

Jeu d'étiquettes Presto_min

Nous avons choisi de distinguer deux niveaux dans le jeu d'étiquettes PRESTO : Presto_min (jeu minimal) et Presto_max (jeu maximal).

Pour Presto_min, les champs utilisés sont <catégorie> <type> <mode>.

Pour chaque catégorie, nous spécifions les modifications apportées par rapport à MULTEXT *english* (2010) (désormais MULTEXT (*en*) : <http://nl.ijs.si/ME/V4/msd/html/msd-en.html>) et GRACE (1997).

CATEGORIE	Valeur	Code
CATEGORIE	Nom	N
CATEGORIE	Verbe	V
CATEGORIE	Adjectif	A
CATEGORIE	Pronom	P
CATEGORIE	Déterminant	D
CATEGORIE	Participe-Adjectif-Gérondif	G
CATEGORIE	Adverbe	R
CATEGORIE	Adposition	S
CATEGORIE	Conjonction	C
CATEGORIE	Numéral	M
CATEGORIE	Interjection	I
CATEGORIE	Résidu	X
CATEGORIE	Ponctuation	F

1. Noms (Nouns)

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Nom	N	
1	Type	commun	c	<i>livre</i>
		propre	p	<i>Jean</i>

Comparaison par / à MULTEXT (*en*) et GRACE

ATTRIBUT : TYPE

- MULTEXT (*en*) propose deux valeurs : « common » (c), « proper » (p).
- GRACE, Outre les valeurs « common » (c), « proper » (p), propose la valeur « cardinal » (k). En effet, GRACE prend le parti de supprimer la catégorie « Numéral » proposée par MULTEXT (*en*) au profit d'une valeur additionnelle « cardinal » (notée : k) à l'attribut type des différentes catégories syntaxiques pouvant intégrer des emplois de numéraux cardinaux.
- PRESTO-MIN propose deux valeurs : « common » (c), « proper » (p) et opte comme MULTEXT (*en*) pour une catégorie « Numéral » (M).

2. Verbes (Verbs)

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Verbe	V	
1	Type	être & avoir	u	<i>ai, suis</i>
		autre	v	<i>pars</i>
2	VForme	V conjugué à un mode personnel	c	<i>avons, étions, partirai</i>
		infinitif	n	<i>être, avoir, partir</i>

Comparaison par / à MULTEXT (*en*) et GRACE

ATTRIBUT : TYPE

- MULTEXT (*en*) propose 4 valeurs : « main » (m), « auxiliary » (a), « modal » (o), base (b).

- GRACE propose 2 valeurs : « main » (m), « auxiliary » (a).
 - PRESTO-MIN propose 2 valeurs : « être / avoir » (u), « Autre verbe » (v). On ne tranche pas entre emplois d’auxiliaires pour *être/avoir* et emplois de verbes pleins. Cette décision est directement liée au traitement appliqué dans Presto_min aux participes passés, pour lesquels nous avons décidé de ne pas trancher entre participes et adjectif (cf. *infra*). Or cette décision implique de ne pas trancher entre les structures du type *NO être Participe / NO être Adj.*, c’est-à-dire entre *être* auxiliaire et *être* verbe copule.

ATTRIBUT : FORME VERBALE (VFORM)

- MULTTEXT (*en*) propose 4 valeurs : « indicative » (i), conditional (c), infinitive (n), participe (p)
- GRACE propose 6 valeurs : « indicative » (i), conditional (c), « subjonctive » (s), « imperative » (m), infinitive (n), participe (p).
- PRESTO-MIN propose deux valeurs pour l’attribut VForme
 - « Verbe conjugué à un mode personnel » (c) : cette valeur a été empruntée à Cattex09min (http://bfm.ens-lyon.fr/article.php3?id_article=176) qui lui a affecté le code (cjk) ;
 - « infinitif » (n).

Rem : Le mode non personnel « participe » ne donne pas lieu à une valeur car il est traité dans la catégorie G qui ne distingue pas entre participes (présent ou passé), adjectifs verbaux et gérondifs.

3. Adjectifs (Adjectives)

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Adjectif	A	
1	Type	général	g	<i>aimable, municipal, futur, tel, ...</i>
		possessif	s	<i>(un) mien (cousin)</i>

Comparaison par / à MULTTEXT (*en*) et GRACE

ATTRIBUT : TYPE

- MULTTEXT (*en*) propose 1 valeur : « qualificative » (q).
- GRACE distingue 5 valeurs: « qualificative » (q), « ordinal » (o), « cardinal » (k), « indéfinie » (i), « possessive » (s).
- PRESTO-MIN propose 2 valeurs.
 - la valeur « général » (g) se substitue à « qualificatif » (q). Cette valeur de type rassemble, outre les traditionnels adjectifs qualificatifs, d’autres sous-catégories qui ne présentent pas les mêmes caractéristiques syntaxiques et distributionnelles que les traditionnels qualificatifs : les adjectifs « relationnels » (*municipal, ...*), les adjectifs du « troisième type » (Schnedeker (éd.) (2002) ; Riegel & al. 2009 : 634).
 - la valeur « possessif » (s) est conservée telle quelle.

Rem 1 : Sont éliminées les valeurs « ordinal » et « cardinal », les « adjectifs » correspondants étant placés dans la catégorie « numéral » (M) ; quant aux traditionnels « adjectifs indéfinis », la plupart sont versés dans la catégorie « Déterminants » (D).

Rem 2 : Est conservée l’étiquette « adjectif possessif » pour les occurrences de *mien, tien, sien*, dans des contextes comme « *un mien cousin* », pour des raisons d’ordre diachronique et distributionnelle. Il s’avère en effet que ces formes sont combinables avec un adjectif qualificatif épithète dans un GN aux XVI^e et au XVII^e e. s. du moins

« *Je propose les fantasies humaines et miennes, simplement comme humaines fantasies* », M. de Montaigne, *Essais : t. 1 (livres 1 et 2)*, 1592

« *tirée de ceste cordiale et mienne bñeficence* » (lettres missives de Henri IV, t VII, p. 623. 23 octobre 1608. Citée dans *Henri IV et sa politique*, Charles Mercier de Lacombe, 1860, p. 814.

4. Pronoms (Pronouns)

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Pronom	P	
1	Type	personnel	p	<i>je, le, en</i>
		démonstratif	d	<i>ce, celui</i>
		indéfini	i	<i>certain, plusieurs</i>
		possessif	s	<i>(le) mien</i>
		interrogatif	t	<i>qui, que</i>
		relatif	r	<i>qui, lequel</i>

Comparaison par / à MULTEXT (en) et GRACE

ATTRIBUT : TYPE

- MULTEXT (en) propose 7 valeurs : « personal » (p), « possessive » (s), « interrogative » (q), « relative » (r), « reflexive » (x), « general » (g), « ex-there » (t).
- GRACE propose 8 valeurs : « personal » (p), « demonstrative » (d), « indefinite » (i), « possessive » (s), « interrogative » (t), « relative » (r), « reflexive » (x), « cardinal » (k).
- PRESTO_MIN propose 6 valeurs : « personal » (p), « demonstrative » (d), « indefinite » (i), « possessive » (s), « interrogative » (t), « relative » (r).

Rem 1 : Sont éliminées les valeurs :

- o « reflexive » (x) : valeur fondue dans la valeur « personnel » (p)
- o « general » (g) qui correspond à un choix propre à MULTEXT que nous ne suivons pas¹
- o « ex-there » (t) : non pertinent pour le français
- o « cardinal » (k) : les pronoms cardinaux sont versés dans la catégorie englobante « numéral » (M)

Rem 2 : pour la valeur « interrogatif », le code (t) adopté est repris de GRACE.

5. Déterminants (Determiners)

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Déterminant	D	
1	Type	article défini	a	<i>le, la, l', les</i>
		démonstratif	d	<i>ce, cet, cette, ...</i>
		possessif	s	<i>mon, ta, leur, ...</i>
		article indéfini	n	<i>un, une, des, de, d'</i>
		article partitif	p	<i>du, de la, de l', des</i>
		indéfini	i	<i>quelque(s) N, tout N, chaque N, ...</i>
		relatif	r	<i>lequel, laquelle, ...</i>
		interrogatif/ exclamatif	t	<i>quel, quelle, ...</i>

Comparaison par / à MULTEXT (en) et GRACE

ATTRIBUT : TYPE

- MULTEXT (en) propose 4 valeurs : « demonstrative » (d), « indefinite » (i), « possessive » (s), « général » (g).
- GRACE propose 7 valeurs : « article » (a), « demonstrative » (d), « possessive » (s), « indefinite » (i), « interr./excl. » (t), « relative » (r), cardinal (k).
- PRESTO_MIN propose 10 valeurs :
 - o Sont retenues les 5 valeurs « demonstrative » (d), « possessive » (s), « indefinite » (i), « interr./excl. »

¹ "General" pronouns are those which are not personal, possessive, demonstrative or reflexive. The choice of these four categories is based on distributional facts, though at a rather high level of abstraction. They enter into anaphoric dependencies which are signalled morphosyntactically and are therefore (in principle) more amenable to automatic detection. Most general pronouns do not, although they too sometimes encode number information.

- (t), « relative » (r)
- o Est modifiée la valeur (« article défini ») associée au code (a)
 - o Sont ajoutées les valeurs :
 - « article indéfini » (n)
 - « article partitif » (p)
 - « négation » (n) : réunit les emplois de *de* sous la portée de la négation : *Je n'ai pas de voiture / Pas de nuages à l'horizon.*
 - déterminants « complémentaires » (c) : réunit l'ensemble des prédéterminants, postdéterminants et identificateurs qui entrent dans la composition des groupes déterminants définis et indéfinis sans en constituer la tête (Riegel & al. 2009 : 304-305).

6. Participes, adjectifs verbaux, gérondifs

Cette catégorie est inexistante dans MULTEXT (*en*) ET dans GRACE.

Ce choix s'explique par le constat que la distinction entre les trois classes de mots : *participes, adjectifs verbaux, gérondifs* pose des problèmes nombreux en synchronie et en diachronie.

- *En synchronie*, la mise au point de procédures de décisions pour les emplois ambigus (nombreux) nécessitent plusieurs tests (Riegel & al. 2009 : 737-738) qui augmentent les chances de divergences entre annotateurs. En outre, ces tests ne garantissent pas la mise à l'écart de toute appréciation subjective : *Selon les cas (le type de verbe, le contexte), ils [les participes] sont sentis comme plus ou moins « verbaux » ou « adjectivaux » (avec une marge appréciable de liberté d'interprétation)* (P. le Goffic 1993, § 134 : 201)
- *La dimension diachronique* ajoute une difficulté supplémentaire car la distinction morphologique entre participe présent, adjectif verbal et gérondif est problématique. *En français classique, la tripartition des formes en –ant ne va pas de soi (...) dans la mesure où la différence syntaxique et sémantique entre les trois catégories ne se marque pas formellement par une morphologie distinctive : le gérondif, invariable, se distingue mal du participe (au masculin singulier) du fait qu'il n'est pas régulièrement précédé de en ; le participe qui peut être variable en genre et en nombre, se distingue mal de l'adjectif verbal.* (N. Fournier, 2002, § 421 : 291-292)

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Participe, adjectif verbal, gérondif	G	
1	Type	participe présent - adjectif verbal - gérondif	a	<i>chantant, (en) chantant</i>
		participe passé - adjectif verbal	e	<i>instruit</i>

7. Adverbes

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Adverbe	R	
1	Type	général	g	<i>fortement, hier, ici</i>
		particule	p	<i>ne, n'</i>
		interro-exclam	t	<i>où, quand comment, pour-quoi</i>

Comparaison par / à MULTEXT (*en*) et GRACE

ATTRIBUT : TYPE

- MULTEXT (*en*) propose 2 valeurs : « modifier » (m), « spécifier » (s),
- GRACE propose 3 valeurs : général (g), particule (p), interro-exclam (x)
- PRESTO_MIN reprend les 3 valeurs proposées par GRACE mais affecte le code (t) aux adverbes interro-exclam. pour conserver une cohérence avec le code utilisés pour les pronoms et les déterminants interro-exclam.

8. Prépositions (Adpositions)

P	Attribut	Valeur	Code	Exemple
0	CATEGORIE	Préposition	S	

Comparaison par / à MULTEXT (*en*) et GRACE

ATTRIBUT : TYPE

- MULTEXT (*en*) propose 2 valeurs : « preposition » (p), « postposition » (t),
- GRACE propose 2 valeurs : « préposition » (p), « déictique » (d)
- PRESTO_MIN propose la seule valeur « préposition ».

9. Conjonctions (conjunction)

P	Attribut	Valeur	Code	Exemple
0	CATEGORY	Conjonction	C	
1	Type	coordination	c	<i>mais</i>
		subordination	s	<i>que</i>

Pas de modifications pour les attributs du « type » par rapport à MULTEXT (*en*) et GRACE.

Rem. : *donc* est traité comme un adverbe

10. Numéral (Numeral)

P	Attribut	Valeur	Code	Exemple
0	CATEGORY	Numeral	M	
1	Type	cardinal	c	<i>deux</i>
		ordinal	o	<i>deuxième</i>

Pas de modifications par rapport à MULTEXT (*en*).

Pour GRACE, voir 1.1.

Rem. : le mot *dernier* est codé Ag (n'est pas un numéral).

11. Interjections (Interjections)

P	Attribut	Valeur	Code	Exemple
0	CATEGORY	Interjection	I	Hep !

Pas de modifications par rapport à MULTEXT (*en*) et GRACE

12. Ponctuations (Punctuations)

P	Attribut	Valeur	Code	Exemple
0	CATEGORY	Numeral	F	
1	Type	forte	s	, ! ?
		faible	w	, : ;
		Autre (trait d'union, tiret, ponctuation parenthétique, ...)	o	- () []

Catégorie absente de MULTEXT (*en*)

Présente dans GRACE, qui ne propose pas de champ <type>.

13. Résidu (Residual)

P	Attribut	Valeur	Code	Exemple
0	CATEGORY	Résidu	X	
1	Type	abréviation	a	<i>Dir.</i>
		mot étranger	e	<i>linguistics</i>

	symbole	s	@
	préfixe	p	<i>hyper-, ex-</i>
	consonne intercalée	i	<i>a-t-on, l'on</i>

Cette catégorie, absente de GRACE [catégorie « unknown » (?)], est empruntée à MULTEXT (*en*), qui ne propose pas de champ <type>.

Rem : Les abréviations que nous faisons figurer comme « valeur » pour l'attribut « type » de la catégorie « résidu » font l'objet d'une catégorie spécifique dans MULTEXT (*en*): « Abbreviation » (Y).

INDEX

Index des attributs

Attribut	Catégorie	Position
Type	Adjectif	1
Type	Adposition	1
Type	Adverbe	1
Type	Conjonction	1
Type	Déterminant	1
Type	Interjection	1
Type	Nom	1
Type	Numéral	1
Type	Participe-Adjectif-Gérondif	1
Type	Ponctuation	1
Type	Pronom	1
Type	Résidu	1
Type	Verbe	1
Vforme	Verbe	2

Index des valeurs

Valeur	Code	Attribut	Catégorie
abréviation	a	type	résidu
article défini	a	type	déterminant
article indéfini	n	type	déterminant
article partitif	p	type	déterminant
autre	o	type	ponctuation
autre	v	type	verbe
cardinal	c	type	numéral
commun	c	type	nom
consonne intercalée	i	type	résidu
coordination	c	type	conjonction
démonstratif	d	type	déterminant
démonstratif	d	type	pronom
être & avoir	u	type	verbe
faible	w	type	ponctuation
forte	s	type	ponctuation
général	g	type	adjectif
général	g	type	adverbe
indéfini	i	type	déterminant
indéfini	i	type	pronom
infinitif	n	Vforme	verbe
interrogatif/ exclamatif	t	type	adverbe
interrogatif/ exclamatif	t	type	déterminant
interrogatif	t	type	pronom
mot étranger	e	type	résidu
ordinal	o	type	numéral

participe passé - adjectif verbal	e	type	Participe, adjectif verbal, gérondif
participe présent - adjectif verbal - gérondif	a	type	Participe, adjectif verbal, gérondif
particule	p	type	adverbe
personnel	p	type	pronom
possessif	s	type	adjectif
possessif	s	type	déterminant
possessif	s	type	pronom
préfixe	p	type	résidu
propre	p	type	nom
relatif	r	type	déterminant
relatif	r	type	pronom
subordination	s	type	conjonction
symbole	s	type	résidu
V conjugué à un mode personnel	c	Vforme	verbe