textit{coudre}.\footnote{Which is \textit{je mourrai} and \textit{je coudrai}.} Moreover, as Abouda and Skrovec (2015) have shown, the use of the periphrastic future tends to surpass the use of the synthetic form in spoken French.

Regarding the etymology of the TMAs, Degraff (2005: 320) clearly explains that “all the preverbal TMA morphemes in Haitian Creole, […] can be straightforwardly traced back to 17th-18th century Fr cognates, some of which still exist in certain contemporary French dialects, including sometimes the ‘standard’ dialect”.

This does not entail that the conjugation of French and Creole are similar (which is clearly not the case), but it questions the deepness of the gap between the two languages.

However, I would like to question the claim that Creole does not show inflectional morphology whereas French does. This leads to discuss the syntactic status of periphrasis.

As it is widely known, Latin shows good examples of periphrastic forms competing with synthetic forms. For instance, an ordinary Latin verb expresses perfect with a synthetic form whereas for passive and deponent verbs, a periphrastic form is mandatory. As clearly stated in Matthews (1991: 219-220):

\begin{itemize}
  \item In Latin, schoolboys learned \textit{amo} ‘I love’ as Present Active, \textit{amor} ‘I am loved’ as Present Passive, \textit{amavi} ‘I loved’ as Perfect Active, but then \textit{amatus sum} (a form consisting of a Masculine Nominative Singular Participle, \textit{amatus}, and the form for ‘I am’, \textit{sum}) as the Perfect Passive. The last is clearly two words, which obey separate syntactic rules (for example, of agreement). Nevertheless, they are taken together as a term in what are otherwise morphological oppositions.
\end{itemize}
Ackerman, Stump, and Webelhuth (2011) argue that “periphrasis (multi-word expression) is as much a mode of morphological realization as synthesis is”. Bonami (2015) proposes that “periphrases are similar to syntactically flexible idioms; the theory of periphrasis is thus embedded within a more general theory of collocation”.

However, Blevins (2008) provides arguments for periphrasis as a syntactic exponent and challenges the fact that periphrasis should be considered as inherently morphological. Yet, he discusses the ‘bottom-up’ approach of syntactic periphrasis where the meaning of the periphrasis is deduced from the meaning of its parts.

The discussions about the morphological or syntactic nature of periphrasis and its typological implications (see for instance Ackerman and Stump 2004; Brown et al. 2012; Bonami and Webelhuth 2013) are too complex to be developed any further here. The challenges here is to adequately describe Guadeloupean Creole in the TAG framework.

The approach I will defend here is based on the idea that the sequence of TMA + V is similar to multi-word expressions. The meaning of a multi-word expression cannot be reduced to the meaning of its parts.

### 3.2 TMAs in Gwadloupéyen

The TMA markers and their uses are described exhaustively in Bernabé (1983), Pfänder (2000), McCrindle (1999) among others.

Table 1 provides a quick overview of the main uses of the TMAs (as proposed in Vaillant 2008).

<table>
<thead>
<tr>
<th>Value</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accomplished /Aoristic</td>
<td>dansé</td>
</tr>
<tr>
<td>Unaccomplished / Present</td>
<td>ka dansé</td>
</tr>
<tr>
<td>Frequentative</td>
<td>ka dansé</td>
</tr>
<tr>
<td>Progressive</td>
<td>ka dansé</td>
</tr>
<tr>
<td>Future</td>
<td>ké dansé</td>
</tr>
<tr>
<td>Unaccomplished Future (seldom)</td>
<td>ké ka dansé</td>
</tr>
<tr>
<td>Accomplished past (pluperfect)</td>
<td>té dansé</td>
</tr>
<tr>
<td>Unaccomplished past</td>
<td>té ka dansé</td>
</tr>
<tr>
<td>Irrealis (Past)</td>
<td>té ké dansé</td>
</tr>
<tr>
<td>Irrealis unaccomplished (extremely rare)</td>
<td>té ké ka dansé</td>
</tr>
<tr>
<td>Conditional / Optative</td>
<td>té dansé</td>
</tr>
</tbody>
</table>

Table 1. TMAs values, adapted from Vaillant (2008)

The problem faced by bottom-up approaches, where the TMAs combine in syntax, is the fact that the meaning of the sequence TMA + V is dependent on the aspectual class of the V (or more generally, on the aspectual class
of the predicate since Creole have nominal and prepositional predicates). This phenomenon has already been described in the literature on Antillean Creoles: Pfänder (2000); McCrindle (1999); Bernabé (1983) and Damoiseau (2012) among others.

For instance, with a stative predicate such as *be at school*, *ka* as in (3) can only be interpreted with an iterative meaning and not as an ongoing event (progressive).

(3) Jan ka lékol
    Jean ipfv school

    ‘Jean is at school (Context: every time I come to see him…)

Let us consider another example. The anterior marker of Gwadloupéyan is *té*. When combined with non-stative verbs, *té* provides a perfective interpretation:

(4) Sofi té palé ba Jan
    Sophie ant speak to Jean

    ‘Sophie had spoken to Jean’

and a past imperfective reading with stative verbs:

(5) Jan té enmé Sofi
    Jean pst love Sophie

    Litt.: ‘(At this time) Jean was loving Sophie’

As a consequence, it has been proposed that there are homonymous markers (e.g. Bernabé 1983 proposes different *ka* morphemes) and many zero positions to account for the differences of interpretation. Since the interpretation of a marker depends on its position on a syntactic node, the more different interpretation we have, the more nodes we need.

While this is descriptively correct, it is problematic for computational models. The cost of having to choose between two (or more) homonymous markers and/or zero positions (hence empty markers) is computationally expensive.5

In the TAG framework, Vaillant (2008) proposes an analysis based on the adjoining of the TMAs as auxiliaries to the main verb. As illustrated in Fig. 10,6 the TMAs anchor their own tree.

---

5 Clearly, one is not obliged to force a theoretical (or descriptive) model to be computationally efficient but this is obviously the choice made here in adopting the TAG framework.

6 For the sake of simplicity, I do not note here the features of the trees which reduce the combinations. The reader is asked to refer to Vaillant (2008) for the complete details.
Figure 10. TMAs in Vaillant (2008a)

Technically, this solution works as it correctly reject undue combination. However, as mentioned in Schang (2013), this violates the semantic well-formedness of the Elementary Trees since the TMAs cannot be interpreted alone.

Let me present briefly the arguments for and against a purely syntactic or morphological analysis of the TMAs.

3.2.1 TMAs as syntactic elements

The TMA markers can combine with adverbs, as shown in (6). This clearly indicates that syntactic nodes are needed to insert the adverbs such as ja ‘already’. Thus, TMAs cannot be considered as clitics contrary to what has been proposed for other creole languages (see Henri and Kihm 2015). 7

(6) Pyè té ja ka vin
Pierre past already ipfv come
‘Pierre was already coming’

3.2.2 TMAs as morphological elements

However TMAs don’t have the freedom expected from purely syntactic elements. They cannot be coordinated (7a) while verbs can; unlike standard verbs they cannot be clefted (predicate cleft) (7b) and they can fuse with other functional elements (such as the negative marker) in certain configurations (7c).

7 I review here briefly the arguments presented in Schang (2013).
(7) a. Jan  ka  (*é ké)  manjé
   Jean  ipfv  and prosp  eat
   ‘Jean is (and will) be eating’

b. *sé  ka  manjé  Jan  ka  manjé.
   it.is  ka  eat  Jean  ipfv  eat
   Intended: ‘Jean is eating’

c. Jan  péké  manjé
   Jean  neg.prosp  eat
   ‘Jean will not eat’ (expected: pa ké)

3.3 TMAs in the metagrammar

I will present briefly here how the metagrammar offers an elegant way to reconcile the relative freedom of the combination of the TMAs with other elements with the fact that the TMAs are not autonomous elements.

As shown in Schang (2013), TMAs can be considered as co-head (co-anchor) of a verb. While Tense inflectional elements combine with the verbal root in French at word level (manger-ai ‘eat-fut.1sg’), the TMAs combine at a syntactic level in Gwadloupéyen and provide syntactic nodes for adjoining.

Fig.11 presents the Elementary Trees corresponding to manjé in (8).

(8) Jean  té  ka  manjé
    Jean  pst  ipfv  eat
    ‘Jean was eating’

This tree can be divided in the metagrammar into different fragments (Fig. 12) that combine to form the elementary trees.
We find in Fig. 12 the fragments that were presented earlier (CanonicalSubject in a., the Intransitive class in d.) and the fragments corresponding to \( \text{té} \) and \( \text{ka} \) (with a feature ‘proj’ (projection) which restrains the combinations). The various inflected forms of a verb (or other predicates) in Gwадloupéyen can be derived by combining the following fragments:

\[
(9) \quad \{ \{ \text{Prospective (ké) | None } \}; \\
\{ \text{Imperfective (ka) | None } \}; \\
\{ \text{Anterior (té) | None } \}; \\
\text{V} \}
\]

As a result, the process of incorporating the TMAs as extended projections of the verb elementary trees is not different from the process of assembling a verb with its arguments requirements (a leaf for every argument). As such, the sequences TMA + V constitute inflectional forms of a verbal lexeme. It is then a morphological process. This process is similar the generation of inflected verbs in Ikota (see Section 2.2.2).

There are also felicitous side effects of treating TMAs as co-anchors. First, just as for multi-word expressions, the meaning of the sequence is the meaning of the entire sequence (as in \textit{to kick the bucket} ‘to die’). The individual fragments of the tree are not the adequate level for interpretation.

Second, if one wishes to compare the form \textit{mangeait} ‘was eating’ in French with its corresponding form in Gwадloupéyen (see examples in (2)),
the derived tree (in (10)) is not the appropriate level Fig.13; however, the derivation trees in both languages are similar Fig. 14

![Diagram of derived trees for Creole (left) and French (right).](image)

Figure 13. Derived Trees for Creole (left) and French (right)

\[
\alpha_1 \text{ manjé  \{past;ipfv\}} \quad \alpha_1 \text{ mangeait  \{past;ipfv\}}
\]

\[
\alpha_2 \text{ Jan} \quad \alpha_2 \text{ Jean}
\]

Figure 14. Derivation Trees for Creole (left) and French (right)

To sum up, TMAs are better analyzed as co-anchors of elementary trees in Gwadloupéyen. This morphological process operates at the level of an Elementary Tree (i.e. the projection of a head) whereas synthetic morphology operates at the word level.

4. **Periphrasis beyond the Verbal Domain**

As it has already been explained by others (see Bonami 2015 for a review and a discussion), periphrasis can be found in the inflection of all major categories. The articles (definite and demonstrative) can be considered as co-anchors of the Noun (see Schang in preparation) for a complete development).

In some languages, such as Albanian for instance, the definite marker is an affix. It seems that it is never the case in Creole languages (Velupillai 2015). In Gwadloupéyen, the definite article is not an affix but a marker placed on the left margin of the NP. I consider it as a functional projection (Fig. 15), i.e. as a co-anchor of the head noun (symbolized here as a diamond). It correctly predicts that it can only occur once in a particular nominal domain.

---

8 The reader can already see the implementation of the articles in the metagrammar here: <https://github.com/eschang/xmg_GC_metagrammar>. 
Again, a multi-word expression is used here to mark inflection. The same applies to the plural marker and to the demonstrative.

But one can also consider the variation in the possessive form $N + (\text{Prep}) + \text{possessive pronoun}$ as a form of morphological variation. Indeed, the presence of the preposition, as shown in (10) and Table 2, depends on the head noun.

(10) a. vwati an mwen
car of me

‘my car’

b. manman mwen
mother me

‘my mother’

<table>
<thead>
<tr>
<th></th>
<th>manman/papa ‘mum/dad’</th>
<th>vwati/biten ‘car/thing’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1sg</td>
<td>mwen</td>
<td>an mwen</td>
</tr>
<tr>
<td>2sg</td>
<td>-w</td>
<td>a-w</td>
</tr>
<tr>
<td>3sg</td>
<td>-y</td>
<td>a-y</td>
</tr>
<tr>
<td>1pl</td>
<td>-nou</td>
<td>an nou</td>
</tr>
<tr>
<td>2pl</td>
<td>-zot</td>
<td>a zot</td>
</tr>
<tr>
<td>3pl</td>
<td>-yo</td>
<td>a yo</td>
</tr>
</tbody>
</table>

Table 2. Possessive forms

Since there is no syntactic motivation for the absence of the preposition in (10b), it is easy to analyze this as two different possessive paradigms.

In the TAG grammar, the possessive form (Fig. 16) of a noun depends on the particular class of the noun and the weak pronoun (wPr) is inserted as a co-anchor.

Figure 16. Elementary Trees for ‘N wPr’ and ‘N a wPr’
5. Conclusion

In this paper, I have presented some arguments in favor of the treatment of functional elements of Gwadloupéyen as multi-word (grammatical) expressions, i.e. periphrasis. Contrary to a syntactic approach of periphrasis, that derives the meaning in a bottom-up manner (syntactic derivation) I have defended an approach which considers the periphrasis as a single syntactic element (a complex tree) which is clearly assembled in morphology. The only difference between synthetic forms and periphrastic forms is the level (or the domain) where the process takes place. I have shown that the TMAs in Gwadloupéyen constitute a clear case of inflectional periphrasis (§3) and that inflectional periphrasis can be found outside the verbal domain (§4). This analysis has been implemented using XMG2 (Petitjean 2014).  

The results presented here contribute to the discussion on the morphology of Creole languages. While some researchers (for instance Seuren and Wekker 1986 and McWhorter 2001) have claimed that creole languages are morphologically poor, the facts presented here (but see also Henri and Kihm 2013) tend to show the contrary.

References


9 The complete metagrammar is freely available for verification or reuse for other languages on GitHub. The link has been mentioned in footnote 8.


