

Une grammaire TAG du créole guadeloupéen

E. Schang

Univ. Orléans, LLL

6 octobre 2016

- Disposer d'une grammaire formelle avec une couverture la plus large possible.

Pourquoi ?

- Disposer d'une grammaire formelle avec une couverture la plus large possible.
- Utiliser les 'règles' de cette grammaire pour annoter syntaxiquement des textes, créer un corpus arboré disponible pour des annotations de 'haut niveau' (TimeML, anaphores, relations de discours...).

Pourquoi ?

- Disposer d'une grammaire formelle avec une couverture la plus large possible.
- Utiliser les 'règles' de cette grammaire pour annoter syntaxiquement des textes, créer un corpus arboré disponible pour des annotations de 'haut niveau' (TimeML, anaphores, relations de discours...).
- Métagrammaire disponible sur GitHub
https://github.com/eschang/xmg_GC_metagramar

Pourquoi ?

- Disposer d'une grammaire formelle avec une couverture la plus large possible.
- Utiliser les 'règles' de cette grammaire pour annoter syntaxiquement des textes, créer un corpus arboré disponible pour des annotations de 'haut niveau' (TimeML, anaphores, relations de discours...).
- Métagrammaire disponible sur GitHub
https://github.com/eschang/xmg_GC_metagrammar
- Travaux existants : [Vaillant, 2008b] , [Vaillant, 2008a]

Pourquoi ?

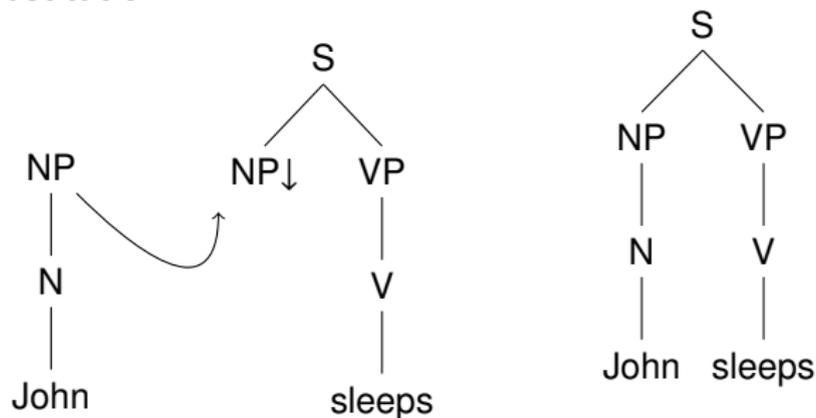
- Disposer d'une grammaire formelle avec une couverture la plus large possible.
- Utiliser les 'règles' de cette grammaire pour annoter syntaxiquement des textes, créer un corpus arboré disponible pour des annotations de 'haut niveau' (TimeML, anaphores, relations de discours...).
- Métagrammaire disponible sur GitHub
https://github.com/eschang/xmg_GC_metagrammar
- Travaux existants : [Vaillant, 2008b] , [Vaillant, 2008a]
- Inspiration majeure : Kriyol Syntax [Kihm, 1994]

Grammaire d'arbres lexicalisée.

Deux opérations: Substitution & Adjonction

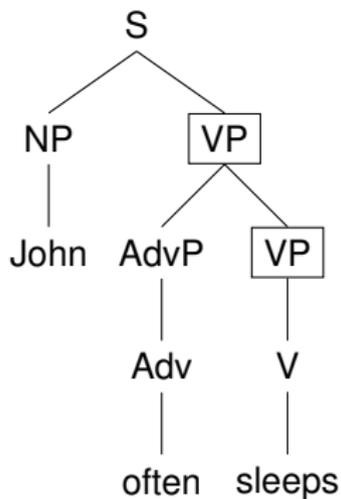
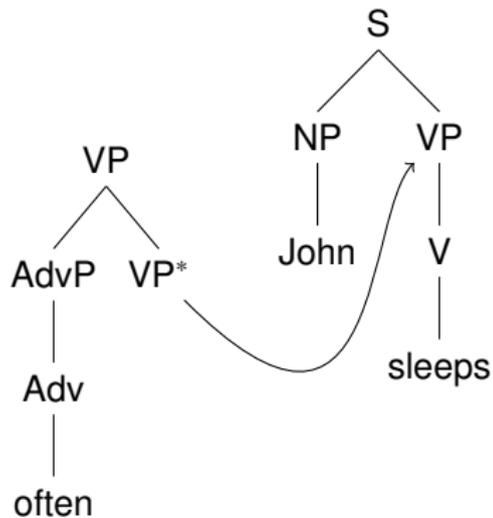
(1) John sleeps.

Substitution:



Adjonction:

Insertion au noeud VP:



Principes (cf. [Abeillé, 2002]) :

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

Principes (cf. [Abeillé, 2002]) :

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

Principes (cf. [Abeillé, 2002]) :

- 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

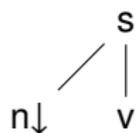
- Alternative aux règles lexicales

- Alternative aux règles lexicales
- Outil de composition des arbres élémentaires : [Candito, 1999] et [Crabbé, 2005]

- Alternative aux règles lexicales
- Outil de composition des arbres élémentaires : [Candito, 1999] et [Crabbé, 2005]
- XMG-NG : compilateur de métagrammaires v. [Petitjean et al., 2016]

Un exemple (adapté d'une présentation de S. Petijean au LIFO en mars 2016) :
Les classes...

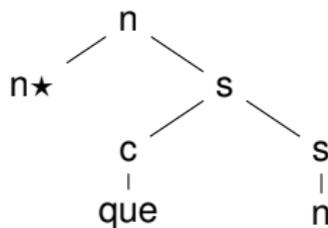
CanSubj



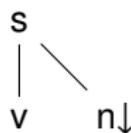
Active



RelObj

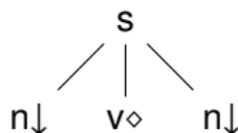
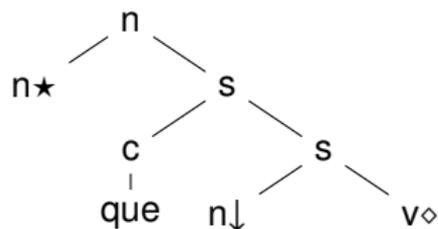


CanObj



se combinent pour former les arbres:

$\text{CanSubj} \wedge \text{Active} \wedge (\text{RelObj} \vee \text{CanObj})$



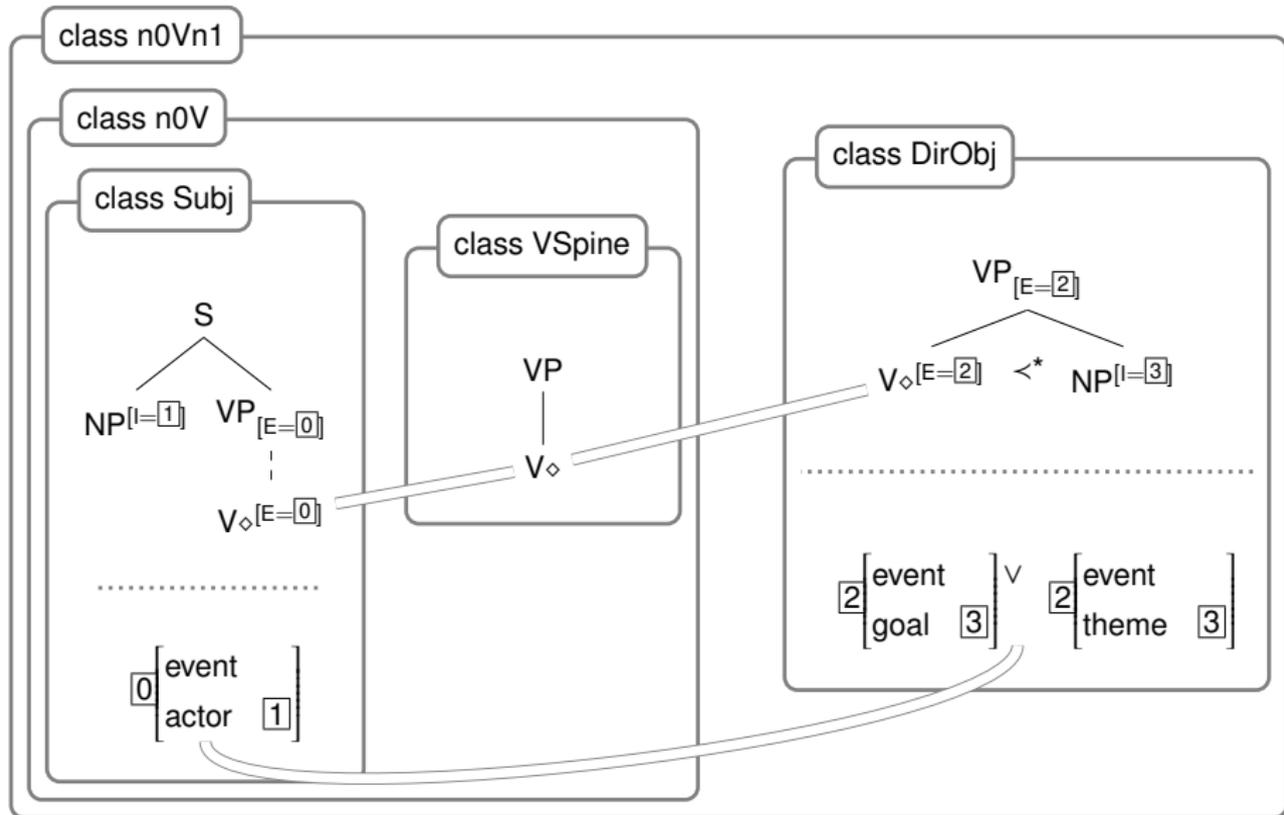


Figure: Factorisation metagrammaticale pour les verbes transitifs. (Merci Simon !)

- Utiliser la combinaison de fragments pour générer des arbres assez complexes.
- Trouver une alternative à l'Adjonction pour les marqueurs TMA notamment (pour une analyse semblable sur le créole santoméen : [Schang et al., 2012]).

Marqueurs de Temps et d'Aspect : symboles de la différence entre les créoles et leur langue lexificatrice.

Marqueurs de Temps et d'Aspect : symboles de la différence entre les créoles et leur langue lexicatrice.

Pour le guadeloupéen : [McCrindle, 1999] et [Pfänder, 2000] notamment.

Pour les créoles en général : [Winford, 2012]

Table: Marqueurs TMA

VALEUR	FORME
Accompli	manjé
Inaccompli (Itératif etc.)	ka manjé
Futur (Prospectif)	ké manjé
Passé accompli	té manjé
Passé inaccompli	té ka manjé
Irréaliste passé	té ké manjé
Irréaliste inaccompli	té ké ka manjé

Que faire de ces marqueurs ?

Problème : difficile de comparer avec le français (cf. [Abeillé and Godard, 2003]),
mais :

Que faire de ces marqueurs ?

Problème : difficile de comparer avec le français (cf. [Abeillé and Godard, 2003]),
mais :

- on ne peut pas les coordonner

(2) *Jan ka é ké manjé.
Jean IPFV et PROSP manger
'Jean mange et mangera'

La répétition du verbe est la norme :

(3) an fouté y, an ka fouté y, an ké fouté y !
'J'ai déjà gagné, je gagne cette fois-ci et je gagnerai encore ! (version polie)'

- Clivage du prédicat possible mais sans les TMA :

- (4) sé monté nou ka monté pou nou rivé la nou ka alé.
c'est monter 1pl IPFV monter pour 1pl arriver là 1pl IPFV aller
'On va grimper, grimper pour arriver là où l'on va' (intensif)
- (5) a. *sé **ka** monté nou ka monté...
b. *sé **té** monté nou té monté...

- la négation également exclue du clivage :

(6) Sé/pou (*pa) monté, nou pa monté
c'est/pour NEG monter 1pl IMPERF monter
'Nous ne sommes PAS monté.'

Mais exception avec les modaux *pé* et *vlé* pour la négation :

(7) Sé vlé pa, i vlé pa !
c'est vouloir NEG 3SG vouloir NEG
'Ah ça, y veut pas, y veut pas !'

- Fusion

(8) Jan péké manjé
Jean NEG-PROSP manger
'Jean ne mangera pas'

- Pas de coordination des TMA
- Pas de clivage
- Fusion entre TMA

Les TMA ne sont pas des verbes, mais des éléments proches des affixes (mais certains adverbes s'intercalent entre les TMA).

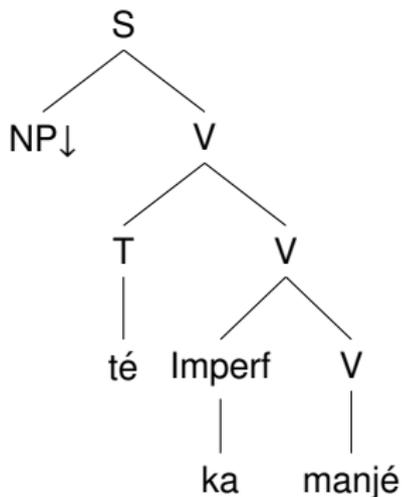
Idée :

- projections fonctionnelles plutôt qu'ancres.
- utiliser la métagrammaire pour générer les différentes formes 'fléchies' des verbes.

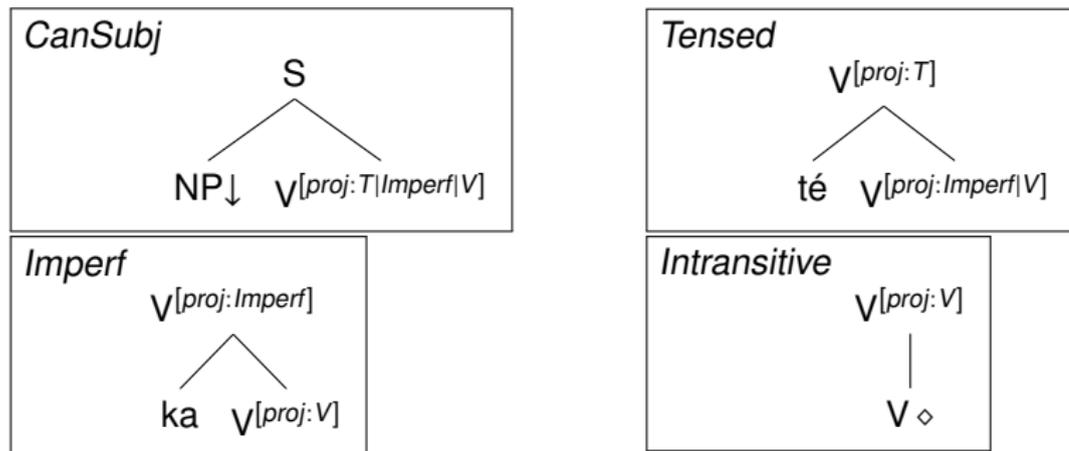
Exemple

- (9) NP té ka manjé
'NP était en train de manger.'

(10)



Quatre *Classes* (trait *proj* pour restreindre les combinaisons) :



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

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

Les combinaisons possibles entre TMA sous forme de Classes indépendantes :

BareV | ImperfV | ProspImperfV | TensProspImperfV
| TensedV | ProspV | TensImperfV | TensProspV

Plus simple : 'factoriser' les TMA pour toutes les classes verbales.

Un peu de code :

```
class Imperf
export ?X ?Y ?Z
declare ?X ?Y ?Z
{ <syn>{
    node ?X (color=white)[cat = v, proj = imperf]{
        node ?Y (mark=flex,color=red)[cat= ka]
        node ?Z (color=black)[cat = @{v}, proj =
            @{v} ]
    }
}
}
class TMA
{
{ Prosp[] | None[] };
{ Imperf[] | None[] };
{ Tensed[] | None[] }
}
```

Encore un peu de code :

```
class n0V
{
    subject[] ; % CanSubject et RelSubject
    Intransitive[];
    TMA[]
}
```

```
class VTrans %n0Vn1 est plus riche
{
    subject[];
    Transitive[] ;
    TMA[]
}
```

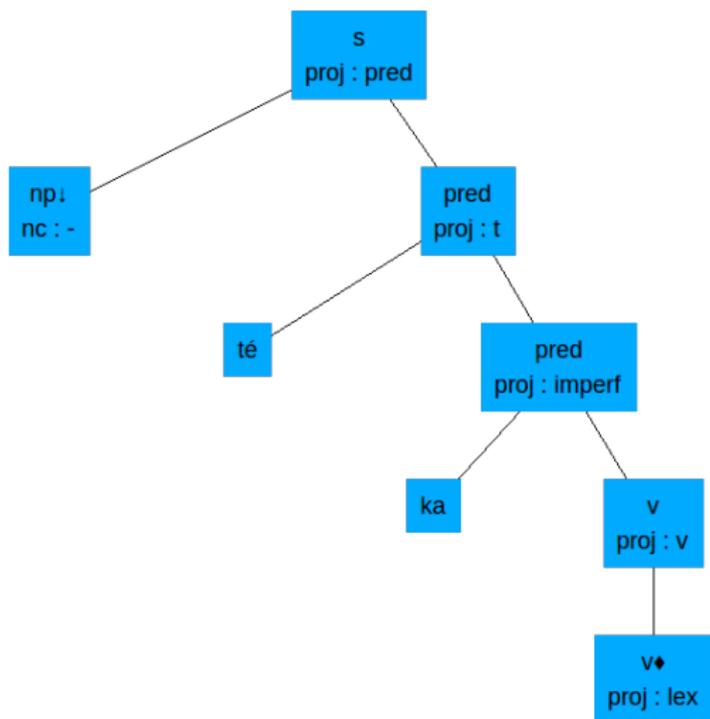


Figure: Résultat de la compilation : NP té ka V

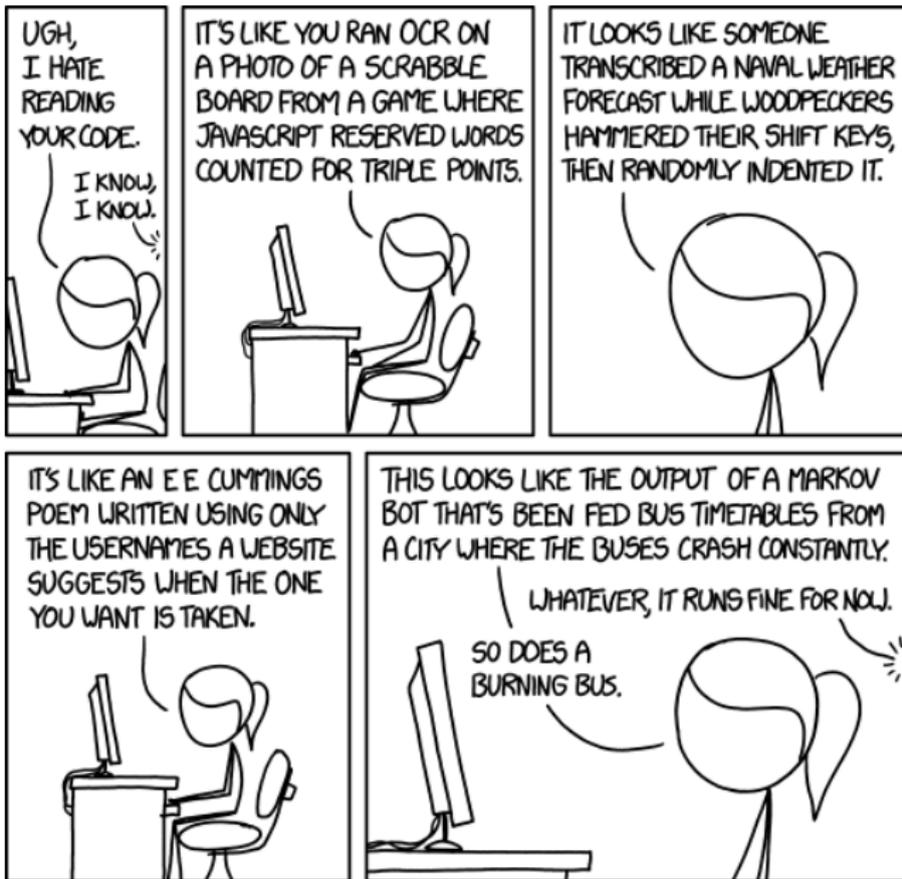


Figure: Permanent link to this comic: <http://xkcd.com/1695/>

Marqueurs préverbaux :

- (11) a. Jan pa manjé
Jean NEG manger
'Jean n'a pas mangé'
- b. Jan poko manjé
Jean NEG.ENCORE manger
'Jean n'a pas encore mangé'
- c. ou pé(ké) pé vin
2SG NEG.PROSP POUVOIR venir
'Tu ne pourras pas venir'
- d. tout sé moun la-sa yo pa jen moli
tout PL personne DEM 3PL NEG jamais mollir
'Toutes ces personnes, elles ne mollissent jamais'

On trouve *pa* 'ne pas', *poko* 'pas encore', *péké* 'pas + PROSP' et *pa jen* 'litt: pas jamais' dans la série des marques pré-verbales, entre le sujet et les TMA.
Pourquoi ne pas les ajouter aux classes TMA ?

- (12) a. pon moun pa vini.
aucune personne NEG venir
'Personne n'est venu'
- b. Jan pa manjé pon bokit
Jean NEG manger aucun bokit
'Jean n'a mangé aucun bokit'

N-words présents si *pa* est présent.

N-words présents si *pa* est présent.

La concordance est locale (arbre élémentaire du prédicat).

- (13) Jan pa di ké ou *(pa) vlé ban mwen anyen.
'Jean n'a pas dit que tu ne veux rien me donner'

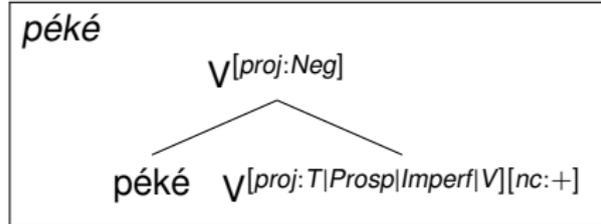
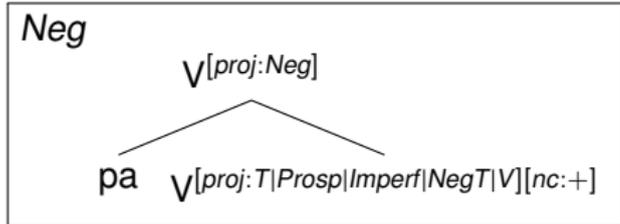
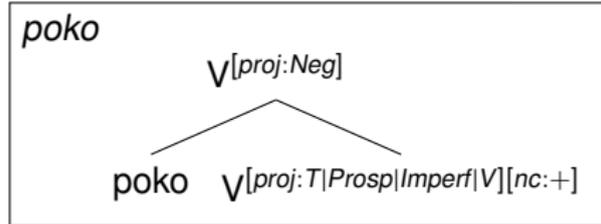
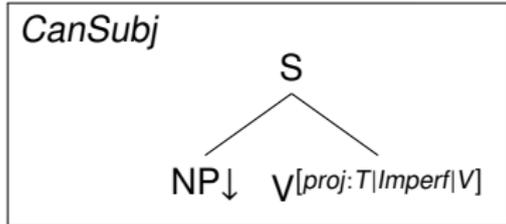
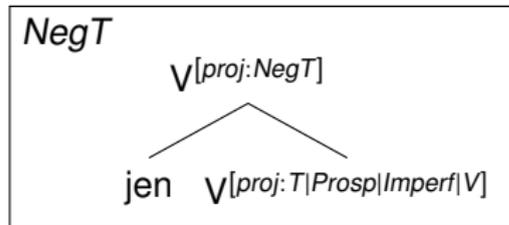
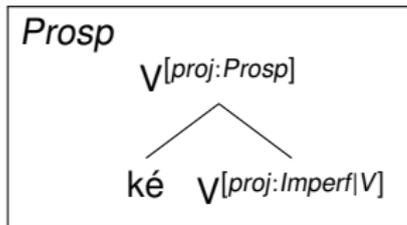
N-words présents si *pa* est présent.

La concordance est locale (arbre élémentaire du prédicat).

- (13) Jan pa di ké ou *(pa) vlé ban mwen anyen.
'Jean n'a pas dit que tu ne veux rien me donner'

Facile à gérer si le marqueur *pa* (la négation phrastique sous ses différentes formes) fait partie de l'arbre élémentaire du V.

Quelques fragments pour la négation :



Encore un peu de code :

```
class TMA
{
  { Prosp[] | None[] };
  { Imperf[] | None[] };
  { Tensed[] | None[] };
  {{ Neg[] | NegT[] | poko[] | peke[] }*=[nc= +] | None
   []*=[nc= -] }
}
class CanSubject
export ?X ?Y ?Z
declare ?X ?Y ?Z ?NC
{ <syn>{
  node ?X (color=black)[cat = s, proj = v]{
    node ?Y (mark=subst,color=black)[cat= np,
      nc=?NC]
    node ?Z (color=black)[cat = @{v}, proj =
      @{t,v,prosp,imperf,neg}]
  }
  }*=[nc=?NC]% on récupère la valeur nc du noeud Y
}
```

Un exemple de compilation de classes avec la négation :

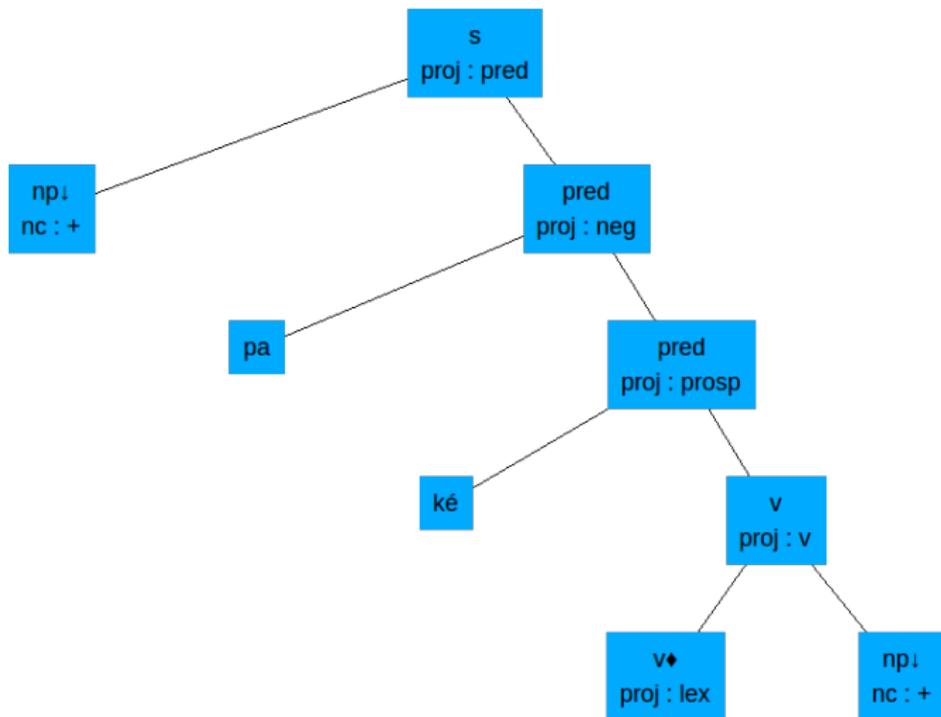


Figure: NP pa ké V NP

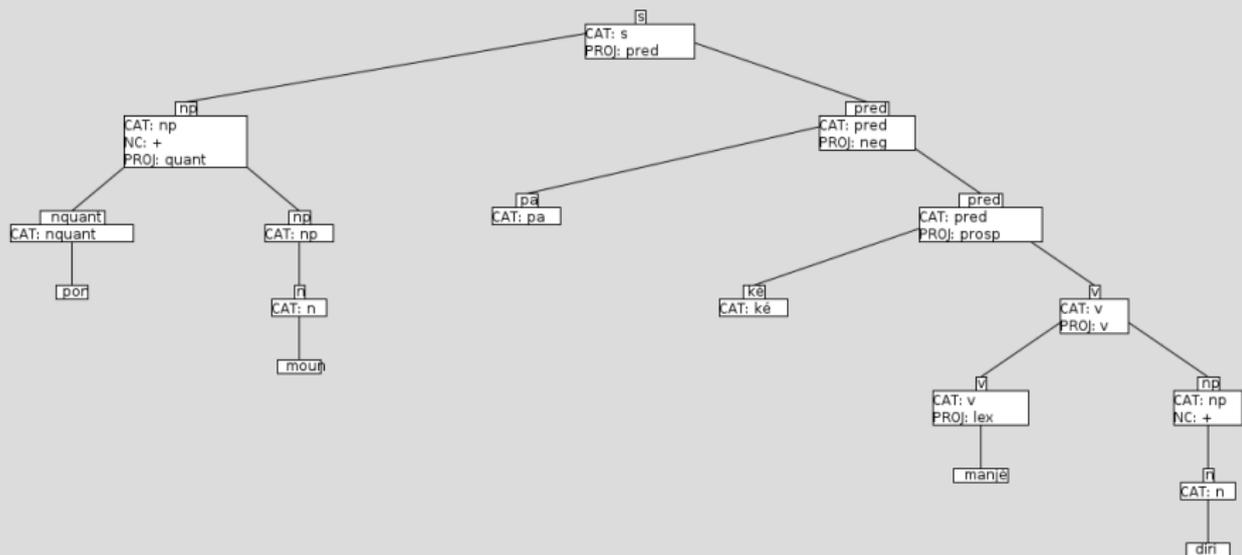


Figure: Arbre dérivé pour 'pon moun pa ké manjé diri'

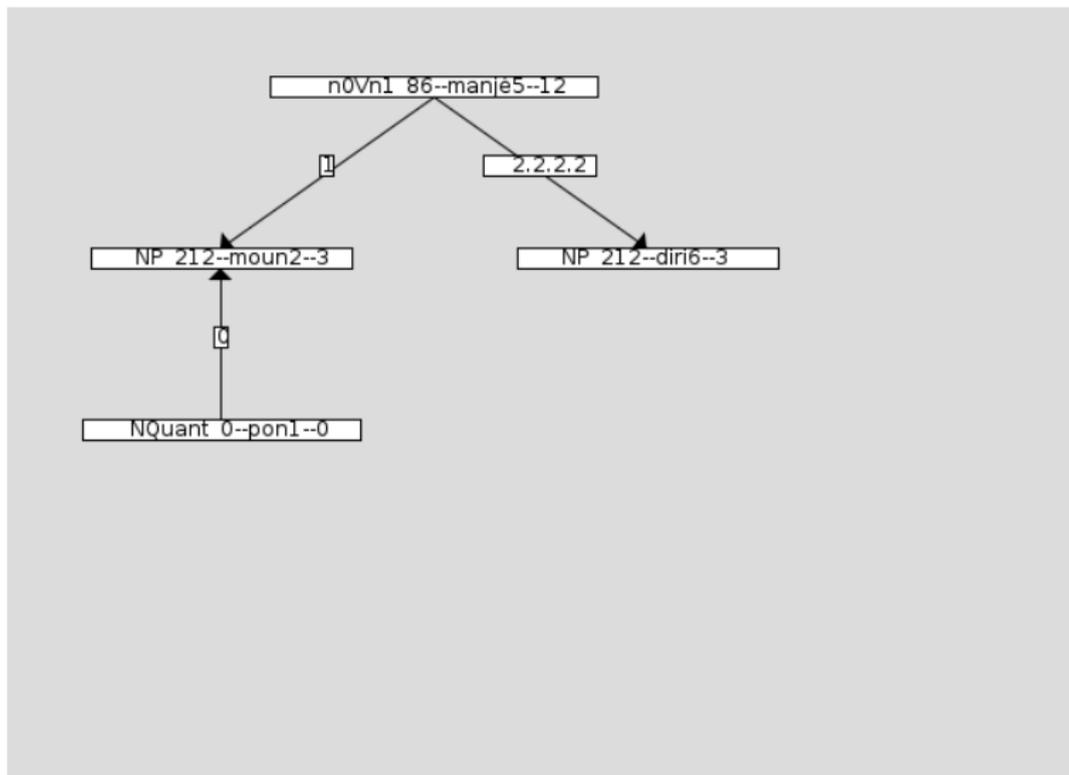


Figure: Arbre de dérivation de 'pon moun pa ké manjé diri'

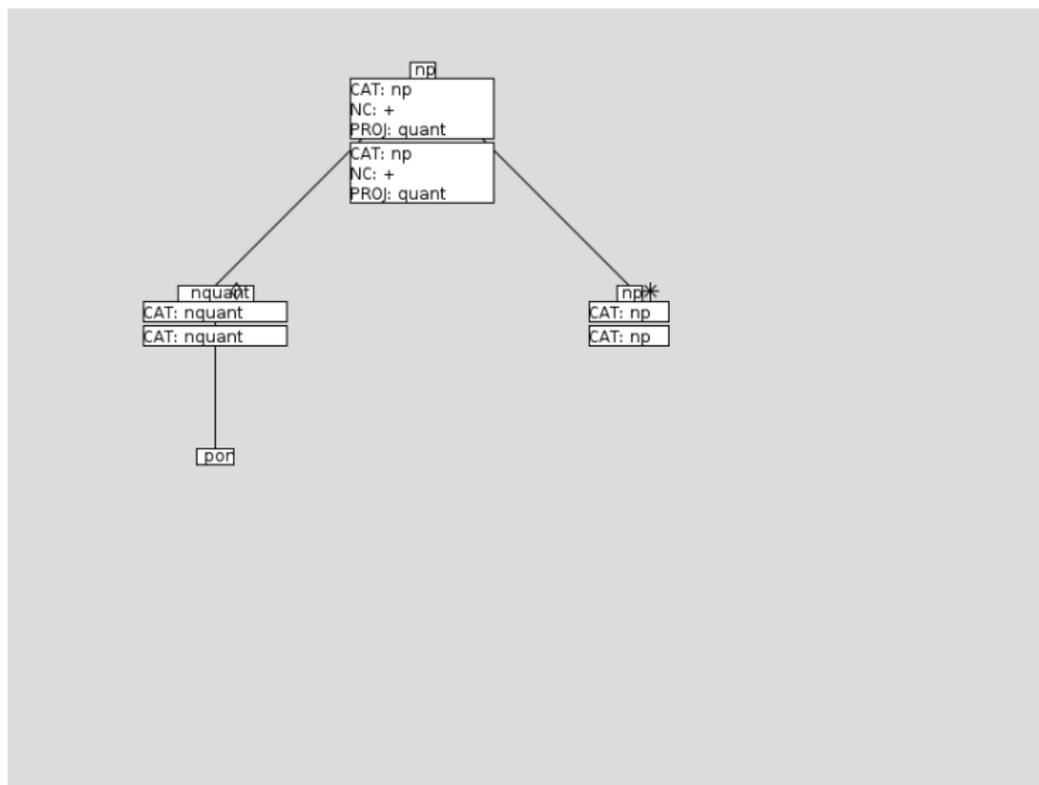


Figure: Arbre élémentaire de *pon* 'aucun'

Même idée avec le défini, le pluriel et le démonstratif :

- (14) a. sé vwati -la
voiture DEF
'les voitures'
- b. on imaj an timoun -la
INDF image de enfant DEF
'une image de l'enfant'

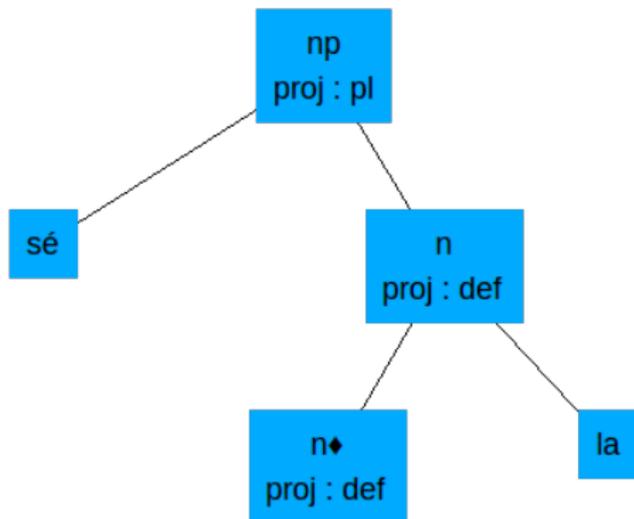


Figure: Compilation des Classes $Def \wedge Pl \wedge N$

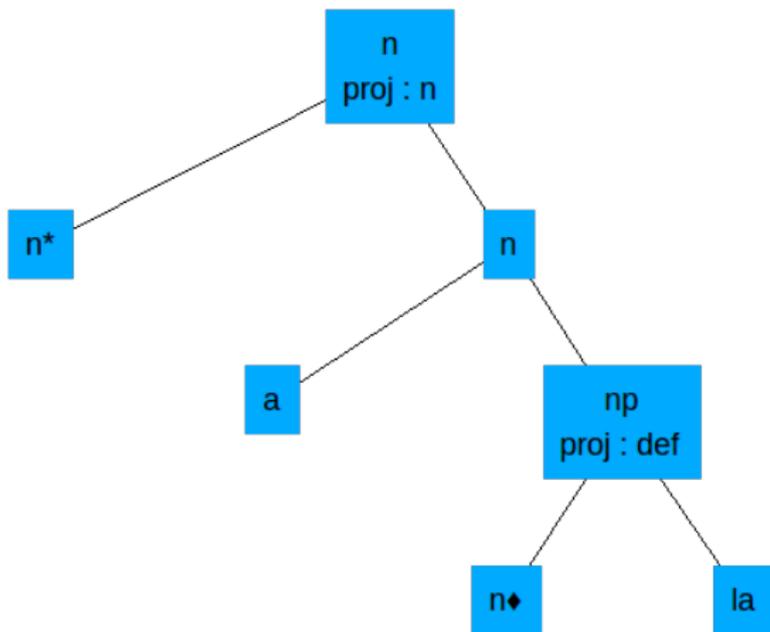


Figure: Compilation des Classes $Def \wedge Genitif \wedge N$

Encore du code ! :

```
class NP
{
    {DemN[] | PlN[] | Nounsat[] | DefN[] | Genitive
      []}
}
```

- Lien avec les 'semantic frames'
- Meilleure articulation de la métagrammaire avec les éléments lexicaux
- Intégrer la variation (et le français ?)

Merci de votre attention

Remerciements à Y. Parmentier, Simon Petitjean et D. Duchier

 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.

 Candito, M.-H. (1999).
Représentation modulaire et paramétrable de grammaires électroniques
lexicalisées: application au français et à l'italien.
These de Doctorat, Université Paris, 7.

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

 Kihm, A. (1994).
Kriyol syntax: the Portuguese-based creole language of Guinea-Bissau,
volume 14.
John Benjamins Publishing.

 McCrindle, K. L. (1999).

Temps, mode et aspect, les créoles des Caraïbes à base lexicale française.
PhD thesis.

 Petitjean, S., Duchier, D., and Parmentier, Y. (2016).

Xmg: Describing description languages.

In Logical Aspects of Computational Linguistics (LACL) 2016.

 Pfänder, S. (2000).

Aspekt und Tempus im Frankokreol.

G. Narr.

 Schang, E., Duchier, D., Ekoukou, B. M., Parmentier, Y., and Petitjean, S. (2012).

Describing Sao Tomense Using a Tree-Adjoining Meta-Grammar.

In 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+ 11), pages 82–89.

 Vaillant, P. (2008a).

Grammaires factorisées pour des dialectes apparentés.

In TALN 2008: Actes de la 15ème conférence annuelle sur le Traitement Automatique des Langues Naturelles.

 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.