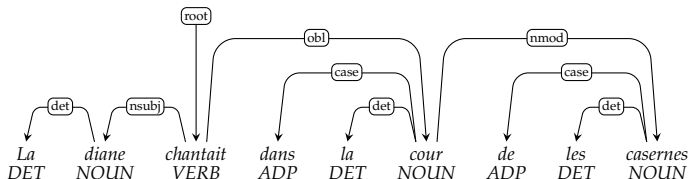# Universal Dependencies

Universal Dependencies

- Annoter des phrases dans plusieurs langues avec les mêmes conventions, en particulier :
  - Les mêmes catégories morpho-syntaxiques (parties de discours, parts of speech)
  - Les mêmes fonctions syntaxiques
- Plus de 60 langues annotées

17 Parties de discours

Classes lexicales		Classes grammaticales		Autre	
ADJ	adjectif	ADP	adposition	PUNCT	ponctuation
ADV	adverbe	AUX	auxiliaire	SYM	symbole
INTJ	interjection	CCONJ	conj. coo.	X	autre
NOUN	nom	DET	déterminant		
PROPN	nom propre	NUM	numéral		
VERB	verbe	PART	particule		
		PRON	pronom		
		SCONJ	conj. sub.		

Correspondance avec les classes traditionnelles

Nom	NOUN
	PROPN
Verbe	VERB
	AUX
Adjectif	ADJ
Adverbe	ADV
Pronom	PRON
Préposition	ADP
Conjonction	SCONJ
	CCONJ
Numéral	NUM
Article	DET
Interjection	INTJ
	PART
	PUNCT
	SYM
	X

NOUN

- Nouns are a part of speech typically denoting a person, place, thing, animal or idea.
- Note that some verb forms such as gerunds and infinitives may share properties and usage of nouns and verbs. Depending on language and context, they may be classified as either VERB or NOUN.
  - **étudiant**
  - **campagne**
  - **feuille**
  - **hyène**
  - **démocratie**
  - **boire** est conseillé dans ces conditions

PROPN : proper noun

- A proper noun is a noun (or nominal content word) that is the name (or part of the name) of a specific individual, place, or object.
    - **Zacharie**
    - **Marseille**
    - **ONU**
    - le **campus de Luminy**

VERB

- A verb is a member of the syntactic class of words that typically signal events and actions, can constitute a minimal predicate in a clause, and govern the number and types of other constituents which may occur in the clause.
- Verbs are often associated with grammatical categories like tense, mood, aspect and voice, which can either be expressed inflectionally or using auxilliary verbs or particles.
    - **penseraient**
    - **mangeront**

AUX : auxiliary

- An auxiliary is a function word that accompanies the lexical verb of a verb phrase and expresses grammatical distinctions not carried by the lexical verb, such as person, number, tense, mood, aspect, voice or evidentiality.

- It is often a verb (which may have non-auxiliary uses as well) but many languages have nonverbal TAME (Tense, Aspect, Modality, Evidentiality) markers and these should also be tagged AUX.

- The class AUX also include copulas (in the narrow sense of pure linking words for nonverbal predication).
  - il **a** dormi
  - il **est** parti
  - il **a été** nommé
  - il **est** grand

ADJ : adjective

- Adjectives are words that typically modify nouns and specify their properties or attributes
- They may also function as predicates
- Numbers vs. Adjectives : In general, cardinal numbers receive the part of speech NUM, while ordinal numbers receive the tag ADJ.
    - un drapeau **blanc**
    - un **grand** bâtiment
    - le **troisième** essai
    - le trajet est **dangereux**

ADV : adverb

- Adverbs are words that typically modify verbs for such categories as time, place, direction or manner. They may also modify adjectives and other adverbs, as in *very briefly* or *arguably wrong*.
- Note that in Germanic languages, some adverbs may also function as verbal particles, as in write down or end up. They are still tagged ADV and not PART.
  - il avance **lentement**
  - il habite **ici**
  - il va **loin**
  - il vient **souvent**
  - **vraisemblablement** faux

ADP : adposition

- Adposition is a cover term for prepositions and postpositions.

- Adpositions belong to a closed set of items that occur before (preposition) or after (postposition) a complement composed of a noun phrase, noun or pronoun, and that form a single structure with the complement to express its grammatical and semantic relation to another unit within a clause.

- In many languages, adpositions can take the form of fixed multiword expressions, such as *in spite of*, *because of*, *thanks to*. The component words are then still tagged according to their basic use (in is ADP, spite is NOUN, etc.) and their status as multiword expressions are accounted for in the syntactic annotation.
  - il est arrivé **avant** la nuit
  - il a mangé **devant** lui
  - il a dit ça **pour** rigoler

CCONJ : coordinating conjunction

- A coordinating conjunction is a word that links words or larger constituents without syntactically subordinating one to the other and expresses a semantic relationship between them.
    - La pluie **et** les nuages
    - il mange **et** il boit
    - il mange **et** boit beaucoup
    - il tape haut **et** fort
    - il rentre calmement **et** gentiment

SCONJ : subordinating conjunction

- A subordinating conjunction is a conjunction that links constructions by making one of them a constituent of the other. The subordinating conjunction typically marks the incorporated constituent which has the status of a (subordinate) clause.
    - je pense **qu'** il va bien
    - je travaille **tandis qu'**il se repose

DET : determiner

- Determiners are words that modify nouns or noun phrases and express the reference of the noun phrase in context. That is, a determiner may indicate whether the noun is referring to a definite or indefinite element of a class, to a closer or more distant element, to an element belonging to a specified person or thing, to a particular number or quantity, etc.
- Also note that the notion of determiners is unknown in traditional grammar of some languages (e.g. Czech); words equivalent to English determiners may be traditionally classified as pronouns and/or numerals in these languages. They should be tagged DET.
- Their inflection (if applicable) is similar to that of adjectives, and distinct from nouns. They agree with the nouns they modify. Especially the ability to inflect for gender is typical for adjectives and determiners.
    - **un** chien
    - **le** chien
    - **ce** chien
    - **mon** chien
    - **quel** chien

NUM : numeral

- A numeral is a word, functioning most typically as a determiner, adjective or pronoun, that expresses a number and a relation to the number, such as quantity, sequence, frequency or fraction.
    - **trois** chiens
    - les **trois** chiens
    - le **tiers** de la quantité

PART : particle

- Particles are function words that must be associated with another word or phrase to get its meaning and that do not satisfy definitions of other universal parts of speech.
- Particles may encode grammatical categories such as negation, mood, tense etc. Particles are normally not inflected, although exceptions may occur.
- In general, the PART tag should be used restrictively and only when no other tag is possible. The the language-specific documentation should list the words classified as PART in the given language.
  - il **ne** viendra pas
  - his father **'s** bike

PRON : pronoun

- Pronouns are words that substitute for nouns or noun phrases, whose meaning is recoverable from the linguistic or extralinguistic context.
- Pronouns under this definition function like nouns. Note that some languages traditionally extend the term pronoun to words that substitute for adjectives. Such words are not tagged PRON under our universal scheme. They are tagged as determiners in order to annotate the same thing the same way across languages.
- It is not always crystal clear where pronouns end and determiners start.
    - **il** mange
    - l'homme **qui** mange

INTJ : Interjection

- An interjection is a word that is used most often as an exclamation or part of an exclamation. It typically expresses an emotional reaction, is not syntactically related to other accompanying expressions, and may include a combination of sounds not otherwise found in the language.
- Note that words primarily belonging to another part of speech retains their original category when used in exclamations. For example, God is a NOUN even in exclamatory uses.
- As a special case of interjections, we recognize feedback particles such as yes, no, uhuh, etc.
    - **bravo**
    - **hello**
    - **ouch**
    - **psst**

PUNCT : punctuation

- Punctuation marks are non-alphabetical characters and character groups used in many languages to delimit linguistic units in printed text.
- Spoken corpora contain symbols representing pauses, laughter and other sounds ; we treat them as punctuation, too.

SYM : Symbol

- A symbol is a word-like entity that differs from ordinary words by form, function, or both.
- Many symbols are or contain special non-alphanumeric characters, similarly to punctuation. What makes them different from punctuation is that they can be substituted by normal words. This involves all currency symbols, e.g. $ 75 is identical to seventy-five dollars.
- Another group of symbols is emoticons and emoji.
    - 130 **min**
    - 40 **%**

X : other

- The tag X is used for words that for some reason cannot be assigned a real part-of-speech category. It should be used very restrictively.
  - mots étrangers
  - puis il y a **xyz**