

Corpus arborés et parsing

Cours 5

Santiago Herrera
s.herrera@parisnanterre.fr

RAPPEL :Grew

- Outil de réécriture de graphes
- Grew-match est seulement une partie de Grew
- Il sert à la transformation et modification de graphes, dont les arbres
 - Par exemple, à la transformation d'UD vers SUD et de SUD vers UD
- Il y a une version en ligne de commande, une autre version en python en cours de développement (mais qui marche déjà)
- Documentation disponible :

<https://web.grew.fr/>



Tutorial

Documentation

Graphs

Requests

Clustering

Command syntax

Rule syntax

GRS syntax

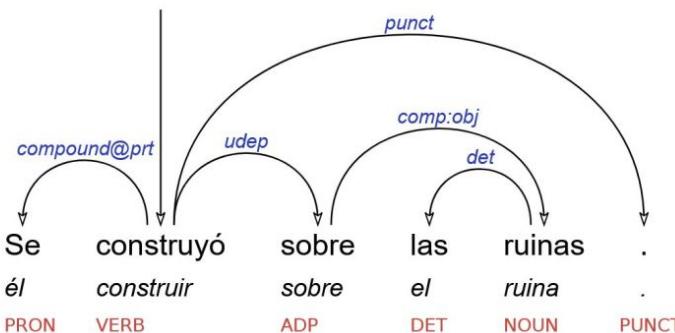
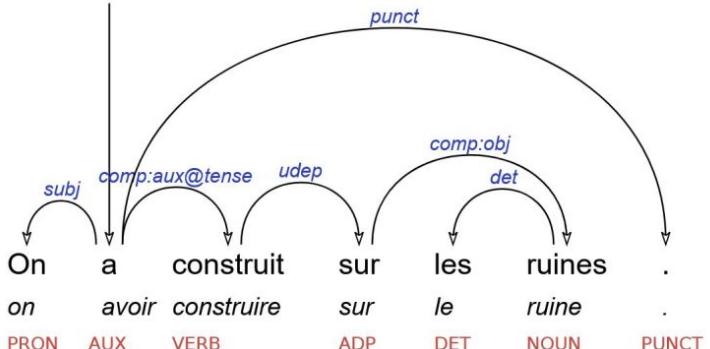
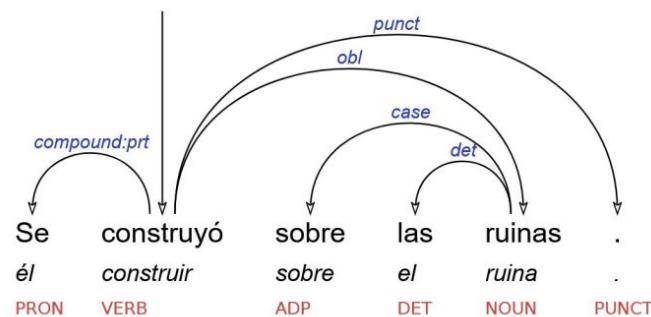
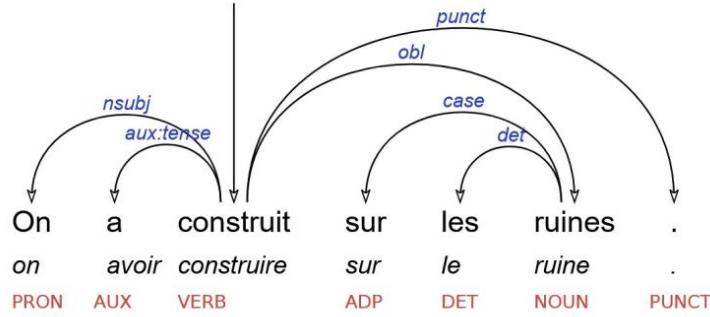
Rewriting

The 'Documentation' section is highlighted with a thin black border. The 'Tutorial' button has a yellow star icon and is also bordered in black. The 'Command syntax' and 'Rule syntax' items are enclosed in a rectangular box with a black border.

CORRECTION : Exos

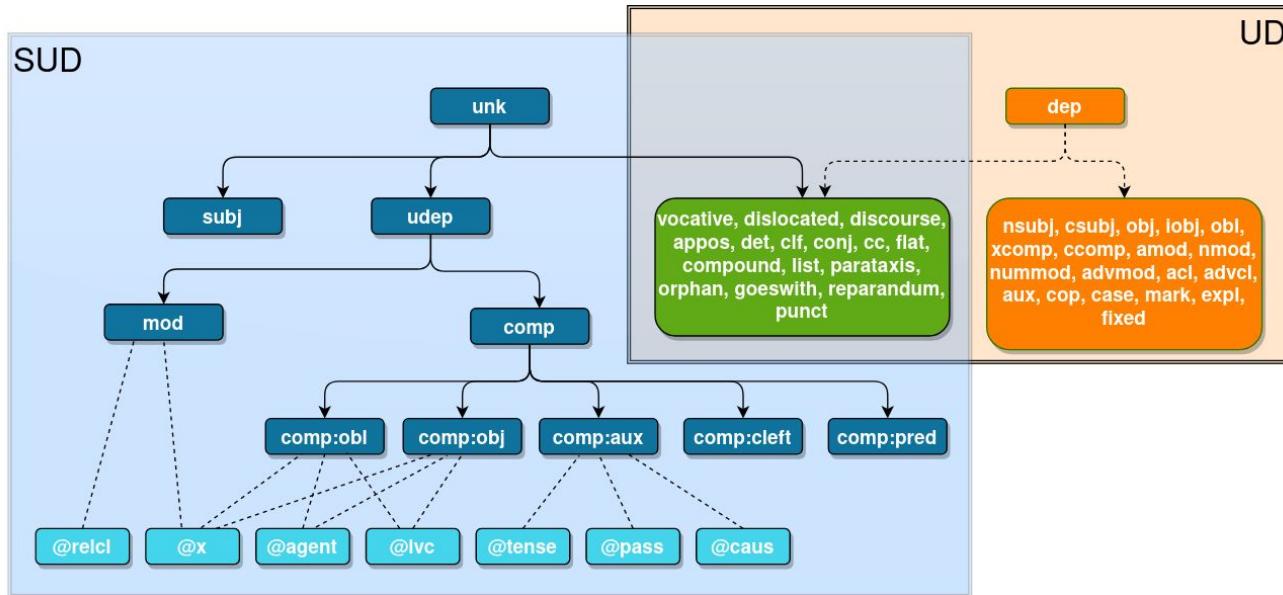
- Changez tous les étiquettes VERB pour V
- Ajoutez aux verbes transitifs un pair trait-valeur, Transitive=Yes
 - Faites pareil pour les verbes transitifs en ajoutant Transitive>No
- Transformez le trait ‘Gender’ de noms dans ‘Genre’, en gardant les mêmes valeurs.
- Transformez les déterminants contractés ‘au’ dans deux nœuds avec les formes ‘à’ et ‘le’
 - Ajoutez les étiquettes UPOS respectives (ADP, DET)
 - Changez les formes des nœuds à ‘à’ et ‘le’.
 - Ajoutez une relation ‘comp:obj’ de la préposition à la tête du groupe nominal et une relation ‘det’ de cette tête au déterminant.

Différences dans l'analyse



Surface Syntactic UD

- Alternative à l'UD
- Basés sur des critères distributionnels
- Les **relations sont définies sur des bases distributionnelles et fonctionnelles.**



Différences entre UD et SUD (explorer les treebanks)

- Regardez comment est-il annoté la coordination dans SUD et dans UD
(visualisation du corpus parallèles)
- Observez comment sont annotés les groupes prépositionnels dans SUD et dans UD
- Comparez les nombres de relations syntaxiques dans SUD et dans UD.
- Trouvez la tête de subordonnées dans SUD et UD
- Et les relatives ?
- Entre autres, différences.

Définition de tête distributionnelle

Critère distributionnel avec effacement positif

- Si A est autonomisable (et donc B est effaçable) et U et A ont la même distribution, alors A est la tête de U

U	
Demain	matin
A	B

**U et A ont la même distribution
→ A est la tête de U**

Définition de tête distributionnelle

Critère distributionnel avec effacement négatif

- Si A est autonomisable, et U et A n'ont pas la même distribution, alors B est probablement la tête de U

	U
Marie	dormait
A	B

A est autonomisable, B non.
U et A n'ont pas la même distribution
→ **B est la tête de U**

	U
à	Claude
A	B

B est autonomisable, A non.
U et B n'ont pas la même distribution
→ **A est la tête de U**

Définition de tête distributionnelle

Critère distributionnel sans effacement

- A commute avec A', et U et U' (A'+B) n'ont pas la même distribution, alors A est probablement la tête de U

_____U_____
à Claude
A _B_

_____U_____
par Claude
A _B_

**U and U' n'ont pas même distribution
→ A est la tête de U**

Règle grew de réécriture

shift : changement de direction des arêtes

- shift A ==> B

Toutes les arêtes qui sortent de A, sont rédigérées pour sortir de B

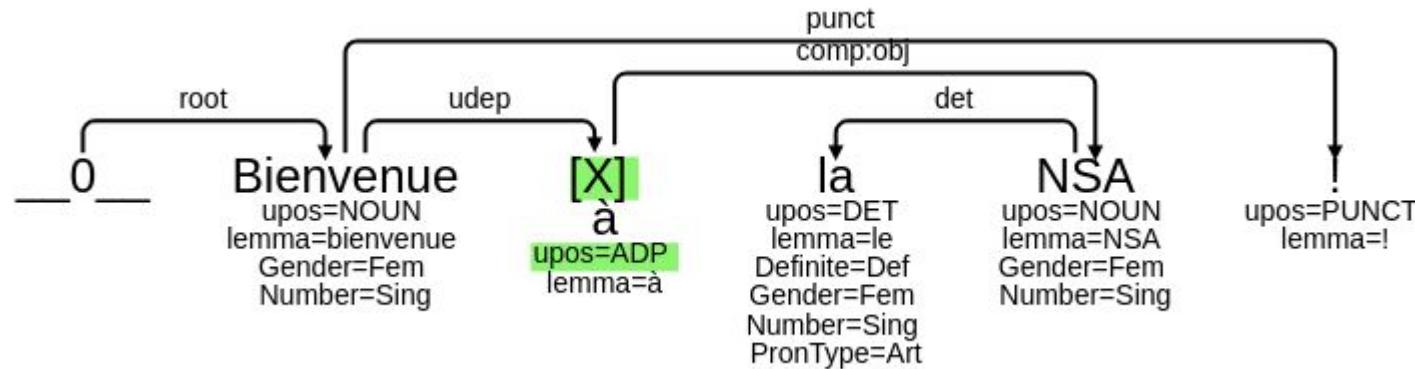
- shift_out B =[subj|obj]=> C

On déplace les arêtes avec les relations subj et obj qui sortent de B pour sortir de C

- shift_in C =[subj|obj]=> D

Les arêtes qui arrivent dans C avec les relations subj et obj sont déplacées pour arriver à D

exo Grew



Convertir l'annotation SUD vers UD : déplacer la relation udep pour qu'elle arrive au nœud 'NSA'.

Principe de commutation

Exclusion mutuelle

Claude parle à voix haute

Pierre vient en voiture

Claude parle à Camille

Pierre vient demain

A et B commutent dans l'environnement X Y s'ils peuvent se remplacer l'un l'autre et qu'en plus ils s'excluent l'un l'autre.

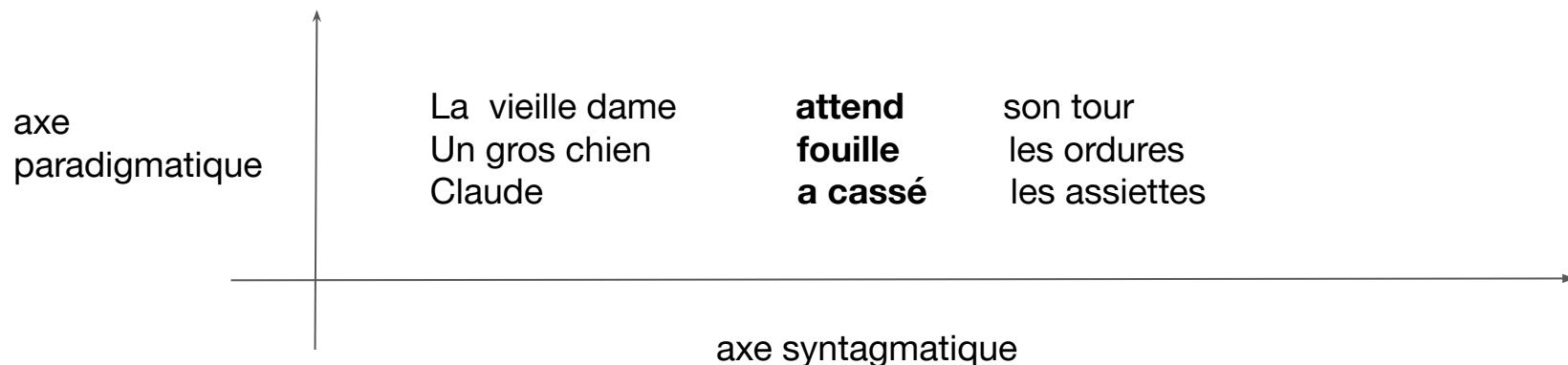
XAY et **XBY** sont acceptables

XABY et **XBAZ** sont inacceptables

Principe de commutation

Paradigme de commutation

L'ensemble de signes qui peuvent se commuter les uns avec les autres est appelé un **paradigme de commutation**



Distribution et classe distributionnelle

La **distribution** d'un signe est l'ensemble des environnements dans lequel ce signe peut apparaître.

La souris courr- -ait

L'oiseau chant- -e

Les enfants saut- -eront

Les élèves écri- -ent

Le vent souffl- -ait

Je veux un chien



Je veux manger

Je veux cela

Un ensemble de signes compatibles avec un ensemble donné d'environnements est une **classe distributionnelle**

Tests

Si le segment A commute avec une unité syntaxique N, alors A est potentiellement une unité syntaxique

- Test de commutation
 - Commutation avec un pronom
 - Repérage de la classe distributionnelle
- Test d'autonomisabilité illocutoire / prosodique
 - **Une partie d'un énoncé qui peut former un énoncé autonome sans modification de sens est une unité syntaxique**
 - Critère suffisant, mais pas nécessaire

Exo

Montrez, pour chaque paire d'exemples, que les mots en gras ont des paradigmes de commutation différents. À quelles parties du discours correspondent les deux emplois ? Justifiez vos réponses par des critères formels.

- La **danse** est une passion pour Marie. / Marie **danse** plusieurs fois par jour.
- Paul est un **bon** ami. / Paul a reçu un **bon** d'achat de cinq euros.
- **Son** ami est parti sans prévenir. / Cette enceinte produit un meilleur **son**.
- **Soit** tu récupères tes affaires, soit tu les jettes, mais elles doivent disparaître.
/ Il faut qu'il **soit** fou pour penser une telle chose.

Exo

Donner une liste des unités ou syntagmes qui peuvent commuter avec les unités en gras :

1. C'est sûrement **lui** qui viendra demain.
2. Pierre **a rangé** sa chambre.
3. **Il** n'est pas commun de se faire insulter par des animaux lors d'une visite au zoo.
4. C'était devenu dingue, **ils** juraiient tous.
5. **Elle** ne sort jamais sans

Curiosité

enfin c'est vraiment, euh, histoire de donner quelque chose.

et du coup, histoire de faire quelque chose, euh,
je me suis inscrit en licence d'espagnol, donc, à Paris 3.

Connexions syntaxiques

Il y a une **connexion** entre A et B si AB sont **une unité syntaxique**.

Marie achète des robes en France

Marie achète des robes en soie

Tests sur l'ordre des mots

Test de déplacement

- Plus le déplacement d'une unité dépendant d'une autre unité, plus il y a des chances que les deux unités soit connectées

Test d'insertion

- Si on ne peut pas insérer entre A et B d'unité qui ne soit pas connectée à A ou B, alors A et B sont probablement connectés

Exo

À l'aide de tests, vérifiez les connexions entre les segments. Justifiez votre réponse

Jeanne a acheté une glace à la fraise

Jeanne a acheté une glace à la plage

La faible déclivité de la vallée de la Seine en Ile-de-France

Marie regardait Pierre

Le chat noir dort paisiblement sur le vieux canapé en cuir

Annotation en ligne : Arborator

- <https://arboratorgrew.elizia.net>
 - Projet = M1_treebanks_exos_2023
 - Besoin d'un ordinateur (tablette ?)
-
- Exos avec une correction automatique
 - Quelques exos notés à venir

