

Guide d'annotation

*pour la création d'un corpus de
référence pour l'étiquetage
morpho-syntaxique*

**Corpus ESLO
2010**

Sommaire

Introduction	5
Les noms	6
➤ Noms communs.....	6
• Ambiguïté	6
• Noms composés	7
• Noms d'origine étrangère	7
• Problème de genre	7
➤ Noms propres	8
Les adjectifs.....	8
• Problèmes de genre.....	9
Les déterminants	10
➤ Déterminants démonstratifs.....	10
➤ Déterminants possessifs	10
○ Un possesseur	11
○ Plusieurs possesseurs	11
➤ Déterminants définis	12
• Article contracté	12
➤ Déterminants indéfinis	12
➤ Déterminants interrogatifs.....	14
Les pronoms	14
➤ Pronoms relatifs.....	14
➤ Pronoms interrogatifs	15
• Ambiguïté entre les pronoms relatifs et interrogatifs	16
➤ Pronoms personnels.....	17
• Ambiguïté de certains pronoms	18
• Répétition.....	19
➤ Pronoms démonstratifs	19
➤ Pronoms possessifs.....	20
• Un possesseur:	20
• Plusieurs possesseurs	21
➤ Pronoms indéfinis.....	21
Les verbes.....	23

➤	Indicatif	23
•	Présent	23
•	Futur.....	23
•	Imparfait	23
•	Passé simple.....	24
•	Conditionnel	24
•	Passé composé	24
➤	Subjonctif	25
➤	Impératif	25
•	Ambiguïté avec les marqueurs discursifs	25
➤	Infinitif	26
➤	Participe présent	26
•	Gérondif.....	26
➤	Voix passive	27
➤	Temps composés	27
•	Etiquette pour les temps composés.....	27
•	L'accord du participe passé	27
➤	Ambiguïté entre les participes et les adjectifs	28
Les adverbes		29
➤	Adverbes de négation	30
•	Constructions discontinues	30
•	Restriction.....	30
Les prépositions.....		31
•	Elision.....	31
Les conjonctions.....		32
➤	Conjonctions de coordination.....	32
•	Ambiguïté avec les marqueurs discursifs	32
➤	Conjonctions de subordination.....	33
•	Bien que / Bien + que	33
Les présentateurs		33
•	« C'est » et toutes ses formes	34
•	« Il y a » et toutes ses formes	35
•	Les problèmes de discontinuité	36

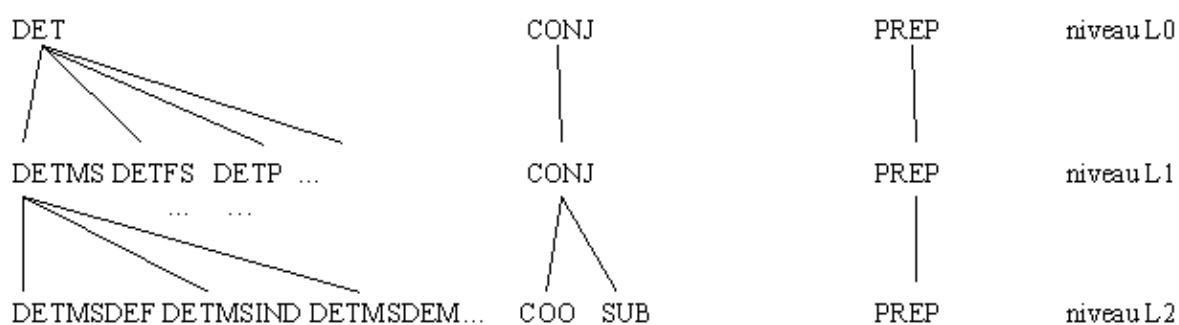
Les marqueurs discursifs	37
➤ Marqueurs discursifs	37
• Bon alors	38
• Ou	38
➤ Marqueurs discursifs « euh »	38
➤ Marqueurs discursifs « interjections »	38
Les expressions de temps	40
➤ Dates	40
➤ Heure	41
Les mots inconnus	42
Les chiffres	43
Les unités euphoniques	44
La ponctuation	44
Cas particuliers :	45
➤ Que	45
➤ Tout et ses variantes	46
➤ Bon : adjectif ou marqueur discursif :	47
➤ Constructions discontinues	47
• Négation	47
• Restriction	48
• Unités composées	48
• Croisement entre les constructions discontinues	48

Introduction

Ce guide présente l'ensemble des étiquettes définies pour l'étiquetage morpho-syntaxique d'un corpus oral ESLO.

L'étiquetage morpho-syntaxique est une étape fondamentale de son analyse, et un préliminaire à tout traitement de plus haut niveau. L'objectif de l'étiquetage est d'attribuer à chacun des mots d'un corpus une étiquette qui récapitule ses informations morpho-syntaxiques : partie du discours, genre, nombre, type, etc. Ce processus d'étiquetage est accompagné de celui de lemmatisation, dont l'objectif est de ramener l'occurrence d'un mot donné à sa forme de base ou lemme.

L'étiquetage proposé structure les étiquettes sur 3 niveaux : L0 (niveau des étiquettes sur les parties du discours), L1 (niveau des variantes morphologiques) et L2 (niveau syntaxico-sémantique) :



Certaines étiquettes restent les mêmes sur les 3 niveaux, d'autres ne changent qu'au deuxième, telle que verbes, et les dernières enfin intègrent chaque fois de nouvelles informations, comme les pronoms et déterminants.

Le résultat de l'étiquetage se présente sous format suivant :

	les	le		DET		DETP		DETPDEF
Unité lexicale présente dans le corpus	Son	lemme	Etiquette définissant la partie du discours :	Adjonction de l'information portant sur la morphologie :	de	Adjonction de l'information concernant la fonction de l'unité lexicale dans son emploi : il s'agit d'un déterminant pluriel défini		
	(forme canonique d'un mot variable)	d'un	déterminant	il s'agit d'un déterminant pluriel	il s'agit d'un déterminant au pluriel	il s'agit d'un déterminant pluriel défini		

Ce guide est un fruit de la correction manuelle de l'étiquetage fait par apprentissage (premier tests mai-juin 2009). L'objectif est construire le corpus de référence pour le développement

d'un étiqueteur par apprentissage avec des CRFs qui tiendra compte du nouveau jeu d'étiquettes établi plus approprié au corpus oral

Le corpus que nous voulons étiqueter est composé des fichiers de transcription. Cette-ci est orthographique mais n'a pas de marques typographiques sauf les tirets et les points d'interrogation. Les majuscules sont mis que pour les noms propres. La segmentation suit le changement du locuteur et des pauses longues. Le jeu d'étiquettes élaboré essaie de suivre les principes suivants :

- Tenir compte des spécificités de l'oral (étiquettes pour les marqueurs discursifs, le présentatif)
- Essayer d'éviter les ambiguïtés
- Essayer de simplifier les étiquettes là où ca peut poser un problème pour l'étiquetage automatique (où l'information n'est pas homogène comme le cas du genre et du nombre pour les noms propres)
- Essayer ne pas augmenter le nombre d'étiquettes (pas d'étiquette pour les articles partitifs, une seule étiquettes pour tous les adjectifs)

Les noms

Nous distinguons deux types de noms.

➤ Noms communs

Les lemmes sont par défaut à la forme du masculin singulier.

NCMS Nom Commun Masculin Singulier
NCMP Nom Commun Masculin Pluriel
NCFS Nom Commun Féminin Singulier
NCFP Nom Commun Féminin Pluriel

l'écriture de mon mari

écriture	écriture	N	NCFS	NCFS
mari	mari	N	NCMS	NCMS

Difficultés de l'étiquetage :

- **Ambiguïté**

Dans le cas d'ambiguïté entre le nom et une autre POS, c'est la possibilité de restituer le déterminant qui permet de résoudre ce problème :

ne pas parler (le) français

français	français	N	NCMS	NCMS
----------	----------	---	------	------

faire latin
latin N NCMS NCMS

Par rajout d'un déterminant, n'importe quel mot peut devenir un nom.

- Des verbes

il n'y a pas de parler patois
parler N NCMS NCMS

- Des adverbes

un peut-être
peut-être N NCMS NCMS

- Des pronoms personnels

mon moi intérieur
moi N NCMS NCMS

- Des conjonctions

avec des si on referait le monde
si N NCMP NCMP

- **Noms composés**

Les noms composés font partie des noms communs et sont étiquetés sous une seule étiquette.

Le goût du porte-plume
porte-plume N NCMS NCMS

- **Noms d'origine étrangère**

Les noms étrangers portent également l'étiquette de nom.

Je vais au snack
Snack N NCMS NCMS

- **Problème de genre**

Pour les noms pouvant être soit masculins, soit féminins, l'étiquette est par défaut au masculin.

➤ Noms propres

Les noms propres sont marqués dans les conventions de transcription par majuscule. Pour simplifier l'étiquetage automatique et éviter les erreurs liées aux problèmes de non homogénéité du genre et du nombre de noms propres (la Loire / Orléans) nous avons décidé ne pas mentionner cette information. Le lemme est identique à la forme prononcée.

NP Nom Propre

Quelques exemples de noms propres :

- Un prénom suivi d'un nom : *Victor Hugo*
- Un nom composé : *Vals les Bains*
- Les marques (avec ou non des références chiffrées antéposées ou postposées) :
Citroën, Peugeot 404 ou *404 Peugeot, Tefal*

[Retour au sommaire](#)

Les adjectifs

L'adjectif est un mot qui varie en genre et en nombre. Il s'accorde avec le nom auquel il se rapporte ainsi il va toujours porter cette information. L'adjectif est au masculin dans sa forme de base. Nous n'avons qu'une seule étiquette pour tous les types d'adjectifs.

L'étiquetage :

L'étiquette de l'adjectif est l'abréviation employée par les grammaires françaises *ADJ*.

La marque en genre, correspondant au masculin est notée *M* et celle du féminin *F*.

La marque du singulier est notée *S* et celle du pluriel *P*.

ADJMS Adjectif Masculin Singulier ADJMP Adjectif Masculin Pluriel ADJFS Adjectif Féminin Singulier ADJFP Adjectif Féminin Pluriel
--

Remarque :

Nous ne tiendrons pas compte des adjectifs appelés parfois invariables dont la forme ne varie ni au féminin ni au singulier, car même s'ils sont invariables au niveau morphologique et ne portent pas de marques « visibles » sur leur forme, ils sont toujours employés avec le nom/pronom avec qui ils s'accordent.

L'école primaire est indispensable (ADJFS) [...] c'est indispensable (ADJMS). »

Le premier emploi de l'adjectif « indispensable » s'accorde avec « l'école primaire », alors que le deuxième avec le présentatif « c'est », qui est par défaut masculin singulier (voir ci-dessous).

Difficultés de l'étiquetage :

- **Problèmes de genre**

L'adjectif précédé du présentatif « c'est » :

« C'est étrange », « c'est bizarre », « c'est formidable », etc. Ces adjectifs sont tous au masculin singulier car ils s'accordent avec le sujet de la phrase, « ce » qui est au masculin singulier par défaut (la preuve est le masculin singulier du verbe être qui s'accorde avec lui). Cependant, lorsqu'il y a un nom qui vient s'ajouter derrière l'adjectif, c'est avec celui-ci que l'adjectif s'accorde : « C'est encore autre chose », « c'est autre chose »

C'est autre (ADJFS) chose (NCFS)

Le genre du locuteur

Comme l'étiquetage a été corrigé manuellement, nous pouvions tenir compte des informations dans le contexte sur le sexe du locuteur. Ainsi, si nous avons observé dans le contexte l'énoncé « Quand j'étais jeune fille » qui dévoile qu'il s'agit d'une locutrice, dans la question lui posée « Seriez-vous favorable à une réforme de l'orthographe ? » : l'adjectif « favorable » est étiqueté au féminin car la question est posée à une femme. Sans ces informations, nous avons étiqueté les adjectifs employés dans les questions aux locuteurs comme dans « Etes-vous favorable ? », au Masculin Singulier même si le verbe est au pluriel.

L'accord avec le Nom Propre

Dans la phrase « Orléans est calme », comme le nom de ville est par défaut au masculin singulier l'adjectif sera étiqueté pareil.

Les déterminants

Nous distinguons cinq classes de déterminants.

➤ Déterminants démonstratifs

Ils sont variables en genre et en nombre.

DETMSEDEM	Déterminant Masculin Singulier Démonstratif (ce, cet)
DETFSEDEM	Déterminant Féminin Singulier Démonstratif (cette)
DETPDEM	Déterminant Pluriel Démonstratif (ces)

Le lemme de toutes ces déterminants démonstratifs est la forme au masculin singulier *ce*.

*elle était bien muette pendant **ce** temps-là*

ce ce DET DETMS DETMSEDEM

*il a une culture artistique formidable **cet** homme*

cet ce DET DETMS DETMSEDEM

*ça a été mal géré toute **cette** affaire*

cette ce DET DETFS DETFSEDEM

*tous **ces** mots sont francisés*

ces ce DET DETMP DETMPDEM

➤ Déterminants possessifs

Dans l'étiquetage de ce type de déterminants nous avons tenu compte du possesseur (celui qui possède) d'où sa variabilité en personne et en nombre, et de l'objet possédé, le nom qui suit, et avec qui il s'accorde également en genre et en nombre.

Qui possède	Objet possédé		
	SINGULIER		PLURIEL
	MASC	FEM	
1s	Mon	ma	Mes
2s	Ton	Ta	Tes
3s	Son	Sa	Ses
1p	Notre		Nos

2p	Votre	Vos
3p	Leur	Leurs

○ **Un possesseur:**

Le lemme est toujours à la forme du masculin singulier

▪ **Un objet :**

DETMS1SPOSS	Déterminant Masculin Singulier 1 ^{ère} Personne Singulier Possessif (mon)
DETMS2SPOSS	Déterminant Masculin Singulier 2 ^{ème} Personne Singulier Possessif (ton)
DETMS3SPOSS	Déterminant Masculin Singulier 3 ^{ème} Personne Singulier Possessif (son)
DETFMS1SPOSS	Déterminant Féminin Singulier 1 ^{ère} Personne Singulier Possessif (ma)
DETFMS2SPOSS	Déterminant Féminin Singulier 2 ^{ème} Personne Singulier Possessif (ta)
DETFMS3SPOSS	Déterminant Féminin Singulier 3 ^{ème} Personne Singulier Possessif (sa)

▪ **Plusieurs objets :**

DETP1SPOSS	Déterminant Pluriel 1 ^{ère} Personne Singulier Possessif (mes)
DETP2SPOSS	Déterminant Pluriel 2 ^{ème} Personne Singulier Possessif (tes)
DETP3SPOSS	Déterminant Pluriel 3 ^{ème} Personne Singulier Possessif (ses)

○ **Plusieurs possesseurs :**

Le lemme est toujours à la forme du masculin singulier.

▪ **Un objet :**

DETS1PPOSS	Déterminant Singulier 1 ^{ère} Personne Pluriel Possessif (notre)
DETS2PPOSS	Déterminant Singulier 2 ^{ème} Personne Pluriel Possessif (votre)
DETS3PPOSS	Déterminant Singulier 3 ^{ème} Personne Pluriel Possessif (leur)

▪ **Plusieurs objets :**

DETP1PPOSS	Déterminant Pluriel 1 ^{ère} Personne Pluriel Possessif (nos)
DETP2PPOSS	Déterminant Pluriel 2 ^{ème} Personne Pluriel Possessif (vos)
DETP3PPOSS	Déterminant Pluriel 3 ^{ème} Personne Pluriel Possessif (leurs)

*je m'occupe de **mon** intérieur*

mon mon DET DETMS1S DETMS1SPOSS

ses études primaires

ses son DET DETP3S DETP3SPOSS

notre vie

notre notre DET DETS1P DETS1POSS

vous avez terminé vos études

vos votre DET DETP2P DETP2POSS

➤ Déterminants définis

DETMSDEF	Déterminant Masculin Singulier Défini (le, l')
DETFMSDEF	Déterminant Féminin Singulier Défini (la, l')
DETPDEF	Déterminant Pluriel Défini (les)

- **Article contracté**

Certains déterminants définis sont contractés avec la préposition « à » ou « de ». Dans ce cas, on sépare les deux termes que l'on étiquette séparément.

mon mari travaillait au tribunal (→ mon mari travaillait à le tribunal)

à à PREP PREP PREP
le le DET DETMS DETMSDEF

l'affaire des enseignants (→ l'affaire de les enseignants)

de de PREP PREP PREP
les le DET DETP DETPDEF

➤ Déterminants indéfinis

Ils déterminent « soit une quantité non chiffrée, soit une identification imprécise ou même un refus d'identification » (Le Bon Usage).

Il est difficile de les lister tous.

On trouvera parmi eux les articles indéfinis (un, une, des), les articles partitifs (du, de la, des), les unités lexicales appelés par certains grammairiens « adjectifs indéfinis » (aucun, certain, quelque, etc.), et des locutions incluant souvent la quantité (beaucoup de, pas mal de, etc.).

Certains sont variables en genre et en nombre (dans ce cas le lemme sera la forme du masculin singulier), d'autres sont invariables car d'origine adverbiale (le lemme sera donc identique). Dans ce dernier cas, ils se différencient de l'emploi adverbial par le fait qu'ils sont suivis de la préposition « de » et d'un nom (« beaucoup de », « peu de », « assez de », « tellement de », « pas mal de », « trop de » etc).

DETMSIND	Déterminant Masculin Singulier Indéfini
DETFSIND	Déterminant Féminin Singulier Indéfini
DETPIND	Déterminant Masculin Pluriel Indéfini
DETIIND	Déterminant Invariable Indéfini (déterminants d'origine adverbiale)

*est-ce que vous faites **un** brouillon*
un un DET DETMS DETMSIND

***des** amis*
des un DET DETP DETMPIND

*on prend n'importe quoi oui **n'importe quel** papier*
n'importe_quel n'importe_quel DET DETMS DETMSIND

→ Il faut différencier « n'importe quoi » qui est le pronom indéfini, pour nous (PMSIND) du déterminant indéfini « n'importe quel ».

*on prend **une** poêle Téfal*
une un DET DETFS DETFSIND

***quelques** jours peut-être*
quelques quelque DET DETMP DETMPIND

*ils nous dépassent en quantité dans **certaines** choses*
certaines certain DET DETFP DETFPIND

*on a **pas mal de** bouquins*
pas_mal_de pas_mal_de DET DETI DETIIND

*il n'y a pas **assez d'**étudiants qui cherchent des places*
assez_d' assez_de DET DETI DETIIND

*il n'y a pas vraiment **beaucoup de** choix*
beaucoup_de beaucoup_de DET DETI DETIIND

*y a **trop de** circulation*
trop_de trop_de DET DETI DETIIND

Parmi les déterminants invariables figure le « de » après la négation.

<i>on</i>	<i>on</i>	<i>P</i>	<i>P3MS</i>	<i>P3MSPERSUJ</i>
<i>n'</i>	<i>ne</i>	<i>ADV</i>	<i>ADV</i>	<i>ADVNEGDISC1</i>
<i>a</i>	<i>avoir</i>	<i>V</i>	<i>V3SINDP</i>	<i>V3SINDPAUX</i>
<i>jamais</i>	<i>jamais</i>	<i>ADV</i>	<i>ADV</i>	<i>ADVNEGDISC2</i>
<i>pris</i>	<i>prendre</i>	<i>V</i>	<i>VMSPP</i>	<i>VMSPP</i>
<i>de</i>	<i>de</i>	<i>DET</i>	<i>DETI</i>	<i>DETIIND</i>
<i>vacances</i>	<i>vacances</i>	<i>N</i>	<i>NCFP</i>	<i>NCFP</i>

ainsi que ce qu'on appelle les articles partitifs :

<i>vo</i>	<i>vo</i>	<i>DET</i>	<i>DETS2P</i>	<i>DETS2PPOSS</i>
<i>mari</i>	<i>mari</i>	<i>N</i>	<i>NCMS</i>	<i>NCMS</i>
<i>a</i>	<i>avoir</i>	<i>V</i>	<i>V3SINDP</i>	<i>V3SINDP</i>
<i>de la</i>	<i>de la</i>	<i>DET</i>	<i>DETFS</i>	<i>DETFSIND</i>
<i>famille</i>	<i>famille</i>	<i>N</i>	<i>NCFS</i>	<i>NCFS</i>

➤ Déterminants interrogatifs

Ils sont au nombre de quatre : « quel », « quelle », « quels » et « quelles ».

Le lemme est la forme du masculin singulier « quel ».

DETMSINT Déterminant Masculin Singulier Interrogatif
DETFSINT Déterminant Féminin Singulier Interrogatif
DETMPINT Déterminant Masculin Pluriel Interrogatif
DETFPINT Déterminant Féminin Pluriel Interrogatif

quel genre de papier

quel quel DET DETMS DETMSINT

quels genres de choses

quels quel DET DETMP DETMPINT

à la suite de *quelles* circonstances

quelles quel DET DETFP DETFPINT

[Retour au sommaire](#)

Les pronoms

Il existe plusieurs catégories de pronoms. Nous les avons divisés en six.

➤ Pronoms relatifs

Les pronoms relatifs s'accordent en genre, en nombre et en personne avec leur antécédent. Cependant, certains ne portent pas eux-mêmes la marque de cet accord et restent donc invariables.

Nous les appellerons donc *pronoms invariables relatifs*. Ils sont au nombre de cinq :

- *qui*
- *que*

- *quoi*
- *dont*
- *où*

Le pronom porte dans chacun des cas son initiale comme étiquette *P*.

Il n'y a pas de marque en genre et en nombre mais le pronom reste toujours invariable dans sa forme *I*.

Ce pronom est dit relatif, nous ajoutons donc à son étiquette cette information identifiée par les trois premières lettres *REL*.

PIREL	Pronom Invariable relatif
-------	---------------------------

Les autres pronoms relatifs peuvent au contraire porter la marque de l'accord en genre et en nombre.

Nous les appellerons également *pronoms relatifs* mais nous ajoutons les indications de genre et de nombre. Ces pronoms sont au nombre de quatre :

- lequel
- lesquels
- laquelle
- lesquelles

Les pronoms variables en genre et en nombre sont étiquetés différemment :

On remplace la marque de l'invariable par les suivantes :

- La marque en genre, correspondant au masculin est notée *M* et celle du féminin *F*.
- La marque du singulier est notée *S* et celle du pluriel *P*.

PMSREL Pronom Masculin Singulier Relatif
PMPREL Pronom Masculin Pluriel Relatif
PFSREL Pronom Féminin Singulier Relatif
PFPREL Pronom Féminin Pluriel Relatif

les milieux dans lesquels ils sont
 lesquels lequel P PMP PMPREL

Remarque : Les pronoms composés tels que « auquel, etc. » doivent être décomposés en :

à PREP+lequel

➤ Pronoms interrogatifs

Les pronoms interrogatifs varient en genre et en nombre mais aussi d'après leur fonction syntaxique dans la phrase (sujet, objet). Cependant, comme dans le cas des pronoms relatifs, certains ne portent pas sur eux la marque de cet accord et restent donc invariables.

Nous les appellerons *pronoms invariables interrogatifs* :

- qui, que, quoi
- qui est-ce que, qui est-ce qui
- qu'est-ce que, qu'est-ce qui, qu'est-ce qu'

L'étiquette est identique à celle des *pronoms invariables relatifs* sauf qu'il s'agit des *pronoms invariables interrogatifs*.

PIINT	Pronom Invariable Interrogatif
-------	--------------------------------

Comme les pronoms relatifs, les autres pronoms interrogatifs portent la marque de l'accord en genre et en nombre.

Nous ajoutons également les indications de genre et de nombre pour marquer la variabilité. Ces pronoms sont les mêmes que les pronoms relatifs (attention donc à l'ambiguïté !!!) :

- lequel
- lesquels
- laquelle
- lesquelles

Les pronoms variables en genre et en nombre sont étiquetés différemment :

On remplace la marque de l'invariable par les suivantes :

- La marque en genre, correspondant au masculin est notée *M* et celle du féminin *F*.
- La marque du singulier est notée *S* et celle du pluriel *P*.

PMSINT Pronom Masculin Singulier Interrogatif
PMPINT Pronom Masculin Pluriel Interrogatif
PFSINT Pronom Féminin Singulier Interrogatif
PFPIINT Pronom Féminin Pluriel Interrogatif

- **Ambiguïté entre les pronoms relatifs et interrogatifs**

Il faut faire attention à l'ambiguïté des formes « qui, que, quoi, lequel et ses variantes ».

On sait qu'à l'écrit le pronom interrogatif est placé souvent en début de phrase « *Lequel* de ces plats veux-tu ? ». On peut également le retrouver seul, en mot phrase dans une conversation : « Je possède déjà l'un de ces tableaux. – *Lequel* ? ».

Cependant, à l'oral, où la notion de phrase est mise déjà en question et où les marques typographiques (comme le point ou la majuscule pour les débuts de phrases) sont absentes, les deux emplois sont plus difficiles à distinguer. Ces pronoms interrogatifs peuvent être employés n'importe où dans l'énoncé « C'est dans une boîte. D'accord, mais dans *laquelle* ? » ou « dans *laquelle* c'est ? ».

Nous sommes conscients que cette ambiguïté peut poser des problèmes à l'étiquetage automatique mais nous avons voulu quand même de garder cette distinction.

➤ Pronoms personnels

Les pronoms personnels varient en personne, en genre, en nombre mais aussi selon leur fonction syntaxique dans l'énoncé (sujet, complément direct, indirect, circonstanciel, etc.). Au début, nous avons voulu tenir compte de toutes ces informations dans les étiquettes, mais en nous apercevant de nombreuses ambiguïtés que ça pourrait provoquer

Nous regardons (sujet)

Nous nous regardons (complément direct, réfléchi)

Ils nous regardent (complément direct)

Ils parlent avec nous (complément indirect)

etc.

nous avons renoncé, nous nous contentons de la distinction sujet / complément / adverbial pour les formes clitiques. Voilà le tableau récapitulatif des informations dont nous avons tenu compte :

Tableau des pronoms personnels :

		SUJET	COMPLEMENT	ADVERBIAL	TONIQUE
1s		Je	me, m'		moi, moi-même
2s		Tu	te, t'		toi, toi-même
3s	masc	il, on	le		lui, lui-même
			l', lui		
	fem	Elle	la		elle, elle-même
			se		soi, soi-même
1p		Nous	nous		nous, nous-mêmes
2p		Vous	vous		vous, vous-mêmes
3p	masc	Ils			eux, eux-mêmes
	fem	Elles			elles, elles-mêmes
			les, leur		
			se		
			en, y	en, y	

Les pronoms personnels sujets :

P1SPERSUJ	Pronom 1 ^{ère} Personne Singulier Personnel Sujet (je)
P2SPERSUJ	Pronom 2 ^{ème} Personne Singulier Personnel Sujet (tu, t')
P3MSPERSUJ	Pronom 3 ^{ème} Personne Masculin Singulier Personnel Sujet (il, on)
P3FSPERSUJ	Pronom 3 ^{ème} Personne Féminin Singulier Personnel Sujet (elle)
P1PPERSUJ	Pronom 1 ^{ère} Personne Pluriel Personnel Sujet (nous)
P2PPERSUJ	Pronom 2 ^{ème} Personne Pluriel Personnel Sujet (vous)
P3MPPERSUJ	Pronom 3 ^{ème} Personne Masculin Pluriel Personnel Sujet (ils)

P3FPPERSUJ	Pronom 3 ^{ème} Personne Féminin Pluriel Personnel Sujet (elles)
------------	--

Les pronoms personnels compléments :

P1SPERCOMPL	Pronom 1 ^{ère} Personne Singulier Personnel Complément (me, m')
P2SPERCOMPL	Pronom 2 ^{ème} Personne Singulier Personnel Complément (te, t')
P3MSPERCOMPL	Pronom 3 ^{ème} Personne Masculin Singulier Personnel Complément (le)
P3FSPERCOMPL	Pronom 3 ^{ème} Personne Féminin Singulier Personnel Complément (la)
P3SPERCOMPL	Pronom 3 ^{ème} Personne Singulier Personnel Complément (lui, l')
P1PPERCOMPL	Pronom 1 ^{ère} Personne Pluriel Personnel Complément (nous)
P2PPERCOMPL	Pronom 2 ^{ème} Personne Pluriel Personnel Complément (vous)
P3PPERCOMPL	Pronom 3 ^{ème} Personne Pluriel Personnel Complément (les, leur)
P3SPERCOMPL	Pronom 3 ^{ème} Personne Singulier Personnel Complément (en, y)

Remarque : Dans le cas des pronoms personnels compléments « les, lui et leur », nous avons décidé de retirer l'information du genre masculin indiquée à l'origine dans l'étiquette car il peut aussi s'agir d'un féminin. En effet, le pronom ne porte pas la marque du genre sur lui.

Les pronoms personnels adverbiaux :

P3SPERADV	Pronom 3 ^{ème} Personne Singulier Personnel Adverbial (en, y)
-----------	--

Les pronoms personnels toniques :

P1SPERTON	Pronom 1 ^{ère} Personne Singulier Personnel Tonique (moi, moi-même)
P2SPERTON	Pronom 2 ^{ème} Personne Singulier Personnel Tonique (toi, toi-même)
P3MSPERTON	Pronom 3 ^{ème} Personne Masculin Singulier Personnel Tonique (lui, lui-même)
P3FSPERTON	Pronom 3 ^{ème} Personne Féminin Singulier Personnel Tonique (elle, elle-même)
P3SPERTON	Pronom 3 ^{ème} Personne Singulier Personnel Tonique (soi, soi-même)
P1PPERTON	Pronom 1 ^{ère} Personne Pluriel Personnel Tonique (nous, nous-mêmes)
P2PPERTON	Pronom 2 ^{ème} Personne Pluriel Personnel Tonique (vous, vous-mêmes)
P3MPPERTON	Pronom 3 ^{ème} Personne Masculin Pluriel Personnel Tonique (eux, eux-mêmes)
P3FPPERTON	Pronom 3 ^{ème} Personne Féminin Pluriel Personnel Tonique (elles, elles-mêmes)

- **Ambiguïté de certains pronoms :**

Comme nous l'avons remarqué ci-dessus, certains pronoms ont plusieurs étiquettes possibles. En effet, les pronoms tels que « nous, vous » peuvent être sujet, complément ou tonique, ce qui peut poser un problème lors de leur étiquetage.

Lorsqu'ils sont utilisés après une préposition, cela ne pose pas de problème, ils sont forcément toniques, « est-ce que vous aviez un dictionnaire *chez vous* », mais malheureusement le contexte ne peut pas toujours être si tranchant.

- **Répétition**

Cela devient un vrai problème dans les cas de répétition des pronoms :

« ça traduit toute une euh *nous nous* n'étions pas comme ça »

ou encore

« *vous vous* savez »

Il est difficile d'identifier la catégorie des deux termes identiques qui se suivent : s'agit-il d'une simple répétition du mot qui traduirait une hésitation ou bien d'un pronom personnel tonique suivi d'un pronom personnel sujet. Pour résoudre partiellement ce problème, il faut écouter les fichiers son, mais ce n'est pas possible avec l'étiquetage automatique. Le problème pourrait peut-être être résolu avec la post-correction manuelle, en ne recherchant que ces répétitions.

➤ **Pronoms démonstratifs**

Il existe dans la liste des pronoms démonstratifs, des pronoms invariables :

- *ça*
- *ce*
- *cela*
- *ceci*

Ils portent l'étiquette du pronom *P*, *I* pour invariable et *DEM* pour démonstratif.

PIDEM	Pronom Invariable Démonstratif
-------	--------------------------------

Remarque : « *ce* » ne sera pas un pronom démonstratif dans le cas de « *ce sont* » qui, dans ce cas, correspond au présentateur « *c'est* ».

Les autres pronoms démonstratifs peuvent au contraire porter la marque de l'accord en genre et en nombre.

Nous les appellerons également *pronoms démonstratifs* mais nous ajoutons les indications de genre et de nombre. Ces pronoms sont au nombre de douze :

- celui, celui-ci, celui-là
- ceux, ceux-ci, ceux-là
- celle, celle-ci, celle-là
- celles, celles-ci, celles-là

Les pronoms variables en genre et en nombre sont étiquetés différemment :

On remplace la marque de l'invariable par les suivantes :

- La marque en genre, correspondant au masculin est notée *M* et celle du féminin *F*.

- La marque du singulier est notée *S* et celle du pluriel *P*.

PMSDEM	Pronom Masculin Singulier Démonstratif
PMPDEM	Pronom Masculin Pluriel Démonstratif
PFSDEM	Pronom Féminin Singulier Démonstratif
PPFDEM	Pronom Féminin Pluriel Démonstratif

c'était celui qui vivait largement quoi
celui celui P PMS PMSDEM

➤ Pronoms possessifs

Les pronoms possessifs comme les déterminants possessifs varient en personne et en nombre du/des possesseur(s) et en genre et en nombre de l'objet possédé.

Tableau des pronoms possessifs :

Qui possède 2	Objet possédé 1			
	SINGULIER		PLURIEL	
	MASC	FEM	MASC	FEM
1s	le mien	la mienne	les miens	les miennes
2s	le tien	la tienne	les tiens	les tiennes
3s	le sien	la sienne	les siens	les siennes
1p	le nôtre	la nôtre	les nôtres	
2p	le vôtre	la vôtre	les vôtres	
3p	le leur	la leur	les leurs	

- **Un possesseur:**

- **Un objet :**

PMS1SPOSS	Pronom Masculin Singulier 1 ^{ère} Personne Singulier Possessif (le mien)
PFS1SPOSS	Pronom Féminin Singulier 1 ^{ère} Personne Singulier Possessif (la mienne)
PMS2SPOSS	Pronom Masculin Singulier 2 ^{ème} Personne Singulier Possessif (le tien)
PFS2SPOSS	Pronom Féminin Singulier 2 ^{ème} Personne Singulier Possessif (la tienne)
PMS3SPOSS	Pronom Masculin Singulier 3 ^{ème} Personne Singulier Possessif (le sien)
PFS3SPOSS	Pronom Féminin Singulier 3 ^{ème} Personne Singulier Possessif (la sienne)

- **Plusieurs objets :**

PMP1SPOSS	Pronom Masculin Pluriel 1 ^{ère} Personne Singulier Possessif (les miens)
PFP1SPOSS	Pronom Féminin Pluriel 1 ^{ère} Personne Singulier Possessif (les miennes)
PMP2SPOSS	Pronom Masculin Pluriel 2 ^{ème} Personne Singulier Possessif (les tiens)
PFP2SPOSS	Pronom Féminin Pluriel 2 ^{ème} Personne Singulier Possessif (les tiennes)
PMP3SPOSS	Pronom Masculin Pluriel 3 ^{ème} Personne Singulier Possessif (les siens)
PFP3SPOSS	Pronom Féminin Pluriel 3 ^{ème} Personne Singulier Possessif (les siennes)

- **Plusieurs possesseurs :**

- **Un objet :**

PMS1PPOSS	Pronom Masculin Singulier 1 ^{ère} Personne Pluriel Possessif (le nôtre)
PFS1PPOSS	Pronom Féminin Singulier 1 ^{ère} Personne Pluriel Possessif (la nôtre)
PMS2PPOSS	Pronom Masculin Singulier 2 ^{ème} Personne Pluriel Possessif (le vôtre)
PFS2PPOSS	Pronom Féminin Singulier 2 ^{ème} Personne Pluriel Possessif (la vôtre)
PMS3PPOSS	Pronom Masculin Singulier 3 ^{ème} Personne Pluriel Possessif (le leur)
PFS3PPOSS	Pronom Féminin Singulier 3 ^{ème} Personne Pluriel Possessif (la leur)

- **Plusieurs objets :**

PP1PPOSS	Pronom Pluriel 1 ^{ère} Personne Pluriel Possessif (les nôtres)
PP2PPOSS	Pronom Pluriel 2 ^{ème} Personne Pluriel Possessif (les vôtres)
PP3PPOSS	Pronom Pluriel 3 ^{ème} Personne Pluriel Possessif (les leurs)

➤ **Pronoms indéfinis**

Les pronoms indéfinis sont nombreux. En voici une liste réalisée à partir des fichiers de corpus :

Les pronoms masculins singuliers :

- | | | |
|-----------------|----------------------|--------------------|
| - certain | - tout | - autre |
| - aucun | - un autre | - un |
| - nul | - quiconque | - grand-chose |
| - le même | - tel | - tout le monde |
| - personne | - chacun | - n'importe quoi |
| - quelque chose | - l'un l'autre | - quoi que ce soit |
| - quelqu'un | - ni l'un ni l'autre | |
| - rien | - l'un et l'autre | |

« on peut faire *quelque chose* »

« c'est pas *grand-chose* »

« *tout* est très bien le printemps »

« on est allé à Fays-aux-Loges là tout à côté voir *quelqu'un* »

« *rien* mais *rien* »

Les pronoms masculins pluriels :

- certains
- les uns les autres
- tous
- autres
- quelques-uns
- plusieurs
- beaucoup

« c'est *tous* des enfants catholiques »
 « y'a *quelques-uns* de nos étudiants »

Remarque : le pronom « beaucoup » est à différencier du déterminant « beaucoup de ».

Les pronoms féminins singuliers :

- certaine
- chacune
- aucune
- une

Les pronoms féminins pluriels :

- certaines
- aucunes
- quelques-unes
- les unes les autres
- toute

Remarque : quelques-uns / quelques-unes ont comme lemme « quelqu'un »

Le pronom porte dans chacun des cas son initiale comme étiquette *P*.

Les autres pronoms indéfinis portent la marque de l'accord en genre et en nombre indiquées par *MS*, *MP*, *FS* et *FP*.

Ce pronom est dit indéfini, nous ajoutons donc à son étiquette cette information identifiée par les trois premières lettres *IND*.

PMSIND	Pronom Masculin Singulier Indéfini
PMPIND	Pronom Masculin Pluriel Indéfini
PFSIND	Pronom Féminin Singulier Indéfini
PFPIIND	Pronom Féminin Pluriel Indéfini

on prend n'importe quoi

n'importe_quoi n'importe_quoi P PMS PMSIND

je les garde toutes

toutes tout P PFP PFPIND

[Retour au sommaire](#)

Les verbes

Les verbes varient en mode, en temps, en personne et en nombre. Leurs lemmes sont à l'infinitif.

➤ Indicatif

Tous les temps de ce mode possèdent leurs propres étiquettes sauf les temps composés et surcomposés autres que le passé composé.

- **Présent**

V1SINDP Verbe 1ere personne du Singulier Indicatif Présent
V2SINDP Verbe 2e personne du Singulier Indicatif Présent
V3SINDP Verbe 3e personne du Singulier Indicatif Présent
V1PINDP Verbe 1ere personne du Pluriel Indicatif Présent
V2PINDP Verbe 2e personne du Pluriel Indicatif Présent
V3PINDP Verbe 3e personne du Pluriel Indicatif Présent

on parle plus de latin

parle parler V V3SINDP V3SINDP

- **Futur**

V1SINDF Verbe 1ere personne du Singulier Indicatif Futur
V2SINDF Verbe 2e personne du Singulier Indicatif Futur
V3SINDF Verbe 3e personne du Singulier Indicatif Futur
V1PINDF Verbe 1ere personne du Pluriel Indicatif Futur
V2PINDF Verbe 2e personne du Pluriel Indicatif Futur
V3PINDF Verbe 3e personne du Pluriel Indicatif Futur

ça n'existera plus

existera exister V V3SINDF V3SINDF

- **Imparfait**

V1SINDI Verbe 1ere personne du Singulier Indicatif Imparfait
V2SINDI Verbe 2e personne du Singulier Indicatif Imparfait
V3SINDI Verbe 3e personne du Singulier Indicatif Imparfait
V1PINDI Verbe 1ere personne du Pluriel Indicatif Imparfait

V2PINDI Verbe 2e personne du Pluriel Indicatif Imparfait
V3PINDI Verbe 3e personne du Pluriel Indicatif Imparfait

j'étais très bonne en maths
étais être V V1SINDI V1SINDI

- **Passé simple**

V1SINDPS Verbe 1ere personne du Singulier Indicatif Passé Simple
V2SINDPS Verbe 2e personne du Singulier Indicatif Passé Simple
V3SINDPS Verbe 3e personne du Singulier Indicatif Passé Simple
V1PINDPS Verbe 1ere personne du Pluriel Indicatif Passé Simple
V2PINDPS Verbe 2e personne du Pluriel Indicatif Passé Simple
V3PINDPS Verbe 3e personne du Pluriel Indicatif Passé Simple

Pas d'exemples dans le corpus étiqueté.

- **Conditionnel**

Souvent considéré comme un mode, nous le présentons ici en tant que temps de l'indicatif.

V1SINDCON Verbe 1ere personne du Singulier Indicatif Conditionnel
V2SINDCON Verbe 2e personne du Singulier Indicatif Conditionnel
V3SINDCON Verbe 3e personne du Singulier Indicatif Conditionnel
V1PINDCON Verbe 1ere personne du Pluriel Indicatif Conditionnel
V2PINDCON Verbe 2e personne du Pluriel Indicatif Conditionnel
V3PINDCON Verbe 3e personne du Pluriel Indicatif Conditionnel

qu'est-ce qu'on devrait apprendre surtout aux enfants
devrait devoir V V3SINDCON V3SINDCON

- **Passé composé**

Ce temps est composé d'un auxiliaire (« avoir » ou « être ») conjugué au présent de l'indicatif suivi d'un participe passé, les deux étant étiquetés sur deux lignes différentes.

V1SINDPAUX Verbe 1ere personne du Singulier Indicatif Présent Auxiliaire
V2SINDPAUX Verbe 2e personne du Singulier Indicatif Présent Auxiliaire
V3SINDPAUX Verbe 3e personne du Singulier Indicatif Présent Auxiliaire
V1PINDPAUX Verbe 1ere personne du Pluriel Indicatif Présent Auxiliaire
V2PINDPAUX Verbe 2e personne du Pluriel Indicatif Présent Auxiliaire
V3PINDPAUX Verbe 3e personne du Pluriel Indicatif Présent Auxiliaire
+
VMSPP Verbe Masculin Singulier Participe Passé

VFSP	Verbe Féminin Singulier Participe Passé
VMPPP	Verbe Masculin Pluriel Participe Passé
VFPPP	Verbe Féminin Pluriel Participe Passé

*la langue française est **dérivée** du latin*

est	être	V	V3SINDP	V3SINDPAUX
dérivée	dériver	V	VFSP	VFSP

Remarque : On parlera de l'accord du participe passé plus loin.

➤ **Subjonctif**

V1SSUB	Verbe 1ere personne du Singulier Subjonctif
V2SSUB	Verbe 2e personne du Singulier Subjonctif
V3SSUB	Verbe 3e personne du Singulier Subjonctif
V1PSUB	Verbe 1ere personne du Pluriel Subjonctif
V2PSUB	Verbe 2e personne du Pluriel Subjonctif
V3PSUB	Verbe 3e personne du Pluriel Subjonctif

*faut pas qu'on **dise** ça*

dise	dire	V	V3SSUB	V3SSUB
------	------	---	--------	--------

Les temps comme l'imparfait ou le plus-que-parfait du subjonctif sont absents dans notre corpus.

➤ **Impératif**

Il ne s'emploie qu'à la deuxième personne du singulier et aux première et deuxième personnes du pluriel.

V2SIMP	Verbe 2e personne du Singulier Impératif
V1PIMP	Verbe 1ere personne du Pluriel Impératif
V2PIMP	Verbe 2e personne du Pluriel Impératif

***vas** le chercher*

vas	aller	V	V2SIMP	V2SIMP
-----	-------	---	--------	--------

- **Ambiguïté avec les marqueurs discursifs**

A l'oral, certains verbes comme « voir, tenir, comprendre, etc ». peuvent être employés dans leurs forme impérative non pour exprimer un ordre ou un conseil mais pour attirer l'attention d'interlocuteur ou juste reprendre une idée. Dans ce cas, pour nous, ce sont des marqueurs

discursifs (à voir l'explication plus loin) qui peuvent facilement être supprimés ou modifiés sans que la syntaxe et le sens de l'énoncé soit atteint.

c'est devenu français ça voyez ce sont des mots qui sont passés dans la langue
voyez voyez MD MD MD

→ Ici, il n'est pas question du sens du verbe *voir* mais plutôt du verbe *comprendre*.

➤ Infinitif

VINF Verbe à l'Infinitif

j'ai pas l'occasion d'écrire
écrire écrire V VINF VINF

Si le verbe à l'infinitif est précédé d'un déterminant, il est considéré comme un nom

il n'y a pas de parler patois
parler N NCMS NCMS

→ Ici, c'est l'énoncé « il y a un parler patois » qui est dans la forme négative. « parler » est précédé donc de déterminant indéfini.

➤ Participe présent

VPPRES Verbe au Participe Présent

il était là avec son dictionnaire ne reconnaissant pas l'alphabet
reconnaissant reconnaître V VPPRES VPPRES

• Gérondif

Cette forme permettra également de retrouver le gérondif qui est composé de la préposition « en » suivie d'un participe présent.

en relisant quelques fois j'en retrouve une
en en PREP PREP PREP
relisant relire V VPPRES VPPRES

➤ Voix passive

A la voix passive, le verbe au participe passé est précédé de l'auxiliaire « être ». L'information de la forme passive se greffe sur l'étiquette du participe passé.

VMSPPPAS Verbe Masculin Singulier Participe Passé Passif
VMPPPPAS Verbe Masculin Pluriel Participe Passé Passif
VFSPPPAS Verbe Féminin Singulier Participe Passé Passif
VFPPPPAS Verbe Féminin Pluriel Participe Passé Passif

j'ai été mutée

j'	je	P	P1S	P1SPERSUJ
ai	avoir	V	V1SINDP	V1SINDPAUX
été	être	V	VMSP	VMSP
mutée	muter	V	VFSP	VFSPPPAS

ils sont pas un peu dirigés d'un côté ou de l'autre
dirigés diriger V VMPP VMPPPPAS

L'information sur le passif nous semblait être importante à distinguer pour retrouver plus facilement son emploi dans le corpus. Sans cela, le problème d'ambiguïté aurait pu se créer en ce qui concerne les énoncés au passé composé avec l'auxiliaire être et au passif :

Je suis prise / je suis partie

Ainsi, sans l'annotation du passif, il faudrait rechercher le verbe auxiliaire « être » suivi d'un ou de deux participes passés ce qui nous aurait extrait les énoncés avec le passé composé.

➤ Temps composés

• Etiquette pour les temps composés

Comme on voit sur l'exemple du passé composé, nous n'avons pas créé des étiquettes pour les temps composés (passé composé, plus-que-parfait, futur antérieur, etc.) en considérant que dans le cas de besoin, on pourrait les retrouver en cherchant l'auxiliaire à un temps précis (présent, imparfait, futur, etc.) suivi d'un participe passé.

• L'accord du participe passé

Le participe passé est marqué au masculin singulier par défaut sauf les cas de l'accord avec le sujet où il est employé avec le verbe auxiliaire « être » ou avec le COD qui précède si le verbe auxiliaire est « avoir ».

elle a disparu

a	avoir	V	V3SINDP	V3SINDPAUX
disparu	disparaître	V	VMSP	VMSP

nous avons une Citroën on l'a vendue

vendue	vendre	V	VFSP	VFSP
--------	--------	---	------	------

les académiciens ce sont eux qui sont chargés de faire le dictionnaire

chargés	charger	V	VMPP	VMPP
---------	---------	---	------	------

l'orthographe est moins bien enseignée de nos jours

enseignée	enseigner	V	VFSP	VFSPPPAS
-----------	-----------	---	------	----------

Quand le sujet est le pronom « on », le participe passé est toujours au masculin singulier.

Attention, quand le sujet du verbe est le pronom personnel de vouvoiement « vous », le participe passé reste au singulier quand il s'agit d'une seule personne.

quand vous êtes pas connu

connu	connaître	V	VMSP	VMSPPPAS
-------	-----------	---	------	----------

➤ Ambiguïté entre les participes et les adjectifs

Les participes présents et passés peuvent être employés en tant qu'adjectif :

vous êtes toujours prête/reconnaisante.

prête	prêt	ADJ	ADJFS	ADJFS
-------	------	-----	-------	-------

Pour distinguer les deux formes voilà quelques indices non exhaustifs :

Il s'agit d'un adjectif :

- lorsque le participe passé est précédé d'un verbe être (alors qu'il se conjugue d'habitude avec le verbe « avoir ») et ce n'est pas le cas du passif
- lorsque le participe est précédé d'un verbe attributif comme « être, devenir, rester, paraître, apparaître, sembler, etc. ». L'adjectif joue le rôle de l'attribut du sujet.

Il est décevant/ému.

- Lorsque le participe présent est précédé d'un nom

Je trouve ce livre décevant

Ces indices sont loin d'être satisfaisantes mais elles peuvent néanmoins aider un petit peu les futurs correcteurs. Elles restent néanmoins difficilement applicables pour la désambiguïsation automatique.

[Retour au sommaire](#)

Les adverbes

La classe d'adverbes est une classe très hétérogène. Nous avons distingué au sein de cette classe qu'un seul sous-groupe, les adverbes de négation. Tout d'abord, c'est la classe finie qui contient un nombre fini d'unités lexicales. D'autre part, ça permettrait de repérer plus facilement les énoncés négatifs (ceux où la négation est exprimée à travers les adverbes).

Les adverbes porteront une étiquette unique sur tous les niveaux (sauf les adverbes de négation) :

ADV ADV ADV

oui évidemment il y a une tendance à revenir

oui oui ADV ADV ADV

évidemment évidemment ADV ADV ADV

Parmi les adverbes, on peut retrouver les mots-phrases comme :

- oui / ouais
- si
- non

ou des locutions et des groupes prépositionnels comme :

- d'abord
- tout de suite
- dans le temps
- à droite, à gauche
- au fond
- bien sûr
- tout le temps
- tout à coup
- en ce moment
- etc.

➤ Adverbes de négation

On distingue leur étiquette par l'ajout de l'information *NEG* sur la dernière colonne de l'étiquetage.

ADV	ADV	ADVNEG
-----	-----	--------

Voilà la liste de ces adverbes :

- non (nan), ne, pas, plus, rien, jamais, rien du tout, personne, plus du tout, rien du tout, etc.

Ces adverbes peuvent être employés seuls, il s'agira dans ce cas de mots-phrases. D'habitude, ils sont accompagnés la particule de négation « ne » ou sa contraction « n' » qui disparaît très souvent à l'oral.

• Constructions discontinues

La négation en français est formée généralement par deux adverbes négatifs ne ... pas/rien, personne ne ..., etc.. séparés souvent par d'autres unités lexicales (verbe,, pronom, adverbe, etc.). Il s'agit donc d'un constituant discontinu. Nous avons décidé de marquer les frontières de ce constituant pour pouvoir l'opérer plus facilement par la suite :

Je ne pense pas

ne	ne	ADV	ADV	ADVNEGDISC1
pense	penser	V	V1SINDP	V1SINDP
pas	pas	ADV	ADV	ADVNEGDISC2

• Restriction

Ne pas confondre la négation avec la restriction ne...que où l'adverbe « ne » n'est pas négatif et où la forme adverbiale « ne ... que » peut être remplacée par l'adverbe « seulement ».

« il ne vous reste qu'un enfant »

ne	ne	ADV	ADV	ADVDISC1
qu'	que	ADV	ADV	ADVDISC2

• L'absence de la particule « ne »

Comme nous l'avons mentionné plus haut, la particule « ne » de négation ou de restriction peut être omis :

*c'est **qu'**un Larousse »*
que que ADV ADV ADV

*il a **rien** vu venir*
rien rien ADV ADV ADVNEG

[Retour au sommaire](#)

Les prépositions

PREP Préposition

*on vous parle **de** la retraite*
de de PREP PREP PREP

*on est parti **en** vacances*
en en PREP PREP PREP

*j'ai mal **dans** les bras*
dans dans PREP PREP PREP

Dans le cas de prépositions composées, l'étiquetage se fait sur une seule ligne.

*on était **loin de** s'attendre à ça*
loin_de loin_de PREP PREP PREP

***au début du** siècle (→ **au début de** le siècle)*
au_début_de au_début_de PREP PREP PREP
le le DET DETMS DETMSDEF

- **Elision**

Dans le cas d'élision de la voyelle finale, celle-ci est rajoutée dans le lemme.

*leur façon de **d'**écouter*
d' de PREP PREP PREP

c'est à côté d'Aubenas
à_côté_d' à_côté_de PREP PREP PREP

[Retour au sommaire](#)

Les conjonctions

Il existe deux types de conjonctions : conjonctions de coordination et de subordination. Dans tous les cas, le lemme sera identique à la forme employée.

➤ Conjonctions de coordination

Les conjonctions de coordination permettent de relier deux éléments de même statut ou de même fonction dans une phrase. Il n'y a pas de dépendance entre ces deux éléments. Leur liste est finie : « mais », « ou », « et », « donc » (dans certains cas), « or », « ni », « car ».

CONJCOO Conjonction de Coordination

*à Orléans **ou** Blois*
ou ou CONJ CONJ CONJCOO

*c'est un petit bol comme un vase **mais** c'est rond*
mais mais CONJ CONJ CONJCOO

- **Ambiguïté avec les marqueurs discursifs**

La conjonction comme « donc » n'est pas toujours employée pour la coordination. Son but initial est de relier deux propositions dont la seconde est la conséquence de la première. Cependant dans certains cas, surtout à l'oral, il a un autre emploi où il n'apporte pas d'informations particulières et sera classé parmi des marqueurs discursifs.

*c'est tout ce que je connais **donc** je m'y plais*
donc donc CONJ CONJ CONJCOO

*aujourd'hui c'est plus ça et puis regardez **donc** autrefois on apprenait l'alphabet*
donc donc MD MD MD

On retrouve souvent ce type d'emploi en fin de questions :

*comment s'appelle-t-il **donc***
donc donc MD MD MD

➤ Conjonctions de subordination

Les conjonctions de subordination permettent de relier deux éléments de statuts ou fonctions inégaux. Un élément mis en relation avec un autre lui est dépendant, il n'existe que parce qu'il complète le premier élément.

Les conjonctions de subordination peuvent être constituées d'un seul mot (« lorsque », « comme », « puisque », « que », « quoique », « quand », « si ») ou de plusieurs mots (« parce que », « ainsi que », « bien que », etc).

CONJSUB Conjonction de Subordination

*on s'en doutait **parce qu'**on était loin de s'attendre à ça*
parce_que parce_que CONJ CONJ CONJSUB

*ils meurent tous **avant que** [...] ce soit fini*
avant_que avant_que CONJ CONJ CONJSUB

*faut que ce soit une lettre un peu particulière **pour que** je fasse un brouillon*
pour_que pour_que CONJ CONJ CONJSUB

Difficultés de l'étiquetage :

- **Bien que / Bien + que**

Il faut différencier les emplois de la conjonction « bien que » des emplois du type :

je crois bien que le gouvernement

Dans cet exemple, il ne s'agit pas d'une conjonction de subordination mais d'un adverbe « bien » suivi de la conjonction « que ». Il n'y a pas de règles qui permettent de différencier ces deux emplois, mis à part le sens et l'intonation.

[Retour au sommaire](#)

Les présentateurs

Les présentateurs sont ce qu'on appelle plus communément *introduceurs*. Il s'agit de mots invariables utilisés pour « introduire un mot, un syntagme ou une phrase » (Le Bon Usage). Ils sont notés *PRES* chez nous.

Cette définition ne s'applique pas entièrement à ce que nous appelons *présentateur* puisque nous constatons que deux types de présentateurs existent dans les corpus oraux. Ainsi, on retrouve :

- les présentateurs dits « invariables » où le lemme reste identique à l'emploi fait à l'oral ;

Les présentateurs « invariables » :

- | | |
|---------------|----------------------|
| ▪ Voici | ▪ t'as |
| ▪ Voilà | ▪ à savoir |
| ▪ O | ▪ c'est pourquoi |
| ▪ quant à | ▪ il est certain que |
| ▪ soit | ▪ il s'agit |
| ▪ vive | ▪ ça y est |
| ▪ par exemple | ▪ n'est-ce pas |

- les présentateurs qui peuvent se décliner et ayant ainsi une forme différente de leur lemme (c'est, il y a)

PRES PRES PRES

« *voilà* je veux que les français sachent le français »
 « ils ne trouvent pas de travail alors *c'est pourquoi* ils font des maths »
 « *il est certain que* le résultat c'est pas pareil »

Ils peuvent se suivre dans la phrase :

« *voilà c'est ça* qui a changé »

Les présentateurs et leurs « déclinaisons » :

- « **C'est** » et toutes ses formes

« *c'est* trop tard »
 c'est c'est PRES PRES PRES

On le retrouve :

- Au pluriel : « ce sont »
ce sont des choses tout à fait ridicules
 ce_sont c'est PRES PRES PRES
- A l'imparfait : « c'était »
c'était des gens parfaits
 c'était c'est PRES PRES PRES

- Au conditionnel : « ce serait »

*la vie va si vite peut-être **ce serait** mieux les langues et les sciences*
 ce_serait c'est PRES PRES PRES

Remarque : « ce serait » (PRES) ne doit pas être confondu avec « ça serait » (Pronom + « être »).

- A la forme négative :

***ce n'est plus** la même vie*
 ce c'est PRES PRES PRESDISC1
 n' ne ADV ADV ADVNEGDISC1
 est c'est PRES PRES PRESDISC2
 plus plus ADV ADV ADVNEGDISC2

➔ Il s'agit de la forme *ce + est* combinée à la forme négative « ne ... plus ».

- « **Il y a** » et toutes ses formes :

***il y a** des programmes*
 il_y_a il_y_a PRES PRES PRES

On le retrouve :

- A la forme contractée : « y'a »

***y'a** rien à dire à l'école laïque*
 y'a il_y_a PRES PRES PRES

- A l'imparfait : « il y avait »

***il y avait** des lacunes dans l'instruction des écoles*
 il_y_avait il_y_a PRES PRES PRES

- Au futur : « y'en aura »

***y'en aura** plus c'est fini*
 y il_y_a PRES PRES PRESDISC1
 en en P P3S P3SPERCOMPL
 aura il_y_a PRES PRES PRESDISC2

- A la forme négative : « il n'y a pas »

***il n'y a pas** vraiment beaucoup de choix*
 il il_y_a PRES PRES PRESDISC1
 n' ne ADV ADV ADVNEGDISC1
 y_a il_y_a PRES PRES PRESDISC2
 pas pas ADV ADV ADVNEGDISC2

- A la forme interrogative : « y'a-t-il »

y a-t-il chez vous madame des dictionnaires
 y_a-t-il il_y_a PRES PRES PRES

Difficultés de l'étiquetage :

- **Les problèmes de discontinuité**

« Il y a » / « c'est »

- Le présentateur « il y a » peut être modifié par l'ajout de pronoms ou d'adverbes de négation.

il y en a

il y il_y_a PRES PRES PRESDISC1
 en en P P3S P3SPERCOMPL
 a il_y_a PRES PRES PRESDISC2

y en a eu

y il_y_a PRES PRES PRESDISC1
 en en P P3S P3SPERCOMPL
 a_eu il_y_a PRES PRES PRESDISC2

Remarque : les temps composés sont étiquetés ensemble (a eu, aurait eu, avait eu, ...).

y en a

y il_y_a PRES PRES PRESDISC1
 en en P P3S P3SPERCOMPL
 a il_y_a PRES PRES PRESDISC2

Remarque : le pronom sujet « il » qui disparaît souvent à l'oral est reconstitué dans le lemme.

Même système pour la négation ou restriction :

« *il n'y a pas / plus / que* »

il il_y_a PRES PRES PRESDISC1
 n' ne ADV ADV ADVNEGDISC1
 y_a il_y_a PRES PRES PRESDISC2
 pas pas ADV ADV ADVNEGDISC2
 (plus) plus ADV ADV ADVNEGDISC2)

« *y_a_t'il* » prend comme lemme « il y a » car la particule euphonique n'est utilisée que dans le cas de la question. Elle peut donc disparaître dans le lemme.

○ « *C'est ... pas* »

Il en est de même pour le présentateur « c'est » qui voit sa forme se modifier à cause de divers phénomènes comme la négation ou le changement de temps :

« *Ce n'est pas / plus* »

ce	c'est	PRES	PRES	PRES	DISC1
n'	ne	ADV	ADV	ADV	NEGDISC1
est	c'est	PRES	PRES	PRES	DISC2
pas	pas	ADV	ADV	ADV	NEGDISC2

[Retour au sommaire](#)

Les marqueurs discursifs

Les marqueurs discursifs se classés chez nous en trois catégories : marqueurs discursifs propres, interjections et marqueurs discursifs de la forme « euh ». Les lemmes sont à chaque fois identiques à la forme prononcée.

➤ Marqueurs discursifs

Ce sont des unités lexicales, souvent utilisées à l'oral, qui ne sont pas employées dans leur sens et leur catégorie grammaticale d'origine. Ces éléments comme *bon, bien, quoi, comment dire, etc.* apparaissent avec une fréquence élevée dans les corpus oraux. Qu'on les désigne comme des *phatiques*, des *particules*, des *marqueurs discursifs* ou des *inserts*, ces formes figées ou invariables, peuvent constituer des énoncés à elles seules ou se manifester à différentes places d'un énoncé sans intégrer sa structure, (c'est-à-dire sans entrer en relation syntaxique avec un autre élément) et peuvent être remplacées à l'écrit par des signes de ponctuation. Leurs fonctions sont très hétérogènes. Ils permettent de temporiser le discours, d'exprimer une marque d'hésitation, ils peuvent avoir une valeur argumentative. Ces mots n'ont pas de liens syntaxiques avec d'autres mots de l'énoncé et peuvent donc être supprimés ou modifiés sans que le sens de l'énoncé soit touché. Une liste exhaustive est difficilement réalisable puisque ces mots dépendent de la façon dont ils sont employés et n'importe quelle unité lexicale pourrait potentiellement changer de statut à l'oral et être employés en tant que marqueur discursifs.

MD Marqueur Discursif

*ça a pu arriver mais je me souviens plus **vous savez***

vous_savez vous_savez MD MD MD

*une fois qu'on la sait on la sait **quoi***

quoi quoi MD MD MD

*j'habite Orléans depuis vingt-sept **hm** dix-neuf cent vingt-sept*

hm hm MD MD MD

- **Bon alors**

Les unités « bon » et « alors » peuvent être des marqueurs discursifs lorsqu'ils sont employés séparément. En revanche, lorsqu'ils se suivent « alors bon », ils n'ont pas la même fonction : « bon » clôt ce qui vient d'être dit alors que « alors » introduit ce qui va suivre. On les étiquette donc sur deux lignes séparées.

- **Ou**

Nous avons remarqué que la forme « ou » est suivie souvent de « bien » ou de « alors ». Dans le premier cas, il peut s'agir soit du marqueur discursif « ou_bien » soit de la conjonction de coordination « ou » et du marqueur « bien ». « Ou_alors » est étiqueté comme marqueur discursif.

➤ **Marqueurs discursifs « euh »**

Le « euh » d'hésitation est une particule neutre qui indique que le locuteur cherche ses mots. Caractéristique de l'oral, elle reste au centre de nombreuses études linguistiques ce qui nous a conduits de la distinguer parmi les autres marqueurs.

MDEUH Marqueur Discursif « euh »

*les problèmes **euh** littéraires*

euh euh MD MD MDEUH

➤ **Marqueurs discursifs « interjections »**

Ils permettent au locuteur d'exprimer un sentiment ou d'imiter phonétiquement une chose.

Il s'agit de mots invariables dont la liste est ici restreinte : « ah », « oh », « hein », « eh », « hè », « hm », « bah », etc. Ils sont tous constitués souvent d'un seul phonème vocalique. Des onomatopées peuvent également y être ajoutées comme « pan » ainsi que d'autres qui ne sont pas dans le corpus.

Étiquetage :

MDINT Marqueur Discursif Interjection

*un an et demi **ah** peut-être deux ans*
 ah ah MD MD MDINT

***oh** ça je peux pas vous dire*
 oh oh MD MD MDINT

*les temps changent vous savez c'est formidable **hein** ça va vite*
 hein hein MD MD MDINT

*j'habite Orléans depuis vingt-sept **hm** dix-neuf cent vingt-sept*
 hm hm MD MD MD

Difficultés de l'étiquetage

Parfois, un marqueur discursif de type interjection peut être joint à un marqueur discursif « simple » ou à un adverbe pour ne former qu'une seule et même unité. L'ensemble permet notamment d'introduire une explication mais sera étiquetée en tant que MD seulement.

***ah oui** j'écris beaucoup*
 ah_oui ah_oui MD MD MD

Parfois, on peut retrouver les deux interjections employés ensemble :

***eh ben** vous avez le Val de Loire*
 eh_ben eh_ben MD MD MD

Les unités pouvant être ainsi jointes sont : « ah oui », « ah non », « bah non », « ah si », « oh oui », « oh non ».

[Retour au sommaire](#)

Les expressions de temps

En corrigeant l'étiquetage automatique, nous avons décidé d'annoter également quelques types d'entités nommées. Il nous semblait d'être utile de marquer les dates et les heures vu qu'à l'oral ils n'ont pas toujours la structure classique. Dans les deux cas, le lemme est le même que la forme prononcée.

➤ Dates

(dates, années, mois, jours de la semaine)

ECD	Expression Calendaire de Dates
-----	--------------------------------

Elles sont généralement précédées d'une préposition, ce qui permet de les différencier de la catégorie des chiffres.

J'habite Orléans depuis vingt-sept hm dix-neuf cent vingt-sept

j'	je	P	P1S	P1SPERSUJ					
habite	habiter	V	V1SINDP	V1SINDP					
Orléans	Orléans	N	NP	NP					
depuis	depuis	PREP	PREP	PREP					
vingt-sept	vingt-sept	EC	EC	ECD					
hm	hm	MD	MD	MD					
dix-neuf_cent_vingt-sept	dix-neuf_cent_vingt-sept	EC	EC	ECD					

le plus jeune maintenant a vingt et un ans

le	le	DET	DETMS	DETMSDEF					
plus	plus	ADV	ADV	ADV					
jeune	jeune	N	NCMS	NCMS					
maintenant	maintenant	ADV	ADV	ADV					
a	avoir	V	V3SINDP	V3SINDP					
vingt_et_un	vingt_et_un	CH	CH	CH					
ans	an	N	NCMP	NCMP					

Difficultés de l'étiquetage :

alors le huit (mai)

Il s'agit ici d'une date, mais il n'y a pas d'autres moyens de le savoir mis à part le contexte :

*Jeanne d'Arc avait délivré Orléans le sept mai alors le **huit** il y a eu une grande procession*

Le fait qu'il y ait un déterminant n'indique pas qu'il s'agit nécessairement d'une date puisque ça peut tout simplement être un chiffre

j'ai choisi le [numéro] huit.

➤ **Heure**

ECH Expression Calendaire des Heures

*on travaillait d'abord à **huit heures***

huit_heures huit_heures EC EC ECH

Difficultés de l'étiquetage :

Le problème concernant ce type d'étiquette est le fait qu'il faille différencier les expressions horaires des noms communs.

*si on vous donnait **deux heures** par jour*

deux deux CH CH CH
heures heure N NCFP NCFP

*il est **deux heures***

deux_heures deux_heures EC EC ECH

Le fait que l'expression soit précédée du verbe « être » n'est pas un critère de reconnaissance puisque nous pouvons avoir des exemples du type :

*on commençait à midi jusqu'à **sept heures***

on on P P3MS P3MSPERSU
commençait commencer V V3SINDI V3SINDI
à à PREP PREP PREP
midi midi EC EC ECH
jusqu'à jusqu'à PREP PREP PREP
sept_heures sept_heures EC EC ECH

[Retour au sommaire](#)

Les mots inconnus

MI	Mot Inconnu
----	-------------

Les mots inconnus dans notre étiquetage sont de quatre types. Dans chaque cas, le lemme est identique à la forme prononcée.

- Des phonèmes et morphèmes indéterminés ou ne faisant pas partie du lexique. Il peut s'agir d'amorces de mots connus, ou de phonèmes incompréhensibles. Ils seront reconnaissables à l'écrit (dans la transcription) par un tiret.

i- ils parlent comme nous

i-	i-	MI	MI	MI
ils	il	P	P3MP	P3MPPERSUJ
parlent	parler	V	V3PINDP	V3PINDP

*les **magi-** les magistrats*

les	le	DET	DETP	DETPDEF
magi-	magi-	MI	MI	MI
les	le	DET	DETP	DETPDEF
magistrats	magistrat	N	NCMP	NCMP

- Des mots entiers inconnus. Ceux-ci ne possèdent pas de tiret. Nous ne savons pas leur origine.

*des avocats qui **parlement** très bien*

des	un	DET	DETMP	DETMPIND
avocats	avocat	N	NCMP	NCMP
qui	qui	P	PI	PIREL
parlement	parlement	MI	MI	MI
très	très	ADV	ADV	ADV
bien	bien	ADV	ADV	ADV

- Des locutions non terminées. Elles ne présentent pas de tirets mais sont sur une seule et même ligne

*y **en** y en a beaucoup*

y_en	y_en	MI	MI	MI
y	il_y_a	PRES	PRES	PRESDISC1
en	en	P	P3S	P3SPERCOMPL
a	il_y_a	PRES	PRES	PRESDISC2

- Des mots comportant des fautes d'orthographe.

*faut repasser faut **racommoder***
 racommoder racommoder MI MI MI

[Retour au sommaire](#)

Les chiffres

Tous les chiffres et nombres (écrits en lettres) ordinaux font partie de la catégorie des chiffres sauf « un » qui sera toujours traité comme un déterminant indéfini. Les cardinaux font partie de la classe des adjectifs

CH	Chiffre
----	---------

*si on vous donnait **deux** heures par jour*
 deux deux CH CH CH
 heures heure N NCFP NCFP

Les nombres composés forment un tout et sont écrits sur une seule ligne.

*Elle avait **vingt-sept** ans*

elle	Elle	P	P3FS	P3FSPERSUJ
avait	Avoir	V	V3SINDI	V3SINDI
vingt-sept	vingt-sept	CH	CH	CH
ans	An	N	NCMP	NCMP

X et N, minuscule ou majuscule, représentent un nombre quelconque qu'il parait inutile d'expliciter. Ils seront étiquetés CH.

*On leur donnait de l'andouillette **X** fois par jour*
 X X CH CH CH
 fois fois N NCFP NCFP

Difficultés de l'étiquetage :

Nous constatons que « un » est toujours un déterminant sauf dans certains cas où il peut s'agir d'un chiffre (« un seul volume »). Dans ce cas, on peut l'étiqueter de cette manière :

un	un	CH	CH	CH
seul	seul	ADJ	ADJMS	ADJMS
volume	volume	N	NCMS	NCMS

➔ En effet, il est repérable et différenciable du déterminant grâce à son entourage, « un seul ». C'est bien la quantité que l'on veut mettre en évidence dans ce syntagme.

Dans l'exemple suivant, « il ne vous reste qu'un enfant », il s'agit également d'un chiffre :

il	il	P	P3MS	P3MSPERSUJ
ne	ne	ADV	ADV	ADVDISC1
vous	vous	P	P2P	P2PPERCOMPL
reste	rester	V	V3SINDP	V3SINDP
qu'	que	ADV	ADV	ADVDISC2
un	un	CH	CH	CH
enfant	enfant	N	NCMS	NCMS

Puisqu'il est difficile pour le logiciel de déterminer s'il s'agit d'un DET ou d'un CH, on décide cependant de noter DET par défaut.

[Retour au sommaire](#)

Les unités euphoniques

Ces unités sont au nombre de deux : « l' » et « t' »

UEUPH Unité Euphonique

comment s'appelle-t-il

t	t	UEUPH	UEUPH	UEUPH
---	---	-------	-------	-------

[Retour au sommaire](#)

La ponctuation

PCT Ponctuation.

Dans la transcription, les marques de ponctuation n'ont pas été volontairement mises. Cependant, il subsiste quelques signes de ponctuation présents dans :

- Les inversions sujet-verbe : verbe + ponctuation + sujet

madame que pensez-vous du français
 pensez penser V V2PINDP V2PINDP
 - - PCT PCT PCT
 vous vous P P2P P2PPERSUJ

- Les cas de renforcement de démonstratifs : nom + ponctuation + adverbe

ces moments-là
 moments moment N NCMPNCMP
 - - PCT PCT PCT
 là là ADV ADV ADV

Remarque :

Le pronom « celui-là » est étiqueté sur la même ligne car l'ensemble du groupe correspond à un pronom démonstratif.

On peut retrouver aussi le point d'interrogation qui sera étiqueté pareil.

[Retour au sommaire](#)

Cas particuliers :

➤ Que

L'unité lexicale « que » peut avoir plusieurs catégories.

- Quand l'emploi est restrictif, il s'agira d'un adverbe.

c'est une ville tranquille mais y a pas qu'Orléans
 qu' que ADV ADV ADV

- Quand il introduit une proposition subordonnée, c'est une conjonction de subordination.

supposons que quelqu'un frappe à la porte
 que que CONJ CONJ CONJSUB

- Quand il remplace le groupe nominal, c'est un pronom relatif ou interrogatif.

où est-ce qu'on parle le mieux français c'est ça que vous voulez savoir
 que que P PI PIREL

que voulez-vous savoir

que que P PI PIINT

- Quand il fait partie de l'expression « est-ce que » qui est omis :
Comment *qu'on* a dit → comment *est-ce qu'on* a dit
C'est le pays où *qu'on* parle → c'est le pays où *est-ce qu'on* parle

➤ Tout et ses variantes

Le mot « tout » a aussi une multitude d'emplois en tant que

- Adjectif :

toute tout ADJ ADJFS ADJFS
la la DET DETFS DETFSDEF
famille famille N NCFS NCFS

- déterminant,

pour pour PREP PREP PREP
faire faire V VINF VINF
toute_sorte_de toute_sorte_de DET DETI DETIIND
travaux travail N NCMP NCMP

- adverbe

qui qui P PI PIREL
est être V V3SINDP V3SINDP
tout_aussi tout_aussi ADV ADV ADV
euh euh MD MD MDEUH
valablevalableADJ ADJMS ADJMS

- pronom indéfini :

je les aime tous
tous tout P P3P P3SIND

tirer_parti tirer_parti V VINF VINF
de de PREP PREP PREP
tout tout P PMS PMSIND

Le mot « tout » comme adjectif suit les mêmes règles d'accord en genre et en nombre que tous les autres adjectifs. Il est adjectif lorsqu'il est suivi d'un déterminant et d'un nom commun. :

toutes les villes
tout le centre de la ville
tous ces métier
tous les ans

➤ **Bon : adjectif ou marqueur discursif :**

L'adjectif « bon » peut changer de catégorie grammaticale et être employé à l'oral comme un marqueur discursif :

*oui **bon** alors pour revenir un peu à l'enseignement madame*
bon bon MD MD MD

*on n'était pas toujours **bon** en français*
bon bon ADJ ADJMS ADJMS

Cependant, on peut les distinguer grâce au contexte de l'énoncé :

En tant qu'adjectif :

- il détermine le nom qui le suit ou précède ;
- il suit le verbe attributif (« il est bon »).

En tant que marqueur discursif, il est souvent accompagné d'un adverbe avant ou après (« bon oui ») ou d'un autre MD (« ah_oui bon », « bon alors »).

➤ **Constructions discontinues**

On constate qu'à l'oral on trouve de nombreux cas de discontinuité. Nous appelons « éléments discontinus » des structures unies au sein desquels des mots étrangers viennent s'insérer à l'intérieur pour les scinder en deux parties voire en trois. Les éléments discontinus posent donc des problèmes d'étiquetage majeurs. Nous les étiquetons en reconstituant le lemme et en ajoutant l'information de la discontinuité *DISC*.

• **Négation**

La négation en français est formée généralement par deux adverbes négatifs ne ... pas/rien, personne ne ..., etc. qui permettent l'insertion d'autres unités lexicales entre eux (verbe, pronom, adverbe, etc.). Nous avons décidé de marquer les frontières de ce constituant pour pouvoir l'opérer plus facilement par la suite :

Je ne pense pas

ne	ne	ADV	ADV	ADVNEGDISC1
pense	penser	V	V1SINDP	V1SINDP
pas	pas	ADV	ADV	ADVNEGDISC2

- **Restriction**

La discontinuité liée à la négation est valable aussi pour la restriction marquée par le constituant ne...que :

Il ne vous reste qu'un enfant

il	il	P	P3MS	P3MSPERSUJ
ne	ne	ADV	ADV	ADVDISC1
vous	vous	P	P2P	P2PPERCOMPL
reste	rester	V	V3SINDP	V3SINDP
qu'	que	ADV	ADV	ADVDISC2
un	un	CH	CH	CH
enfant	enfant	N	NCMS	NCMS

- **Unités composées**

Cette discontinuité peut être montrée dans les cas des unités composées si elles sont séparées :

J'ai besoin de ça

ai_besoin	avoir_besoin	V	V1SINDP	V1SINDP
-----------	--------------	---	---------	---------

alors que

J'ai vraiment besoin

ai	avoir_besoin	V	V1SINDP	V1SINDPDISC1
vraiment	vraiment	ADV	ADV	ADV
besoin	avoir_besoin	V	V1SINDP	V1SINDPDISC2

Val c'est de Loire

Val	Val_de_Loire	N	NP	NPDISC1
c'est	c'est	PRES	PRES	PRES
de_Loire	Val_de_Loire	N	NP	NPDISC2

- **Croisement entre les constructions discontinues**

Les constructions discontinues peuvent se croiser. Il s'agit des cas où la structure discontinue porte la négation :

ce n'est plus la même vie

ce	c'est	PRES	PRES	PRES	DISC1
n'	ne	ADV	ADV	ADV	NEGDISC1
est	c'est	PRES	PRES	PRES	DISC2
plus	plus	ADV	ADV	ADV	NEGDISC2

je n'ai pas besoin

n'	ne	ADV	ADV	ADV	NEGDISC1
ai	avoir_besoin	V	V1SINDP	V1SINDP	
pas	pas	ADV	ADV	ADV	NEGDISC2
besoin	avoir_besoin	V	V1SINDP	V1SINDP	