Manual d'annotacion morfosintactica del projècte RESTAURE per l'occitan

Myriam Bras, 10 novembre 2017 (modifié 13/11/2017 ; 02-08/02/2018)
CLLE-ERSS

Ce document présente la deuxième campagne d'annotation morphosyntaxique du projet RESTAURE pour l'occitan (section 1). Les textes à annoter sont des textes déjà pré-annotés avec l'analyseur morpho-syntaxique TALISMANE de Assaf Urielli, entrainé par Marianne Vergez-Couret dans le cadre du projet RESTAURE. Il va s'agir de corriger les annotations proposées en utilisant le logiciel AnaLog de Marie-Hélène Lay. Le jeu d'étiquettes à utiliser est fourni en section 3, à la suite des instructions techniques en section 2, une explicitation détaillée des étiquettes est donnée en annexe. Le manuel de prise en main guidée du logiciel AnaLog peut également être consulté.

1. Campagne d'annotation : objectifs et déroulement

Nous visons la constitution d'un corpus annoté d'environ 12000 tokens ou mots-formes. Plus d'un quart du corpus (environ 4000 tokens) a déjà été déjà annoté dans le cadre de la première campagne pour l'entrainement de TALISMANE pour l'occitan par Marianne Vergez-Couret. Il reste donc environ 8000 mots-formes à annoter, à trois annotateurs (Myriam Bras, Louise Esher, Jean Sibille). Le corpus comporte des extraits de textes de la base textuelle BaTelòc (60 formes maximum pour respecter le droit de citation) et de courts articles du journal en ligne Lo Jornalet (de 200 à 900 formes).

La campagne se déroulera de la façon suivante :

- Etape 1 : annotation d'un extrait de 350 tokens par les 3 annotateurs pour harmonisation
- Etape 2 : annotation de 2250 tokens environ par chaque annotateur
    - Etape 2.1 : premier lot de textes (en languedocien, gascon, provençal)
    - Etape 2.2 : second lot de textes (langedocien, gascon, provençal, vivaro-alpin, auvergnat et limousin)
    - Une triple annotation d'un extrait de 200 tokens sera effectuée au milieu de l'étape 2 (entre 2.1 et 2.2)
- Etape 3 : annotation d'un extrait de 200 tokens par les 3 annotateurs pour calcul des accords par paires d'annotateurs

La liste des textes à annoter, et, pour chacun d'eux, l'état d'avancement de l'annotation, est consultable (et modifiable) par les annotateurs.

Pour évaluer les ressources nécessaires à des campagnes d'annotation ultérieures, il est recommandé aux annotateurs :

- de compter le temps passé à l'annotation de chaque extrait
- de noter toute difficulté rencontrée

2. Instructions techniques

- Mettre les fichiers .csv et .csv.meta dans le répertoire AnaLog\DATA\OUT\TABLE_CSV\TABLEAU_VALIDATION

- Ouvrir Analog (fichier *ANALOG v…. .jar* sous Mac, *LANCEMENT-Analog.bat* sous PC)

- Dans Analog, cliquer sur « Analyser Texte », et ensuite sur « Choisir une analyse antérieure ».

- Là, sélectionner le fichier .csv

Toutes les colonnes « parties du discours » s'affichent dans un tableau contenant le texte annoté. L'information à vérifier se trouve dans les colonnes bleues sur votre gauche (situées juste à droite de la colonne « forme rencontrée »).

Il est possible de n'afficher que les colonnes « parties du discours » qui sont instanciées dans le texte en allant dans « choix pour l'affichage » et en cliquant sur « voir le tableau »

- Se mettre en mode validation et se placer dans le tableau

L'annotation consiste à modifier les informations situées dans les pavés bleus.

Les corrections possibles sont les suivantes :

- changer un pavé de colonne
- corriger le lemme
- ajouter un lemme dans la bonne colonne

Pour valider la modification, soit appuyer 2 fois sur la touche entrée, soit sélectionner la ligne et cliquer sur entrée, soit recliquer ailleurs puis sur la ligne (pas besoin d'enlever le tiret dans la mauvaise colonne, ce qui importe est ce qui est dans les colonnes de gauche à côté de la forme). Les informations modifiées s'affichent en rouge.

Pour trier le tableau, il suffit de cliquer sur le haut de la colonne selon laquelle on veut trier.

On peut par exemple cliquer sur « Forme rencontrée » : ça met par ordre alphabétique les formes et permet d'avoir à côté des lignes à modifier de la même façon. On peut alors valider les modifications en sélectionnant toutes les lignes et en appuyant sur la touche entrée

Pour revenir à l'ordre du texte, il suffit de cliquer sur « Mot n° » en haut à gauche.

Si on souhaite afficher une colonne pour une catégorie « partie du discours » qui n'est pas encore instanciée, aller dans « choix pour l'affichage », cocher la colonne choisie puis sur « voir le tableau »

- Avant de quitter la session : exporter et choisir le même nom de fichier

option « exporter dans un fichier csv », puis « exporter le tableau » puis soit donner un nouveau au fichier si on veut sauvegarder une nouvelle version, soit sélectionner le nom du fichier actuel et la nouvelle version écrasera l'ancienne

NB : comme on part de fichiers déjà annotés, on n'utilise pas les dictionnaires, mais il faudrait explorer la possibilité d'en utiliser un pour les formes inconnues qui pourraient ainsi être annotées automatiquement dans tout le fichier une fois ajoutées dans le dictionnaire (grâce à la fonction « Re-annoter » = « annoter à nouveau le texte en utilisant les informations de ce tableau »).

3. Jòc d'etiquetas complet per l'annotacion morfosintactica (38 etiquetas)

A = Adjectius

Af	Adj qualificatiu
Ao	Adj ordinal
Ak	Adj cardinal
Ai	Adj indefinit
As	Adj possessiu

C = Conjonccions

Cc	Conj de coordinacion
Cs	Conj de subordinacion

D = Determinants

Da	Det article
Dd	Det demonstratiu
Di	Det indefinit
Ds	Det possessiu
Dt	Det interrogatiu/exclamatiu
Dr	Det relatiu
Dk	Det cardinal
Dp	Det partitiu

F = ponctuacion

F

I = Interjeccions

I

N = Noms

Nc	Nom comun
Np	Nom pròpri
Nk	Nom cardinal (k)

P = Pronoms

Pp	Pronom personal
Pd	Pronom demonstratiu
Pi	Pronom indefinit
Ps	Pronom possessiu
Pt	Pronom interrogatiu
Pr	Pronom relatiu
Px	Pronom reflexiu
Pk	Pronom cardinal

R = advèrbis

Rg	Adv general
Rx	Adv interrogatiu/exclamatiu
Rp	Adv particula
Rq	Adv intensiu/quantitatiu

S = preposicions

Sp	Preposicion
Spda	Preposicion + article
Sd	Deictic

V = Vèrbes

Vm	Màger
Va	Auxiliari

X = demai, çò autre

X

Exemples de lemma per cada etiqueta

A = Adjectius

Af	Adj qualificatiu	polit, triste, …
Ao	Adj ordinal	primièr, segond, tresen, …
Ak	Adj cardinal	dos, tres, …
Ai	Adj indefinit	cèrt
As	Adj possessiu	miu, teuna, seu, sieuna, …

C = Conjonccions

Cc	Conj de coordinacion	e, mas, que …
Cs	Conj de subordinacion	quand, coma, que, se…

D = Determinants

Da	Det article	lo, la, los, las, un, una, de
Dd	Det demonstratiu	aquel, aqueste, aiceste, este
Di	Det indefinit	cada, qualque, mantun, mai d'un, tot
Ds	Det possessiu	mon, ton, son, ma, ta, sa, nòstre, vòstre, lor, nòstra, vòstra, lors…
Dt	Det interrogatiu/exclamatiu	quin, qual, quun, quane
Dr	Det relatiu	lo qual (??)
Dk	Det cardinal	un, dos, tres
Dp	Det partitiu	de

F = ponctuacion : , ; . ? ! (pas dans Loflòc)

I = Interjeccions : zo, i, a, o, òu, flica-flaca, pam ….

N = Noms

Nc	Nom comun : ostal, dròlla, …
Np	Nom pròpri : Maria, Aran, Tolosa, …
Nk	Nom cardinal (k) : dos (dins « un parelh de dos ») mas pas etiqueta Nk dins Loflòc

P = Pronoms

Pp	Pronom personal	ieu, tu, el, nosautres, eles, ne, …
Pd	Pronom demonstratiu	aquò, aquel, aqueste, çò, …
Pi	Pronom indefinit	pauc, qualques unes, mantuns, cadun, quicòm, degun, totòm, res, …
Ps	Pronom possessiu	miu, teu, vòstra…
Pt	Pronom interrogatiu	quant, quin, qual, que…
Pr	Pronom relatiu	Ont, dont, que, …
Px	Pronom reflexiu	Me, te, se, lor, …
Pk	Pronom cardinal	dos, tres, trenta-cinc, …

R = advèrbis

Rg	Adv general	pas, aisidament, bravament, ara, uèi, puèi, …
Rx	Adv interrogatiu/exclamatiu	quant
Rp	Adv particula	ne, non (quand ils accompagnent le « pas » negatif) ; *Enonciatifs du Gasc :* que, be, e, ja, si
Rq	Adv intensiu/quantitatiu	fòrça, plan, cap, pus, brica, mai, tot, tròp, gaire…

S = preposicions

Sp	Preposicion	per, de, coma, dins, abans, dempuèi, a, sus, jós, en …
Spda	Preposicion + article	del, dels, al, als, pel, pels …
Sd	Deictic	

V = Vèrbes

Vm	Màger	dansar, manjar, poder, èsser, aver
Va	Auxiliari	èsser, aver

X = demai, çò autre : n-, -n-, -z-

Décisions d'annotation à l'issue des 2 premières harmonisations

Quand la correction de l'annotation exige l'introduction d'une ligne supplémentaire[1]
Nous avons rencontré ce problème pour « amb de » catégorisé à tort comme une locution prépositionnelle (Sp) alors qu'il s'agissait d'une prep (Sp) suivie d'un article (Da).
Consigne :

- dans la ligne x mettre par exemple ici « amb » catégorie Sp
- noter dans un fichier séparé qu'il faut ajouter une ligne x+1 pour inscrire « de » catégorisé Da

Quand la correction de l'annotation exige la suppression d'une ligne
Par exemple « rendètz-vos » analizat coma Vm Pp, alara que lo calriá analizar coma un N compausat
Consigne :

- dans la ligne x mettre le lemme du nom composé, ici « rendètz-vos » et l'annoter Nc
- noter dans un fichier séparé qu'il faut supprimer la ligne x+1

Quand il y a une faute d'orthographe ou une faute de frappe dans le texte : corriger la faute
Idem dans le cas d'erreurs de syntaxe manifestes

Annotation des dates et des expressions temporelles
Pour les jours de la semaine, plutôt que de leur affecter 2 étiquettes (adverbe et nom), on les étiquette tout le temps Nc (en fait Nt = Nom de temps)
NB : l'emploi adverbial correspond à un SP avec Prep et Det vides (« dimars » = prep vide det vide Nc)

Idem pour les noms de mois et de saisons

Les horaires : « 18h » sont traités aussi comme des Noms de temps, donc des Nc
Mais quand on écrit en séparant le det du N « a 10 oras » ou « a dètz oras », 10 est un Dk.
Exemples : *Dissabte que ven, 23 d'abril* : Nc Pr Vm, Nk Sp Nc
Dijòus 21 : Nc Nk
A 18h : Sp Nc
A 10 oras : Sp Dk Nc
lo dijòus 23 de març : Da Nc Nk Sp Nc

Annotation des nombres
Même lemme pour formes « vint-e-tres » e « 23 » = « 23 » (contrairement aux alsaciens)
Les nombres ont 3 étiquettes possibles :

- déterminant : Dk : trenta euròs
- adjectif : Ak : los tres enfants
- pronom : Pk : ne vòli tres

[1] J'ai posé la question à Marie-Hélène Lay : pas possible pour l'instant d'ajouter une ligne en cours d'annotation dans Analog.

Annotations des composés en général
Face à un syntagme ou à une séquence de mots potentiellement liés entre eux par des relations de figement, plusieurs cas de figure peuvent se présenter :

- la séquence est analysée comme une forme composée, et c'est effectivement une forme composée → on accepte l'annotation
- la séquence est analysée comme une forme composée alors qu'elle a une lecture libre dans le cas présent → on annote le premier élément de la séquence et on indique dans le fichier annexe qu'il faut ajouter des lignes (cf. ci-dessus)
- la séquence est analysée comme une forme libre alors qu'elle a fonctionne comme une séquence figée dans le cas présent → on laisse l'annotation telle quelle, même si le sens est opaque, le repérage des composés se fera dans une autre phase.

Annotation des Noms propres composés
Choix d'analyser les séquences comportant un Np de type la *Lei d'Aran, l'Estanquet de la Robina* renvoyant à un lieu comme des séquences libres. Exemples :

Lei d'Aran : *lei* Nc (lemma *lei* sens majuscula), *d* Pr, *Aran* Np

Generalitat de Catalonha : *Generalitat* Nc (lemma *parlament* sens majuscula) , *de* Pr, *Catalonha* Np.

Conselh General d'Aran : *conselh* Nc, *general* Aj, *d* Pr, *Aran* Np

La justificacion es que, emai se *Lei d'Aran*, *Parlament de Catalonha* e *Conselh General d'Aran* se pòdon benlèu considerar coma de Np (compausat), aquò relèva d'un segond nivel d'analisi e que al primièr nivèl, dins *Lei d'Aran*, *lei* es un Nc e *Aran* un Np etc.

NB : lo mot « sant » dabans lo nom d'un sant, coma « Sant Miquèu » : es un nom comun avec una majuscula, coma Conselh, mas dins un nom de villatge coma per exemple « Sant Jòrdi de Lusençon » serà partida d'un Nom Pròpi compausat.
Doncas dins :
« Avèm rendètz-vos Plaça Sant Jòrdi » → Nc
« Vau a Sant Miquèl de Lanas « → Np

Cas de coalescences de + de
Dans certains cas, « de » correspond à « de (Sp) + de (Da) » est-ce un SpDa ou un Sp ?
Cf. : « s'emplis de taulièrs de libres, d'escribans que signan … »
En teoria seriá SpDa = de + de (coalescence entre los dos « de »)
mas fin finala, es pus simple de dire qu'es una prep sola (Sp) e qu'après aquela prep i a pas de det
→ Décision : on met Sp

Participes passés
Certains PP peuvent jouer le rôle d'adjectifs, dans ce cas on peut hésiter entre Vm et Af.
Tous les PP sont annotés Vm dans Loflòc parce qu'issus de la flexion de verbòc, certains sont aussi annotés Adj.
La présence d'un complément de temps ou de lieu rattaché au PP va dans le sens de Vm.
Exemple : *le café associatif ouvert chaque jeudi à partir de 18h*

Décision : laisser Vm sauf si indices clairs que c'est un Adj (modification par adv intensif, coord avec un autre adj, etc.).

NB : on pourrait utiliser une étiquette PP pour régler le pb, mais la stratégie utilisée pour Loflòc et dans Restaure est de ne pas multiplier les étiquettes

Infinitiu emplegat coma nom

« lo picar del jorn », « lo manjar e lo beure » Nc

Distinction Va/Vm et tornar + infinitiu / far + infinitiu
Va : être et avoir comme auxilaires pour conjugaison des temps composés + passif
Les « semi-auxiliaires » (factititifs, causatifs, aspect lexical) sont étiquetés Vm : leur sens n'est pas vide
tornar + INFINITIU : Vm Vm
far + INFINITIU : Vm Vm

Annotation de « que »
« que » pòt èsser analisat coma :
- un Pro relatiu (=ont)
- una conj de sub (=per çò que)
- una conjonccion de coordinacion (=car)

Parfois, plusieurs analyses sont possibles :
- si le choix de l'analyseur est une de ces analyses, on laisse la solution de l'analyseur
- sinon, on choisit celle qui nous semble le plus pertinente

Exemple : « una jornada fòrça especiala en Catalonha, **que** lo nòstre país fraire s'emplís de taulièrs de libres » → 3 analyses sont possibles, analyseur choisi Cs, on laisse cette étiquette

Cf. Remarca Joan sus tèxte Barsòti : *que* dins lo sens del francés *car* : *car* es tradicionalament considerat coma una Cc, s'òm considèra qu'es vertat, caldrá logicament etiquetar aquel *que* coma Cc.
→ per aquesta campanha, daissam aqueles « que » Cs

Annotation de « tot »
« tot » peut avoir plusieurs étiquettes : Ai, Nc, Pi, Rq, Rg

en tot festejar/ tot festejant → Rg

totas las filhas → 2 solutions :
- un Di compausat « tot lo »
- una seguida Di Da : tot Di + lo Da → choix de la solution 2 : Di Da

tota la vila → Di
tot un hormatge → Di

la vila tota → Ai

tot plen original → rôle de « tot » ici : Rq ; et de « plen » : Af , o « tot plen » coma un Rq compausat
tot en doçor → Rq
una tota pichona rota → Rq

Annotation de « i » e de « çò »
« i » dans « i a un arbre » : Pp (parce que c'est un pronom clitique)
« i » dans « Pèire es a la cantina. I vau tanben » : Pp (parce que pronom clitique et anaphorique) en situation de parole quelqu'un dit « bon, i vau », on met aussi Pp, parce que c'est un pronom clitique, qui a la même place devant le verbe que les autres
Loflòc propose deux possibilités : Rg et Pp, mas aici metrem Pp pertot
NB : dans les grammaires traditionnelles, on dit que c'est un pronom adverbial (c'est pour ça qu'on a les 2)

Çò : Pd (démonstratif/déictique)

Cò

en cò de, a cò de, a cò meu → « cò » : Nc (sòrta de substantiu calhat, perque Pd nos agrada pas)
en çò de, a çò de → Pd (coma endacòm mai)

Annotation des formes négatives discontinues

pas …. cap/pus/brica/mai/fòrça/plan/res/degun

pas → Rg
cap/pus/brica/mai/fòrça/plan → Rq
res/degun → Pi

Articles gascons

eth → lemme « eth », et on reliera les lemmes « eth », « le », « lo » après, a un autre nivèl (cf discussions sur Lofl̀oc)

Articles lemosins

Pel lemosin e los parlars qu'an *dels*, *de las* article indefinit plural

de las → Di compausat (la lectura Sp Da pòt existir tanben)
dels → Di compausat (la lectura SpDa pòt existir tanben)
del pan → Di
de la sopa → Di

Annexe : description détaillée du jeu d'étiquettes complet

Les étiquettes utilisées pour l'annotation sont celles choisies pour le lexique LOFLÒC. Ce jeu d'étiquettes est issu du standard de description GRACE (Rajman *et al.*, 1997), lui-même adapté du jeu d'étiquettes MULTEXT (Ide & Véronis, 1994) et EAGLES (von Reckowski, 1996).

Les étiquettes GRACE sont utilisées comme des étiquettes à 3 niveaux.

- Le premier niveau donne principalement la catégorie grammaticale (part-of-speech, POS) mais il y a également deux étiquettes, une pour la ponctuation (F) et l'autre pour des formes attestées dont la classification n'a pas encore été réalisée (X) :

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
Ponctuation	(Punctuation)	F

- Le deuxième niveau propose une classification sémantique ou fonctionnelle : dans la présente annotation on s'arrêtera au second niveau, avec les 38 étiquettes présentées dans la section 3.

- Le troisième niveau apporte des informations morphosyntaxiques. Quand un trait de description d'une étiquette est non pertinent, il est indiqué avec le symbole tiret '-' et quand l'information est manquante, cela est signalé avec un point d'interrogation '?'. IL est indiqué ici à titre indicatif.

Les modifications apportées par rapport au jeu d'étiquettes GRACE sont signalées au fur et à mesure de la description du jeu d'étiquettes ci-dessous.

1.1.1 Noms (N)

Niv 1	Nom (N)	**Valeur**	**Code**	**Exemple**	**Glose**
Niv 2	Type	common proper cardinal[1]	c p k	abelha Pèir tres	abeille Pierre trois
Niv 3	Genre	masculine feminine	m f	libre abelha	livre abeille
	Nombre	singular plural	s p	abelha abelhas	abeille abeilles

[1] Les cardinaux peuvent être employés comme nom :

a) un parelh de **dos** (une paire de **deux**).

Ils peuvent également être employés comme adjectifs (*cf.* 1.1.3), pronoms (cf. 0) et déterminants (*cf.* 1.1.6). Nous avons suivi les recommandations de GRACE sur ce point (et non celles de EAGLES).

Le patron de formation des étiquettes pour tous les noms peut être schématisé de la façon suivante :

Tous les noms	N[cpk][mf][sp][2]

[2] Les codes entre crochets constituent toutes les valeurs possibles pour chaque attribut.

1.1.2 Verbes (V)

Niv 1	Verbe (V)	**Valeur**	**Code**	**Exemple**	**Glose**
Niv 2	Type	main auxiliary[3]	m a	manjar aver	manger avoir
Niv 3	Mood/Vform	indicative subjunctive imperative conditional infinitive participle	i s m c n p	mangi mange manja manjariái manjar manjat	je mange que je mange mange ! je mangerais manger mangé
	Form[1]	positive negative	a n	manja manges	mange mange
	Tense	present imperfect future past	p i f s	mangi manjavi manjarai mangèri	mange mangeais mangerai mangea
	Person	first second third	1 2 3	mangi manjas manja	mange mange mange
	Number	singular plural	s p	mangi manjam	mange mangeons
	Gender	masculine feminine	m f	manjat manjada	mangé mangée

[1] Par rapport à GRACE, est ajouté l'attribut "Form" qui peut prendre deux valeurs "positive" ou "negative". Ce trait est seulement pertinent pour la valeur "imperative" de l'attribut "Mood". Il sert à annoter l'impératif en occitan qui peut prendre deux formes différentes selon qu'il est employé dans une phrase positive ou négative :

a) Manja ! (Mange)
b) Manges pas ! (Ne mange pas).

Les patrons de formation des étiquettes pour les verbes peuvent être schématisés de la façon suivante en tenant compte du fait que certains attributs peuvent être non pertinents (-) :

Infinitif	V[ma]n-----
Participe présent	V[ma]p-p---
Participe passé	V[ma]p-s-[sp][mf]
Indicatif	V[ma]i-[pifs][123][sp]-
Subjonctif	V[ma]s-[pi][123][sp]-
Conditionnel	V[ma]c-p[123][sp]-
Impératif	V[ma]m[an]p[12][sp]-

[3] Dans Loflòc v 1.0, les auxiliaires ne sont pas annotés. Tous les verbes sont annotés avec l'attribut "main".

1.1.3 Adjectifs (A)

Niv	Adjectif (A)	Valeur	Code	Exemple	Glose
Niv 2	Type	qualificative ordinal cardinal[1] indefinite[2] possessive[3]	f o k i s	bon centen dos cèrt meu	bon centaine deux certain mien
Niv 3	Degree[4]	positive comparative[5]	p c	bon melhor	bon meilleur
	Genre	masculine feminine	m f	bon bona	bon bonne
	Nombre	singular plural	s p	bon bons	bon bons

[1] La valeur "cardinal" pour le trait "Type" permet de rendre compte des emplois adjectivaux des numéraux cardinaux :

a) los **dos** amics (les **deux** amis).

[2] Ne pas confondre les emplois adjectivaux :

b) un **cèrt** monsen Martin (un **certain** Monsieur Martin)

et les déterminants :

c) **mantuns** còps (**plusieurs** fois)

[3] La valeur "possessive" pour le trait "Type" correspond aux adjectifs possessifs dans des emplois comme

d) lo **meu** libre (mon livre).

La forme simple du possessif ***mon*** *libre (mon livre)* est classée comme déterminant possessif. Les formes comme (e) sont classées comme pronom possessif :

e) Aquel libre es lo **meu** (Ce livre est le mien)

[4] La valeur "Degree" n'est pertinente que pour les adjectifs qualificatifs.

[5] Les adjectifs comparatifs de l'occitan sont *màger (plus grand)*, *melhor (meilleur)*, *mendre (moindre)* et *pièger (pire)*. Ils peuvent être variables ou invariables selon les parlers. Il faut donc systématiquement annoter les 3 niveaux.

Les patrons de formation des étiquettes pour les adjectifs peuvent être schématisés de la façon suivante :

Adjectifs qualificatifs	Af[pc][mf][sp]
Adjectifs indéfinis, cardinaux, ordinaux et possessifs	A[ikos]-[mf][sp]

1.1.4 Adverbes (R)

Niv 1	Adverbe (R)	**Valeur**	**Code**	**Exemple**	**Glose**
Niv 2	Type	general particle[1] interrogative/exclamative intensive/quantitative[2]	g p t q	aisidament ne quant plan	facilement ne combien beaucoup
Niv 3	Degree	positive comparative[3] negative[4]	p c n	aisidament melhor pas	facilement meilleur pas
	Genre	masculine feminine	m f	plan plana	beaucoup beaucoup
	Nombre	singular plural	s p	plana planas	beaucoup beaucoup

[1] Le trait "particle" s'applique spécifiquement à la particule *ne*, *non (ne)* dans les parlers où cette dernière accompagne l'adverbe négatif *pas*.

[2] Le trait "intensive/quantitative" est ajouté par rapport au modèle GRACE. Il s'applique aux adverbes tels que *plan (bien), tant (tant)...* qui ont la particularité de pouvoir dans certains parlers s'accorder en genre et en nombre : *de pomas, n'i a planas (Il y a beaucoup de pommes)*.

[3] Les formes comparatives sont par exemple *melhor (meilleur), mens (moins)*. Elles ont la particularité de ne pas pouvoir se combiner avec *mai (plus): *mai melhor (*plus meilleur), *mai mens (*plus moins)*.

[4] Les formes négatives des adverbes sont par exemple *jamai (jamais), sonque (seulement)*. Elles sont souvent combinées avec l'adverbe négatif *pas (pas).*

Les patrons de formation des étiquettes pour les adverbes peuvent être schématisés de la façon suivante :

Adverbes généraux et interro-exclamatifs	R[g][pnc]--
Adverbes quantifieurs	R[qx]-[mf][sp]
Particules négatives	Rpn--

1.1.5 Pronoms (P)

Niv 1	Pronom (P)	**Valeur**	**Code**	**Exemple**	**Glose**
Niv 2	Type	personal demonstrative indefinite[2] possessive interrogative relative reflexive cardinal[3]	p d i s t r x k	ieu aquò mantuns meu quin qui se dos	moi ça plusieurs mien quel qui se deux
Niv 3	Person	first second third	1 2 3	ieu tu el	moi toi lui
	Genre	masculine feminine	m f	quin quina	quel quelle
	Nombre	singular plural	s p	quin quines	quel quels
	Case[1]	accusative dative oblique	a d o	lo li el	le lui lui
	Possessor	singular plural	s p	meu nòstre	mien nôtre

[1] Par rapport au tagset GRACE, l'attribut "nominative" du Trait "Case" a été supprimé étant donné que l'occitan n'a pas de pronom personnel sujet.

[2] Les pronoms indéfinis couvrent les cas d'identificateurs :

a) **D'unes** cresián vertadièrament a una galejada (Viaule) (**Certains** croyaient vraiment à une blague)

et les quantificateurs non cardinaux :

b) **Mantuns** avián fachas tres o quatre sasons (Delèris) (**Plusieurs** avaient fait trois ou quatre saisons)

[3] La valeur "cardinal" de l'attribut "Type" permet de représenter les pronoms cardinaux :

c) - Cinc ostals dins lo Causse (...). **Dos** son barrats. (Gairal) (5 maisons dans le Causse (...). **Deux** sont fermées).

Les patrons de formation des étiquettes pour les pronoms peuvent être schématisés de la façon suivante :

Pronoms possessifs	Ps[123][mf][sp]-[sp]
Pronoms personnels	Pp[123][mf][sp][nado]-
Pronoms démonstratifs, indéfinis, relatifs, interrogatifs, cardinaux	P[dirtk]-[mf][sp]--
Pronoms réflexifs	Px[123][mf][sp]--

1.1.6 Déterminants (D)

Niv	Déterminants (D)	Valeur	Code	Exemple	Glose
Niv 2	Type	article demonstrative possessive indefinite[1] inter./exclam. relative[2] cardinal[3] partitive[4]	a d s i t r k p	lo/un aquel mon mantuns quin lo qual dos de	le/un ce mon plusieurs quel lequel deux du
Niv 3	Person	first second third	1 2 3	mon ton son	mon ton son
	Genre	masculine feminine	m f	mon ma	mon ma
	Nombre	singular plural	s p	mon mos	mon mes
	Possessor	singular plural	s p	mon nòstre	mon nôtre
	Nature	definite indefinite[5]	d i	lo un	le un

[1] La valeur "indefinite" du trait "Type" correspond aux emplois des indéfinis comme déterminants :

a) **mantuns** còps (**plusieurs** fois).

[2] La valeur "relative" du trait "Type" correspond aux emplois partitifs des déterminants :

b) qu'aimava asagada d'un veiròt de riquiquí (**lo qual** veiròt demorava al fons del bòc) (Escafit) (qu'il aimait arrosée d'un petit verre de ratafia (**lequel** petit verre restait au fond de la bouche))

[3] La valeur "cardinal" de l'attribut "Type" permet de représenter les déterminants cardinaux :

c) **Dos** amics (**deux** amis).

[4] Par rapport au tagset GRACE, l'attribut "partitive" a été ajouté au trait "Type". Il existe un déterminant partitif simple : *de*. Pour les autres formes possibles (dans les dialectes du nord) : *del*, *de la,* deux étiquettes seront sollicitées : Dp+Da.

d) que vòl far amb **de** sucre. (Landièr) (que veut-il faire avec **du** sucre) : Dp-ms—

e) A la pouncha d'un pueg, un faus, e de la moussa (Roux) (au sommet d'un mont, un hêtre et de la mousse) : Dp-ms-- + Da-fs-d

[5] "Un" peut être codé Da-ms-i (article indéfini) ou Dk-ms— (cardinal). EAGLES propose arbitrairement de toujours choisir le type "Article". Nous choisissons d'intégrer dans le lexique les deux codes. La désambiguïsation pourra alors être faite lors de l'annotation des corpus en fonction du contexte. En cas de doute, une préférence pour le type "Article" sera de mise.

Les patrons de formation des étiquettes pour les déterminants peuvent être schématisés de la façon suivante :

Déterminants possessifs	Ds[123][mf][sp][sp] -
Déterminants démonstratifs, indéfinis, interro-exclamatifs, relatifs, cardinaux et partitifs	D[ditrkp]-[mf][sp]--
Articles définis et indefinis	Da-[mf][sp]-[di]

1.1.7 Prépositions (S)

Niv 1	Préposition (P)	**Valeur**	**Code**	**Exemple**	**Glose**
Niv 2	Type[1]	preposition déictique	p d	de vaquí	de voilà

[1] Les amalgames d'une préposition et d'un déterminant sont codés avec les deux étiquettes correspondant à la forme non amalgamée. Par exemple *del (du)*, étant équivalent à *de lo (de le)* sera codé *Sp+Da-ms--d* dans Loflòc mais SpDa pour l'annotation morphosyntaxique dans le projet RESTAURE.

Le patron de formation des étiquettes pour toutes les prépositions peut se schématiser de la façon suivante :

Prépositions et déictiques	S[pd]

1.1.8 Conjonctions (C)

Niv 1	Conjonction (C)	**Valeur**	**Exemple**	**Code**
Niv 2	Type	coordinating subordinating	e que	c s

Le patron de formation des étiquettes pour toutes les conjonctions peut se schématiser de la façon suivante :

Conjonctions de coordination et de subordination	C[cs]

1.1.9 Interjections (I)

Exemple	**Glose**
hou oh	oh oh

1.1.10 Résidus (X)

Exemple	**Glose**
-n-	(voyelle épenthétique)

1.1.11 Ponctuation (F)

Exemple
. ; - …