

Éléments pour un étiqueteur morpho- syntaxique

Violaine Prince
Université de Montpellier et LIRMM-CNRS

Jeu d'étiquettes grammaticales

CATEGORIE DES DETERMINANTS

1. DET :

- Déterminant
- Regroupe les sous-catégories suivantes :
 - Article : un, une, des, le, la, les, l'
 - Article partitif : de la, du, de l'
 - Exemple : je bois **de** l'eau, je mange **du** gratin et **de la** soupe.
 - Adjectif possessif : mon, ma, mes, ton, ta, tes, son, sa, ses, nos, vos, leur
 - Adjectif démonstratif : ce, cet, cette, ces
 - Adjectif numéral : les nombres utilisés en déterminant, comme dans « trois hommes dans un bateau »
 - Adjectif indéfini : aucun, aucune
 - Adjectif exclamatif : quel(s), quelle(s) comme dans « quel désastre ! »

CATEGORIE DES SUBSTANTIFS

2. NOMC

- Nom commun

3. NOMP

- Nom propre
- Peut servir aussi aux sigles ou abréviations.
- La distinction est faite parce que le second peut ne pas être précédé d'un déterminant. Les cas où le premier n'est pas précédé d'un déterminant sont soit :
 - Des expressions figées (proverbes), à traiter comme des chaînes lexicales à part éventuellement. Ex : « pierre qui roule n'amasse pas mousse ».
 - Une utilisation en adjectif du nom. Ex: « un gâteau maison ».

CATEGORIE DES PRONOMS

4. PRNSUJ

- Pronom sujet neutre : c', cela. « c'est génial »
- Pronom personnel sujet : je, tu, il, elle, nous, vous, ils, elles
- Pronom interrogatif : quel(s), quelle(s), qui, qu', lequel, lesquels, laquelle, lesquelles.
- Certains pronoms personnels accentués : lui, eux

5. PRNOBJ

- Pronom personnel objet : le, la, les.
- Pronom réfléchi : me, te, se, lui, leur
- Pronom neutre : en
- Pronoms personnels accentués : moi, toi, lui, elle(s), eux (après des prépositions) ou des verbes à l'impératif (moi, toi, lui, leur).

6. PRNREL

- Pronom introduisant une proposition relative: qui, que, qu', quoi, dont, où
- Locutions pronominales introduisant une proposition sujet ou objet :
ce que, ce qui, ce pourquoi, ce à quoi, au(x)quel(s), à laquelle, auxquelles, duquel, desquels, desquelles, de laquelle.

7. PRNPNM

- Pronom personnel accentué : moi, toi, lui, elle(s), eux.
 - Exemple : moi, je suis le professeur.
- Pronom possessif : mien(s), mienne(s), tien(s), tienne(s), sien (s), sienne (s), leur(s).
- Se comporte comme un nom.

La distinction est faite parce que le comportement syntaxique n'est pas le même.

Explication sur les pronoms

- Les pronoms sont des pointeurs. Les objets sur lesquels ils pointent sont appelés **antécédents**.
- Ils peuvent pointer
 - Sur un nom ou un groupe nominal :
 - Pronoms relatifs
 - L'enfant **qui** est assis à côté de moi.
 - La fille **dont** tu es le fiancé.
 - Le village **où** je suis né.
 - Locutions pronominales relatives
 - Le texte **auquel** je me réfère
 - Sur une personne (singulière ou plurielle) :
 - Pronoms personnels => anaphore pronominale.
 - Pronoms possessifs.
 - Sur une idée (groupe ou proposition) :
 - Pronom neutre : « un homme a marché sur la Lune c'est génial ».

CATEGORIE DES VERBES

8. VRB

- Verbe. On gardera cette étiquette pour les verbes à l'infinitif et pour les verbes conjugués à des temps simples.

9. AUX

- Verbe auxiliaire : avoir, être dans les formes composées.

10. PPAS

- Verbe au participe passé.
- Cela permet d'avoir des règles de réécriture de la forme
 - GV -> AUX PPAS
 - Et de reconnaître des verbes dans les temps composés.

CATEGORIE DES MODIFIEURS

11.ADJ

- Adjectifs qualificatifs. Associés aux noms
- Comprend aussi de nombreux participes présents.

12.ATT

- Adjectifs attributs. Associés au verbe « être ».

13.ADV

- Adverbes. Associés aux verbes ou aux phrases.
- Locutions adverbiales
- Associés aux autres verbes qu' être.

CATEGORIE DES PREPOSITIONS

14. PREP

- Prépositions : à, de, sur, vers, avec, sous, dans, sans, en, par, parmi, devant, derrière, après, avant, dès...
- Locutions prépositionnelles : grâce à, à cause de, afin de, en dessous de , au dessus de, en bas de, en haut de, à droite de, à gauche de, etc...
- On peut cependant utiliser un étiquetage plus fin pour certaines locutions : forme [PREP] NOMC PREP.
- Cas particulier des contractions :
 - Au, aux : « à le(s) »
 - Peuvent être traitées soit en expansion (on remplace au, aux par « à le, à les » => modification des chaînes terminales ou grammaire contexte dépendante !
 - Soit comme des prépositions et accepter des formes PREP NOM (ce qui est le cas pour « médecin de famille »).
 - Soit comme des déterminants
 - La MEILLEURE SOLUTION : proposer une étiquette DETPREP

CATEGORIE DES CONJONCTIONS

15.COOR

- Conjonctions de coordination :et, mais, ou, or, ni, car, donc

16.SUB

- Conjonction de subordination (d' une proposition à une autre) :
que, si
 - Je pense **que** tu as raison
 - **Que** tu aies raison est une évidence.
 - **Si** tu étais resté tu aurais pu voir le spectacle.
- Locution de subordination : afin que, pour que, de façon que, sans que, parce que, si tant est que, tant que, à moins que, attendu que, attendu que....

RECAPITULATIF

CATEGORIE	ETIQUETTE
DETERMINANTS	DET
SUBSTANTIFS	NOMC, NOMP
PRONOMS	PRNSUJ, PRNOBJ, PRNREL, PRNPNM
VERBES	VRB, AUX, PPAS
MODIFIEURS	ADJ, ATT, ADV
PREPOSITIONS	PREP, DETPREP
CONJONCTIONS	COOR, SUB

Etiquettes Tree-tagger pour le français (Achim Stein 2003)

- ABR abbreviation
- ADJ adjective
- ADV adverb
- DET:ART article
- DET:POS possessive pronoun (ma, ta, ...)
- INT interjection
- KON conjunction
- NAM proper name
- NOM noun
- NUM numeral

PRO	pronoun
PRO:DEM	demonstrative pronoun
PRO:IND	indefinite pronoun
PRO:PER	personal pronoun
PRO:POS	possessive pronoun (mien, tien, ...)
PRO:REL	relative pronoun
PRP	preposition
PRP:det	preposition plus article (au,du,aux,des)
PUN	punctuation
PUN:cit	punctuation citation
SENT	sentence tag
SYM	symbol
VER:cond	verb conditional
VER:futu	verb futur
VER:impe	verb imperative
VER:impf	verb imperfect
VER:infi	verb infinitive
VER:pper	verb past participle
VER:ppre	verb present participle
VER:pres	verb present
VER:simp	verb simple past
VER:subi	verb subjunctive imperfect
VER:subp	verb subjunctive present

ETIQUETTES DE SEGMENTATION

- Ponctuation
 - PONCTFIN
 - « . », « ! », « ? »
 - PONCTDOUBLE
 - Guillemets, parenthèses (droite, gauche), incises, la présence d'un élément impose celle d'un autre.
 - PONCTSEG
 - « : », « , », « ; »
 - PONCTCOOR
 - La ponctuation sert de coordination
 - « , », « - », « * »
 - PONCTTYPE
 - La ponctuation exprime le type de discours comme le tiret du dialogue.
- Rôle syntaxique de découpage en constituants.
- Dans une grammaire FN-CHOMSKY
 - $S \rightarrow \text{PROP PONCTFIN}$

ETIQUETTES SYNTAXIQUES

- La granularité syntaxique (*chunking*) est discutable d'une grammaire à l'autre.
- On propose les concepts suivants :
 - GROUPE NOMINAL : unité syntaxique dominée par un substantif ou un pronom ayant pour antécédent un groupe nominal avec un substantif ou une personne.
 - GROUPE VERBAL : unité syntaxique dominée par un verbe.
 - PROPOSITION : unité syntaxique correctement constituée. La constitution correcte suppose l'existence d'un groupe verbal et d'un groupe nominal sujet. Une proposition peut être principale ou subordonnée.
 - PHRASE : unité syntaxique marquée comme telle. Une phrase peut être constituée d'une ou plusieurs propositions (au moins une proposition principale) ou d'un groupe nominal. La phrase constitue en même temps une unité sémantique dans le discours.

CONSIGNES

- Définissez les étiquettes syntaxiques à partir des étiquettes grammaticales fournies en les justifiant.
- Vous pouvez utiliser des étiquettes de segmentation si vous le voulez.
- Vous pouvez introduire des notions de dépendance si vous le souhaitez sous forme d'étiquette. Cela dit, la dépendance introduit des contraintes. Ainsi, un pronom personnel objet ne peut pas être un GN dans la structure $S \rightarrow GN\ GV$
- \Rightarrow il existe des grammaires modélisées à l'aide de contraintes (grand domaine de l'I.A)

CONSIGNES (suite)

- Créez une grammaire de réécriture en FN-Chomsky à l'aide du jeu d'étiquettes grammaticales donné, et d'un jeu d'étiquettes syntaxiques que vous définissez
- Implémentez l'algorithme CYK fourni
- Fournissez une analyse avec CYK muni d'une sortie graphique (sous forme de matrice ou d'arbre, à votre choix) pour les cinq phrases suivantes :

PHRASES A ANALYSER

- Le bien qu'il fait, il le fait bien.
- Le fait est que j'ai peu de biens.
- J'ai bien du mal à faire ce que les autres font bien.
- Bien des gens ont fait un peu de bien.
- Bien qu'il ait fait du bien, il ne l' a pas bien fait.

Evaluation et conseils

- Vous serez appréciés pour :
 - La précision de l'affectation des catégories grammaticales
 - La qualité de vos étiquettes syntaxiques
 - La justesse de vos arbres syntaxiques
 - Votre capacité à juger le travail réalisé (pour et contre)
- Je vous conseille de :
 - Chercher sur internet toutes les catégories grammaticales possibles pour chaque mot
 - Consulter les références sur la grammaire pour écrire la grammaire de réécriture
 - Eventuellement, faire des recherches sur ce que feraient d'autres analyseurs superficiels (POS taggers).