

Corpus arborés et parsing

Cours 3

Ioana-Madalina Silai
imsilai@parisnanterre.fr

Universal Dependencies

Guides d'annotation : <https://universaldependencies.org/guidelines.html>

“Tutorial on Universal Dependencies” (de Marneffe, Nivre, Zeeman).

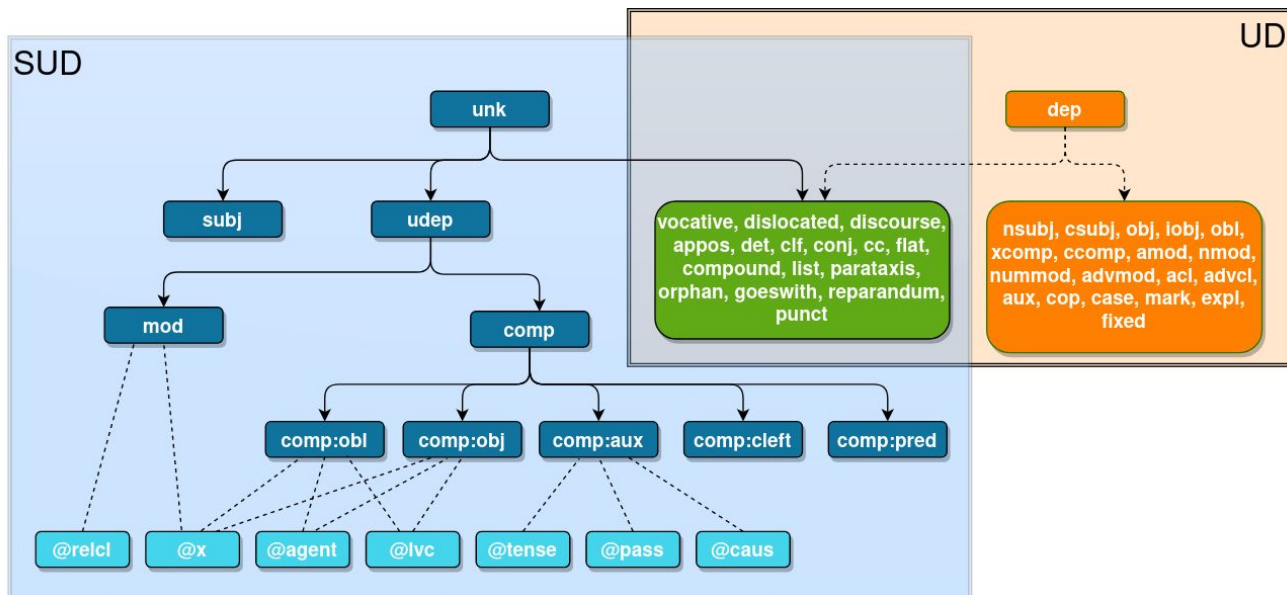
<https://github.com/UniDive/2023-unidive-webinar>

	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

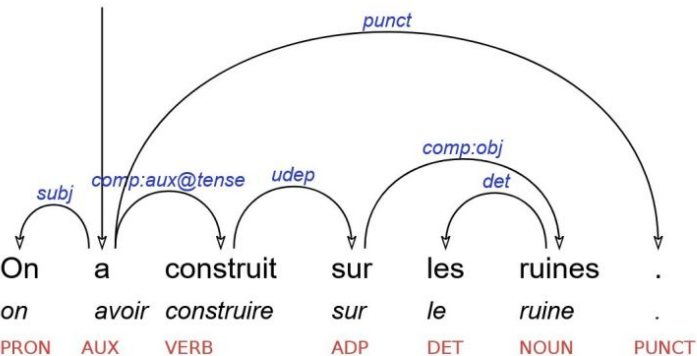
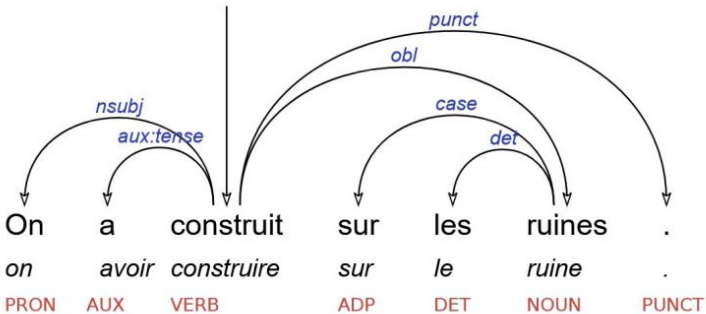
Surface Syntactic UD



- Alternative à l'UD
- Les **relations** sont définies sur des bases distributionnelles et fonctionnelles.
- Guide d'annotation : <https://guidelines.surfacesyntacticud.org/>



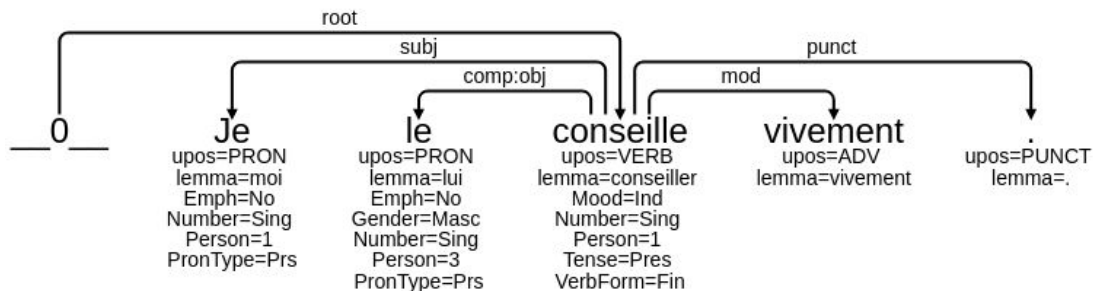
Différences dans l'analyse



Élément	UD	SUD	Difference
Racine	construit	a	UD choisit le verbe sémantique comme tête. SUD le verbe fini.
Relation verbe-aux	a -> construit (aux:tense)	construit -> a (comp:aux @tense)	Direction inversée. En UD l'aux dépend du verbe lexical; en SUD le participe de l'aux
Sujet	On -> construit (nsubj)	On -> a subj	Le sujet dépend du verbe lexical en UD, du verbe fini en SUD.
Prep	ruines -> sur case	construit	ud : dependant du nom sud : dependant verbe a l'infinitif
Nom dans le GP			

Rappel : treebanks aujourd'hui

- Format numérique, requêtable et encodé dans un conll-u



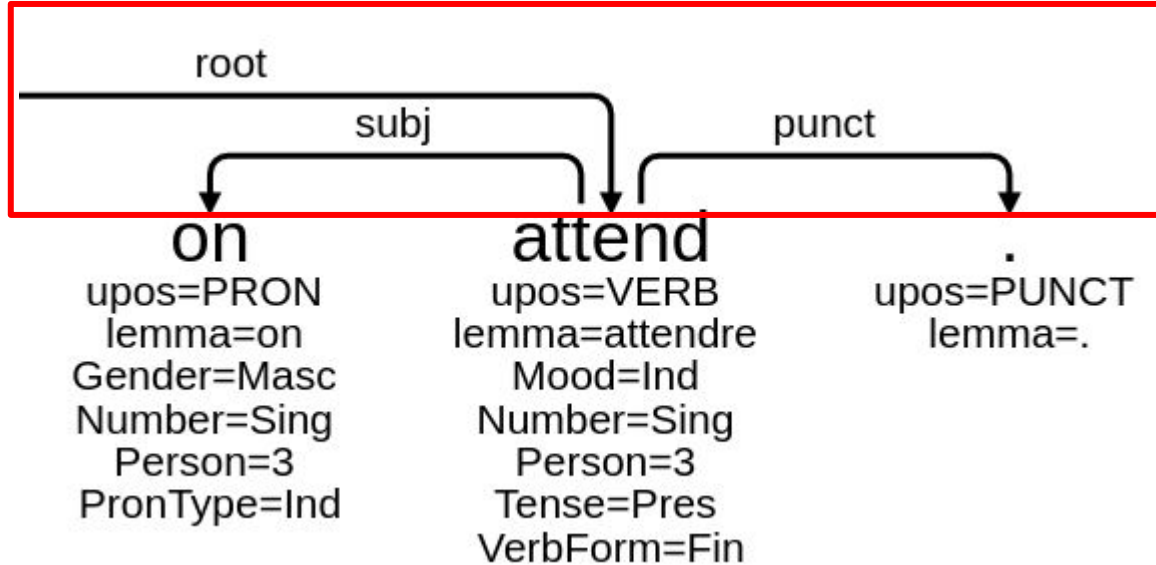
```
# global.columns = ID FORM LEMMA UPOS XPOS FEATS HEAD DEPREL DEPS MISC
```

```
# sent_id = fr-ud-train_06412
```

```
# text = Je le conseille vivement.
```

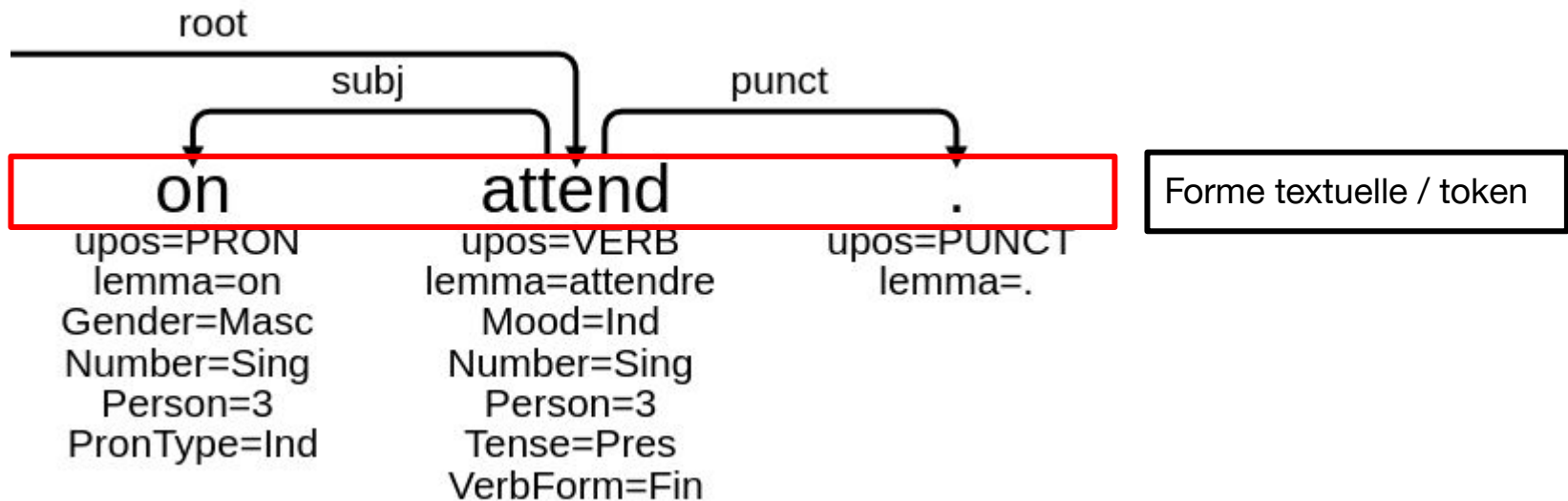
ID	FORM	LEMMA	UPOS	XPOS	FEATS	HEAD	DEPREL	DEPS	MISC
1	Je	moi	PRON		Emph=No Number=Sing Person=1 PronType=Prs	3	subj		wordform=je
2	le	lui	PRON		Emph=No Gender=Masc Number=Sing Person=3 PronType=Prs	3	comp:obj		
3	conseille	conseiller	VERB		Mood=Ind Number=Sing Person=1 Tense=Pres VerbForm=Fin	0	root		
4	vivement	vivement	ADV			3	mod		SpaceAfter=No
5	.	.	PUNCT			3	punct		

Couches d'annotation

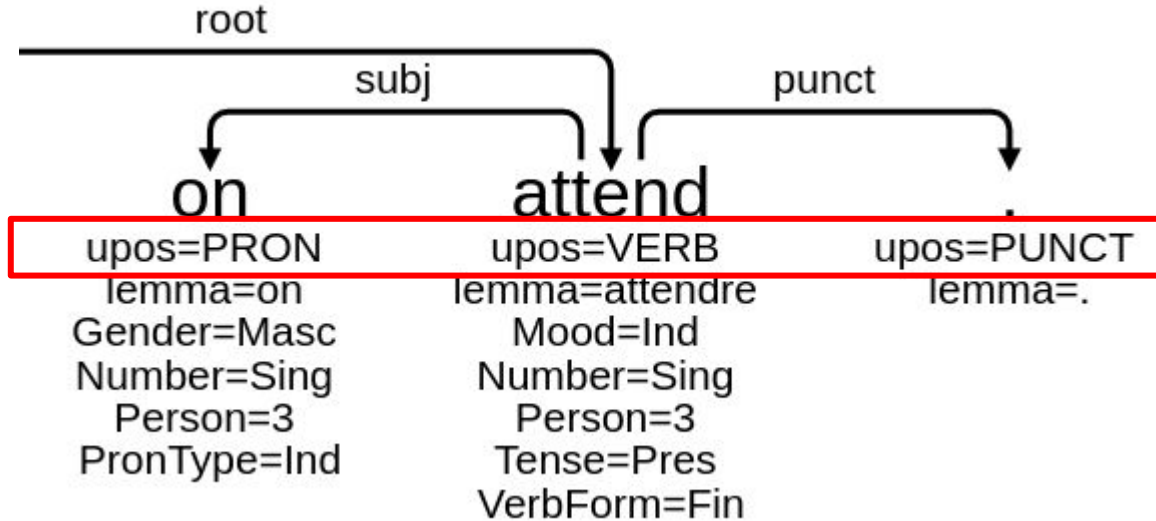


Dépendances syntaxiques

Couches d'annotation

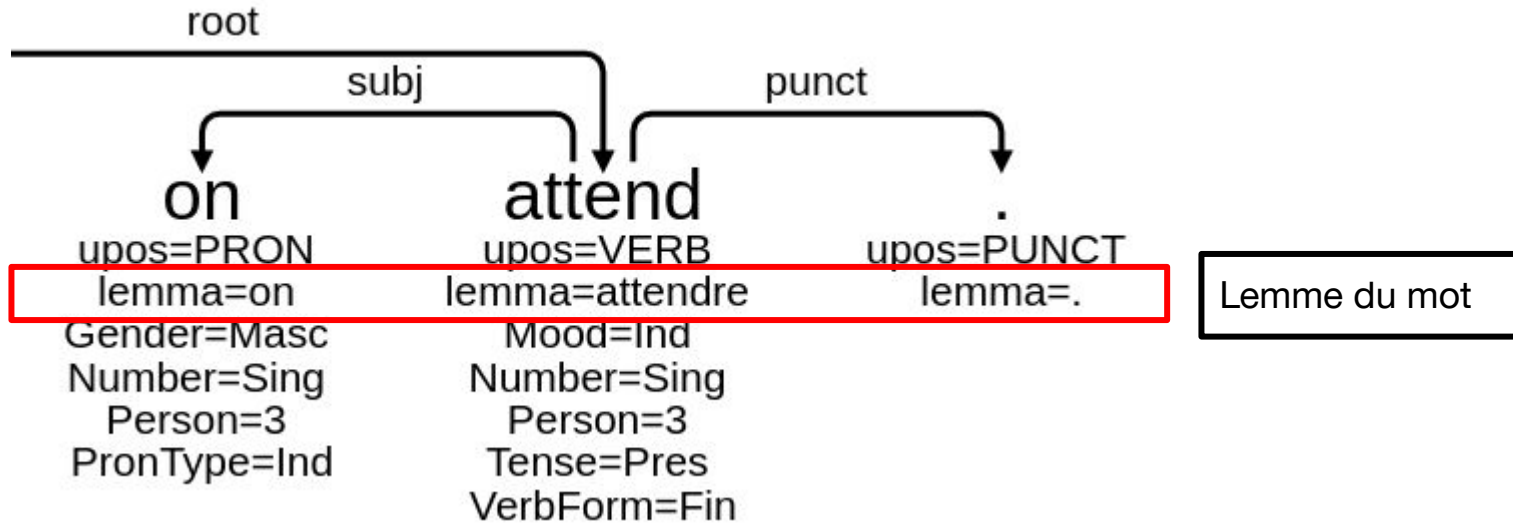


Couches d'annotation

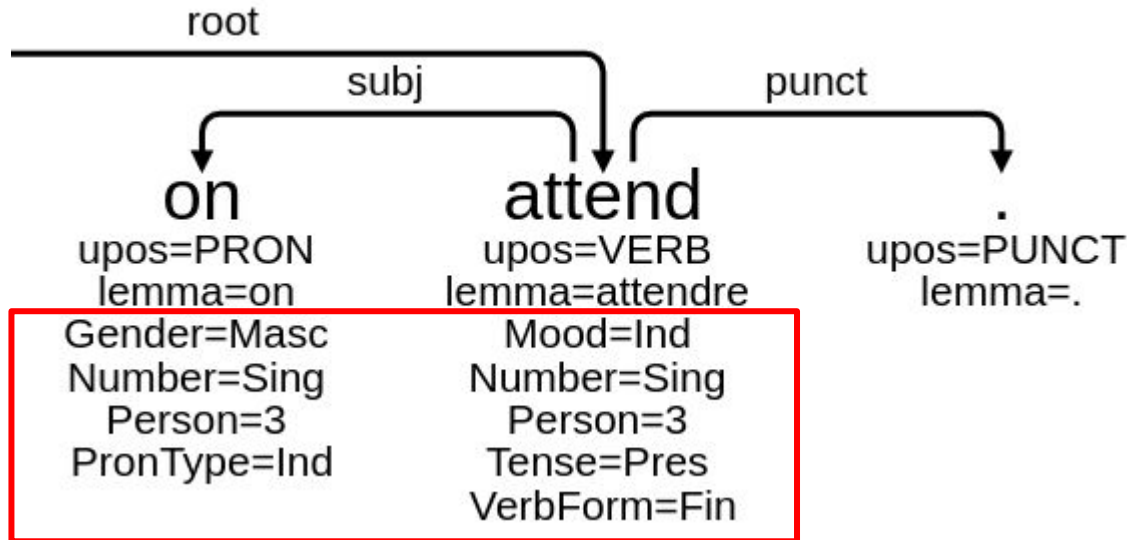


Universal Part of Speech

Couches d'annotation



Couches d'annotation



Annotation
morphosyntaxique /
Traits
morphologiques

Grew-match : outil de requêtage et d'exploration

<https://match.grew.fr>



Grew-match was developed by Bruno Guillaume in the Sémagramme team at LORIA / Inria Nancy Grand-Est.

- outil permettant de formuler des requêtes dans de nombreux corpus
- interface web
- requêtes formulées suivant une syntaxe particulière :
 - expressions sur les noeuds (= mots/tokens)
 - expressions sur les arêtes (= dépendances)
 - contraintes supplémentaires

Expressions concernant les noeuds

- Chaque noeud (=mot) que l'on souhaite repérer doit avoir un identifieur
- L'identifieur est accompagné par des contraintes sur la nature du noeud

{ pattern X [upos=VERB] }

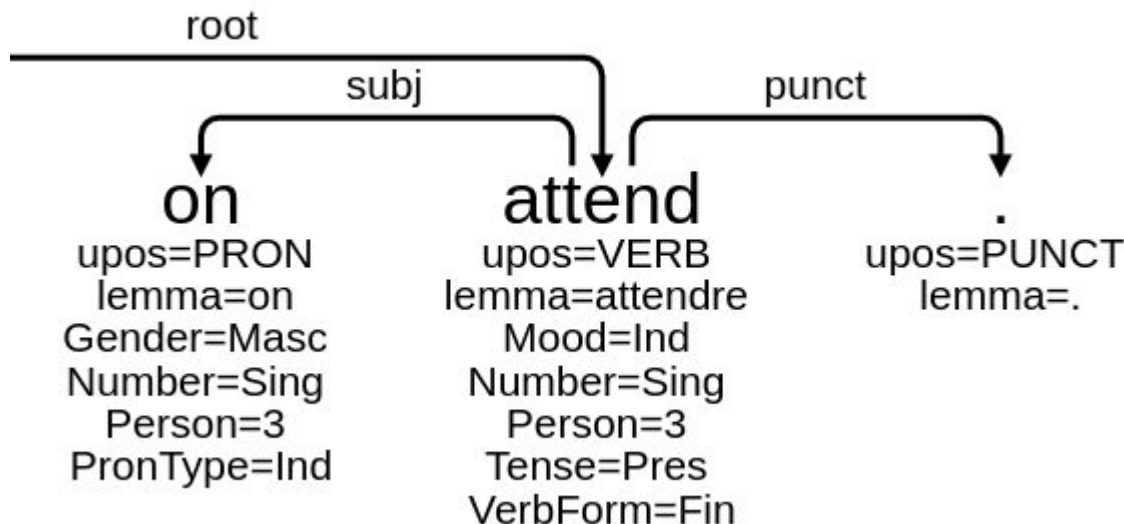
Expressions concernant les arêtes

- Elles expriment l'existence d'une arête (= dépendance) entre deux noeuds (= mots)
- Optionnellement, elles expriment aussi des contraintes concernant l'arête

{ pattern X -> Y }

{ pattern X -[det]-> Y }

Grew-match : requêtes de base



pattern { X [form="attend"] }

pattern { X [upos=VERB] }

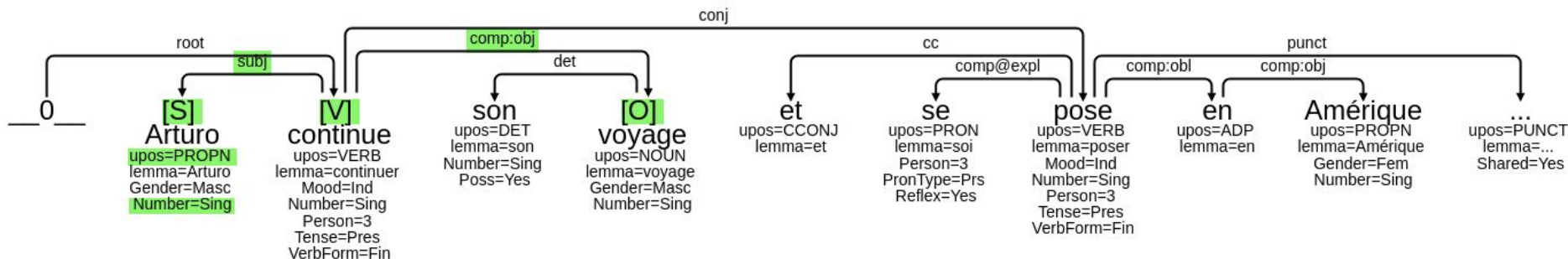
pattern { X [lemma="attendre"] }

pattern { X [Tense=Pres] }

pattern { X -[subj]->Y }

Plusieurs contraintes dans la même requête

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



A vous : testez et observez

pattern { V -[subj]-> S }

pattern { V [form="pensais"] }

pattern { V [lemma="penser"] }

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

pattern { V [upos=VERB] }

pattern { V [Tense=Past] }

pattern { V [upos=VERB | AUX] }

pattern { V [upos=VERB, lemma="penser"]; V -[subj]-> S; V -[comp:obj]-> O }

Corpus à utiliser :
SUD_French_GSD

Exos : À vous

W [root = AUX]

A MAITRISER POUR LE PARTIEL !!!!

Sur le même corpus

- Cherchez tous les sujets (subj) `pattern {X-[subj]->Y}`
- Cherchez toutes les phrases qui ont comme racine (root) un auxiliaire (AUX)
 - Réalisez la même requête en **UD**. Qu'est-ce qui se passe ? `pattern {X-[root]-> Y; Y [upos=AUX]}`
 - 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 ?
- Est-il possible d'avoir deux modifieurs (mod) sur le même verbe ?

https://grew.fr/grew_match/help/

D'autres contraintes

$N1 < N2$

Le mot N1 est immédiatement avant N2

$N1 << N2$

Le mot N1 est avant N2

$N1.feature = N2.feature$

La valeur du trait donné est la même pour N1 et N2

$N1.feature <> N2.feature$

La valeur du trait donné diffère entre N1 et N2

$N1[!Mood]$

Négation : le trait Mood n'est pas présent sur N1

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

Exclusion d'un patron plus général

Exos : À vous

- Y a-t-il des mots qui n'ont pas de trait **upos** ?
- Y a-t-il des paires contiguës **DET-NOUN** qui ne sont pas dans une relation déterminative (**det**) ?
- Trouvez
 - les paires **DET-NOUN** contiguës qui ne s'accordent pas en nombre (**Number**)
 - les sujets (**subj**) qui se trouvent avant le verbe (**VERB**).
 - 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 ?

Exos : À vous

- Y a-t-il des mots qui n'ont pas de trait **upos** ?
- Y a-t-il des paires contiguës **DET-NOUN** qui ne sont pas dans une relation déterminative (**det**) ?
- Trouvez
 - les paires **DET-NOUN** contiguës qui ne s'accordent pas en nombre (**Number**)
 - les sujets (**subj**) qui se trouvent avant le verbe (**VERB**).
 - 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 ?

Exos : À vous

- Y a-t-il des mots qui n'ont pas de trait **upos** ? `pattern {X [!upos]}`
- Y a-t-il des paires contiguës **DET-NOUN** qui ne sont pas dans une relation déterminative (**det**) ? `pattern { N1 < N2 ; N1 [upos = DET]; N2 [upos = NOUN] ; without {N2 -[det] -> N1}`
- Trouvez
 - `pattern {X [upos =DT]; Y[upos=NOUN]; X<Y, X. Number <> Y.Number}`
 - les paires **DET-NOUN** contiguës qui ne s'accordent pas en nombre (**Number**)
 - les sujets (**subj**) qui se trouvent avant le verbe (**VERB**). `pattern {X [upos=VERB]; X-[subj]-> Y; Y <<X}`
 - les phrases verbales sans sujet `pattern {X-[root]-> Y; Y[upos=VERB | AUX] ; without {Y -[subj]-> Z}`
- Comment est-il annoté le trigram "à partir de" ? `#verbeintransitif: pattern {X[upos=VERB] ; without {X-[comp:obj]->Y} ; without {X-[comp:obj]-> Z}`
- Trouvez les verbes intransitifs
- Quels sont les adjectifs sans trait de genre ?

`#trigram: pattern {X1 [form=« à »]; X2[form=« partir »]; X3 [form=« de »]; X1 < X2; X2 < X3}`

`#adj: 1. pattern {X[upos=ADJ, !Gender]}`
 2. `pattern {X[upos=ADJ]}`
`without{X[Gender]}`

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

N1 -[1=subj, deep=expl]-> N2

N1 -[deep=expl]-> N2

- De nouveau : Est-il possible d'avoir deux sujets (**subj**) sur le même verbe ?

Exos : À vous

- Y a-t-il des mots qui n'ont pas de trait **upos** ?
- Y a-t-il des paires contiguës **DET-NOUN** qui ne sont pas dans une relation déterminative (**det**) ?
- Trouvez
 - les paires **DET-NOUN** contiguës qui ne s'accordent pas en nombre (**Number**)
 - les sujets (**subj**) qui se trouvent avant le verbe (**VERB**).
 - 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 ?

Grouper et décompter les résultats dans Grew-match

Clustering

Clustering 1: ☐ No ☒ Key ☐ Whether

X.upos

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

Pour faire appel à l'étiquette d'une dépendance : e.label

(le patron de recherche doit contenir la définition de la dépendance :)

```
1 pattern { e:X->Y }
```


Exos : À vous

- Y a-t-il des mots qui n'ont pas de trait **upos** ?
- Y a-t-il des paires contiguës **DET-NOUN** qui ne sont pas dans une relation déterminative (**det**) ?
- Trouvez
 - les paires **DET-NOUN** contiguës qui ne s'accordent pas en nombre (**Number**)
 - les sujets (**subj**) qui se trouvent avant le verbe (**VERB**).
 - 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 ?

Grouper et décompter les résultats dans Grew-match

Whether:

Clustering 1: ☐ No ☐ Key ☒ Whether

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

Forme du patron habituel (sans 'pattern {}')

2 clusters :

- 1) occurrences qui correspondent au patron
- 2) occurrences qui ne correspondent pas

Exos : À vous

- Trouvez tous les lemmes (**lemma**) des auxiliaires (**AUX**)
- Trouvez toutes les relations de dépendance entre deux verbes (**VERB**)
- Essayez les relations **1=subj** et **subj**. Quelle est la différence?
- Observez la distribution du genre sur les noms français vs les noms allemands (**corpus allemand: SUD_German-GSD**)
- Observez la distribution du POS du **gouverneur** des relatives (**mod@relcl**)
- Observez la distribution du POS de la **tête** des relatives (**mod@relcl**)
- Vérifiez si le pronom relatif (**PronType=Rel**) dépend ou non de la tête de la clause relative

Exploration : adjectif

- Comment les adjectifs sont-ils annotés en SUD ?
 - Trouvez tous les adjectifs (**ADJ**)
 - Identifiez les upos des gouverneurs de tous les adjectifs
 - Quelles sont les dépendance qui relient les adjectifs à leur gouverneur ?
 - Quels sont les lemmes des adjectifs qui se trouvent avant le nom qui les gouverne?
 - Trouvez avec une seule requête la distribution des lemmes des adjectifs selon leur position (avant ou après) par rapport au nom