

# Une grammaire TAG du Gwadeloupéen

---

Emmanuel Schang

GRGC, 15 mai 2018

LLL, UMR7270

# Introduction

---

L'**informatisation des langues** est un enjeu important pour les communautés humaines qui souhaitent conserver cette partie de leur culture. La mondialisation des échanges et les nouvelles technologies de la communication bouleversent le statut de la majorité des langues du monde.

L'existence de ressources électroniques (terminologies, dictionnaires et grammaires) est une condition nécessaire, mais non suffisante, à la survie des langues.

# Langues peu dotées

Difficulté de travailler sur des langues sous-valorisées (patois, parlars. . . ), due en partie à l'attitude sociale des locuteurs :

- dévalorisation de la langue,
- tyrannie des académiciens auto-proclamés ("je ne parle pas bien créole", "mon créole n'est pas le créole pur").

# Langues peu dotées

Difficulté de travailler sur des langues sous-valorisées (patois, parlars. . . ), due en partie à l'attitude sociale des locuteurs :

- dévalorisation de la langue,
- tyrannie des académiciens auto-proclamés ("je ne parle pas bien créole", "mon créole n'est pas le créole pur").

Actions positives :

- le créole est une langue, elle en a tous les 'attributs' modernes (logiciels, sites, dicos électroniques, etc.),
- cela ne peut se faire que de manière collaborative (mais il faut bien démarrer un jour. . . ).

## Absence de standardisation

Dans le cas des créoles antillais (Antilles françaises), le problème n'est pas l'absence de standard (GEREC...) mais trop de standards peu suivis. ➡ Hazaël-Massieux (1993)

## Absence de standardisation

Dans le cas des créoles antillais (Antilles françaises), le problème n'est pas l'absence de standard (GEREC...) mais trop de standards peu suivis. ➡ Hazaël-Massieux (1993)

On trouve sur le net des fils de discussion sur les problèmes graphiques et sur la liberté d'écrire le créole comme on le parle.

# Difficultés

## **Absence de standardisation**

Dans le cas des créoles antillais (Antilles françaises), le problème n'est pas l'absence de standard (GEREC...) mais trop de standards peu suivis. ➡ Hazaël-Massieux (1993)

On trouve sur le net des fils de discussion sur les problèmes graphiques et sur la liberté d'écrire le créole comme on le parle.

## **Variation individuelle**

Travail sur les productions réelles des locuteurs dans des situations idéalisées (entretiens) et sur des questionnaires (élicitation) pour rechercher une couverture 'maximale' des phénomènes.

# Difficultés

## Absence de standardisation

Dans le cas des créoles antillais (Antilles françaises), le problème n'est pas l'absence de standard (GEREC...) mais trop de standards peu suivis. ➡ Hazaël-Massieux (1993)

On trouve sur le net des fils de discussion sur les problèmes graphiques et sur la liberté d'écrire le créole comme on le parle.

## Variation individuelle

Travail sur les productions réelles des locuteurs dans des situations idéalisées (entretiens) et sur des questionnaires (élicitation) pour rechercher une couverture 'maximale' des phénomènes.

Pas de créole folklorique (doudouisme).

# Pour les créolistes

Objectifs :

- disposer d'une grammaire formelle à couverture 'raisonnable'.

# Pour les créolistes

Objectifs :

- disposer d'une grammaire formelle à couverture 'raisonnable'.
- sortir du travail sur le créole basilectal que personne ne parle pour aller vers une grammaire 'couvrante'.

# Pour les créolistes

Objectifs :

- disposer d'une grammaire formelle à couverture 'raisonnable'.
- sortir du travail sur le créole basilectal que personne ne parle pour aller vers une grammaire 'couvrante'.

## Le challenge

Grammaire électronique,

# Pour les créolistes

Objectifs :

- disposer d'une grammaire formelle à couverture 'raisonnable'.
- sortir du travail sur le créole basilectal que personne ne parle pour aller vers une grammaire 'couvrante'.

## **Le challenge**

Grammaire électronique, librement disponible,

# Pour les créolistes

Objectifs :

- disposer d'une grammaire formelle à couverture 'raisonnable'.
- sortir du travail sur le créole basilectal que personne ne parle pour aller vers une grammaire 'couvrante'.

## Le challenge

Grammaire électronique, librement disponible, modifiable et réutilisable.

# Pour les créolistes

## Objectifs :

- disposer d'une grammaire formelle à couverture 'raisonnable'.
- sortir du travail sur le créole basilectal que personne ne parle pour aller vers une grammaire 'couvrante'.

## Le challenge

Grammaire électronique, librement disponible, modifiable et réutilisable.  Accessible sur GitHub :

[https://github.com/eschang/xmg\\_GC\\_metagrammar](https://github.com/eschang/xmg_GC_metagrammar)

# Méthode

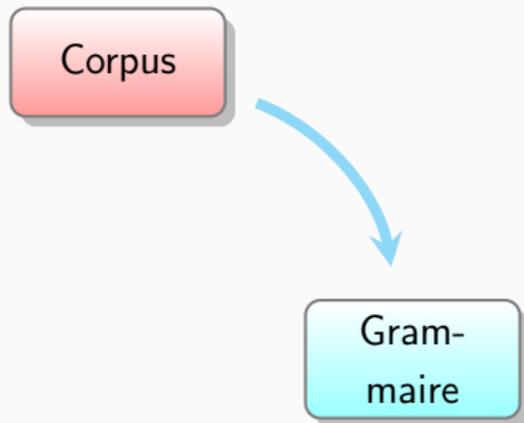
Travail sur des phrases types (questionnaires et corpus) et validation de la grammaire sur un corpus de phrases types.



Corpus

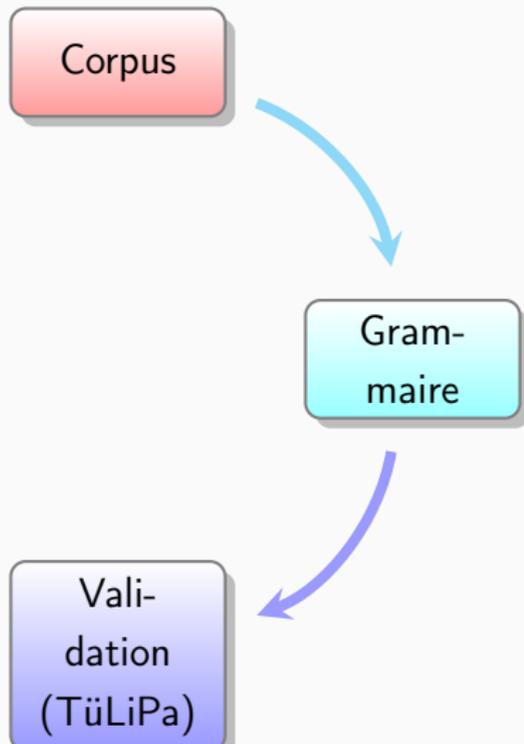
# Méthode

Travail sur des phrases types (questionnaires et corpus) et validation de la grammaire sur un corpus de phrases types.



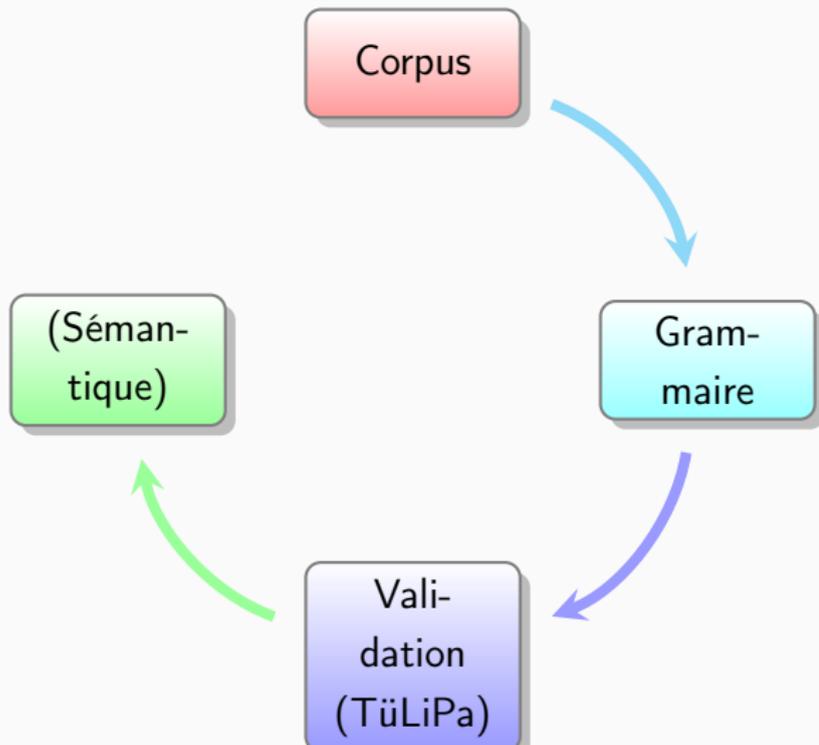
# Méthode

Travail sur des phrases types (questionnaires et corpus) et validation de la grammaire sur un corpus de phrases types.



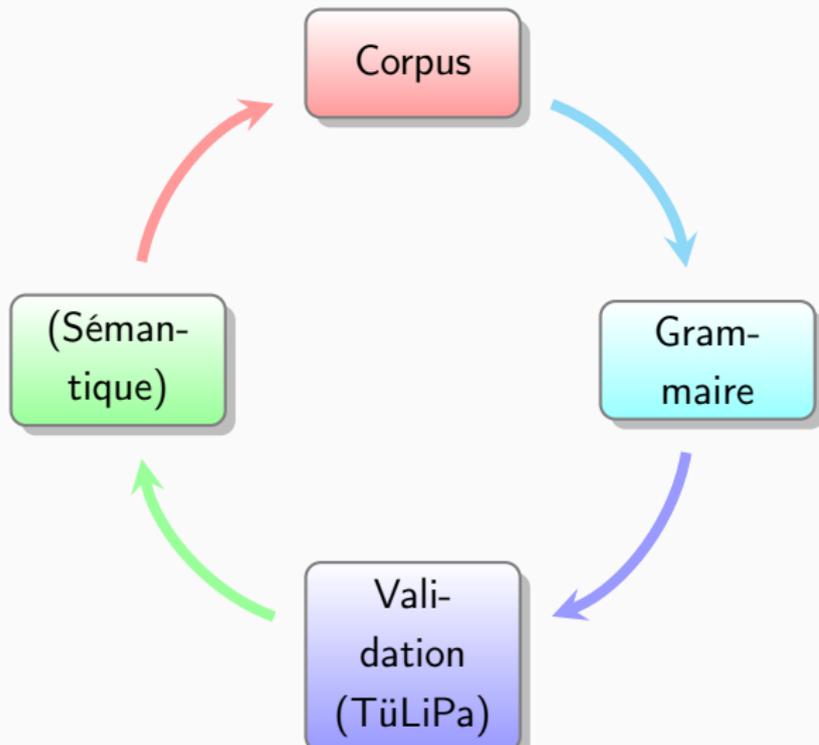
# Méthode

Travail sur des phrases types (questionnaires et corpus) et validation de la grammaire sur un corpus de phrases types.



# Méthode

Travail sur des phrases types (questionnaires et corpus) et validation de la grammaire sur un corpus de phrases types.



# Fichier corpus

---

gloss : Jean mangeait  
text : Jan té ka manjé  
axiom : s  
expected: 1

---

gloss : Jean mangeait  
text : Jan ka té manjé  
axiom : s  
expected: 0

---

gloss : Jean aime Lucette  
text : Jan enmé Lisette  
axiom : s  
expected: 1

---

# Sources principales

- Bernabé (1983)
- Damoiseau (2012)
- Delumeau (2006)
- des dictionnaires : Barbotin et Déterville

# Plan

Introduction

Construire une grammaire TAG

Grammaire TAG du gwadeloupéen

# Construire une grammaire TAG

---

# Grammaire d'unification

La **Grammaire d'Arbres Adjoints** est un type de grammaires d'unification.

# Grammaire d'unification

La **Grammaire d'Arbres Adjoints** est un type de grammaires d'unification. Grammaire d'**arbres** : ce sont des arbres qui se combinent (structures de profondeur  $\geq 1$ ).

# Grammaire d'unification

La **Grammaire d'Arbres Adjoints** est un type de grammaires d'unification. Grammaire d'**arbres** : ce sont des arbres qui se combinent (structures de profondeur  $\geq 1$ ).

Possibilité de construire des arbres avec des co-têtes (expr. idiomatiques, 'prendre une veste'...).

# Grammaire d'unification

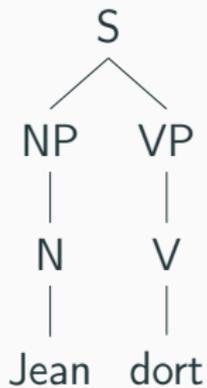
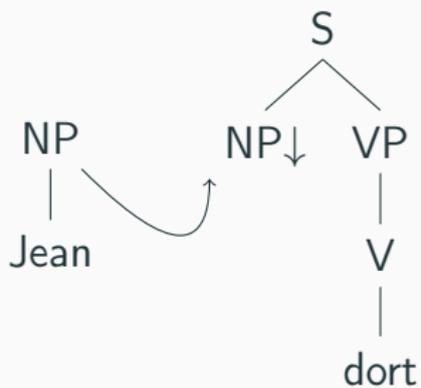
La **Grammaire d'Arbres Adjoints** est un type de grammaires d'unification. Grammaire d'**arbres** : ce sont des arbres qui se combinent (structures de profondeur  $\geq 1$ ).

Possibilité de construire des arbres avec des co-têtes (expr. idiomatiques, 'prendre une veste'...).

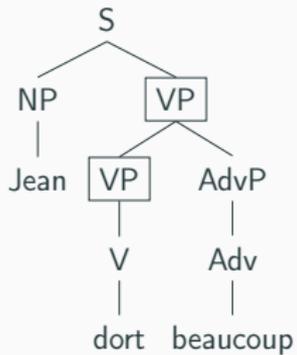
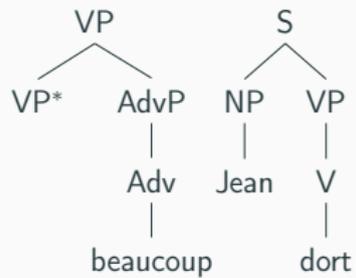
**Principes de fonctionnement contraints :**

2 opérations : substitution et adjonction (adjoining).

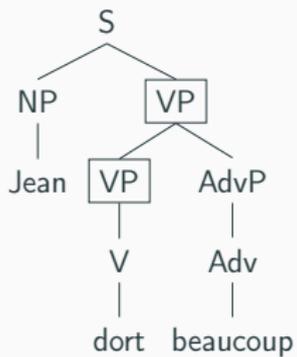
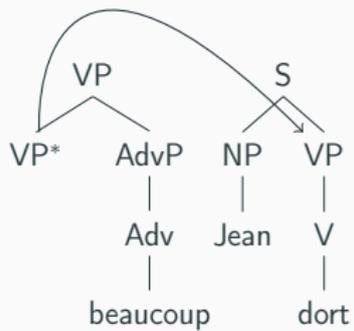
# Substitution



# Adjonction



# Adjonction



# Arbres "ancrés" (LTAG)

Il y a plusieurs étapes dans la constitution de la grammaire :

# Arbres "ancrés" (LTAG)

Il y a plusieurs étapes dans la constitution de la grammaire :

- définition des arbres élémentaires (ET),

# Arbres "ancrés" (LTAG)

Il y a plusieurs étapes dans la constitution de la grammaire :

- définition des arbres élémentaires (ET),
- 'ancrage' des arbres par des items lexicaux,

# Arbres "ancrés" (LTAG)

Il y a plusieurs étapes dans la constitution de la grammaire :

- définition des arbres élémentaires (ET),
- 'ancrage' des arbres par des items lexicaux,
- correspondant sémantique (v. plus tard)

- Ancrage lexical : tout arbre élémentaire a au moins une tête lexicale non vide

## Principes linguistiques, Abeillé (2002)

- Ancrage lexical : tout arbre élémentaire a au moins une tête lexicale non vide
- Cooccurrence prédicat-arguments (un noeud pour chaque argument)

- Ancrage lexical : tout arbre élémentaire a au moins une tête lexicale non vide
- Cooccurrence prédicat-arguments (un noeud pour chaque argument)
- Ancrage et compositionnalité sémantique : un arbre élémentaire = une seule unité sémantique ; les éléments fonctionnels apparaissent comme 'co-ancres' d'arbres élémentaires avec une tête lexicale non-vide

## Analyse en deux étapes

Dans les grammaires TAG "classiques", la description de la combinaison des arbres (diapos précédentes) prend toute la place, mais :

# Analyse en deux étapes

Dans les grammaires TAG "classiques", la description de la combinaison des arbres (diapos précédentes) prend toute la place, mais :

- construire les arbres élémentaires est une étape essentielle,
- c'est une étape coûteuse en temps.

# Analyse en deux étapes

Dans les grammaires TAG "classiques", la description de la combinaison des arbres (diapos précédentes) prend toute la place, mais :

- construire les arbres élémentaires est une étape essentielle,
- c'est une étape coûteuse en temps.

☞ **Utilisation de XMG2 Petitjean (2014) pour la construction des arbres élémentaires**

## Concept de méta-grammaire :

Factoriser les fragments d'arbres (comme factorisation des règles).

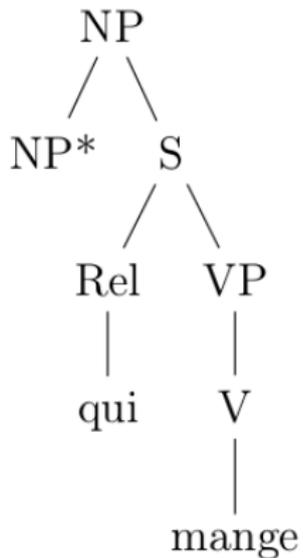
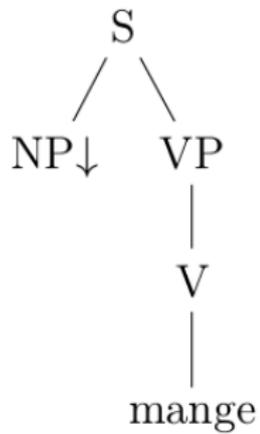
# Un exemple

Lorsqu'on a les phrases :

- *l'homme mange*
- *l'homme qui mange*

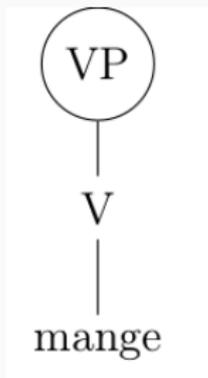
On a deux arbres élémentaires différents associés à *mange*.

## Un exemple (2)



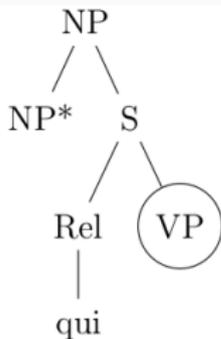
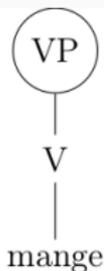
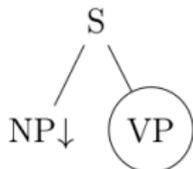
## Un exemple (3)

On peut factoriser un fragment d'arbre :



## Un exemple (4)

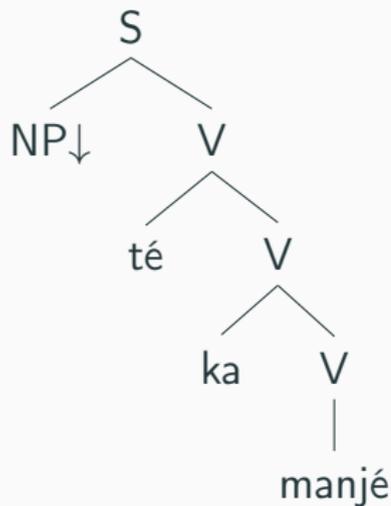
Et définir deux fragments qui, associés au précédent, donneront un arbre élémentaire valide :



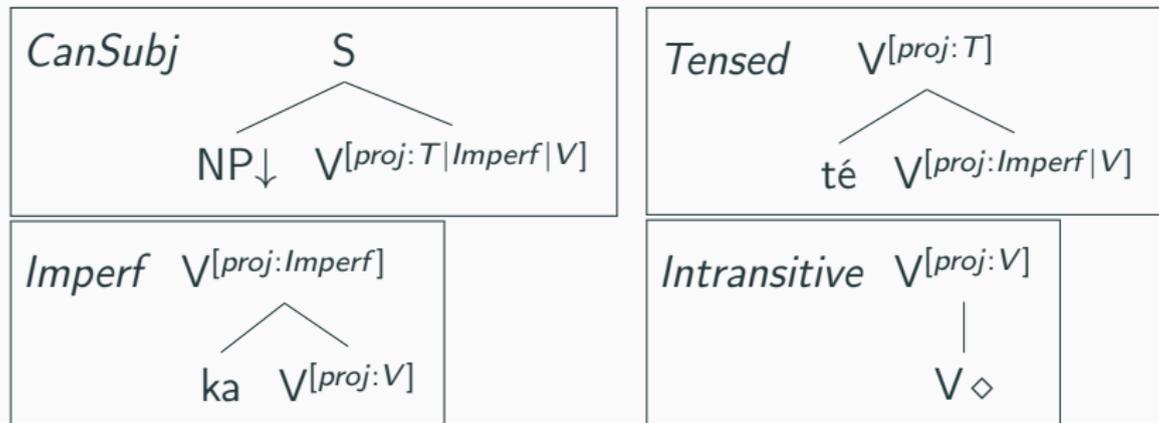
# Un exemple en CG

Prenons *NP* (*Jan par exemple*) *té ka manjé* 'NP était en train de manger'.

On souhaite obtenir l'arbre élémentaire suivant :



## Un exemple en CG (2)



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

**Figure 1:** Les fragments utilisés pour 'NP té ka manjé'

## Utilisation de XMG pour créer une grammaire TAG

- écriture de fragments qui se combinent pour donner des arbres élémentaires,
- les arbres élémentaires se combinent par adjonction et substitution,
- la combinaison de fragment concerne la morphosyntaxe.

## Une famille d'arbres :

- des arbres élémentaires liés par
  1. des arguments communs,
  2. des traits morphosyntaxiques différents (TMA, pluriel, etc.)

## Une famille d'arbres :

- des arbres élémentaires liés par
  1. des arguments communs,
  2. des traits morphosyntaxiques différents (TMA, pluriel, etc.)
- exemple : n0Vn1 (famille des verbes transitifs)

# Grammaire TAG du gwadloupéyen

---

Quelques exemples :

$s(np, pred(té, pred(ka, v(v))))$

$s(np, pred(ka, v(v)))$

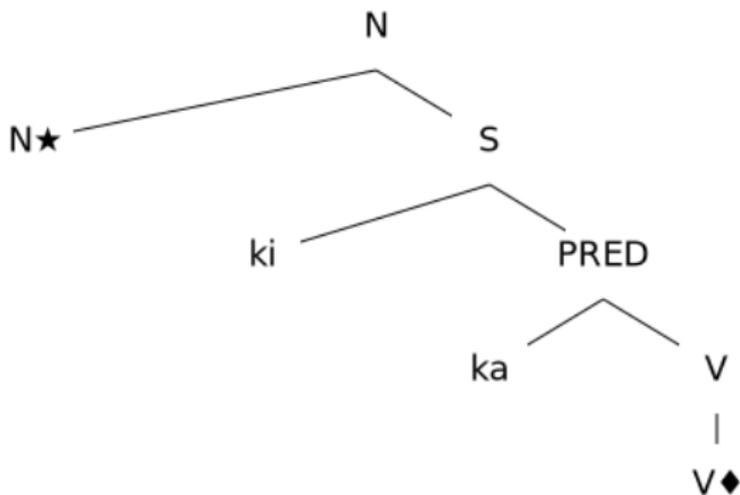
$s(np, pred(pa, pred(jen, pred(té, pred(ka, v(v))))))$

$n(n, s(ki, pred(ka, v(v))))$

$s(pa, v(v))$

[...]

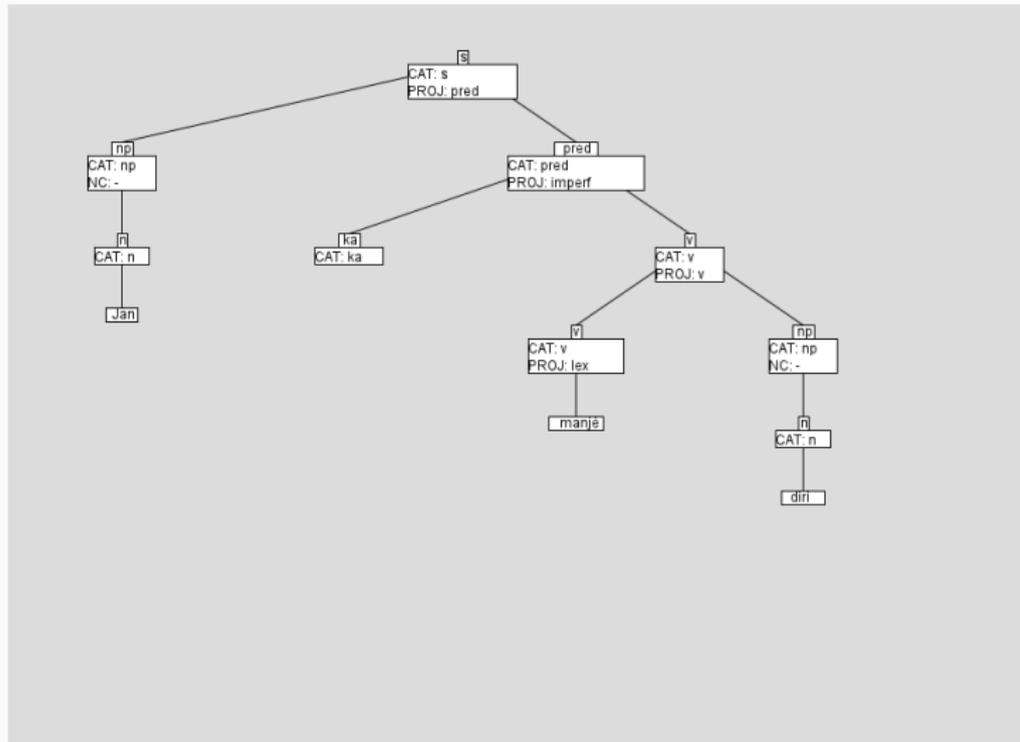
On va mettre sous forme d'arbre  $n(n, s(ki, pred(ka, v(v))))$



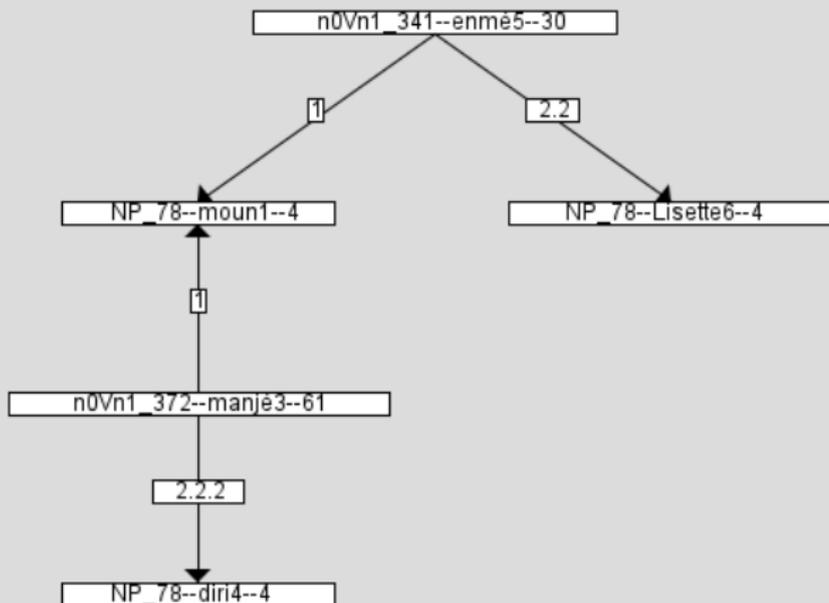
Par exemple, dans :  
timoun ki ka manjé...

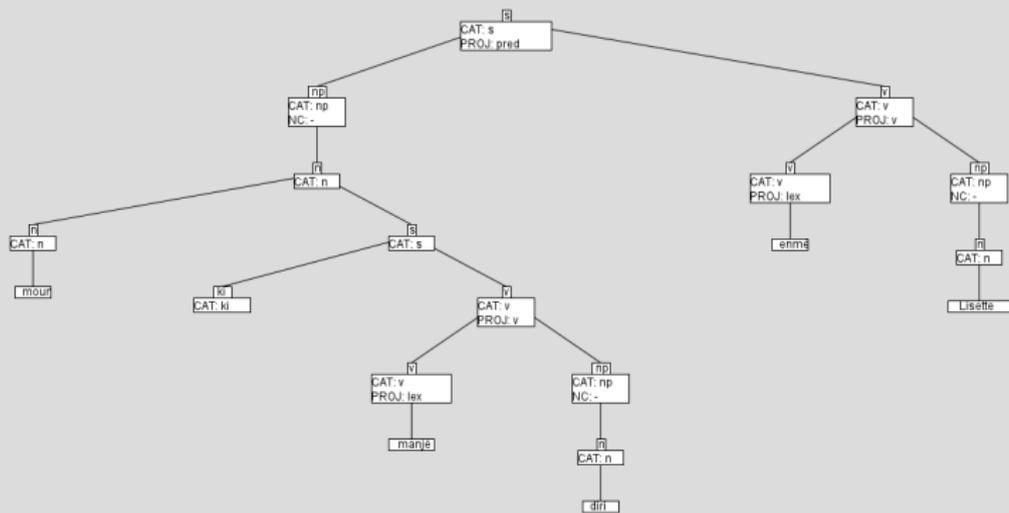
# Famille n0Vn1

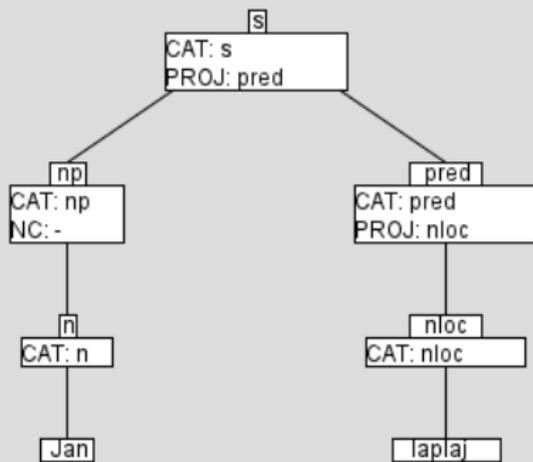
Ex.: *Jan ka manjé diri*



*moun ki manjé diri enmé Lisette*

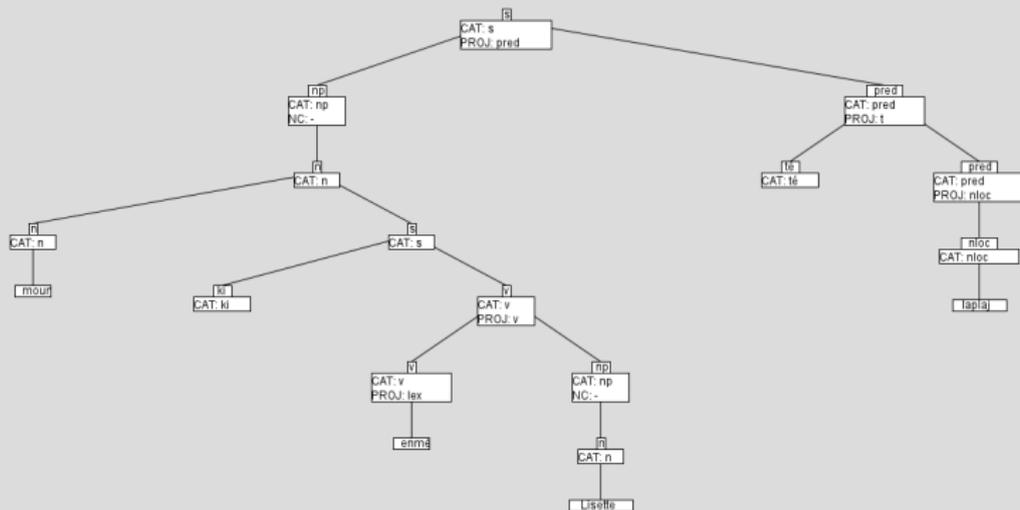




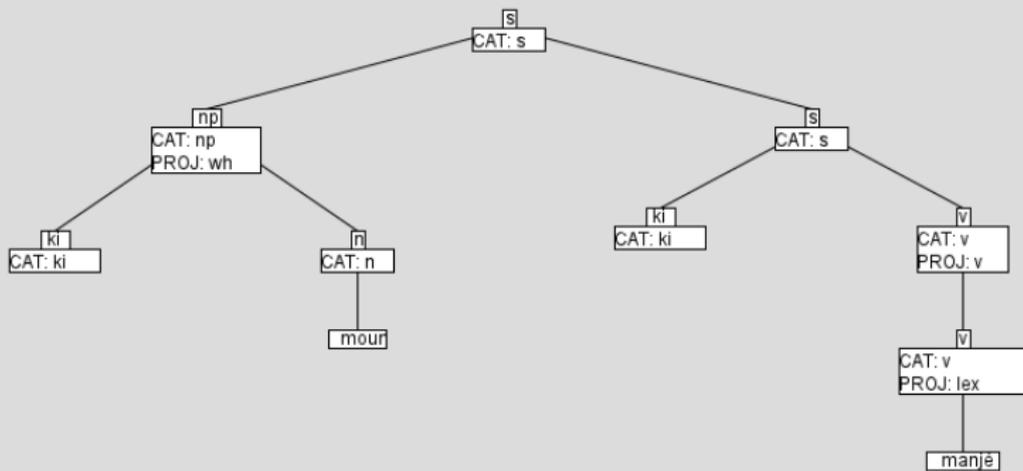


Plus compliqué :

*moun ki enmé Lisette té laplaj*



# Interrogatives



A ce jour, plus de 600 arbres élémentaires.

Les structures les plus courantes sont couvertes.

A ce jour, plus de 600 arbres élémentaires.

Les structures les plus courantes sont couvertes.

Il reste :

- à intégrer du lexique (pour l'instant, un lexique-jouet),
- à relier les structures à une couche sémantique.

# Un peu de code...

```
class subject
{CanSubject[] | RelSubject[] | WhSubject[]}
```

```
class n0V
{
{
  subject[];
  Intransitive[];
  TMA[]
} | {imperative[] | NegImperative[]}
}
```

```
class VTrans
{
{
  subject[];
  Transitive[] ;
  TMA[]
} | {TransImperative[] | NegTransImperative[]}
}
```

# Quelle sémantique ?

Habituellement, les TAG sont couplées à une sémantique logique  
☞ Gardent and Kallmeyer (2003).

# Quelle sémantique ?

Habituellement, les TAG sont couplées à une sémantique logique  
☞ Gardent and Kallmeyer (2003).

Mon choix : sémantique à base de Frames, ☞ Lichte and  
Petitjean (2015, 2014).

# Un exemple

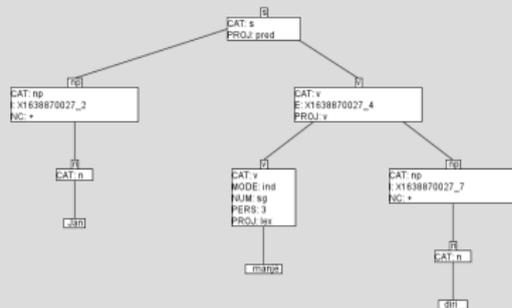
n0Vn1\_435--manjé2--30

elementary trees (no variable renaming)

n0Vn1\_435--manjé2--30

NP\_172--jan1--4

NP\_172--dir13--4



Semantic representation:

X1638870027\_4:  
type: [activity-eat-event]  
target = X628704826\_2:  
type: [entity-food]

X773397364\_1:  
type: [person-entity]  
name = John

X707064390\_1:  
type: [rice-entity-food]

# Codage sémantique

```
frame-types = {event, activity, eat, process, entity, person, food, rice}
frame-constraints = {
activity -> event,
eat -> activity,
eat -> process,
person -> entity,
food -> entity,
rice -> food
}
```

```
class FrameEat
declare ?X0
{
  <frame>{
    ?X0[eat,
      target: [food]]
  };
  <iface>{
    [e=?X0]
  }
}
```

# Conclusion

- Grammaire disponible. Code en libre accès. Facile à enrichir, facile à modifier.

# Conclusion

- Grammaire disponible. Code en libre accès. Facile à enrichir, facile à modifier.
- Possibilité de travailler à partir de ce code pour l'adapter à une langue 'proche'.

## References

---

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

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

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.

## Bibliographie II

- Gardent, C. and Kallmeyer, L. (2003). Semantic construction in feature-based TAG. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - EACL '03*. Association for Computational Linguistics.
- Hazaël-Massieux, M.-C. (1993). *Ecrire en créole(oralité et écriture aux antilles)*.
- Lichte, T. and Petitjean, S. (2014). Adding semantic frames to xmg. *Heinrich-Heine-Universität Düsseldorf. Entwurf*.
- Lichte, T. and Petitjean, S. (2015). Implementing semantic frames as typed feature structures with xmg. *Journal of Language Modelling*, 3(1):185–228.

Petitjean, S. (2014). *Génération modulaire de grammaires formelles*. PhD thesis, Université d'Orléans.