

FORMAT DE DESCRIPTION LEXICALE POUR LE FRANÇAIS

partie 2 : Description morpho-syntaxique

réf. GRACE GTR-3-2.1

M. Rajman (EPFL), J. Lecomte (INaLF), P. Paroubek (INaLF)
révision du GTR-3-2.1 suite à la journée atelier d'avril 97 à Avignon.

30 Juin 1997

La description lexicale (sous forme de traits morpho-syntaxiques) donnée ci-après pour le français repose sur le standard de description préconisé dans les projets MULTEXT [2] et EAGLES [4].

1 Descriptions lexicales

Les catégories définies sont celles indiquées dans le tableau suivant :

Catégorie	(Part-of-Speech)	Code
Nom	(Noun)	N
Verbe	(Verb)	V
Pronom	(Pronoun)	P
Adjectif	(Adjective)	A
Déterminant	(Determiner)	D
Adverbe	(Adverb)	R
Adposition	(Adposition)	S
Conjonction	(Conjunction)	C
Interjection	(Interjection)	I
Résidu	(Residual)	X
Punctuation	(Ponctuation)	F
Unknown	(Forme extra-lexicale)	?

Modifications par rapport à MULTTEXT : suppression de la catégorie « Numeral » qui n'est pas considérée ici comme une catégorie grammaticale particulière mais plutôt une sous-catégorisation de certaines parties du discours. Elle a donc été remplacée par l'adjonction d'une valeur additionnelle « cardinal » (notée : k) à l'attribut « Type » des différentes catégories syntaxiques (nom, pronom, adjectif et déterminant) intégrant des emplois de numéraux cardinaux, ainsi que l'illustrent les exemples suivants :

deux amis	→	<i>déterminant</i>
les deux amis	→	<i>adjectif</i>
le chapitre deux	→	<i>adjectif</i>
deux sont venus	→	<i>pronom</i>
ils sont venus tous les deux	→	<i>pronom</i>
ce soir le programme de la Deux est mauvais	→	<i>nom</i>
celui qui a une paire de deux ouvre la partie	→	<i>nom</i>

Les modifications adoptées soit lors de la journée GRACE d'avril 97 à Avignon avec le concours des participants présents, soit par interaction sur la liste de discussion, sont les suivantes:

- suppression du trait graphie (« *graphic* »),
- modification des motifs pour les cardinaux (ajout des indications de genres au formes plurielles des noms),
- ajout du motif pour les pronoms réflexifs,
- amalgame des motifs des pronoms personnels à la 1^{ère} et 2^{ème} personne avec ceux des pronoms personnels à la 3^{ème} personne,
- précisions concernant :
 - le trait degré pour les adjectifs (remarque 5),
 - la valeur de trait reflexif pour les pronoms (remarque 2),
 - la décomposition des pronoms relatifs contractés (remarque 5),
 - les pronoms relatifs (remarque 6).

ATTENTION ! Dans la suite de ce document, les descriptions lexicales qui seront associées aux formes données en exemple comporteront systématiquement la partie ILEX (identificateur lexical) définie dans [1] comme composée du lemme + la partie invariable de la descriptions morpho-syntaxiques (DMS).

Rappelons que lors de la campagnes d'évaluation, les participants n'auront à fournir que les graphies accompagnées des DMS.

L'information additionnelle apportées par les ILEX que nous mentionnons dans ce document pourra éventuellement être utilisée pour justifier nos choix linguistiques et affiner a posteriori le calcul des mesures de performance par des filtrages sélectifs.

1.1 Noms (Nouns)

Modifications par rapport à MULTTEXT : ajout de la valeur « cardinal » (k)
au trait « Type ».

Attribute	Value	Example	Code
Type	common	livre	c
	proper	Jean	p
	cardinal	deux	k
Gender	masculine	homme	m
	feminine	femme	f
Number	singular	homme	s
	plural	femmes	p

pattern(s) :

N[cp] [mf] [sp]

Nk[mf] [sp]

1.2 Verbes (Verbs)

Modifications par rapport à MULTTEXT : le trait « Clitics » n'est pas repris.

Attribute	Value	Example	Code
Type	main auxiliary	partir avoir	m a
Mood/Vform	indicative subjunctive imperative conditional infinitive participle	viens vienne viens viendrais venir venu	i s m c n p
Tense	present imperfect future past	viens venais viendrai vins	p i f s
Person	first second third	suis es est	1 2 3
Number	singular plural	viens venons	s p
Gender	masculine feminine	venu venue	m f

pattern(s) :

V[ma]n---- (infinitif)
V[ma]pp--- (participe présent)
V[ma]ps-[sp] [mf] (participe passé)
V[ma]ps--- (participe passé aux. être, verbes d'état et impersonnels)
V[ma]i[pifs] [123] [sp]- (indicatif)
V[ma]s[pi] [123] [sp]- (subjunctif)
V[ma]cp [123] [sp]- (conditionnel)
V[ma]mp1p- (impératif 1e pers.)
V[ma]mp2[sp]- (impératif 2e pers.)

Rem1: Pour ce qui est du codage du genre des participes passés, on affectera la valeur « - » (i.e. non applicable) pour le trait « Gender » et « Number » des participes passés de l'auxiliaire être, des verbes d'état (sembler, paraître, ...) et des verbes impersonnels (falloir, ...).

1.3 Adjectifs (Adjectives)

Attribute	Value	Example	Code
Type	qualificat.	bon	f
	ordinal	deuxième	o
	cardinal	deux	k
	indefinite	quelconque	i
	possessive	mien	s
Degree	positive	bon	p
	comparative	meilleur	c
Gender	masculine	bon	m
	feminine	bonne	f
Number	singular	bon	s
	plural	bons	p

pattern(s) :

Af[pc] [mf] [sp] (adjectifs qualificatifs: « bon », « mauvais »,...)

A[ikos] - [mf] [sp] (adjectif indéfini, cardinal, ordinal, possessif)

Rem1: La valeur « cardinal » pour le trait « Type » permet de représenter les adjectifs des numéraux cardinaux (ex: les deux amis, le chapitre un, ...). Ainsi, dans « les deux amis », « deux » sera décrit par :

deux → **deux/Ak: Ak-mp**

Rem2: La valeur « indefinite » pour le trait « Type » correspond aux emplois déterminés des adjectifs indéfinis (ex: une quelconque décision, un certain pouvoir, ...). Pour leurs emplois non déterminés (ex: certains hommes, plusieurs choix, ...), les adjectifs indéfinis sont classés comme « déterminants indéfinis ».

Rem3: La valeur « possessive » pour le trait « Type » correspond aux adjectifs possessifs « mien, tien, sien, ... » dans des emplois (rares) comme « le mien cousin » ou « cette œuvre mienne ». Les adjectifs possessifs habituels (mon, ton, son, ...) sont classés comme « déterminants possessifs ». Les formes comme « le mien », « le tien », ... seront classées comme pronoms possessifs composés.

Rem4: Les formes classiquement appelées « adjectifs interrogatifs » (ex: quel homme?) sont classées comme « déterminants interrogatifs ».

Rem5: Pour le trait « Degree » seuls les 3 comparatifs syncrétiques « meilleur », « pire » et « moindre » pourront avoir pour valeur de trait « Comparative ». Pour tous les autres adjectifs qualificatifs, y compris « bon », « mauvais », « petit », le trait « Degree » aura la valeur « Positive ». Pour les autres adjectifs, non qualificatifs, le trait aura la valeur non-applicable : « - ».

Rem6: Notons qu'il aurait pu être intéressant de disposer, pour les adjectifs qualificatifs concernés, de l'information indiquant qu'ils correspondent en fait à une forme adjectivée de participe passé, cependant nous avons décidé d'assimiler ces formes à des adjectifs qualificatifs ordinaires.

Ainsi une forme comme « passé » peut correspondre aux descriptions lexicales suivantes :

passé	→	passé/Ncm:Ncms	(nom)
passé	→	passé/Sp:Sp	(préposition)
passé	→	passer/Vmps:Vmps-sm	(participe passé)
passé	→	passer/Af:Afpms	(adjectif)

1.4 Pronoms (Pronouns)

Modifications par rapport à MULTTEXT : suppression de la valeur neutre pour le trait « Gender », modification des valeurs du trait « Case », ajout de la valeur « cardinal » (k) au trait « Type ».

Attribute	Value	Example	Code
Type	personal	je	p
	demonstrat.	celui	d
	indefinite	certain	i
	possessive	le mien	s
	interrog.	lequel	t
	relative	qui	r
	reflexive	se, soi	x
	cardinal	deux	k
Person	first	je	1
	second	tu	2
	third	il	3
Gender	masculine	il, celui	m
	feminine	elle, celle	f
Number	singular	aucun	s
	plural	certain	p
Case	nominative	il	n
	accusative	me, le	a
	dative	lui, en, y	d
	oblique	moi	o
Possessor	singular	le mien	s
	plural	le nôtre	p

pattern(s) :

Ps[123][mf][sp]-[sp] (pronoms possessifs)

Pp[123][mf][sp][nado]- (pronoms personnels)

P[dirtk]-[mf][sp]-- (pronoms démonstratifs, indéfinis, relatifs, interrogatifs, cardinaux)

Px[123][mf][sp]-- pronoms réflexifs

Rem1: Le système de cas proposé dans EAGLES (nominative, accusative, dative, genitive, ...) ne correspond pas directement à la typologie usuellement utilisée en français pour les pronoms « personnels » :

– formes conjointes :

sujet : je, tu, il, elle, on, nous, vous, ils, elles

objet direct : me, te, (se), le, la, nous, vous, les

oblique (obj indir ou CC) : me, te, (se), lui, en, y, nous, vous, leur

Rem: en toute rigueur, en et y ne peuvent remplacer des êtres animés et seraient plutôt des pronoms adverbiaux. Comme ils sont conjoints

au verbe, nous les assimilons cependant à des personnels obliques ou accusatifs.

ex: il parle de la voisine → il en parle
il pense à la voisine → il y pense
il y a du pain → donne m'en

– formes disjointes (d'insistance ou après préposition) :
moi, toi, (soi), lui, elle, nous, vous, eux, elles

On pourra utiliser la mise en correspondance suivante :

Nominative → sujet
Accusative → objet direct
Dative → objet indirect et CC
Oblique → formes disjointes

Rem2: La valeur « reflexive » du trait « Type » sera utilisée pour identifier les pronoms personnels réfléchis conjoints au verbe uniquement (me, te, se, nous, vous); les réfléchis disjointes sont considérés comme des personnels disjointes (à moi, pour soi-même...).

Rem3: La valeur « indefinite » du trait « Type » permet de représenter les pronoms indéfinis identificateurs et les pronoms indéfinis quantificateurs non cardinaux. Exemples :

<i>identificateurs</i>	« l'un », « l'une », « les uns », « les unes », « le même », « la même », « les autres », ...
<i>quantificateurs non cardinaux</i>	« beaucoup », « plusieurs », « peu », « nul », ...

Rem4: La valeur « cardinal » (k) du trait « Type » permet d'identifier les pronoms cardinaux, (ex: *trois* sont venus, *deux* de ses amis, il en a *un*, ...). Ainsi, dans « trois sont venus », « trois » sera décrit par :

trois → **trois/Pk:Pk-[mf]p--**

Rem5: Les pronoms relatifs composés sont décomposés en 2 constituants ex: desquels = de + lesquels.

Rem6: La personne n'est pertinente que pour les pronoms personnels, les réfléchis et les possessifs. Le cas n'est pertinent que pour les pronoms personnels.

1.5 Déterminants (Determiners)

Modifications par rapport à MULTEXT : modification de la valeur « interrogative » du trait « Type » en « interrogative/exclamative », ajout des valeurs « relative » (r) et « cardinal » (k) au trait « Type ».

Attribute	Value	Example	Code
Type	article	le, un	a
	demonstrat.	ce	d
	possessive	mon	s
	indefinite	certain	i
	inter/excl.	quel	t
	relative	lequel	r
	cardinal	un, deux	k
Person	first	ma	1
	second	ta	2
	third	sa	3
Gender	masculine	le	m
	feminine	la	f
Number	singular	le	s
	plural	les	p
Possessor	singular	mon	s
	plural	nos	p
Nature	definite	le	d
	indefinite	un	i

pattern(s) :

Ds[123] [mf] [sp] [sp]- (déterminants possessifs)
D[ditrk]- [mf] [sp]-- (déterminants démonstratifs, indéfinis, interro-exclamatifs, relatifs, cardinaux)
Da- [mf] [sp]- [di] (déterminants articles)
Da---i marqueur dissocié du partitif (de + la, du = de + la)

Rem1: La valeur « indefinite » du trait « Type » correspond aux emplois déterminatifs des adjectifs indéfinis (ex: *certain*s hommes, *plusieurs* solutions, ...).

Rem2: La valeur « relative » du trait « Type » correspond aux déterminants relatifs, définis dans EAGLES [4], comme « lequel » dans « le numéro de son bureau, lequel bureau est en travaux, a été modifié ».

Rem3: Pour le codage de « un », qui peut être codé soit **Da-ms-i** (article indéfini), soit **Dk-ms--** (cardinal), EAGLES [4] propose arbitrairement de *toujours* choisir le type « Article ». Nous proposons de suivre également cette règle pour le codage du corpus de référence GRACE, tout en se réservant la possibilité de corriger les mesures qui seront faites en filtrant a posteriori les occurrences de « un » et « une ».

Rem4: articles partitifs et articles contractés

Les articles ***partitifs*** (i.e. indiquant la partie d'un tout) sont en français formés par l'association d'un marqueur partitif « de » et d'un déterminant article ou autre. Cette association peut avoir pour résultat :

1. une forme contractée, composée du marqueur partitif + l'un des deux article définis suivants : « le » ou « les »), par exemple :

du (de le) : manger du pain, avoir du courage, faire du vélo
des (de les) : manger des épinards, avoir des remords
(ne pas confondre l'emploi partitif de « des » dans
« manger des épinards » avec son emploi indéfini dans
« manger des biscuits »)

2. une forme non contractée, composée soit du marqueur partitif + l'article défini « la », soit du marqueur partitif + un autre déterminant, ou soit encore du marqueur partitif seul; par exemples :

de la : boire de la bière, éprouver de la répulsion
de ce : ne pas manger de ce pain-là
de mon : boire de mon eau
de : ne pas manger de pain
de : ne pas avoir de conscience,
de : ne pas prendre de précautions

...

Les ***articles contractés*** (encore appelés prépositions contractées) sont en français formés par la contraction d'une préposition (« à », « de » ou « en ») avec les articles définis « le » ou « les »

Les articles contractés (i.e. formés de deux éléments réunis en un seul) sont, en français, les formes suivantes :

au (à le) : aller au Creuzot
aux (à les) : aller aux Etats-Unis
du (de le) : revenir du Creuzot
des (de les) : revenir des Etats-Unis
ès (en les) : docteur ès sciences

Compte-tenu de ces différents éléments, on aura donc les interprétations lexicales (éventuellement ambiguës) suivantes :

de	: préposition ou marqueur partitif ou article indéfini pluriel
du (de le)	: séquence « marqueur partitif + article défini » contractée ou séquence « préposition + article défini » contractée
des (de les)	: article indéfini ou séquence « marqueur partitif + article défini » contractée ou séquence « préposition + article défini » contractée
au (à le)	: séquence « préposition + article défini » contractée
aux (à les)	: séquence « préposition + article défini » contractée
ès (en les)	: séquence « préposition + article défini » contractée
de la, de ce, de mon, ...	: article partitif composé ou séquence « préposition + article défini »

Du fait que la distinction entre « article partitif » et séquence « préposition + article défini » nécessite des connaissances syntaxiques fines (la connaissance des schémas verbaux, par exemple, pour discriminer entre « il va parler des soupes » et « il a mangé des soupes »), voire des connaissances sémantiques (pour discriminer entre « il n'a plus d'épinards » et « il n'a plus de biscuits »), le choix (en contexte) entre les différentes interprétations possibles des formes « de », « du » et « des » est délicate. Elle n'est cependant pas impossible et nous avons donc décidé de conserver la possibilité d'exprimer, dans les descriptions morpho-syntaxiques GRACE, la distinction entre ces différentes possibilités. Ce choix nous permettra en effet, lors de l'évaluation, une mesure plus fine des différentes options implémentées dans les systèmes (fusion/pas fusion des différents emplois des formes partitives et contractées).

A l'aide de la notation proposée dans [1] pour les formes contractées, il est possible, pour une forme comme « des » par exemple, de distinguer les emplois suivants :

- article contracté :
ex: je viens *des* Etats-Unis

des → de/Sp:Sp+le/Dad:Da-mp-d
- article partitif :
ex: il mange *des* épinards

des → de/Dai:Da----i+le/Dad:Da-mp-d
- article indéfini :
ex: il apporte *des* livres

des → un/Dai:Da-mp-i

1.6 Adverbes (Adverbs)

Modifications par rapport à MULTTEXT : valeur interro-exclamative pour le trait « Type ».

Attribute	Value	Example	Code
Type	general	fortement	g
	particle	ne	p
	interro-excl	pourquoi	x
Degree	positive	fortement	p
	comparative	davantage	c
	negative	ne, pas	n

pattern(s) :

Rg[pcn] (adverbe)

Rx[pcn] (adverbe interro-exclamatif)

Rpn (ne, n')

Rem1: Les valeurs du trait « Degree » seront interprétées de la façon suivante :

- comparative : identifie les formes adverbiales *intrinséquement* comparatives (et qui par conséquent ne peuvent former un comparatif en « plus + adv »); ne s'applique qu'aux adverbes « pis » (« plus mal »), « davantage » (« plus beaucoup »), « mieux », « plus », et « moins ».
- negative : identifie les adverbes pouvant se combiner avec la particule adverbiale « ne »;
- positive : identifie les adverbes n'entrant dans aucune des deux catégories précédentes;

Compte-tenu des ces interprétations, dans une phrase comme « il est venu plus rapidement que Pierre », « rapidement » devra être décrit par :

rapidement → rapidement/Rgp:Rgp

Rem2: Le trait « particle » identifie le comportement « proclitique » de certaines formes adverbiales et s'applique donc spécifiquement à « ne » .

1.7 Prépositions (Adpositions)

Modifications par rapport à MULTTEXT : adjonction d'une valeur déictique pour le trait « Type » (voir aussi [3]).

Attribute	Value	Example	Code
Type	preposition	en, de	p
Type	déictique	voici, voilà	d

pattern(s) :

Sp (préposition)

Sd (déictique)

1.8 Conjonctions (Conjunctions)

Pas de modifications par rapport à MULTTEXT.

Attribute	Value	Example	Code
Type	coordinat.	et	c
	subordinat.	que	s

pattern(s) :

C[cs]

1.9 Interjections (Interjections)

Pas de modifications par rapport à MULTEXT.

Tag	Example
I	eh

pattern(s) :
I

Rem: On classera les onomatopées comme « chut », « bzzz » parmi interjections.

1.10 Résidus X (Residuals)

Modifications par rapport à MULTEXT : sémantique différente.

Attribute	Value	Example	Code
-----------	-------	---------	------

pattern(s) :
X

Rem: Les éléments de la classe des résidus devront être interprétés comme des formes lexicales attestées dont la classification définitive n'a pas encore été réalisée. Comme exemples de telles formes, nous pouvons citer les mots (ou abréviations) latins (*quid* dans « quid de la décision concernant... », *cf* dans « cf le chapitre sur... ») ou des formes comme le « t euphonique » par exemple.

2 Descriptions additionnelles

2.1 Ponctuations (Punctuations)

La description proposée dans [3] distingue les ponctuations.

Tag	Example
F	.:?

pattern(s) :

F

Rem: La classe « Ponctuation » sera utilisée pour représenter toutes les formes assimilables à des ponctuations, parmi lesquelles on peut citer :

- les ponctuations au sens strict, comme « . », « ! », ...
- les séparateurs, comme le tiret, l’apostrophe, les guillemets, ...

2.2 Formes extra-lexicales

Tag	Example
?	juornaliste

pattern(s) :

?

Rem: Seront considérées comme formes extra-lexicales strictes (i.e. non recensées dans le lexique) des formes erronées comme « juornalistes » ou inconnues comme « Xlws » par exemple. « juornalistes » et « Xlws » seront ainsi décrits (hors contexte) par :

```
juornalistes juornalistes/? : ?  
Xlws Xlws/? : ?
```

Leur description en contexte pourra se faire par référence aux catégories définies dans ce document. Par exemple, dans « les journalistes ont été convoqués à la revue de presse », « journalistes » pourra être décrit par :

```
juornalistes journaliste/Ncm:Ncmp
```

Références

- [1] Rajman M.,
« Format de description lexicale pour le français - partie 1 : Format de description lexicale pour le français - Concepts généraux », Martin Rajman, rapport GRACE GTR-2-1.2, Mars 1996

- [2] Ide, N., Véronis, J.,
« MULTEXT (Multilingual Text Tools and Corpora) ». Proceedings of the
14th International Conference on Computational Linguistics, COLING'94,
Kyoto, Japan, 1994.

- [3] Lecomte, J.,
« Morphosyntactic Annotation - French Contribution », Deliverable (report)
for Workpackage 4-1-4a, Project PAROLE LRE 63-368, Février 1996.

- [4] von Reckowski, Ursula,
« ELM-FR: A typed French incarnation of the EAGLES-TS - Definition of
Lexical Specification and Classification Guidelines », GSI-Erli, Janver 1996.