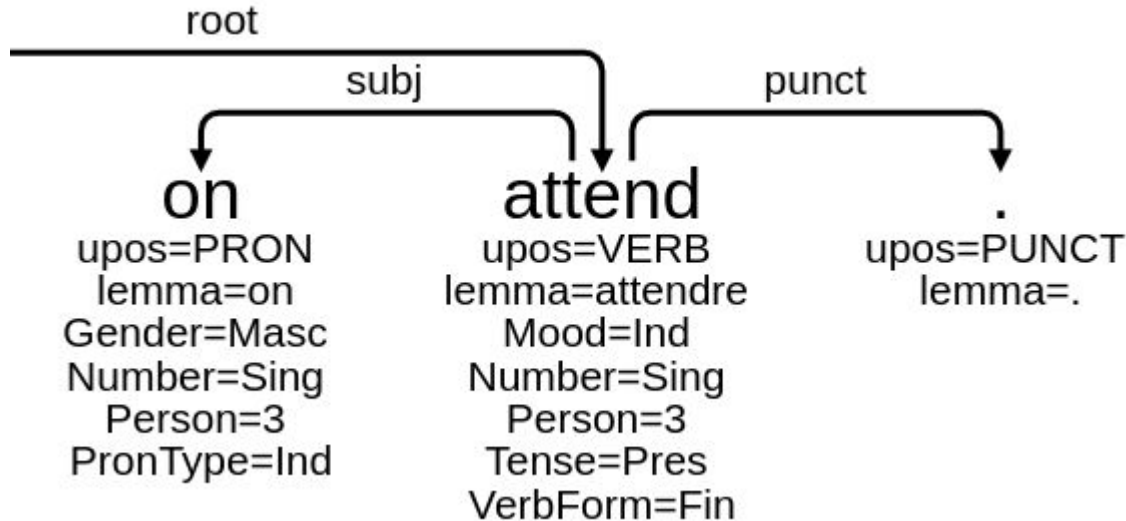


# Corpus arborés et parsing

## Cours 3

**Santiago Herrera**  
[s.herrera@parisnanterre.fr](mailto:s.herrera@parisnanterre.fr)

## Rappel : Grew-match



```
pattern { X -[subj]->Y }
```

```
pattern { X [form="attend"] }
```

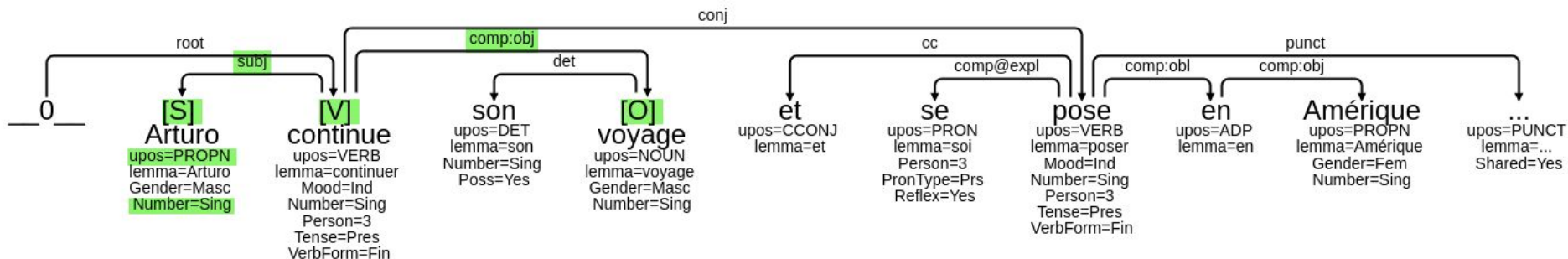
```
pattern { X [upos=VERB] }
```

```
pattern { X [lemma="attendre"] }
```

```
pattern { X[Tense=Pres] }
```

# Rappel : plusieurs contraintes

```
pattern {  
  V - [subj] -> S;  
  V - [comp:obj] -> 0;  
  S [upos=PROPN, Number=Sing]  
}
```



# Syntaxe de Grew-match

pattern { V [lemma <> "manger"] }

pattern { V [upos=VERB | AUX] }

Le lemme est different à "see"

Le POS est soit un VERB soit un AUX

# Exos : À vous (correction ?)

Sur le treebank GSD en SUD du français (sauf indication)

- Cherchez tous les sujets (subj)
- Trouvez tous les lemmes (lemma) des auxiliaires (AUX)
- Cherchez toutes les phrases qui ont comme racine (root) un auxiliaire (AUX)
  - Réalisez la même requête en **UD**. Qu'est-ce qui se passe ?
  - Cherchez les phrases qui ont un auxiliaire **ou** un verbe comme racine
- Est-il possible d'avoir deux sujets (subj) sur le même verbe ?
- Est-il possible d'avoir deux objets (comp:obj) sur le même verbe ?

## Rappel : encore Grew-match

$N1 < N2$

Le nœud N1 est immédiatement avant N2

$N1 << N2$

Le nœud N1 est avant N2

$N1.feature = N2.feature$

Égalité entre les valeurs des features

$N1.feature <> N2.feature$

Inégalité entre les valeurs des features

$N1[upos=Verb, !Mood]$

Négation d'une feature

$without \{ N1[upos=NOUN] \}$

Application de contraintes plus générales

## À vous (correction ?)

- Y a-t-il des nœuds qui n'ont pas d'upos ?
- Y a-t-il des paires contigus DET-NOUN qui ne soit pas dans une relation déterminative (det)
- Trouvez
  - les paires DET-NOUN contigus qui ne s'accordent pas en genre.
  - les sujets qui se placent avant le verbe.
  - les phrases verbales sans sujet
- Comment est-il annoté le trigram “à partir de” ?
- Trouvez les verbes intransitifs
- Quels sont les adjectifs sans trait de genre ?

# Des structures de traits dans les dépendances

N1 -[comp:obj]-> N2

N1 -[1=comp, 2=obj]-> N2

N1 -[1=comp, !2, !deep]-> N2

N1 -[subj@expl]-> N2

N2 -[1=subj, deep=expl]-> N2

N2 -[deep=expl]-> N2

- De nouveau : Est-il possible d'avoir deux sujets (subj) sur le même verbe ?
- Comment sont-ils annotés les passives ?



# Clustering dans Grew-match

- **Clustering**

Clustering 1: ☐ No ☒ Key ☐ Whether

X.upos

Syntaxe : X.upos, X.Number, X.NumType, X.PronType

1 pattern { e:X->Y }

 e.label

- **Whether** (oui/non réponse)

Clustering 1: ☐ No ☐ Key ☒ Whether

1 X[upos=NOUN], X << Y

Syntaxe du pattern normal (sans 'pattern {}')

la réponse négative est similaire à un without

# Exploration : adjectif

- Comment sont-ils annotés les adjectifs en SUD ?
  - Trouvez tous les adjectifs
  - Recherchez les UPOS de gouverneurs.
  - Quel est le lien de dépendance qui les unit ?
  - Quels sont les lemmes des adjectifs qui se trouvent avant leur nom.
  - Pouvez-vous trouver une feature qui détermine l'antéposition de l'adjectif par rapport à son nom ?
  - Trouvez avec une seule requête la distribution de lemmes des adjectifs selon leur position (avant ou après) par rapport au nom

## À vous

- Trouvez tous les lemmes (lemma) des auxiliares (AUX)
- Trouvez les relations entre deux verbes
- Essayez les relations 1=subj et subj, quelle est la différence?
- Observez la distribution du genre sur les noms français, sur les noms allemands (**IMPORTANT : N'UTILISEZ PAS LE TREEBANK SUD\_German-HDT**)
- Observez la distribution du POS de la tête des relatives (mod@relcl)
- Vérifiez si le pronom relatif dépend ou non de la tête de la clause relative

# 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

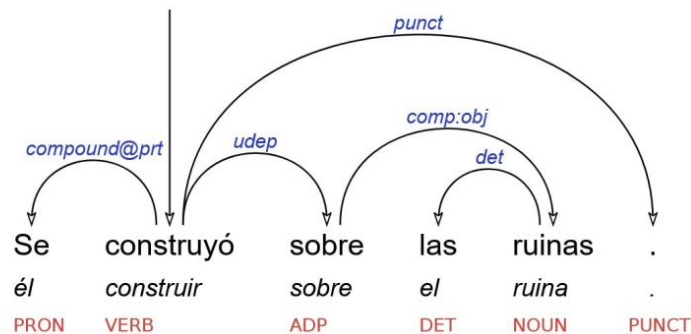
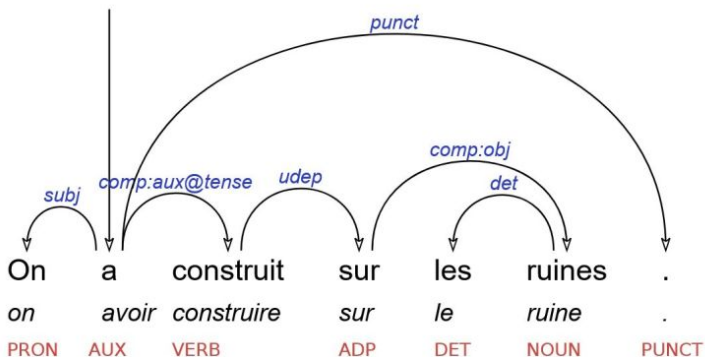
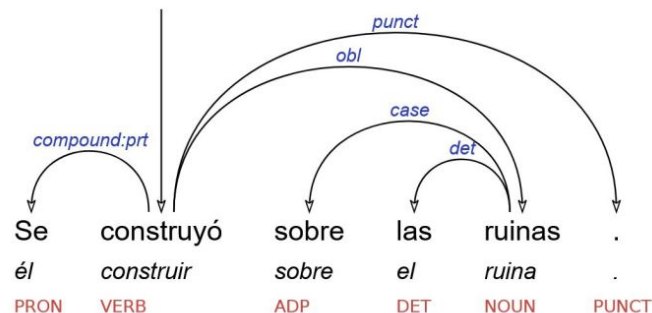
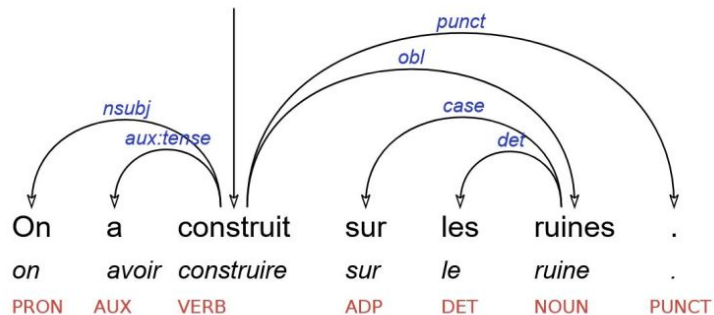
- A commute avec A', et U et U' (A'+B) n'ont pas la même distribution, alors A est probablement la tête de U

U	
à	Claude
A	B

U	
par	Claude
A	B

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

# Différences dans l'analyse



# Universal Dependencies

- “Tutorial on Universal Dependencies” (de Marneffe, Nivre, Zeeman).  
<https://github.com/UniDive/2023-unidive-webinar>

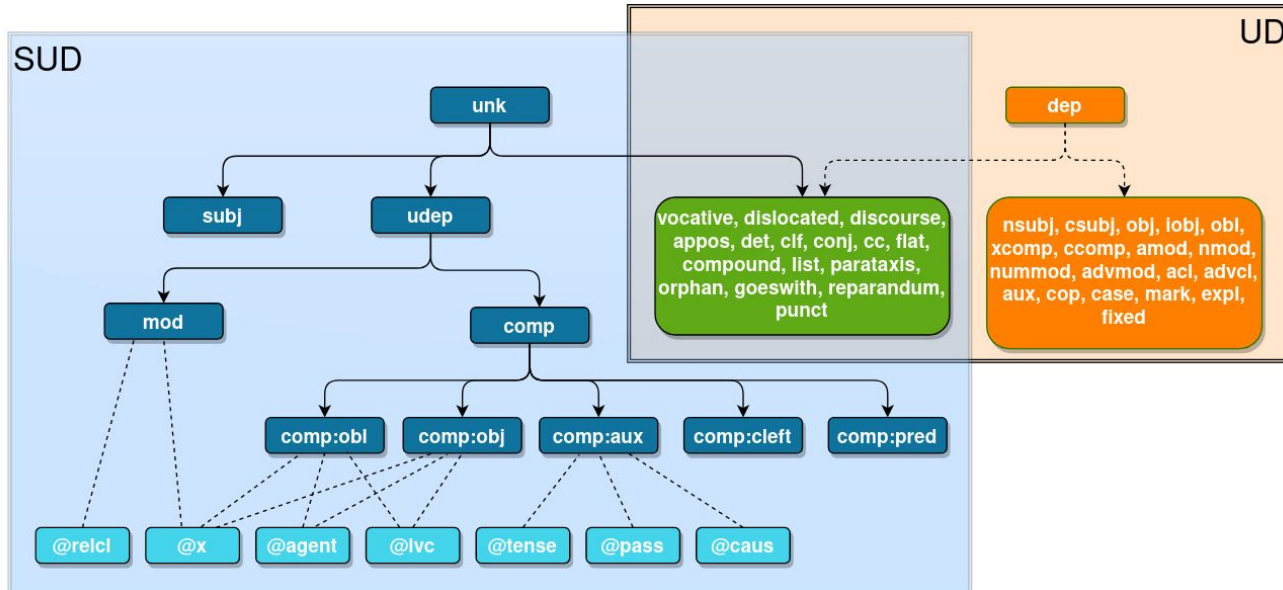
	Nominal	Clause	Modifier Word	Function Word
Core Predicate Dep	nsubj obj iobj	csubj ccomp xcomp		
Non-Core Predicate Dep	obl vocative expl dislocated	advcl	advmod* discourse	aux cop mark
Nominal Dep	nmod appos nummod	acl	amod	det clf case
Coordination	MWE	Loose	Special	Other
conj cc	fixed flat compound	parataxis list	orphan goeswith reparandum	punct root dep



# Surface Syntactic UD



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



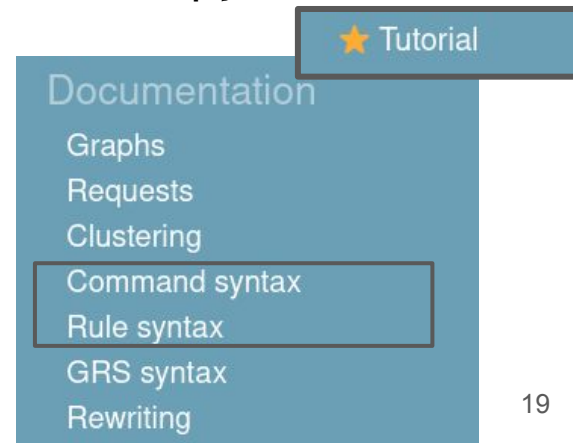
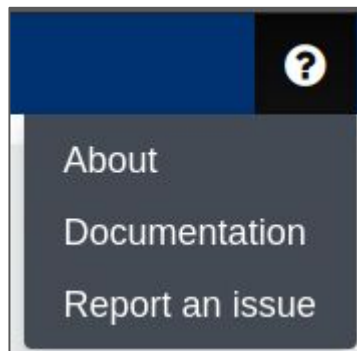
# Différences entre UD et SUD

- 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?
- Etc.

# 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/>



# Syntaxe de Grew

```
rule {  
  pattern {} Forme du motif qu'on cherche ✓  
  without {} Forme du motif qu'on cherche ✓  
  commands {} Transformations qu'on va réaliser sur les motifs requêtés  
}
```

Commands : add node X, del node X, add edge, del edge, del feat X.Feature, etc.

Doc : <https://grew.fr/doc/commands/>

# Règles de transformation

```
rule {  
    pattern {X[form=et}  
    commands {X.upos = CCONJ}  
}
```

Si le trait n'existe pas, on le crée.

# Règles de transformation

```
rule {  
    pattern {X[form=facile, Gender}  
    commands {del X.Gender}  
}
```

On élimine le trait de genre.

# Règles de transformation

```
rule {  
    pattern {e:X-[nsubj]->Y}  
    commands {del_edge e; add_edge X-[subj]->Y}  
}
```

On transforme l'arête nsubj entre X et Y dans subj. Il s'agit d'une règle pour passer de UD à SUD

# Règles de transformation

```
rule {  
    pattern {X[upos=NOUN]}  
    commands {add_node N:>X, X.form="empty_node"}  
}
```

On ajoute un nœud “vide” juste après

`add_node N :> X`

Juste après

`add_node N :< X`

Juste avant




# Grew

- Vous pouvez télécharger les résultats de requêtes et les utiliser sur Grew web.

More than 1000 results found in 0.29% of the corpus [0.710s]

Save 

TSV 

CoNLL 

No GRS loaded

1

GRS:

File

No GRS loaded

Browse



Dep

Close

Upload

# 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 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.
- Transformez le trait 'Gender' de noms dans 'Genre', en gardant les mêmes valeurs.