Formalismes pour l'analyse syntaxique

Cours ENST – TALN
Vendredi 1er Mars 2002

Éric de la Clergerie
Eric.De_La_Clergerie@inria.fr

Transparents disponibles sur http://atoll.inria.fr/~clerger ▶ Enseignement

La syntaxe existe-t-elle ?

A priori oui :

Un schtroumpf sachant schtroumpfer

"And hast thou slain the Jabberwock ?
Come to my arms, my beamish boy !
O frabjous day ! Callooh ! Callay !"
He chortled in his joy.

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe ;
All mimsy were the borogoves,
And the mome raths outgrabe.

« Jabberwocky » (Lewis Caroll)

Colorless green ideas sleep furiously [**Noam Chomsky**]

Mais débat sur

- la nature des constituants : Nom, Verbe, Determinant Syntagme Nominal, Syntagme Verbal
- les relations entre constituants : arborescentes, graphe, . . .
- les fonctions linguistiques : sujet, objet, objet indirect, . . .

Théories linguistiques

Explication des phénomèmes linguistiques

ordre des mots SOV VOS

passif le livre est donné à Paul par Anne

extraction Jean demande quel livre Paul est en train de lire

- **Relatives** Jean connais l'homme **que** Marie regarde ε.
- **Interrogative** Jean demande **quel homme** Marie regarde ε.
- **Clivées** C'est Jean **que** Marie regarde ε.
- **Topicalisées** **Le chocolat amer**, Marie aime ε.

coordination Jean mange une pomme et Paul une poire

Existence d'une théorie « universelle » (couvrant tout les language) ?
Liens avec la sémantique ?

Formalismes

Cadre mathématique pour faire fonctionner les théories

Compromis entre expressivité et complexité (pour un traitement automatique)

Hiérarchie de Chomsky :

Type 3 grammaires régulieres (FSA) ⇒ linéaire

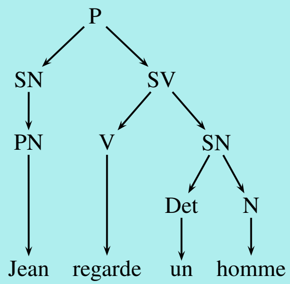
Type 2 grammaires hors-contextes (PDA) ⇒ polynomiale

Type 1 grammaires contextuelles (Context-sensitive grammars)
`uXv-->uyv` y séquence non vide. ⇒ exponentielle

Type 0 grammaires non contraintes (language récursivement énumérable) ⇒ indécidable

CFG

```
P  --> SN SV
SN --> NP
SN --> Det N
SV --> V SN
NP --> 'Jean'
V  --> 'regarde'
Det --> 'un'
N   --> 'homme'
```



Difficultés de gestion des accords (nombre, genre, . . .) : *Jean mangent une homme

Mais possible :

```
SN_masc_sing --> Det_masc_sing N_masc_sing
SN_masc_plur --> Det_masc_plur N_masc_plur
```


Autres limites

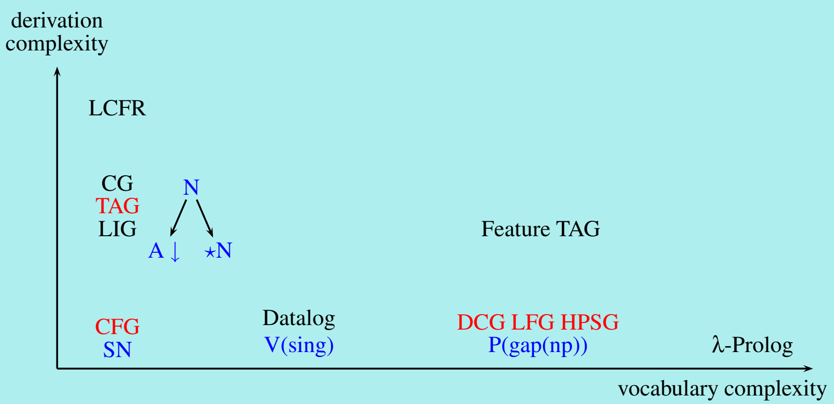
Les CFG manquent de puissance générative, par exemple pour les **croisements**

(1) omdat ik Cecillia de nijlpaarden zag voeren
because I Cecilia the hippopotamuses saw feed
because I saw Cecilia feed the hippopotamuses

Autres difficultés :

- dépendances non bornées, extraction (topicalisation, relative, . . .)
- ordre « libre » des mots (word scrambling)

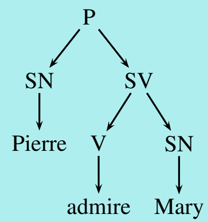
Nature des formalismes grammaticaux

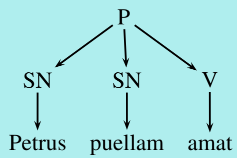


Lexical Functional Grammars [LFG]

Bresnam et Kaplan (1982) The mental representation of grammatical representation

Théorie : Associer structures de constituants (c-structures) et structures fonctionnelles (f-structure) :





$$\begin{bmatrix} \text{Suj} & \begin{bmatrix} \text{Pred} & \text{pierre} \end{bmatrix} \\ \text{Pred} & \text{admirer} \\ \text{Obj} & \begin{bmatrix} \text{Pred} & \text{mary} \end{bmatrix} \end{bmatrix} \qquad \begin{bmatrix} \text{Suj} & \begin{bmatrix} \text{Pred} & \text{pierre} \end{bmatrix} \\ \text{Pred} & \text{admirer} \\ \text{Obj} & \begin{bmatrix} \text{Pred} & \text{mary} \end{bmatrix} \end{bmatrix}$$

Formalisme : Structures de traits

$$\begin{bmatrix} \text{Det} & \begin{bmatrix} \text{Accord } \boxed{1} & \begin{bmatrix}\text{Num sing}\\ \text{Genr masc}\end{bmatrix} \\ \text{Cat Det} & \end{bmatrix} \\ \text{Nom} & \begin{bmatrix}\text{Accord } \boxed{1} \\ \text{Cat N}\end{bmatrix}\end{bmatrix}$$

Unification de structures de traits

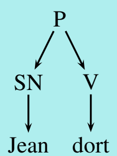
$$\begin{bmatrix} \text{Det} & \begin{bmatrix} \text{Accord} & \begin{bmatrix}\text{Num sing}\end{bmatrix}\end{bmatrix} \\ \text{Nom} & \begin{bmatrix}\text{Accord} & \begin{bmatrix}\text{Num sing}\end{bmatrix} \\ \text{Cat N} & \end{bmatrix}\end{bmatrix} \sqcup \begin{bmatrix}\text{Det} & \begin{bmatrix}\text{Accord} & \begin{bmatrix}\text{Genre Masc}\end{bmatrix}\end{bmatrix}\end{bmatrix} = \begin{bmatrix} \text{Det} & \begin{bmatrix} \text{Accord} & \begin{bmatrix}\text{Num sing}\\ \text{Genre masc}\end{bmatrix}\end{bmatrix} \\ \text{Nom} & \begin{bmatrix}\text{Accord} & \begin{bmatrix}\text{Num sing}\end{bmatrix} \\ \text{Cat N} & \end{bmatrix}\end{bmatrix}$$

$$\begin{bmatrix} \text{Det} & \begin{bmatrix} \text{Accord } \boxed{1} & \begin{bmatrix}\text{Num sing}\end{bmatrix}\end{bmatrix} \\ \text{Nom} & \begin{bmatrix}\text{Accord } \boxed{1} \\ \text{Cat N}\end{bmatrix}\end{bmatrix} \sqcup \begin{bmatrix}\text{Det} & \begin{bmatrix}\text{Accord} & \begin{bmatrix}\text{Genre Masc}\end{bmatrix}\end{bmatrix}\end{bmatrix} = \begin{bmatrix} \text{Det} & \begin{bmatrix} \text{Accord } \boxed{1} & \begin{bmatrix}\text{Num sing}\\ \text{Genre masc}\end{bmatrix}\end{bmatrix} \\ \text{Nom} & \begin{bmatrix}\text{Accord } \boxed{1} \\ \text{Cat N}\end{bmatrix}\end{bmatrix}$$

Productions et équations fonctionnelles

Une grammaire est donnée par des productions CFG décorées par des équations fonctionnelles.

P ⟶	SN	V	SN ⟶	Jean		V ⟶	dort
	(↑Suj)=↓	↑=↓		(↑Num)= sing			(↑Suj Num)= sing
				(↑Genre)= masc			(↑Suj Pers)= 3
				(↑Pred)= 'Jean'			(↑Mode)= indicatif
							(↑Pred)= 'dormir<Suj>'



$$\begin{bmatrix} \text{Suj} & \begin{bmatrix}\text{Pred 'Jean'}\\ \text{Genre masc}\\ \text{Num sing}\end{bmatrix} \\ \text{Pred 'dormir}\left\langle \text{Suj} \right\rangle\text{'} & \\ \text{Mode indicatif} & \end{bmatrix}$$

Variations sur les équations

Contraintes existentielles :

SN ⟶	(Det)	N	N ⟶	Jean	N ⟶	chien	Det ⟶	le
	↑=↓	↑=↓		∼(↑Det)		(↑Det)		(↑Det)=le

Equations « contraintes »

P' ⟶	SN	P
	(↓Qu)$=_c$+	(↑Qu)=+
		↑=↓

Equations ensemblistes

SV ⟶	V	(SN)	(SP)*
	↑=↓	↑Obj=↓ ou ↑Ajout∋↓	↑Ajout∋↓

(Jean dort le matin. Jean mange le gateau Jean mange ce gateau avec Anne)

Fonctions grammaticales

Fonctions catégorisables : Sujet, Objet, Comp(létive), XComp (infinitives et participiales), Prep-Obj (compléments prépositionnels, obliques)

Fonctions non-catégorisables : XAjout, X étant la catégorie

- **Vcomp** Jean veut **partir à Rio**.
- **Acomp** Jean devient **fou**.
- **Ncomp** Ils ont élu Jean **président**
- **Vajout** **Partant en voyage**, Marie se prépare
- **Aajout** Paul est parti **content**
- **Prep-Obj** Paul ressemble **à Jean**

Sous-catégorisation

Le champ Pred indique les fonctions attendues par un mot.

- **manger** (↑Pred)='manger<Suj,Obj>'
- **donner** (↑Pred)='donner<Suj,Obj,A-Obj>'
- **falloir** (↑Pred)='falloir<Obj>Suj' et (↑Suj Form) $=_c$ il
- **vouloir** (↑Pred)='vouloir<Suj,Vcomp>' et (↑Suj)=(↑Vcomp Suj) Jean veut venir
- **proposer** (↑Pred)='proposer<Suj,A-Obj,Vcomp>' et (↑Vcomp Suj)=(↑Suj)/(↑A-Obj)
  Jean propose à Jean de venir
- **destruction** (↑Pred)='destruction<De-Obj,Par-Obj>' Destruction de la maison par les promoteurs

Sous-catégorisation (2)

SP ⟶	Prep	SN	SP ⟶	Prep	SV	Prep ⟶	à
	↑=↓	(↑Obj)=↓		↑=↓	(↑Vcomp)=↓		(↑Pcas)=A
					(↓Mode=inf		

SV ⟶	V	(SN)	(SP)*	(SP)*
	↑=↓	(↑Obj)=↓	(↑(↓PCas)-Obj)=↓	(↑(↓PCas)-Vcomp)=↓
			(↓Obj)	(↓Vcomp)

Dérivations lexicales : Analyse du passif

Certaines « transformations » syntaxiques analysables par des entrées lexicales dérivées (approche lexicale, alternative aux approches transformationnelles)

$$\textbf{Passif :}\quad \begin{array}{lcl} Suj & \Rightarrow & \emptyset/Par-Obj \\ Obj & \Rightarrow & Suj \\ Vinf & \Rightarrow & Vppart \end{array}$$

donner,V : ↑Pred = 'donner<Suj,Obj,A-Obj>'
↑Mode = inf

donné,V : ↑Pred = 'donner<∅/Par-Obj,Suj,A-Obj>'
↑Mode = ppart
↑Passif = +

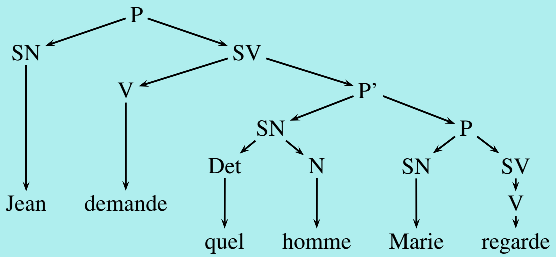
Non passivable : coûter : ↑Pred = 'couter<Suj,Ncomp>'

Extractions

P' $\longrightarrow$	SN	P
	$(\downarrow \text{Qu}) =_c +$	$\uparrow = \downarrow$
	$(\uparrow \text{Focus}) = \uparrow$	$(\downarrow \text{Qu}) = +$
	$(\uparrow \text{Focus}) = (\uparrow \text{Obj})$	

demande,V : $(\uparrow \text{Pred})$='demander<Suj,Comp>'
$(\uparrow \text{Comp Qu}) =_c +$

quel,Det : $(\uparrow \text{Det})$='quel'
$(\uparrow \text{Qu}) = +$



$$\begin{bmatrix} \text{Suj} & \begin{bmatrix}\text{Pred 'Jean'}\end{bmatrix} \\ \text{Pred} & \text{'demander}\langle\text{Suj,Comp}\rangle\text{'} \\ \text{Comp} & \begin{bmatrix} \text{Focus } \boxed{1} & \begin{bmatrix}\text{Pred 'homme'} \\ \text{Det 'quel'}\end{bmatrix} \\ \text{Suj} & \begin{bmatrix}\text{Pred 'Marie'}\end{bmatrix} \\ \text{Qu} & + \\ \text{Pred} & \text{'regarder}\langle\text{Suj,Obj}\rangle\text{'} \\ \text{Obj} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

dépendances non bornées

Nombre arbitraire d'enchâssements entre un constituant extrait et le prédicat dont il dépend :

Jean demande [quel homme Paul pense [que Marie regarde ε]]

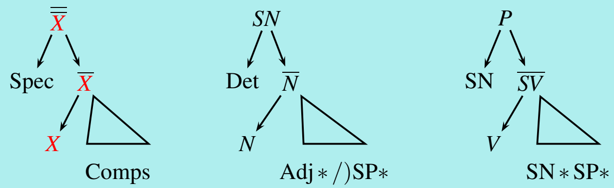
P' $\longrightarrow$	SN	P
	$(\downarrow \text{Qu}) =_c +$	$\uparrow = \downarrow$
	$(\uparrow \text{Focus}) = \uparrow$	$(\downarrow \text{Qu}) = +$
	$(\uparrow \text{Focus}) = \uparrow (\text{Comp})^\star \text{Obj}$	

$$\begin{bmatrix} \text{Suj} & \begin{bmatrix}\text{Pred 'Jean'}\end{bmatrix} \\ \text{Pred} & \text{'demander}\langle\text{Suj,Comp}\rangle\text{'} \\ \text{Comp} & \begin{bmatrix} \text{Focus } \boxed{1} & \begin{bmatrix}\text{Pred 'homme'} \\ \text{Det 'quel'}\end{bmatrix} \\ \text{Qu} & + \\ \text{Pred} & \text{'penser}\langle\text{Suj,Comp}\rangle\text{'} \\ \text{Suj} & \begin{bmatrix}\text{Pred 'Paul'}\end{bmatrix} \\ \text{Comp} & \begin{bmatrix} \text{Pred} & \text{'regarder}\langle\text{Suj,Obj}\rangle\text{'} \\ \text{Suj} & \begin{bmatrix}\text{Pred 'Marie'}\end{bmatrix} \\ \text{Obj} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Head-driven Phrase Structure Grammars [HPSG]

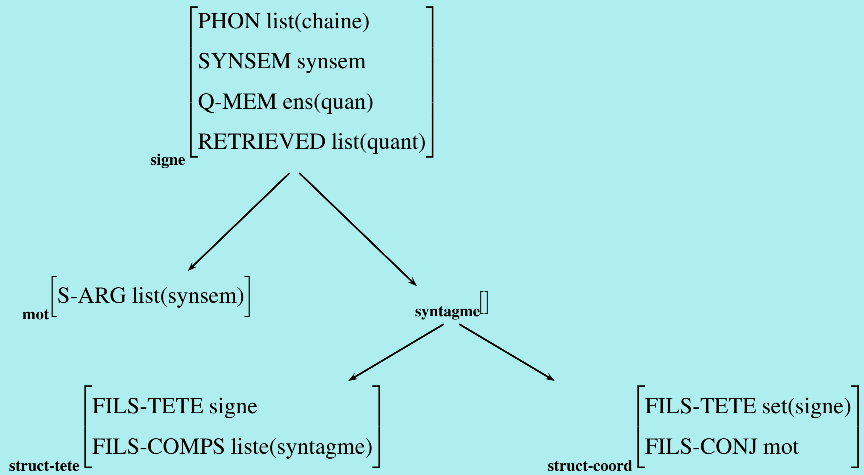
Pollard et Sag (1987) Information-based syntax and semantic

- Théorie linguistique complète : phonologie, lexical, syntaxe, sémantique et pragmatique
- Approche lexicale, avec importance de la sous-catégorisation
- Notion de tête linguistique (Théorie X-Bar)



- Théorie en évolution (HPSG III)
- S'appuie sur un formalisme puissant et homogène : structures de traits typées [TFS]
  Bob Carpenter (1992) The logic of typed feature structures
  Structures de traits possédant un type défini dans une hiérachie

Hiérachie de type



Hiérachie de type

Notation à la Prolog, utilisée dans ALE et DyALog.

```
bot sub [ string , list , cat , synsem].
    string escape symbol.
    cat sub [ s , np, vp, det , n ].
        s sub []. np sub []. vp sub [].
        det sub []. n sub [].
    synsem sub [phrase , lexeme] intro [ cat : cat ] .
        phrase sub [ root ] intro [ args : list ] .
            root sub [] intro [ cat : s ].
        lexeme sub [] intro [ orth : string ] .
    list sub [ ne_list , e_list ].
        ne_list sub [] intro [ hd:bot , tl : list ].
        e_list sub [].
```

- Un type peut introduire des sous-types
- Un type peut introduire ou préciser un trait et le type de sa valeur
- héritage multiple possible
- les types introduisant un trait f ont un type plus général

Architecture HPSG

- entrées lexicales
  exploitant la notion de sous-catégorisation
- règles lexicales de dérivation de nouvelles entrées (ex. le passif)
- des schémas de construction des constituants
- des principes de bonne construction

Nom

$$\begin{bmatrix} \text{PHON} & \langle \text{livre} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & {}_{\textbf{nom}}\begin{bmatrix} \text{TETE} & \begin{bmatrix} \text{MAJ} & \text{N} \\ \text{NFORM} & \text{ref} \end{bmatrix} \\ \text{VALENCE} & \begin{bmatrix} \text{SPR} & \text{Det} \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{INDEX}\ \boxed{1} & \begin{bmatrix} \text{GENR} & \text{masc} \\ \text{NOMB} & \text{sing} \end{bmatrix} \\ \text{RESTR} & \left\{ \begin{bmatrix} \text{QUANTS} & \langle\rangle \\ \text{NUCLEUS} & {}_{\textbf{livre}}\begin{bmatrix} \text{INST} & \boxed{1} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Article

$$\begin{bmatrix} \text{PHON} & \langle \text{un} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & {}_{\textbf{adj}}\begin{bmatrix} \text{TETE} & \begin{bmatrix} \text{MAJ} & \text{Det} \\ \text{SPEC} & \text{N'} : \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{CONT}\ \boxed{2} & \begin{bmatrix} \text{DET} & \text{existe} \\ \text{RESTIND}\ \boxed{1} & \begin{bmatrix} \text{GENR} & \text{masc} \\ \text{NOMB} & \text{sing} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{QMEM} & \boxed{2} \end{bmatrix}$$

Verbe

$$\begin{bmatrix} \text{PHON} & \langle \text{lit} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & {}_{\textbf{verbe}}\begin{bmatrix}\text{TETE} & \begin{bmatrix}\text{MAJ} & \text{V}\\ \text{VFORM} & \text{tensé}\end{bmatrix} \\ \text{VALENCE} & \begin{bmatrix}\text{SUJ} & \left\langle \boxed{1}\text{SN}_{\boxed{3}}[\text{sing},3]\right\rangle \\ \text{COMPS} & \langle\boxed{2}\text{SN}_{\boxed{4}}\rangle\end{bmatrix} \\ \text{S-ARG} & \langle\boxed{1},\boxed{2}\rangle\end{bmatrix} \\ \text{CONT} & \begin{bmatrix}\text{QUANTS} & \langle\rangle \\ \text{NUCLEUS} & {}_{\textbf{lire}}\begin{bmatrix}\text{LECTEUR} & \boxed{3}\\ \text{LU} & \boxed{4}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Adjectif

$$\begin{bmatrix} \text{PHON} & \langle \text{rouge} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix}\text{TETE} & {}_{\textbf{adj}}\begin{bmatrix}\text{MAJ} & \text{Adj}\\ \text{PRD} & -\\ \text{MOD} & \begin{bmatrix}\text{N' :} \begin{bmatrix}\text{INDEX} \boxed{1}\\ \text{RESTR} \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix} \\ \text{VALENCE} & \langle\rangle\end{bmatrix} \\ \text{CONT} & \begin{bmatrix}\text{INDEX} & \boxed{1}\begin{bmatrix}\text{NOMB} & \text{sing}\end{bmatrix} \\ \text{RESTR} & \left\{\begin{bmatrix}\text{QUANTS} & \langle\rangle \\ \text{NUCLEUS} & {}_{\textbf{rouge}}\begin{bmatrix}\text{ARG} & \boxed{1}\end{bmatrix}\end{bmatrix}\right\} \cup \boxed{2}\end{bmatrix}\end{bmatrix}\end{bmatrix}\end{bmatrix}$$

Fragment de phrase (Syntagme Verbal)

$$\begin{bmatrix} \text{PHON} \left\langle \text{lit un livre} \right\rangle \\ \text{SYNSEM} \begin{bmatrix} \text{LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{TETE} \boxed{V} \\ \text{VALENCE} \begin{bmatrix} \text{SUJ} \boxed{S} \left\langle \boxed{1}\text{SN}_{\boxed{3}}[\text{sing},3] \right\rangle \\ \text{COMP} \left\langle \boxed{1} \right\rangle \end{bmatrix} \\ \text{S-ARG} \left\langle \boxed{1} \right\rangle \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \text{QUANTS} \left\langle \begin{bmatrix} \text{DET existe} \\ \text{RESTIND} \boxed{4} \begin{bmatrix} \text{INDEX} \boxed{2}[\text{masc,sing}] \\ \text{RESTR} \left\{ {}_{\textbf{livre}}\begin{bmatrix} \text{QUANTS} \langle\rangle \\ \text{NUCLEUS} [\boxed{2}] \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \right\rangle \\ \text{NUCLEUS} \boxed{sem} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{FILS-TETE} \boxed{verbe} \\ \text{FILS-COMP} \left\langle \boxed{5} \right\rangle \\ \text{Q-MEM} \langle\rangle \\ \text{Q-ACQUIT} \left\langle \begin{bmatrix} \text{DET existe} \\ \text{RESTIND} \boxed{4} \end{bmatrix} \right\rangle \end{bmatrix}$$

$$\boxed{verbe} \begin{bmatrix} \text{PHON} \langle \text{lit} \rangle \\ \text{SYNSEM} \mid \text{LOC} \begin{bmatrix} \text{CAT} \; {}_{\textbf{verbe}}\begin{bmatrix} \text{TETE} \boxed{V} \begin{bmatrix} \text{MAJ V} \\ \text{VFORM tensé} \end{bmatrix} \\ \text{VALENCE} \begin{bmatrix} \text{SUJ} \left\langle \boxed{1}\text{SN}_{\boxed{3}}[\text{sing},3] \right\rangle \\ \text{COMPS} \left\langle \boxed{5}\text{SN}_{\boxed{2}} \right\rangle \end{bmatrix} \\ \text{S-ARG} \left\langle \boxed{1},\boxed{5} \right\rangle \end{bmatrix} \\ \text{CONT} \begin{bmatrix} \text{QUANTS} \langle\rangle \\ \text{NUCLEUS} \boxed{sem} \; {}_{\textbf{lire}}\begin{bmatrix} \text{LECTEUR} \boxed{3} \\ \text{LU} \quad \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Schéma 1 : Constituants saturés

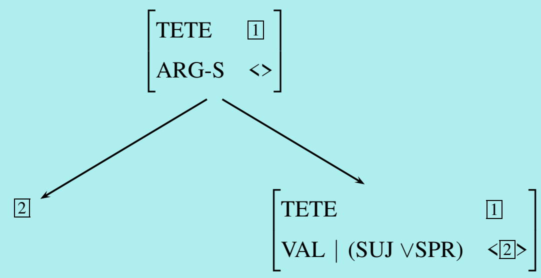


Schéma 2 : Constituants non spécifiés

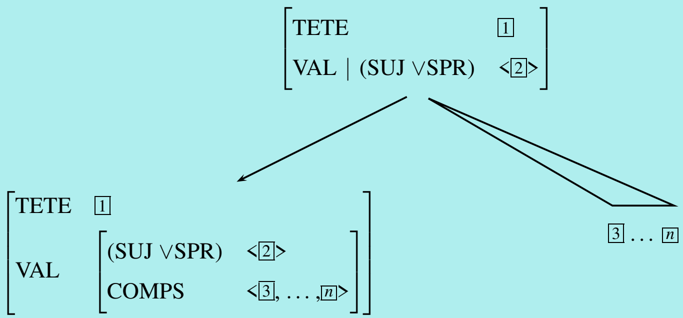


Schéma 3 : Constituants avec compléments catégorisés

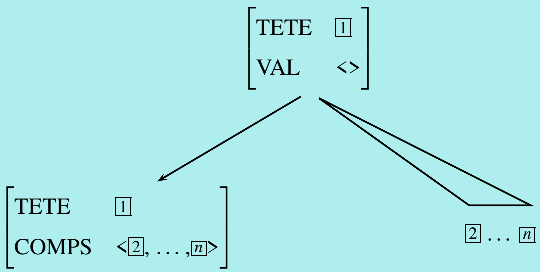
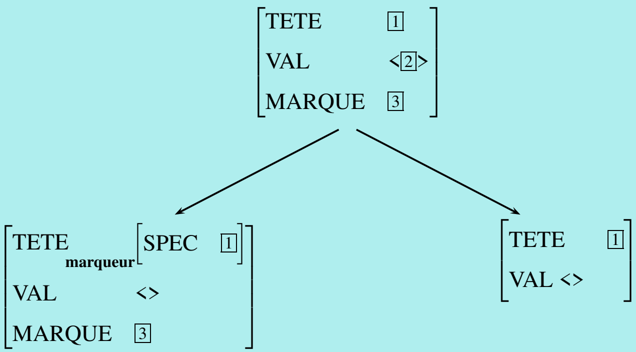
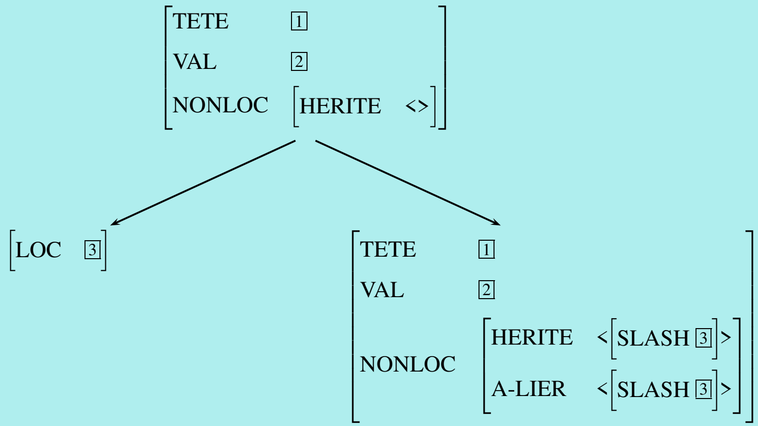


Schéma 4 : Constituants marqués



Permet le traitement des prépositions et de phrases telles que « [que Jean parte] m'inquiète ».

Schéma 5 : Extraction



$$[\text{Règle lexicale}] \begin{bmatrix} \text{COMPS} < \ldots, \boxed{1}, \ldots > \\ \text{HERITE} \mid \text{SLASH} <> \end{bmatrix} \Rightarrow \begin{bmatrix} \text{COMPS} < \ldots > \\ \text{HERITE} \mid \text{SLASH} <\boxed{1}> \end{bmatrix}$$

Les principes

Principe de tête Les valeurs des traits de tête entre un syntagme et sa tête sont identiques

Principe de valence La valence d'un syntagme est la valence de sa tête moins la liste des fils (non tête).

Principes des traits non locaux La valeur des traits non locaux hérités est l'union des valeurs des traits hérités des fils, moins la valeur à lier du syntagme racine.

Règle lexicale : le passif

Construction d'une entrée lexicale pour les verbes transitifs : manger ⇒ mangé

$$\begin{bmatrix} \text{PHON}\ \langle \text{p} \rangle \\ \text{SYNSEM} \mid \text{LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{TETE} \begin{bmatrix} \text{MAJ V} \end{bmatrix} \\ \text{VALENCE} \begin{bmatrix} \text{SUJ}\ \langle \boxed{1}\text{SN} \rangle \\ \text{COMPS}\ \langle \boxed{2}\text{SN}, \ldots \rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT}\ \boxed{3} \end{bmatrix} \end{bmatrix}_{\textbf{trans}}$$

$$\Rightarrow \begin{bmatrix} \text{PHON}\ \langle \text{p} \rangle \\ \text{SYNSEM} \mid \text{LOC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{TETE} \begin{bmatrix} \text{MAJ V} \\ \text{VFORM ppart} \end{bmatrix} \\ \text{VALENCE} \begin{bmatrix} \text{SUJ}\ \langle \boxed{2}\text{SN} \rangle \\ \text{COMPS}\ \left\langle \ldots, \text{SP} \begin{bmatrix} \text{PFORM=par}, \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT}\ \boxed{3} \end{bmatrix} \end{bmatrix}_{\textbf{intrans}}$$

Syntaxe ALE : entrée lexicale

```
persuade ---->                          % object  equi
  word,
  synsem:loc :( cat :( head :( verb ,
                          mod:none, vform:bse , aux:minus , inv :minus),
                     subcat :[( @ np(Ind1 )),
                            ( @ np(Ind2 ), @ case(acc )),
                            ( @ vp(VCont),
                             loc : cat :( head:vform: inf ,
                                        subcat :[( @ np(Ind2 ))]))],
                     marking:unmarked),
               cont :( nucleus :( persuading , persuader :Ind1 , persuaded:Ind2 , soa_arg : VCont),
                      quants :[]),
               conx:backgr: e_set ),
  ( @ empty_non_loc),
  qstore : e_set .
```


Syntaxe ALE : schéma

```
schema1 rule ( Mother,phrase , synsem:loc : cat : subcat :[]) ===>
cat > ( SubjDtr , non_word,synsem:SubjSyn),
cat > ( HeadDtr,phrase ),
goal > ( head_feature_principle (Mother,HeadDtr),
       inv_minus_principle (Mother),
       subcat_principle (Mother,HeadDtr,[SubjSyn ]),
       marking_principle (Mother,HeadDtr),
       spec_principle (SubjDtr , HeadDtr),
       semantics_principle (Mother,HeadDtr,[SubjDtr ]),
        parochial_trace_principle (SubjDtr ),
        nonlocal_feature_principle (Mother,HeadDtr,[SubjDtr ]),
        single_rel_constraint (Mother),
        clausal_rel_prohibition (Mother),
        relative_uniqueness_principle (Mother,[SubjDtr , HeadDtr]),
       conx_consistency_principle (Mother,[SubjDtr , HeadDtr ]),
        deictic_cindices_principle (Mother,[SubjDtr , HeadDtr ])).
```

Grammaires Catégorielles [CG]

Calcul de **Lambek** (1958)
Steedman (1998) Combinatory Categorial Grammars

- Théorie plus linguistique, non centrée sur une notion de constituant $\Rightarrow$ traitement de la coordination et constituants incomplet
- Formalisme : λ-calcul (et combinateurs)

Applications

$$\frac{X/Y : f \quad Y : a}{X : fa} > \qquad\qquad \frac{Y : a \quad X/Y : f}{X : fa} <$$

$$\begin{array}{ccc}
\text{Pierre} & \text{mange} & \text{une pomme} \\
\hline
\text{SN :pierre'} & (P\backslash SN)/SN : \lambda x.\lambda y.manger'xy & \text{SN :pomme'} \\
\hline
 & \multicolumn{2}{c}{(P\backslash SN) : \lambda y.manger'\,pomme'y \quad >} \\
\hline
\multicolumn{3}{c}{P : manger'\,pomme'\,pierre' \quad <}
\end{array}$$

- Puissance générative des CFG $\Rightarrow$ nécessaire d'ajouter des combinateurs supplémentaires
- Possibilité de décorer les non-terminaux avec des structures de traits

Coordination (simple)

$$\frac{X \quad CONJ \quad X}{X} < \Phi >$$

$$\begin{array}{ccccc}
\text{Sophie} & \text{achete} & \text{et} & \text{mange} & \text{une pomme} \\
\hline
\text{SN :s'} & (P\backslash SN)/SN : \lambda x.\lambda y.acheter'xy & \text{CONJ :et'} & (P\backslash SN)/SN : \lambda x.\lambda y.manger'xy & \text{SN :p'} \\
\hline
 & \multicolumn{3}{c}{(P\backslash SN)/SN : \lambda x.\lambda y.et'(acheter'xy)(manger'xy) \quad <\Phi>} & \\
\hline
 & \multicolumn{4}{c}{(P\backslash SN) : \lambda y.et'(acheter'p'y)(manger'p'y) \quad >} \\
\hline
\multicolumn{5}{c}{P : et'(acheter'p's')(manger'p's') \quad <}
\end{array}$$

Mais ne permet pas de traiter toutes les coordinations

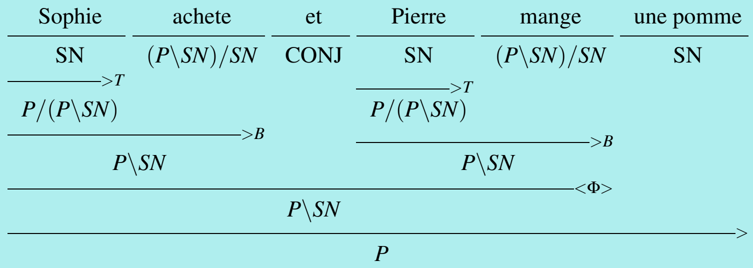
Composition

$$\frac{X/Y : f \quad Y/Z : g}{X/Z : \lambda x.f(gx)} > B$$

$$\begin{array}{cccccc}
\text{Sophie} & \text{achète} & \text{et} & \text{veut} & \text{manger} & \text{une pomme} \\
\hline
SN & (P\backslash SN)/SN & CONJ & (P\backslash SN)/SV & SV/SN & SN \\
\hline
 & & & \multicolumn{2}{c}{(P\backslash SN)/SN \quad >B} & \\
\hline
 & \multicolumn{4}{c}{(P\backslash SN)/SN \quad <\Phi>} & \\
\hline
 & \multicolumn{5}{c}{(P\backslash SN) \quad >} \\
\hline
\multicolumn{6}{c}{P \quad <}
\end{array}$$

Montée de type (type-raising)

$$\frac{SN : a}{T/(T\backslash SN) : \lambda f.fa} > T$$



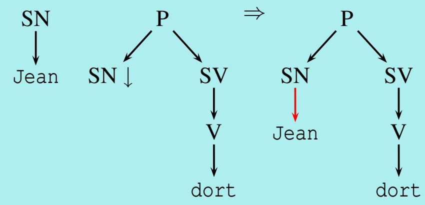
Restriction des montées de type.

Grammaires d'arbres adjoints [TAG]

Joshi (1975) Tree Adjunct Grammars

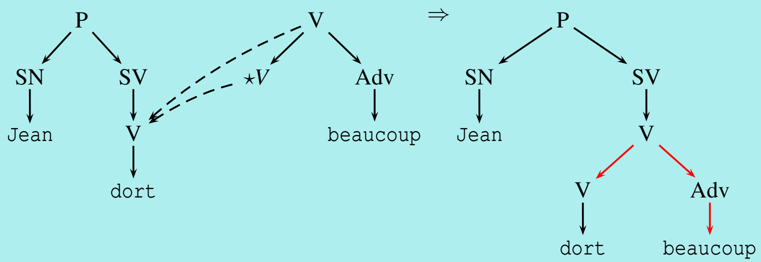
Théorie lexicale, catégorisation, tête, constituants

Formalisme manipulation d'arbres, avec les opérations de substitution et d'adjonction



Adjonction

Permet de gérer les ajouts



Principes de bonne formation

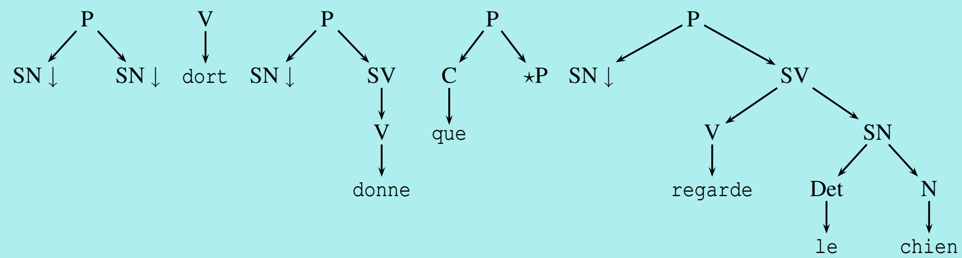
Pour tout arbre élémentaire :

ancrage lexical au moins une tête lexicale (non vide)

catégorisation un noeud pour chaque argument catégorisé par la tête (domaine de localité)

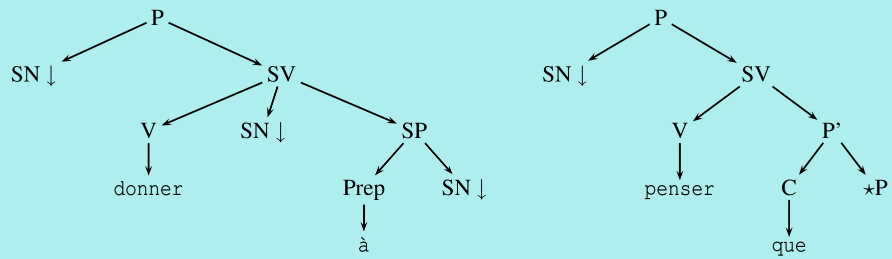
Consistence sémantique une association sémantique

non-composition une seule unité sémantique



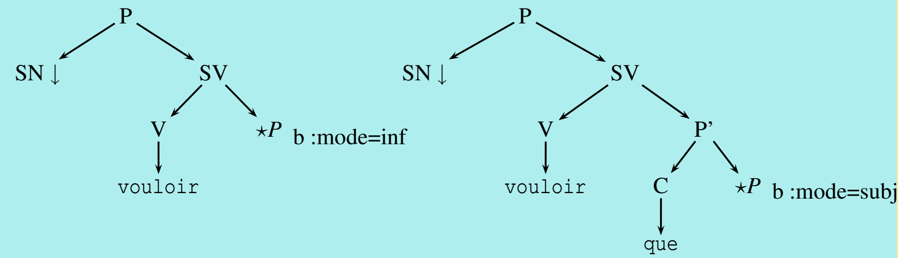
Catégorisation

Immédiate dans la structure des arbres



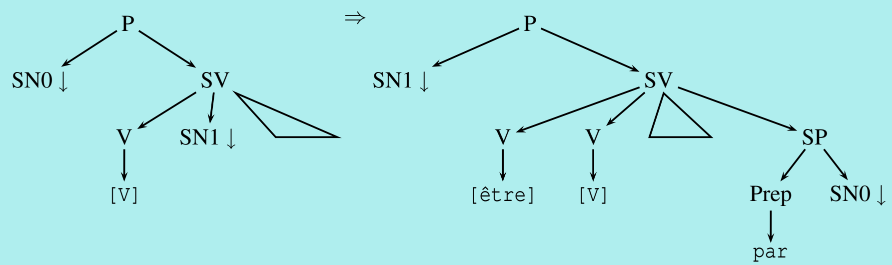
TAG avec traits

Structures de traits top et bot associables aux noeuds :



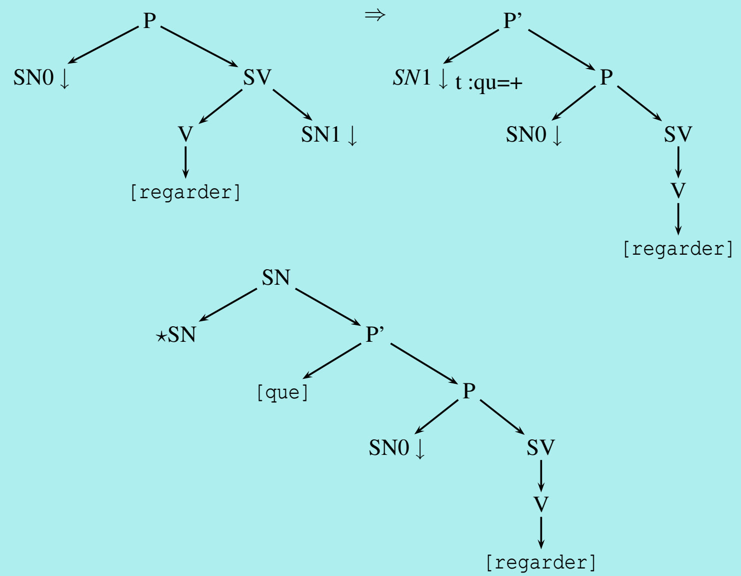
Familles

Transformations lexicales pour gérer le passif :



L'ensemble des arbres obtenus est une famille

Extraction



Dépendances non bornées

Une succession d'adjonctions peut séparer un constituant extrait de son prédicat tête.

Arbres de dérivation et sémantique

Formalismes TAG dérivés

TIG Adjonction réduite (non enveloppante)

Multi-Component Tag insertion par substitution et/ou adjonction d'un ensemble d'arbres élémentaires dans un arbre.

Dominance-Tree Quasi arbres avec relations parents et ancêtres

Formalismes faiblement dépendant du contexte (MCS)
$\Rightarrow$ complexité polynomiale.

Grammaires de dépendance [DG]

Théorie :

- Remise en cause de la notion de constituant et d'arborescence des constituants
- Relations (dépendances) entre mots ($\Rightarrow$ graphe de dépendance)
- Contraintes sur les dépendances

Bibliographie

Les nouvelles syntaxes – Grammaires d'unification et analyse du français Anne Abeillé (1993).

Une introduction à HPSG Philippe Blache
http://www.lpl.univ-aix.fr/~blache/papers/Intro-hpsg.ps.gz

The combinatory manifesto Mark Steedman (Draft, novembre 2000) et transparents associés.
http://www.cogsci.ed.ac.uk/~steedman/

Ressources HPSG http://lingo.stanford.edu/

Categorial Grammars Home Pages http://www.cs.man.ac.uk/ai/CG/

XTAG Group http://www.cis.upenn.edu/~xtag/

Dependency Grammars Home Page http://ufal.mff.cuni.cz/dg/dgmain.html

Les grammaires de dépendance Revue T.A.L., volume 41 (2000), Sylvain Kahane éditeur.
http://www.atala.org/tal/tal.html