

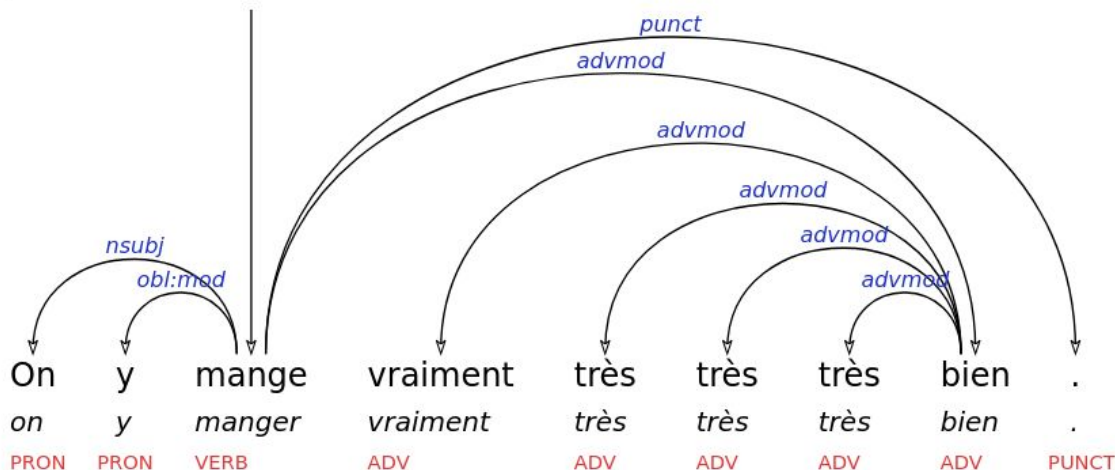
Corpus arborés et parsing

Cours 2

Ioana-Madalina Silai
imsilai@parisnanterre.fr

Syntaxe en dépendances

On y mange vraiment très très très bien.



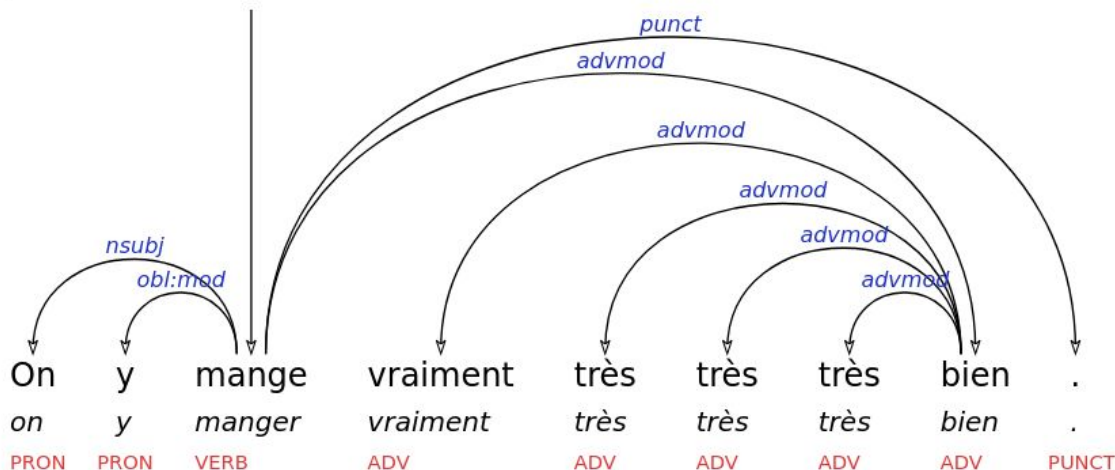
Propriétés principales

- structure **plate**
- racine = **verbe**
- **fonctions** annotées

racine = en générale un verbe, sinon tête

Syntaxe en dépendances

On y mange vraiment très très très bien.



Contraintes

- **racine unique**
- caractère **couvrant**
- caractère **connecté**
- caractère **orienté**
- gouverneur **unique**
- **acyclicité**

connecté : peut pas avoir de sous arbres tout seul
chaque mot à un seul gouverneur

Syntaxe en dépendances

Où sont les constituants
dans un arbre en dépendances ?

*Tu manges **une glace à la fraise***

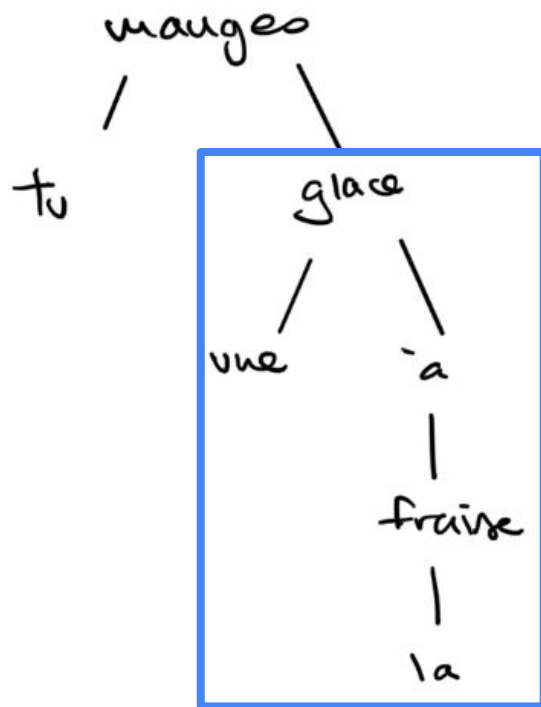


*Tu **la** manges*

c'est un constituant, un sous
arbre, car , par exemple, on
peut le remplacer par un
pronom

« tu la manges »

- comment savoir qui est lié à
qui ?



Syntaxe en dépendances

Analyse d'une phrase :

1. **Existence d'une relation syntaxique** : Est-ce que deux éléments (mots, segments, etc.) sont reliés par une dépendance ?
2. **Orientation de la dépendance** : Si oui, qui est le gouverneur, et qui est le dépendant ?
3. **Type de la dépendance** : Quelle est la relation syntaxique réalisée ?)

Est-ce que deux mots sont reliés par une relation ?

Il y a une **relation syntaxique** entre A et B si AB est **une unité syntaxique**.

Marie achète des robes en France



Marie achète des robes en soie



on peut faire des tests de déplacement,
« En France, marie achète... »

Est-ce que deux mots sont reliés par une relation ?

- critère de **linéarité** : la position d'un mot est définie par rapport à l'autre

Exemple: *C'est ma carte bancaire*

La position de l'adjectif *bancaire* dans la phrase dépend de la position du nom *carte*

- critère d'**unité prosodique** : typiquement, deux formes reliées par une dépendance forment une unité prosodique

Théorie Sens – Texte (Mel'čuk 1988, p. 128-132)

Est-ce que deux mots sont reliés par une dépendance ?

Test de déplacement

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

En France, Marie achète des robes

**En soie, Marie achète des robes*

Test d'insertion

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

Marie achète des robes pour sa mère en France

**Marie achète des robes pour sa mère en soie*

Est-ce que deux mots sont reliés par une dépendance ?

Test de commutation

- Commutation avec un pronom => repérage de la classe distributionnelle
Marie achète des robes en France => Marie en achète en France
Marie achète des robes en soie => ?? Marie en achète en soie

Test d'autonomie illocutoire / prosodique

- Une partie d'un énoncé qui peut former un énoncé autonome sans modification de sens est une unité syntaxique

NB: Critère suffisant, mais pas nécessaire

*(Marie achète des robes en France): Qu'achète Marie ? *Des robes en France*

(Marie achète des robes en soie): Qu'achète Marie ? Des robes en soie.

Principe de commutation

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 **XBAY** sont inacceptables

Dans l'exemple *Claude parle à sa mère* :

à sa mère **commute** avec *à Camille* :

Claude parle à Camille

* *Claude parle à sa mère à Camille*

à sa mère **ne commute pas** avec *de ses vacances* :

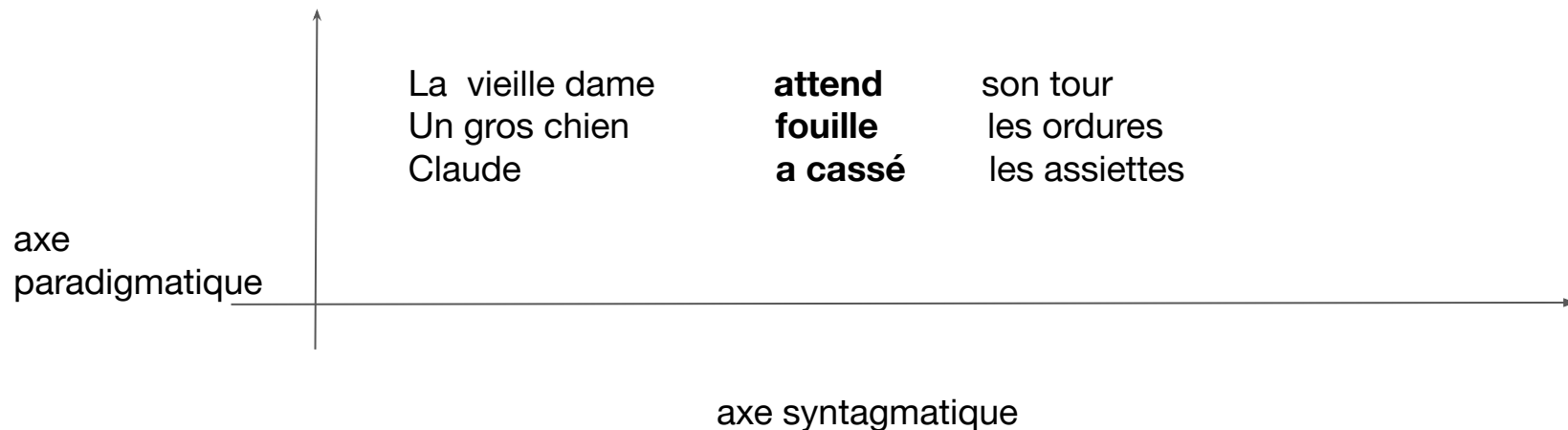
Claude parle de ses vacances

Claude parle à sa mère de ses vacances

Principe de commutation

Paradigme de commutation

L'ensemble de signes qui commutent les uns avec les autres s'appelle un **paradigme de commutation**



Exercice: Qui dépend de qui et pourquoi?

1. Marie lit un livre intéressant.
2. Jean travaille à l'université.
3. Ils voyagent en Espagne.

Exercice: Qui dépend de qui et pourquoi?

1. Marie lit un livre intéressant.

- a. (Marie, lit)
- b. (lit, livre) -> substitution car même classe distrib.
- c. (livre, intéressant)
- d. (livre, un)

2. Jean travaille à l'université.

- a. (travaille, Jean)
- b. (travaille, université)
- c. (à, université)

3. Ils voyagent en Espagne.

- a. (voyagent, Espagne)
- b. (voyagent, ils)
- c. (en, Espagne)

Distribution et classe distributionnelle

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

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

Distribution de l'adjectif :

La **magnifique** peinture est arrivée

C'est une peinture **magnifique**

La peinture est **magnifique**

Il trouve la peinture **magnifique**

Classe distributionnelle du modifieur de nom:

C'est une peinture **magnifique**

C'est une peinture **que j'apprécie**

C'est une peinture **appréciée par tout le monde**

C'est une peinture **d'une valeur inestimable**

Qui est le gouverneur et qui est le dépendant ?

- **valence passive du syntagme** : la forme qui détermine le schéma distributionnel des deux éléments de la dépendance est sa tête
Dans *La belle maison est rouge*, les mots *belle maison* occupent la position d'un nom et non pas d'un adjectif => *maison* gouverne *belle*
- **point de contact morphologique** : la forme qui établit le contact morphologique (accord, rection, etc.) avec le reste de la phrase est la tête d'une dépendance

Théorie Sens – Texte (Mel'čuk 1988, p. 133-135)

Qui est le gouverneur et qui est le dépendant ?

- **omissibilité** : dans certains cas, le dépendant peut être omis sans compromettre la grammaticalité de la phrase, mais pas la tête
Dans *La belle maison est rouge*, *La maison est rouge* est grammatical, mais *La belle est rouge* est questionnable
- **prédictibilité** : la présence du dépendant permet de “prédire” la présence du gouverneur : Si l’on sait que le sujet de la phrase commence par *la belle*, on s’attend à rencontrer un nom ensuite

LE VERBRE EST GOUVERNEUR = CONVENTION

Qui est le gouverneur et qui est le dépendant ?

Critère distributionnel avec effacement positif

- En considérant une unité U composé des éléments A et B :

Si A est autonomisable (et donc B est effaçable), et que 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**

→ **A est le gouverneur de B**

Qui est le gouverneur et qui est le dépendant ?

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
→ B est le gouverneur de A

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
→ A est le gouverneur de B

Qui est le gouverneur et qui est le dépendant ?

Critère distributionnel sans effacement

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

U	
à	Claude
A	B

U'	
par	Claude
A'	B

U et U' n'ont pas la même distribution
→ A est la tête de U
→ A est le gouverneur de B

Exercice: Qui est le gouverneur et le dépendant ?

Pour chaque syntagme identifiez le gouverneur et le dépendant, en utilisant les critères suivants: valence passive du syntagme, point de contact morphologique, ommissibilité, prédictibilité, critères distributionnels (effacement positif, négatif, sans effacement).

1. le chat noir
 2. un verre(d'eau)
 3. (très)rapidement
- critère d'ommissibilité

1. la fille que tu as vue hier
2. manger des pommes
3. une robe rouge éclatante

Est-ce que deux éléments relèvent de la même relation syntaxique ?

- **paires minimales** : mêmes lexèmes, structures syntaxiques qui diffèrent par un seul paramètre : si différence sémantique existe, ce n'est pas la même relation (cf. *the visible stars* vs *the stars visible*)
- **sous-arbres interchangeables** : si les deux sous-arbres représentant le dépendant sont interchangeables sans changement majeur de grammaticalité ou de sens, c'est la même relation
(cf. *la maison de ma mère* vs *la maison que ma mère a acheté*)
- **dépendant répétable ou pas** : dans le cadre de la même relation, le dépendant est soit répétable, soit non. (cf. un verbe peut admettre plusieurs circonstants, mais un seul objet direct)

Est-ce que deux éléments relèvent de la même relation syntaxique ?

Propriétés utiles :

- **catégories morphosyntaxiques** et **lemmes** possibles du G et D
- **traits morphosyntaxiques** spécifiques liés à la relation
(par exemple, l'accusatif pour l'objet direct dans les langues à cas)
- dépendants verbaux : **pronominalisation**
- **accord** : éléments et traits concernés
- **linéarisation**:
 - ordre canonique G - D
 - flexible ou rigide
 - adjacence obligatoire ou non

L'inventaire de relations syntaxiques est **spécifique à la langue**

Est-ce que deux éléments relèvent de la même relation syntaxique ?

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** juraient tous.
5. **Elle** ne sort jamais sans son téléphone

Argument vs modifieur

On parle de **sous-catégorisation** ou de **valence** d'une unité pour faire référence à l'ensemble de ses **arguments**

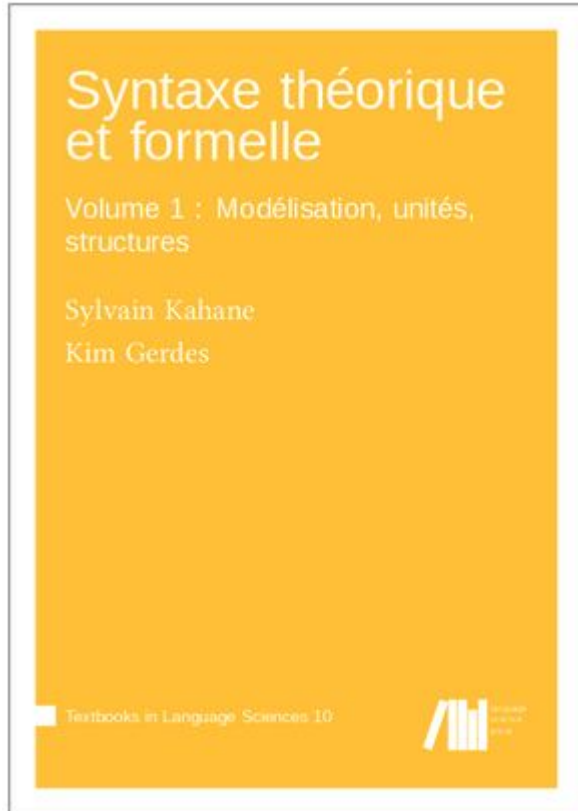
- Les arguments complètent le sens d'un prédicat et sont exigés / prédéfinis
 - C'est pour ça qu'on les appelle compléments
- Les modifieurs sont optionnels

Je bois trop d'eau

J'appelle mon fils tous les soirs

plat copieux

Ouvrage de référence



Disponible en PDF :

<https://langsci-press.org/catalog/book/241>

Schémas d'annotation

Formalisme définissant quelles sont les relations syntaxiques possibles et les conditions de réalisation de chacune d'entre elles

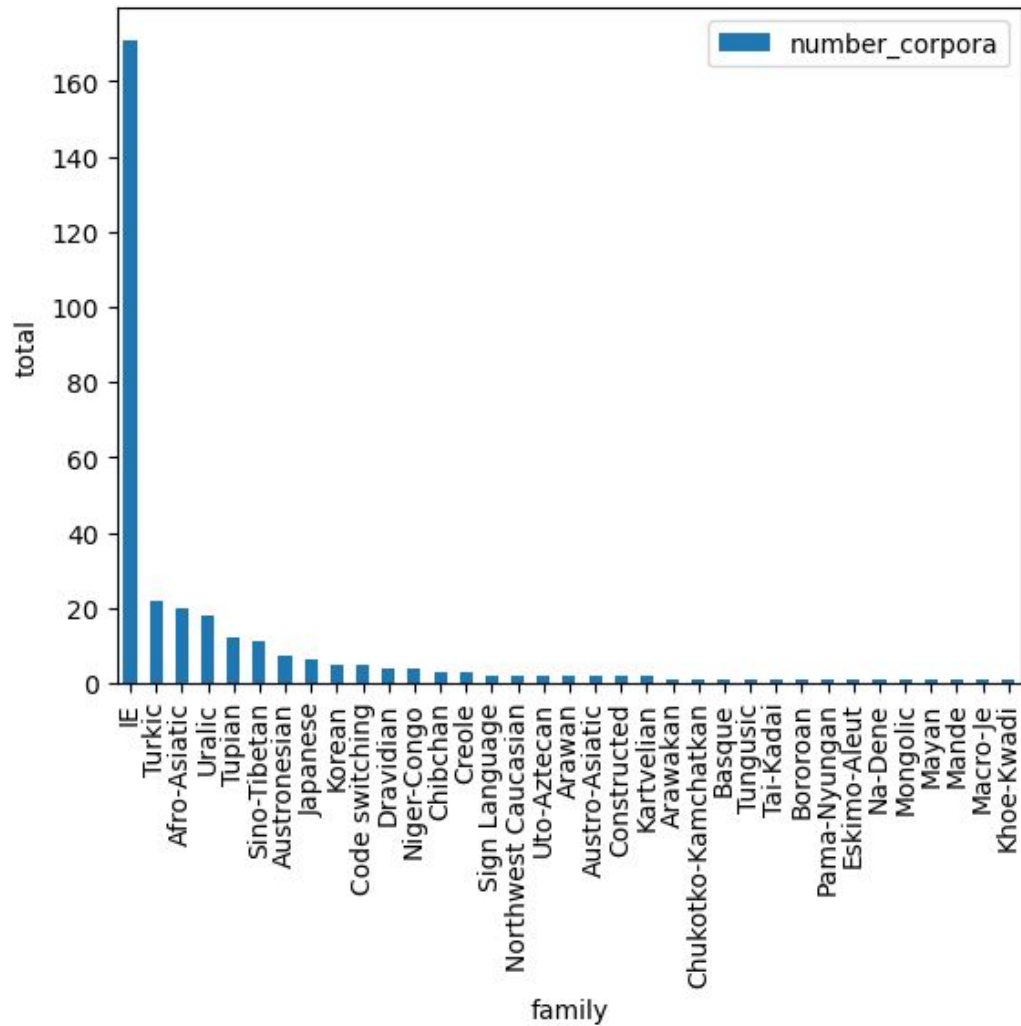
Exemple	French Treebank	Universal Dependencies	Surface Universal Dependencies
Jean mange une pomme	obj	obj	comp:obj
Jean parle à Marie	a_obj	obl	comp:obl
Jean lui parle	a_obj	iobj	comp:obl
Jean va à l'école	p_obj	obl	comp:obl
Jean a vu Marie à l'école	mod	obl	mod

Schémas d'annotation : Universal Dependencies

Développement de treebanks

- Plusieurs projets d'annotation dans plusieurs langues
- Quelques projets multilingues
- Schémas d'annotation spécifiques à chaque projet
=> difficile de comparer
- En 2014, démarrage du projet UD
 - 10 corpus, 10 langues dans la version 1.0 (janvier 2015)
 - 319 corpus, 179 langues dans la version 2.16 (mai 2025)





Schémas d'annotation : Universal Dependencies

Principes de base

- Recherche d'un jeu d'étiquettes applicable à toutes (?) les langues du monde
- Unité de base : mots syntaxiques (*des* → *de* + *les*)
- Niveaux d'annotation :
 - Catégories morphosyntaxiques : ensemble de 17 catégories
 - Lemmes (forme de base de chaque mot)
 - Traits morphosyntaxiques (genre, nombre, temps, aspect...)
 - Dépendances syntaxiques : ensemble de 37 relations de base (extensions possibles)
 - **Primauté des mots lexicaux sur les mots fonctionnels**

Schémas d'annotation : Universal Dependencies

Catégories morphosyntaxiques (Universal Parts of Speech = upos)

- ADJ: adjective
- ADP: adposition
- ADV: adverb
- AUX: auxiliary
- CCONJ: coordinating conjunction
- DET: determiner
- INTJ: interjection
- NOUN: noun
- NUM: numeral
- PART: particle
- PRON: pronoun
- PROPN: proper noun
- PUNCT: punctuation
- SCONJ: subordinating conjunction
- SYM: symbol
- VERB: verb
- X: other

Schémas d'annotation : Universal Dependencies

Traits morphosyntaxiques

Traits flexionnels verbaux

- VerbForm
- Mood
- Tense
- Aspect
- Voice
- Evident
- Polarity
- Person
- Polite
- Clusivity

Traits flexionnels nominaux

- Gender
- Animacy
- NounClass
- Number
- Case
- Definite
- Deixis
- DeixisRef
- Degree

Traits lexicaux

- PronType
- NumType
- Poss
- Reflex

Autres

- Abbr
- Typo
- Foreign
- ExtPos

Schémas d'annotation : Universal Dependencies

Les relations syntaxiques sont organisées selon plusieurs axes :

- **nature du gouverneur** (verbal, nominal ou autre)
- **nature du dépendant** (adverbe / adjectif seul, groupe nominal, groupe prépositionnel, proposition)
- pour les dépendants verbaux, leur statut par rapport au verbe
 - **core dependents** => arguments
 - **non-core dependents** => obliques

Schémas d'annotation : Universal Dependencies

- **root** *root of the annotated element* (racine de la phrase / tête du segment annoté)

verbe principal : je reviendrais avec plaisir !

verbe principal dans une forme composée : elle est arrivée vite

“**attribut de sujet**” avec les verbes attributifs : elle est étudiante

si pas de phrase, tête du groupe : les maisons en carton

Schémas d'annotation : Universal Dependencies

Core arguments (arguments essentiels) : groupes nominaux

- **sujet** : **nsubj** (*nominal subject*)
Colorless green ideas sleep furiously
- **objet direct** : **obj** (*direct object*) = COD
Jean mange une pomme
- **objet indirect** : **iobj** (*indirect object*) != COI
Jean leur parle

Any **nominal phrase** that is a core argument of a verb,
but is **not** its subject or direct object is an indirect object

Schémas d'annotation : Universal Dependencies

Core arguments (arguments essentiels) : propositions

- **csubj : clausal subject** (sujet sous forme d'une proposition)
Qu'il vient si vite est surprenant
- **ccomp : clausal complement** (complétive pleine)
Qu'il vient si vite est surprenant
- **xcomp : open clausal complement** (complétive infinitivale - pas de sujet explicite)
Il est chargé de maintenir la paix
Les femelles peuvent réaliser une à deux portées par an

Schémas d'annotation : Universal Dependencies

Non-core arguments (dépendants non essentiels) : groupes nominaux

- **obl : oblique nominal** (dépendant verbal non essentiel sous la forme d'un GPrep)
Il vient en avril
(vient -[obl]-> avril; avril -[case]-> en)
- **expl : expletive** (élément explétif ; en français typiquement le “sujet apparent”)
Il est surprenant qu’il vient si vite
- **dislocated : dépendant verbal disloqué et repris dans la proposition**
La chose la plus surprenante, c’est son arrivée précoce
*Il faut pas **la** manger, la plasticine*
- **vocative : dialogue participant addressed in text** (‘apostrophe’)
Locataires potentiels, fuyez cette agence !

Schémas d'annotation : Universal Dependencies

Non-core arguments (dépendants non essentiels) : **propositions**

- **advcl : adverbial clause** (subordonnée circonstancielle)

Il vient quand ça lui chante / parce qu'il veut / même s'il ne devrait pas

Non-core arguments (dépendants non essentiels) : **“mots modificateurs”**

- **advmod : modifieur adverbial d'un verbe, adjectif ou autre adverbe**

Je chante mal

Je chante très mal

- **discourse : discourse elements with an expressive function**

Bref, fuyez cette agence !

Schémas d'annotation : Universal Dependencies

Non-core arguments (dépendants non essentiels) : mots fonctionnels

- **aux** : *funct. words associated with main predicate expressing TAMVE* (auxiliaires)
Il est venu ; Je serai là
- **cop** : *funct. word linking a non-verbal predicate to its subject*
Jean est étudiant
- **mark** : *marker that marks a proposition as subordinate* (typ. conjonction de SUB)
Il vient quand ça lui chante

Schémas d'annotation : Universal Dependencies

Nominal dependents (dépendants d'un nom) : groupes nominaux

- **nmod : nominal modifier** (complément de nom)
la maison de ma mère
- **nummod : numeral modifier** (numéral modifiant un nom)
trois pommes rouges
- **appos : apposition**
Le Petit Prince, chef-d'oeuvre de St Exupéry, reste un des livres les plus vendus

Schémas d'annotation : Universal Dependencies

Nominal dependents (dépendants d'un nom) : “mots modifieurs”

- **amod** : *adjectival modifier* (groupe adjectival qui modifie un nom)

la maison rouge

Nominal dependents (dépendants d'un nom) : **propositions**

- **acl** : *adnominal clause* (proposition finie ou non qui modifie un nom)

la maison que ma mère a achetée

la maison achetée par ma mère

Schémas d'annotation : Universal Dependencies

Nominal dependents (dépendants d'un nom) : mots fonctionnels

- **det : *determiner* (déterminant)**
la maison de ma mère
- **case : *case-marking functional word* (typ. prépositions)**
la maison de ma mère
(mère -[case]-> de)
- **clf : *classifier* (élément exprimant la classification d'un nom dans un groupe sémantique)**
 - 三个学生 (三個學生) sān gè xuéshēng = “three [human-classifier] student”
 - 三棵树 (三棵樹) sān kē shù = “three [tree-classifier] tree”
 - 三只鸟 (三隻鳥) sān zhī niǎo = “three [bird-classifier] bird”
 - 三条河 (三條河) sān tiáo hé = “three [long-wavy-classifier] river”

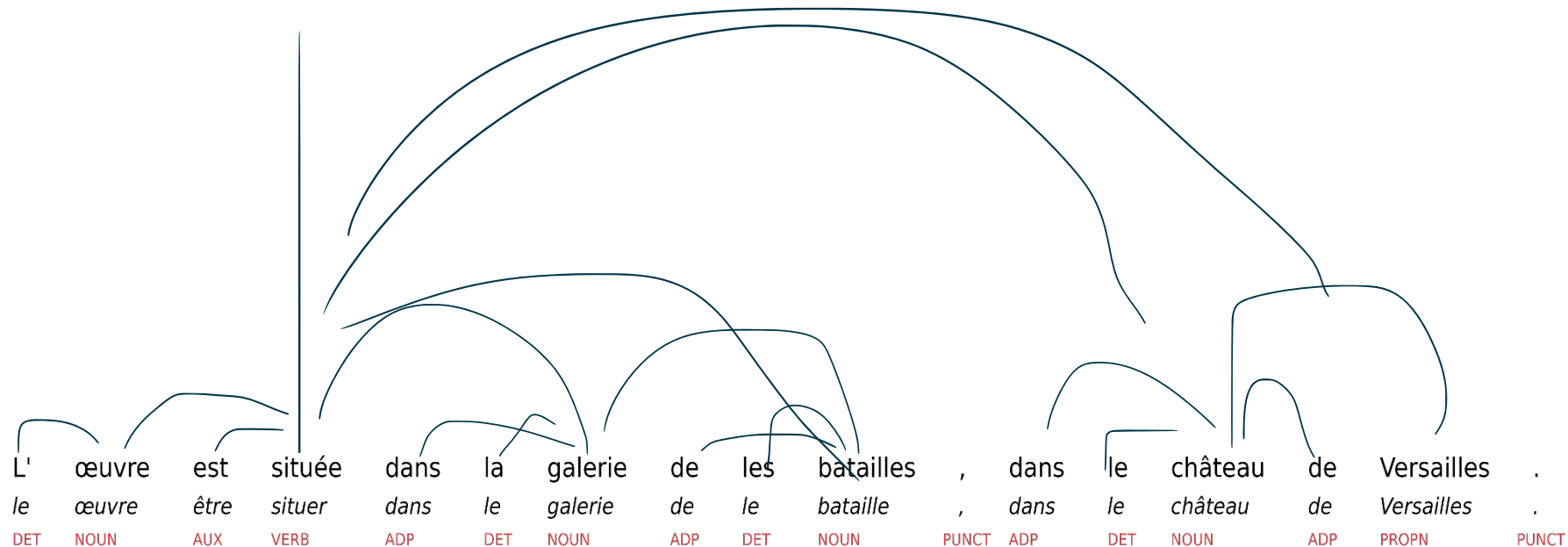
! **clf** est gouverné par le numéral et non pas par le nom

Schémas d'annotation : Universal Dependencies

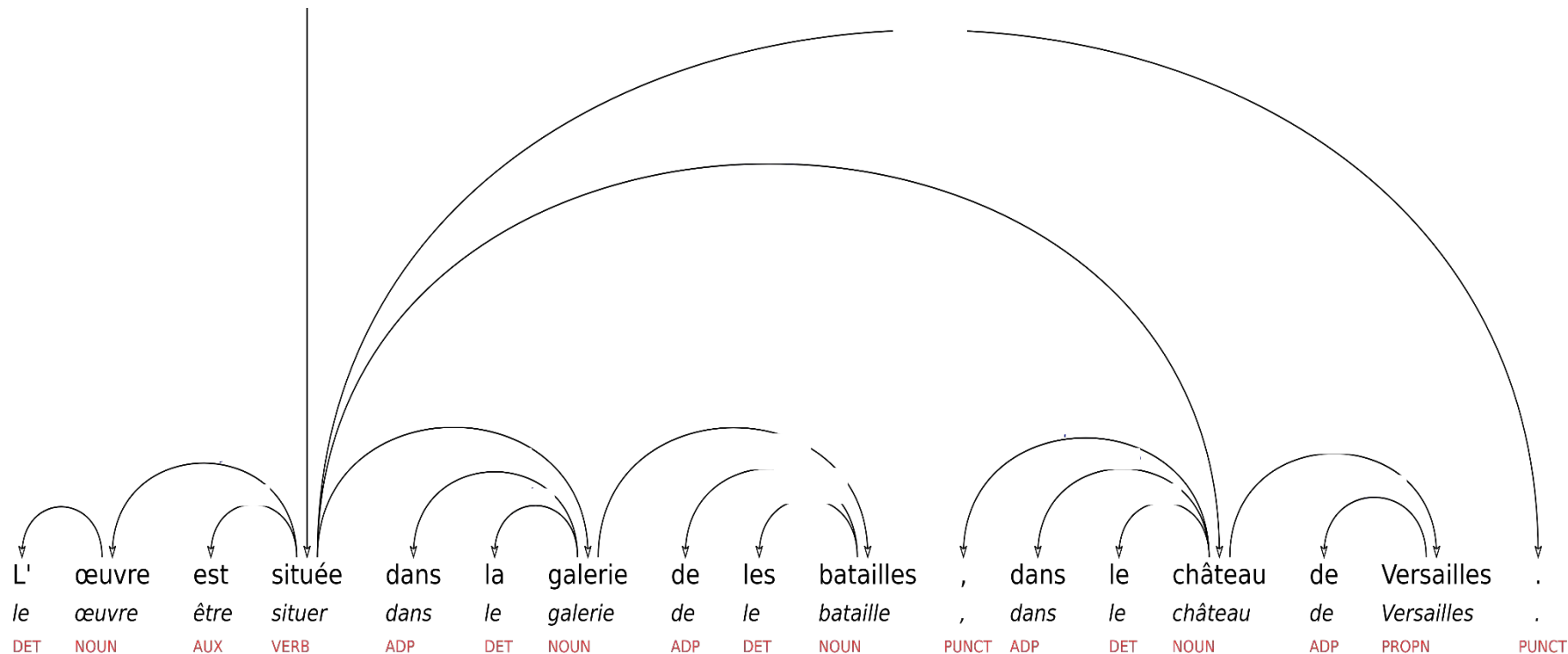
A retenir :

- **Primauté des mots lexicaux** : dans un groupe prépositionnel, le nom est le gouverneur de la préposition ; dans une subordonnée, le verbe est le gouverneur de la conjonction de subordination, etc.
- **Fonction et nature encodées par les relations** :
 - **nsubj vs csubj**
 - **obj vs ccomp**
 - **amod vs acl**
 - **advmod vs advcl**

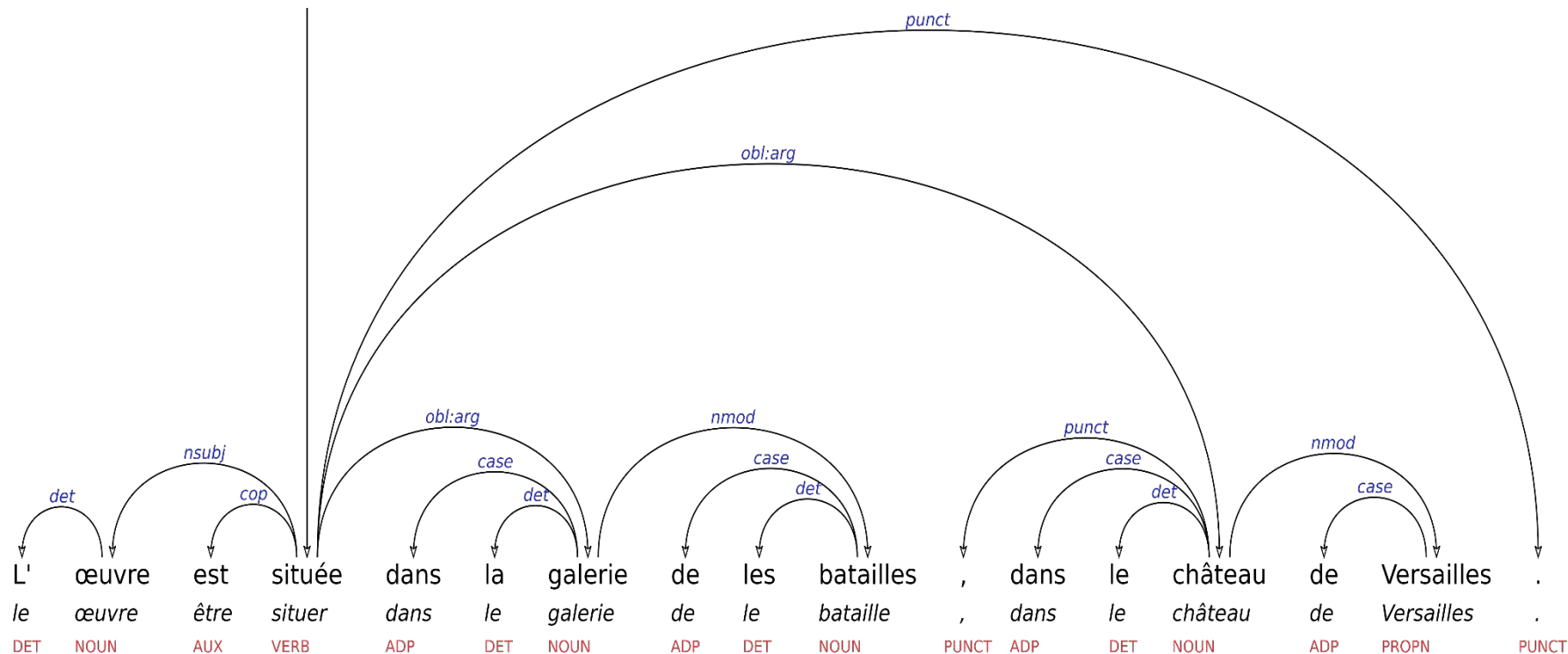
Schémas d'annotation : Universal Dependencies



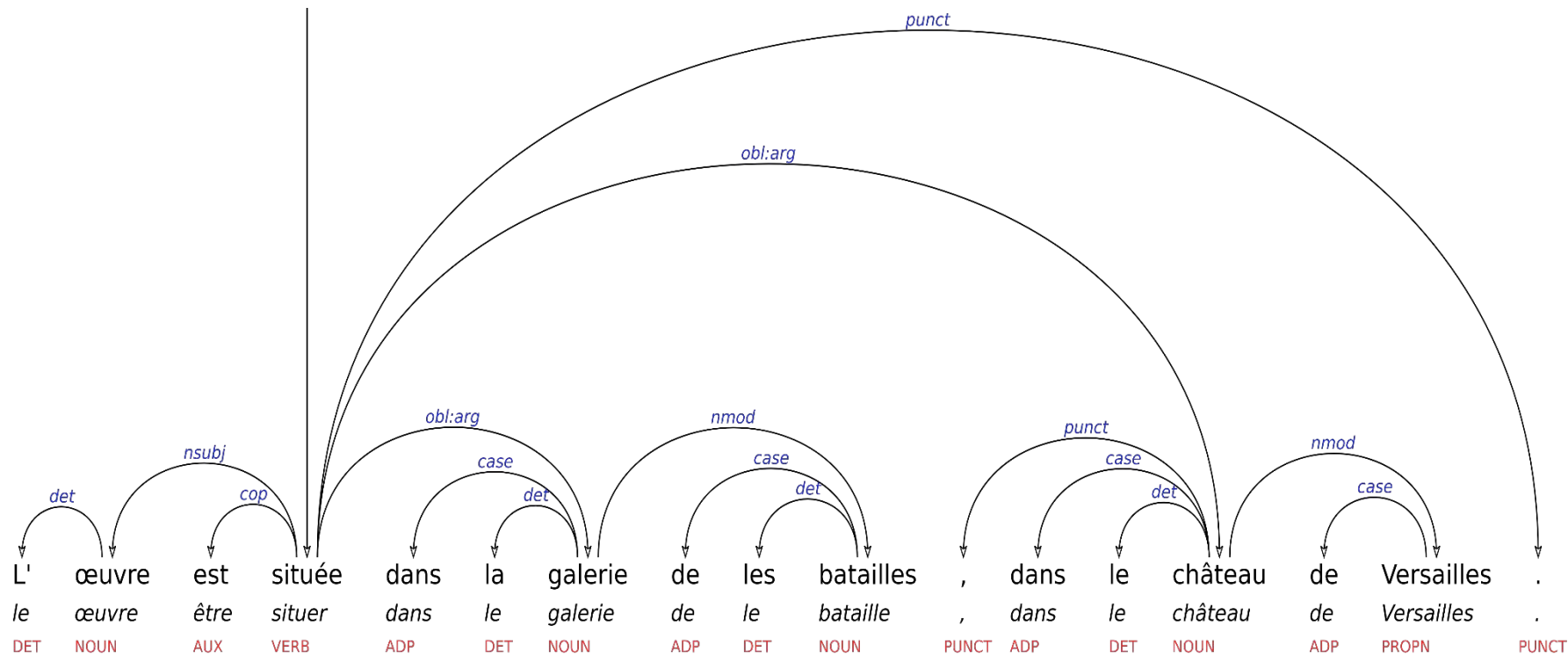
Schémas d'annotation : Universal Dependencies



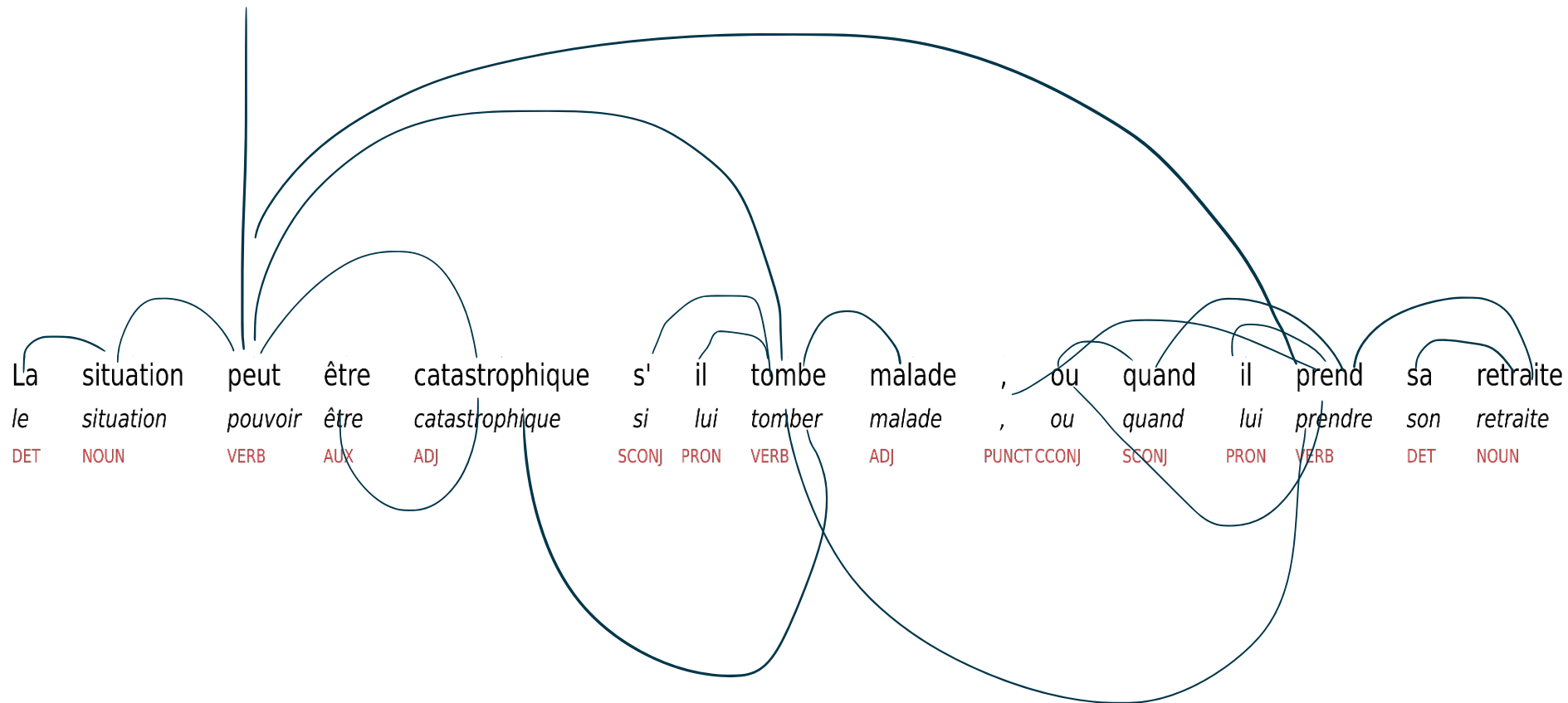
Schémas d'annotation : Universal Dependencies



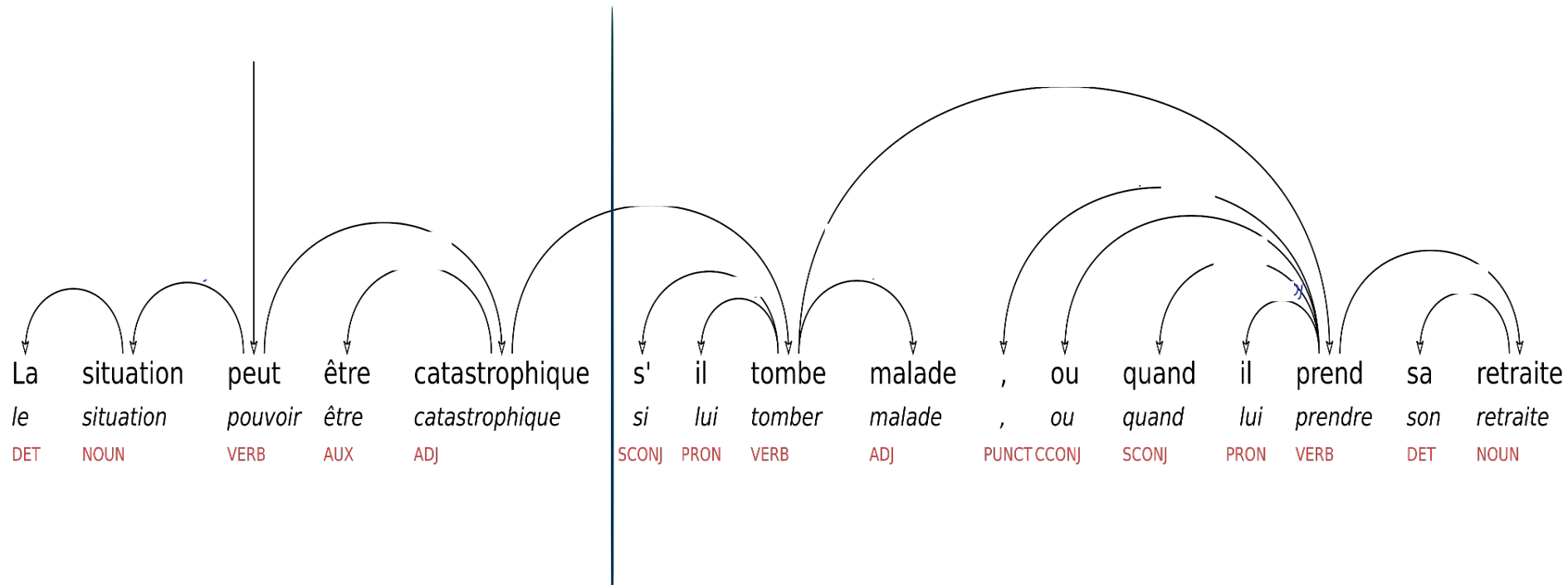
Schémas d'annotation : Universal Dependencies



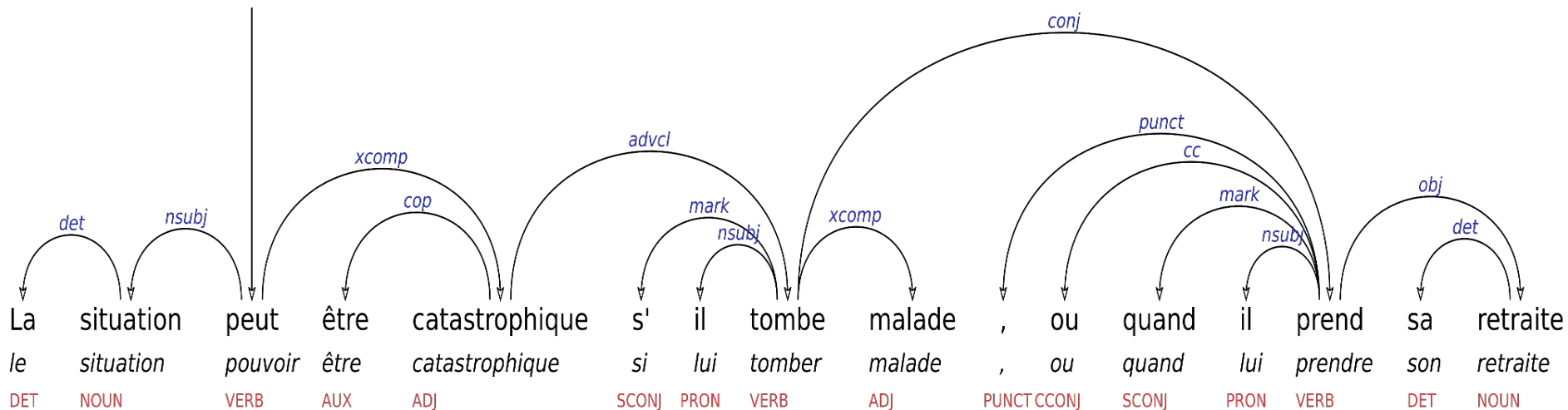
Schémas d'annotation : Universal Dependencies



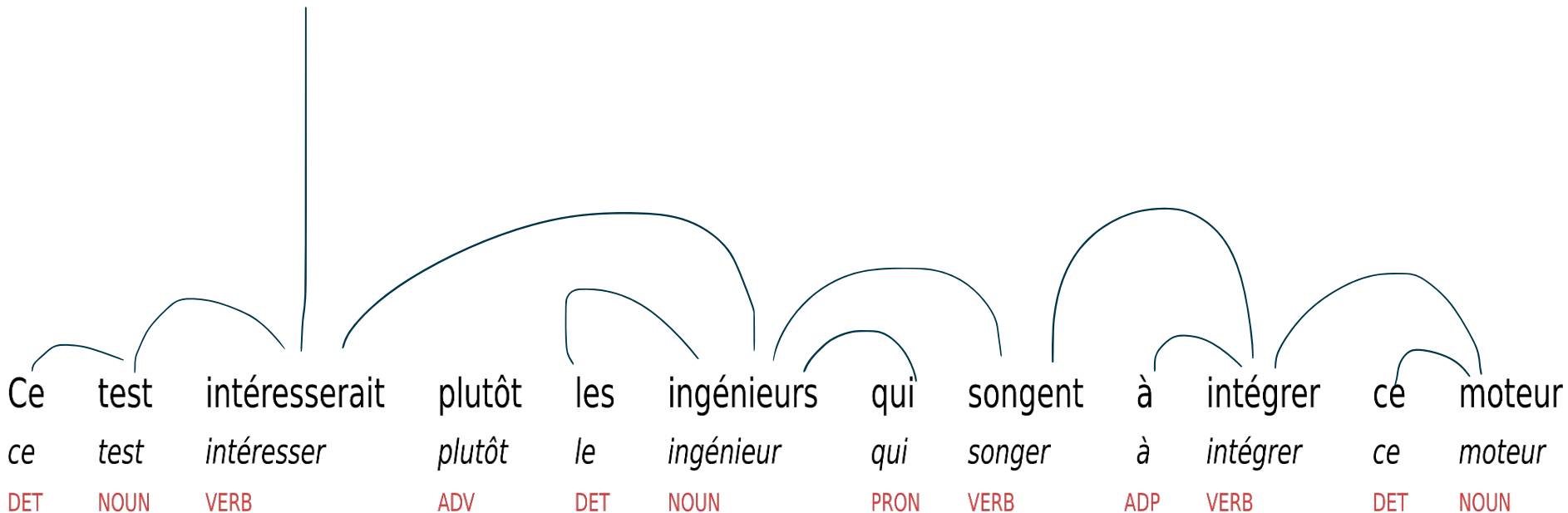
Schémas d'annotation : Universal Dependencies



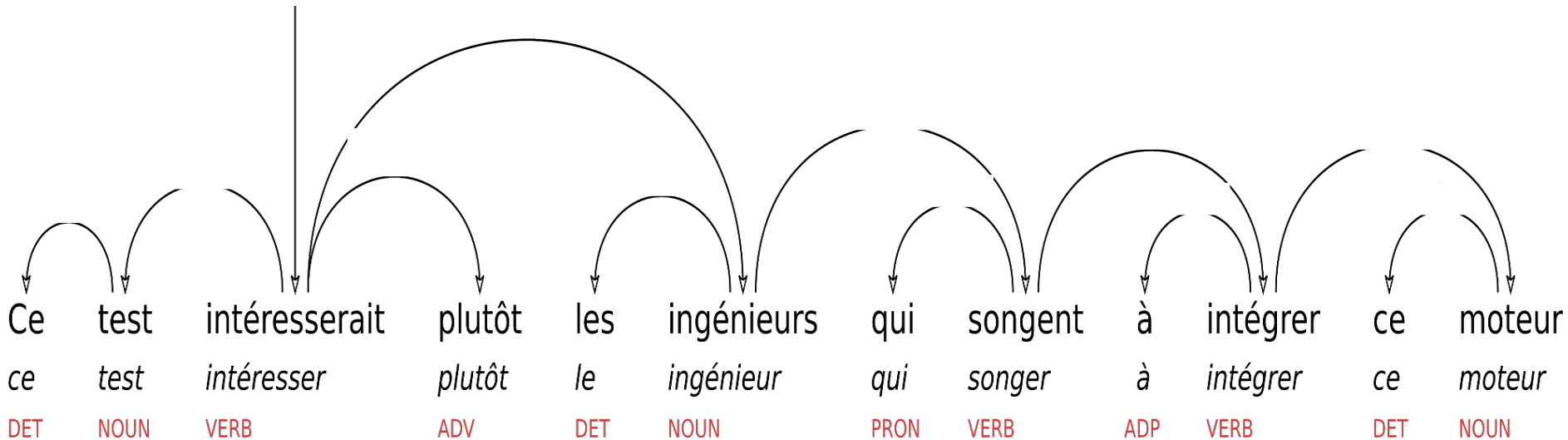
Schémas d'annotation : Universal Dependencies



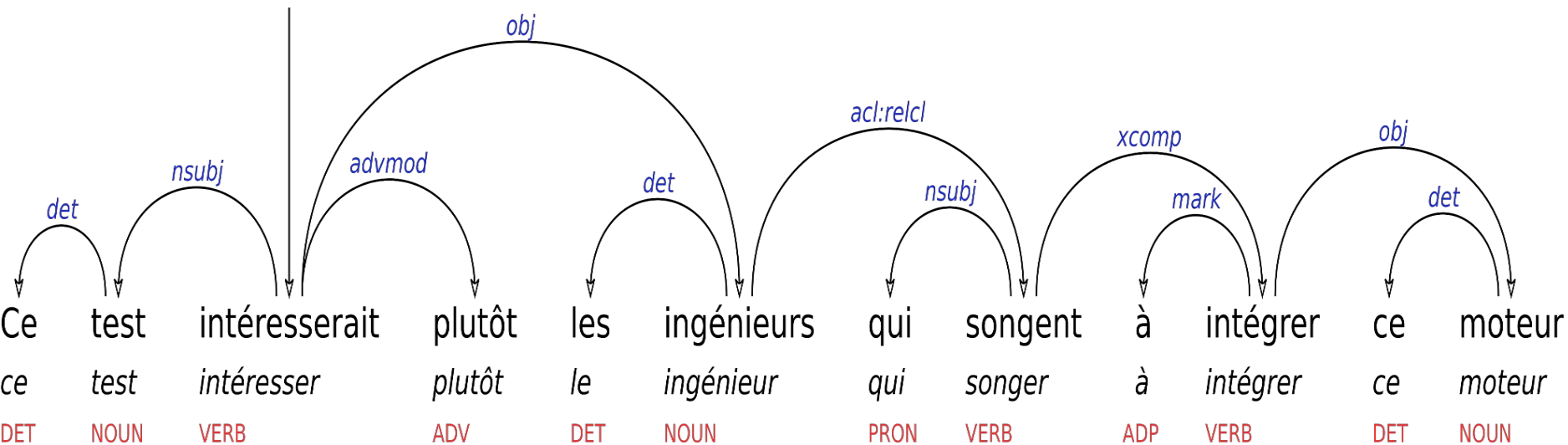
Schémas d'annotation : Universal Dependencies



Schémas d'annotation : Universal Dependencies



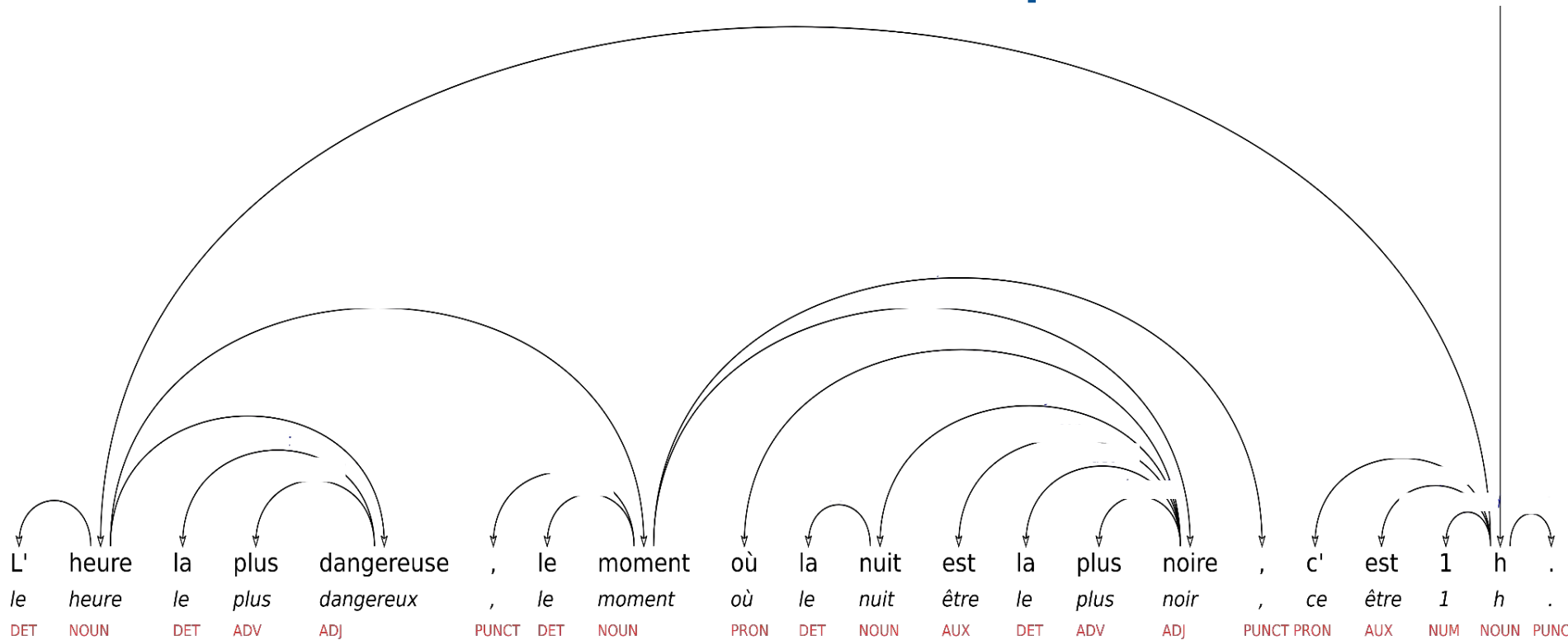
Schémas d'annotation : Universal Dependencies



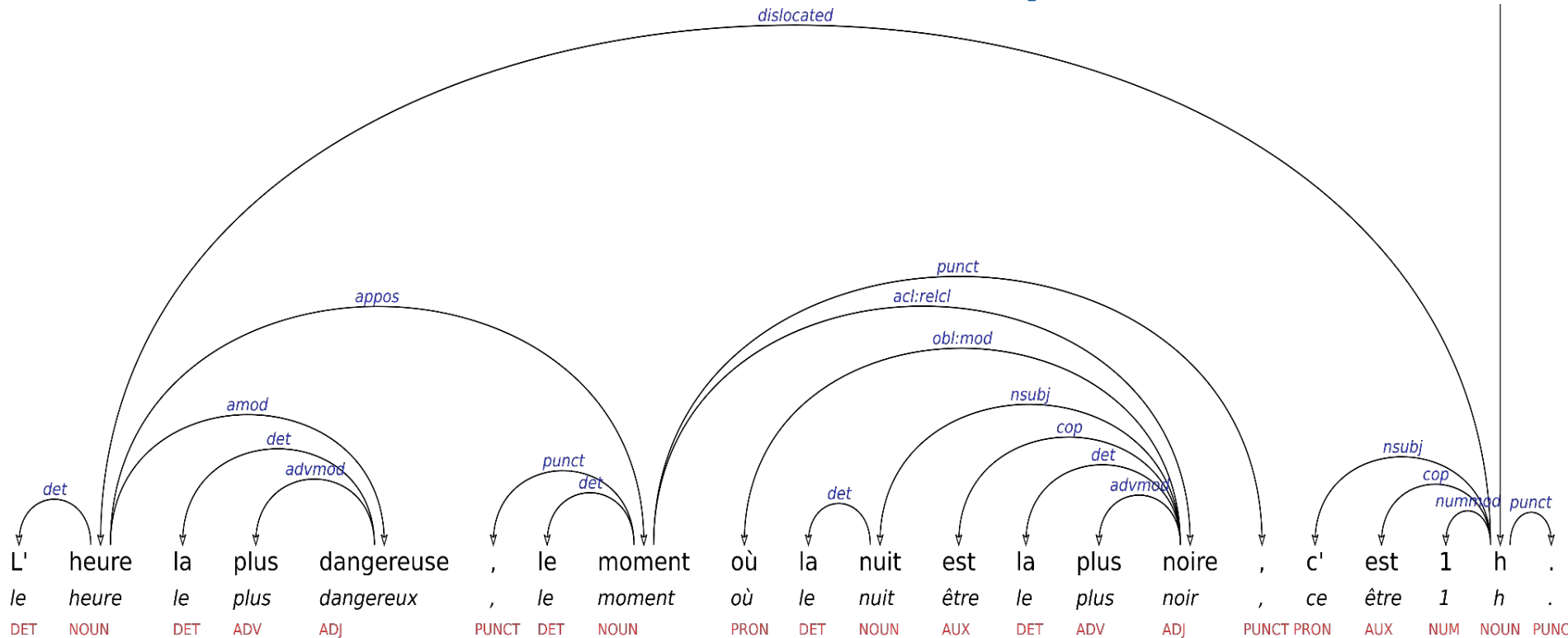
Schémas d'annotation : Universal Dependencies

L' heure la plus dangereuse , le moment où la nuit est la plus noire , c' est 1 h .
le heure le plus dangereux , le moment où le nuit être le plus noir , ce être 1 h .
DET NOUN DET ADV ADJ PUNCT DET NOUN PRON DET NOUN AUX DET ADV ADJ PUNCT PRON AUX NUM NOUN PUNCT

Schémas d'annotation : Universal Dependencies



Schémas d'annotation : Universal Dependencies



Exercice - Corrigez / Faites vos arbres

1. Pierre mange une pomme et boit un café avec Marie.
2. Marie lit des romans et des poèmes dans le jardin.
3. Pierre donne une pomme à Jean qui est fatigué.
4. Paul aime les crêpes parce qu'elles lui rappellent son enfance.
5. Pierre veut acheter et lire un nouveau livre pour ses recherches.

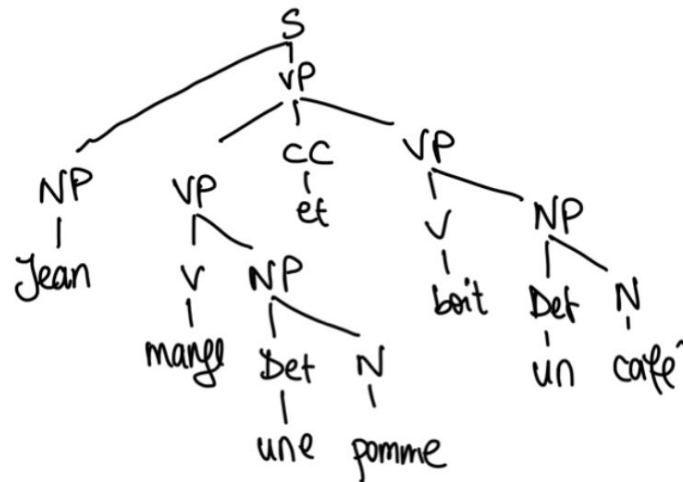
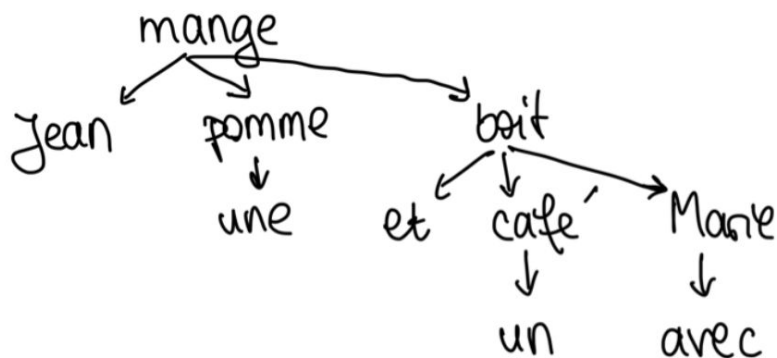
Jean mange une pomme et boit un café avec Marie.

↓
sujet

COB de "mange"
↓
verbes principaux
(coordination de verbes)

COB de "boit"

↓
cc de accompagnement



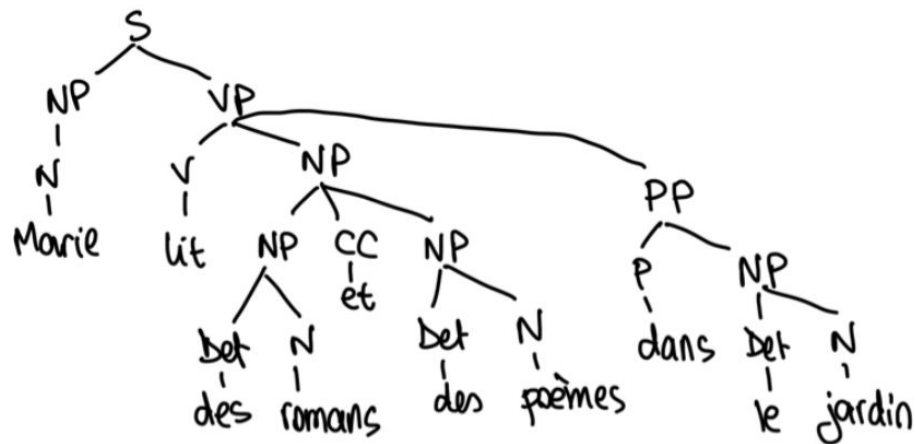
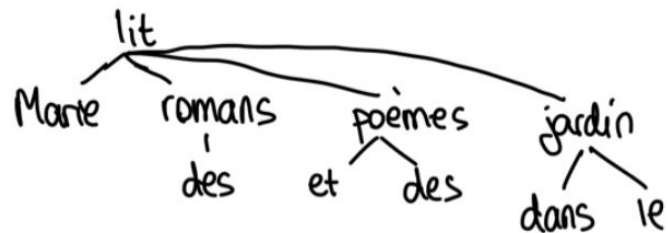
Marie lit des romans et des poèmes dans le jardin.

↓
sujet

↓
verbe principal

↓
cbs coordonnés

↓
cc de lieu



Pierre a donné une pomme à Jean qui était fatigué.

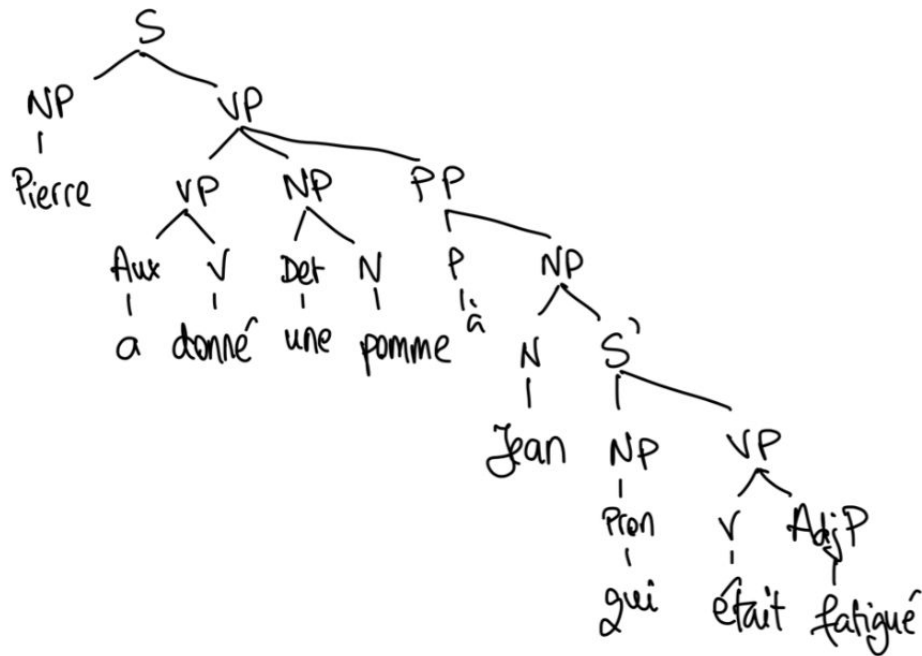
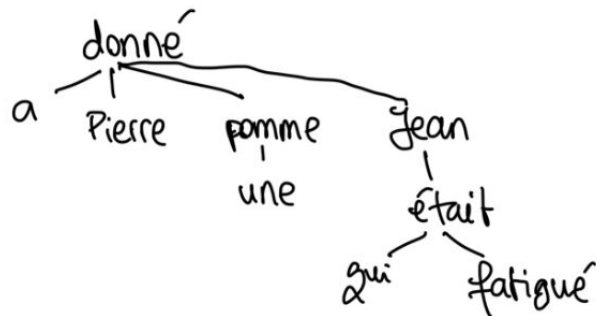
sujet

verbe principal

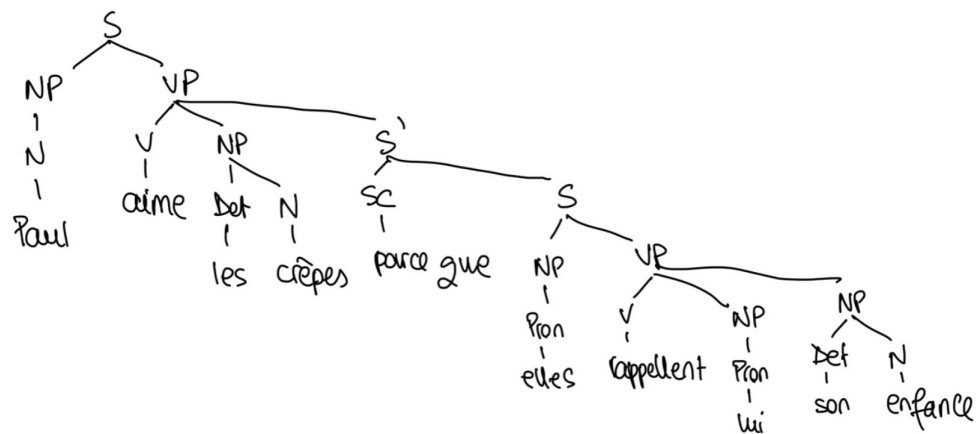
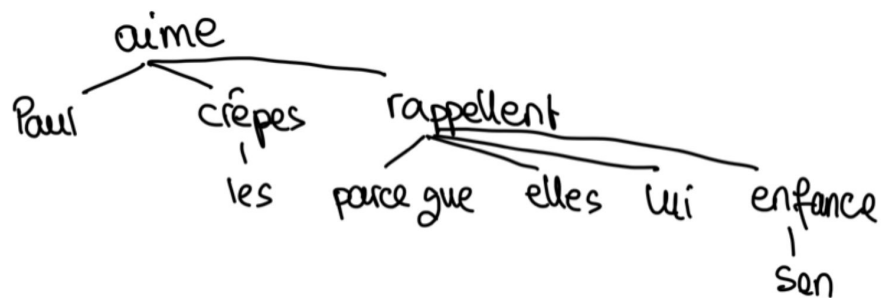
cod

coi

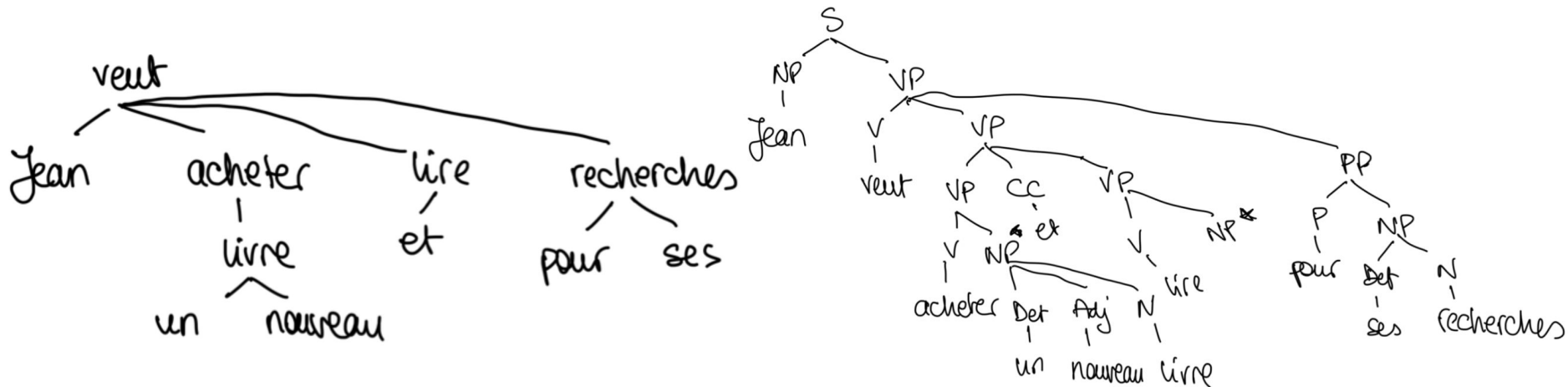
relative - subordination



Paul aime les crêpes parce qu'elles lui rappellent son enfance.



Jean veut acheter et lire un nouveau livre pour ses recherches.



Pour la prochaine fois:

Étudiez le guide d'annotation UD :

guide général : <https://universaldependencies.org/guidelines.html>

POS : <https://universaldependencies.org/u/pos/all.html>

traits morphosyntaxiques : <https://universaldependencies.org/u/feat/all.html>

dépendances : <https://universaldependencies.org/u/dep/all.html>

Approfondir les notions discutés en cours: <https://langsci-press.org/catalog/book/241>
(pages 300-340) - lire le chapitre entier pour ceux qui veulent savoir plus.

Finir exercices commencés en cours