

Guide d'annotation : Questionbank

Création : Décembre 2012/février 2013

Ce guide étend Le guide d'annotation du FTB et du French Social Media Bank .

1. Outils pour annoter ou pour rechercher des motifs

1. Outil pour rechercher des motifs

Tregex : outils de visualisation des graphes. Il permet de rechercher des motifs.
Exemple de requêtes simples :

Ouvrir : File > Load Trees > sélectionner un fichier.mrg

-sur une catégorie morpho-syntaxique : PROWH (cherche tous les arbres qui contiennent un pronom interrogatif)

-sur une suite de catégories morpho-syntaxique : CLS.V (recherche les arbres dont un clitique sujet et suivi d'un verbe)

- sur un lemme : ranger (il renvoi tous les arbres qui contiennent ce lemme)

- sur une suite de deux lemmes : être.rangé (il renvoi tous les arbres qui contiennent ces lemmes dans le même ordre)

Les autres fonctionnalités sont visibles directement dans « Help »

Il est possible d'enregistrer les résultats d'une requête sous forme de phrases ou d'arbres.

NOTE DE DJAME : Virginie et les filles utilisaient une version du treebank avec les lemmes pour pouvoir faire des requêtes plus facilement.

1.2 Outil pour annoter

Wordfreak : outils permettant plusieurs niveaux d'annotation

Phase 1 : annotation des constituants de manière isolé (passage rapide)

Phase 2 : annotation des constituantes au niveau morpho-syntaxique

Phase 3 : annotation des fonctions

Une version de wordfreak différente sera utilisée à chacune de ces étapes.

Enregistrer le nouveau Wordreak.jar dans *arch/i386/bin*

Modifier le nom du fichier dans le script d'exécution Wordfreak dans *arch/i386/bin*

```
#!/bin/sh
```

```
java -jar $PREFIX/bin/wordfreak-+Crap.jar -u virginie -e ISO8859-1 -t FTB-CC_treebank $1
```

1.3 Outil de sauvegarde des données

SVN

Après installation de SVN, les fichiers de travail se trouvent dans
~/arch/SVN/sequoiabank/QuestionBank/

2 Arborescence du répertoire QuestionBank/

ETAPE 1 : Récupérer et enregistrer les fichiers contenant des tags

Chaque fichier est annoté par deux annotatrices différentes. La troisième compare les annotations des deux précédentes et adjudique.

TAGS_A_VERIFIER : l'ensemble des fichiers à annoter au niveau des tags se trouve dans ce dossier

Chacun recherche les noms de fichiers attribués à chacune dans le document excel partagé via le googledoc.

TAGS_TERMINES : Virginie/ Vanessa/ Caroline/

Chaque fichier provenant de **TAGS_A_VERIFIER** annoté est commité sous SVN dans le répertoire de chacune

TAGS_ADJUDICATED : Virginie/ Vanessa/ Caroline/

Les fichiers provenant de **TAGS_TERMINES** sont ensuite soumis à comparaison et adjudication, puis enregistrés dans ce dossier.

Les fichiers annotés par virginie et vanessa sont comparés et adjudiqués à l'aide de Filemerge par caroline.

Les fichiers annotés par vanessa et caroline sont comparés et adjudiqués par virginie.

Le fichier utilisé pour la correction peut être celui de l'une ou de l'autre, avec un nom différent.

ETAPE 2 : Récupérer et enregistrer les fichiers parsés

PARSED_A_VERIFIER : l'ensemble des fichiers parsés se trouvent dans ce dossier.

Les fichiers sont à répartir entre les 3 annotatrices.

PARSED_TERMINES : Virginie/ Vanessa/ Caroline/

Chaque fichier provenant de **PARSED_A_VERIFIER** annoté est commité sous SVN dans le répertoire de chacune dans PARSED_TERMINES/virginie ou /Caroline ou /Vanessa

PARSED_ADJUDICATED : Virginie/ Vanessa/ Caroline/

Les fichiers provenant de **PARSED_TERMINES** sont ensuite soumis à comparaison et adjudication, puis commité dans ce dossier.

ETAPE 3 : Récupérer et enregistrer les fichiers les annotations au niveau des fonctions

FONCTIONS_A_VERIFIER : l'ensemble des fichiers dont il faut vérifier les fonctions sont dans ce dossier

FONCTIONS_TERMINEES : Virginie/ Vanessa/ Caroline/

Chaque fichier provenant de **FONCTIONS_A_VERIFIER** annoté est commité sous SVN dans le répertoire de chacune

FONCTIONS_ADJUDICATED : Virginie/ Vanessa/ Caroline/

Les fichiers provenant de **FONCTIONS_TERMINEES** sont ensuite soumis à comparaison et adjudication, puis enregistrés dans ce dossier.

2 Convention de nommage des fichiers

2.1 Au niveau des tags

questionbank.numero_phrase.nom_annotateur1.nom_annotateur2.mrg

Exemple : questionbank.aa.van.virg.mrg

après adjudication

questionbank.numero_phrase.nom_annotateur1.nom_annotateur2.nom_ADJUDICATED.mrg

Exemple : questionbank.aa.van.virg_ADJUDICATED.mrg

2.2 Niveau des fichiers parsés

questionbank.parsed.numero_phrase.nom_annotateur1.nom_annotateur2.mrg

après adjudication

questionbank.parsed.numero_phrase.nom_annotateur1.nom_annotateur2_ADJUDICATED.mrg

2.3 Niveau des fichiers de fonctions

Questionbank.fct.numero_phrase.nom_annotateur1.nom_annotateur2.mrg

après adjudication

questionbank.fct.numero_phrase.nom_annotateur1.nom_annotateur2_ADJUDICATED.mrg

2 Annotation morpho-syntaxique

MODIFICATION majeure 1 par rapport au guide : Sint sous Ssub

MODIFICATION majeure 2 par rapport au guide : PP-OBJ (P VPinf) dans les tournures impersonnelles du genre « il est difficile d'intéresser les industriels »

Autres MODIFICATION majeures, adaptation au type de corpus :

X : annotation des constituants amalgamés X/jte

Y : annotation des constituants décomposés : Y/ai Y/me

YM pour les cas YM/c YM/a YM/dire

Contractés : CLS+V (jvais)

FRAG : étiquette supplémentaire sur les constituants extra-phrastiques

REPEAT : étiquette supplémentaire sur les constituants répétés (mange REPET/mange)

NO_REALIZED : constituants non réalisés, on fait un constituant vide

CLIV : étiquette supplémentaire permettant d'annoter les clivés, annotation au niveau SENT

META : URL, Hashtags,

Phrases mal découpées :

A redécouper : On met l'étiquette A_DECOUP

A recoller : On met l'étiquette -SUJ sur la SENT N+1

SENT « J'ai mangé »

SENT-SUJ « des pâtes »

Annotation des fonctions sur les contractés :

CS+**SUJ** -> on descend le CS dans la Sint, dans le VN si c'est un PRO « **qil** ». On met la fonction sur le VN

VN-**SUJ**-**ATS** « **cvré** » -> on note la ou les fonctions sur le VN

VN-**SUJ** « **jsais** » -> on note la ou les fonctions sur le VN

CLR « la meuf qui veut pas **sareter** » -> création d'un constituant vide CLR sous VN, et annoter le verbe X

2.3 Constituants, étiquettes morpho-syntaxiques et fonctionnelles

Constituants

NP, PP, VN, AP, AdP, VPinf, Sint, VPpart, COORD, Ssub, Srel

Étiquettes morpho-syntaxiques

ADJ, ADV, NC, NPP, I, P, PRO, PROREL, CS, CC, PREF, ET

V, VIMP, VINP, VPP, VS

P+D, P+PRO,

ADJWH, ADVWH, DET, DETWH, PROWH

CLO, CLR, CLS

X, Y, KK, META

Étiquettes fonctionnelles

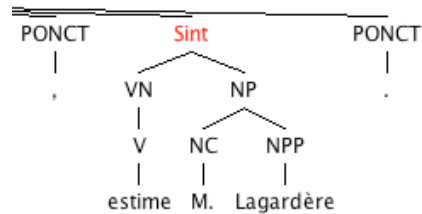
ATO, ATS, A_OBJ, DE_OBJ, MOD, OBJ, SUJ,

P_OBJ ou (P_OBJ_LOC, P_OBJ_DLOC, P_OBJ_AGT)

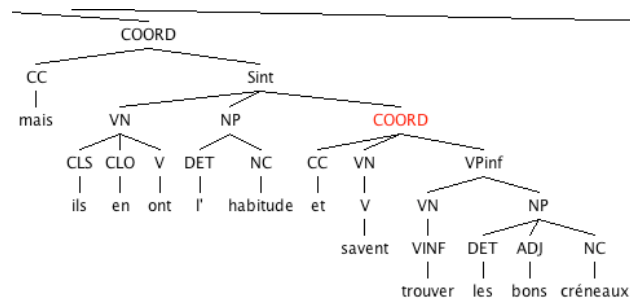
2.3 Points essentiels à connaître par cœur

2.1 Sint sous Ssub

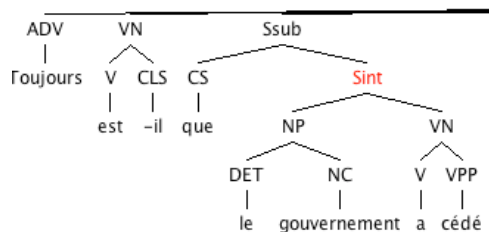
Les Sint (Subordonnées intérieures) permettent d'annoter des phrases internes, des incises.



Sint est incluse sous COORD lorsque celle-ci contient un **sujet** et un **VN**

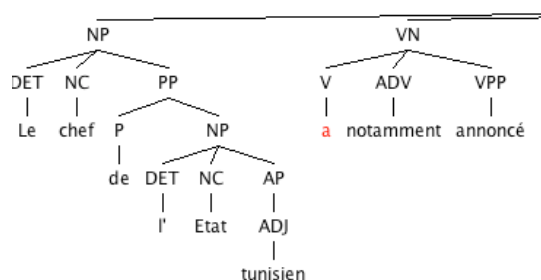


Après une CS (conjonction de subordination) si la subordonnée qui suit contient au moins un sujet et un VN



2.2 Le noyau verbal (VN)

Il comprend le verbe, les clitiques, et éventuellement un adverbe de négation (si il est placé à gauche) ou un adverbe placé entre l'auxiliaire et le verbe.

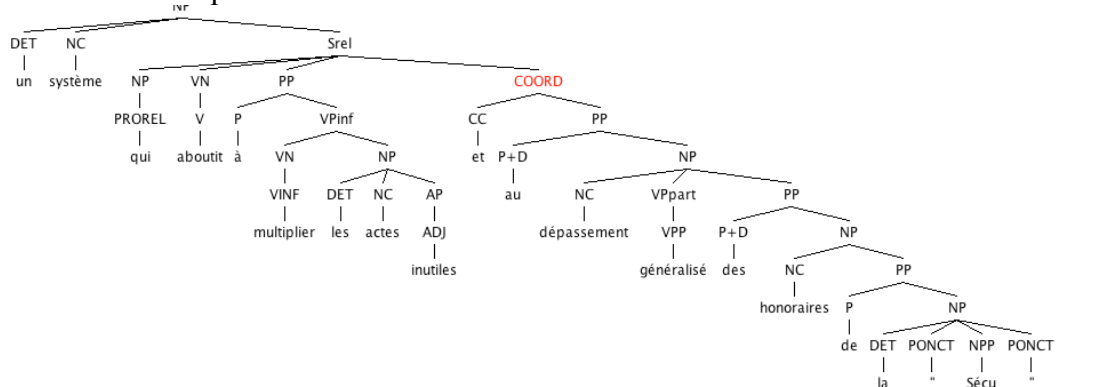


2.3 Les conjonctions de coordination (COORD)

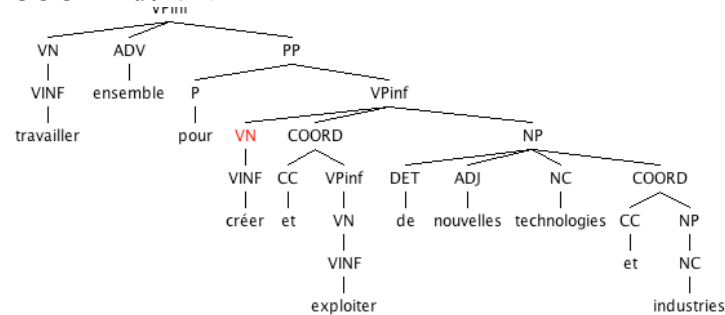
Coordination de VN, coordination de classes différentes

Les CC : *et, ou, mais, donc, or, ni, car, soit, c'est-à-dire, voire, sinon, tantôt, puis*

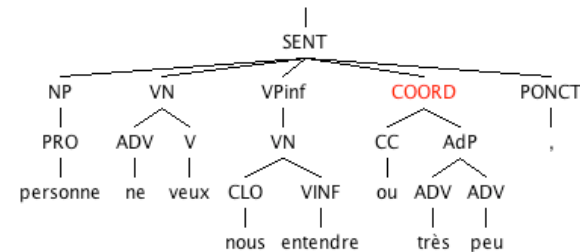
COORD classique



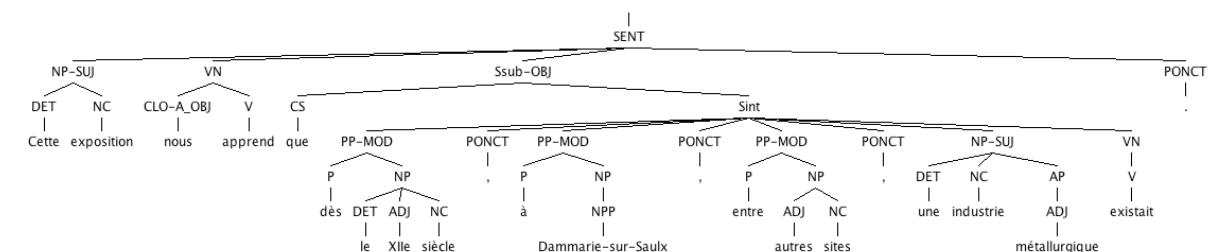
COORD de VN



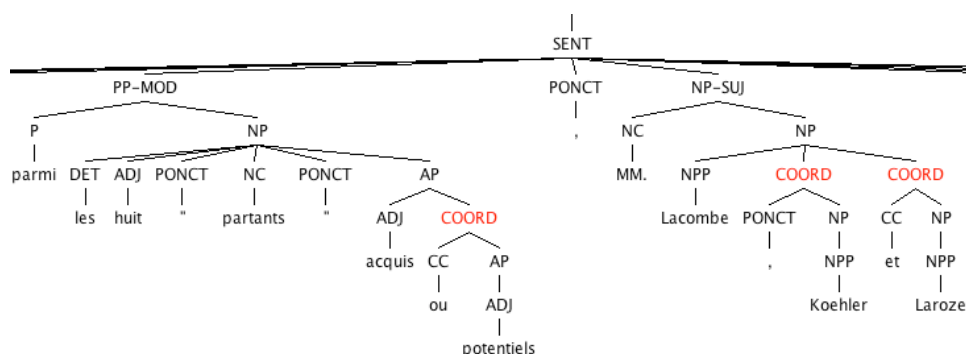
COORD de constituants différents



Les appositions donnent lieu à des structures plates. Les virgules ne sont pas incluses dans les constituants.



Si il y a une conjonction de coordination dans une suite d'appositions, elle est mixte. On traite alors l'ensemble comme une suite de coordination.



2.4 Annotation des Noms propres (NPP)

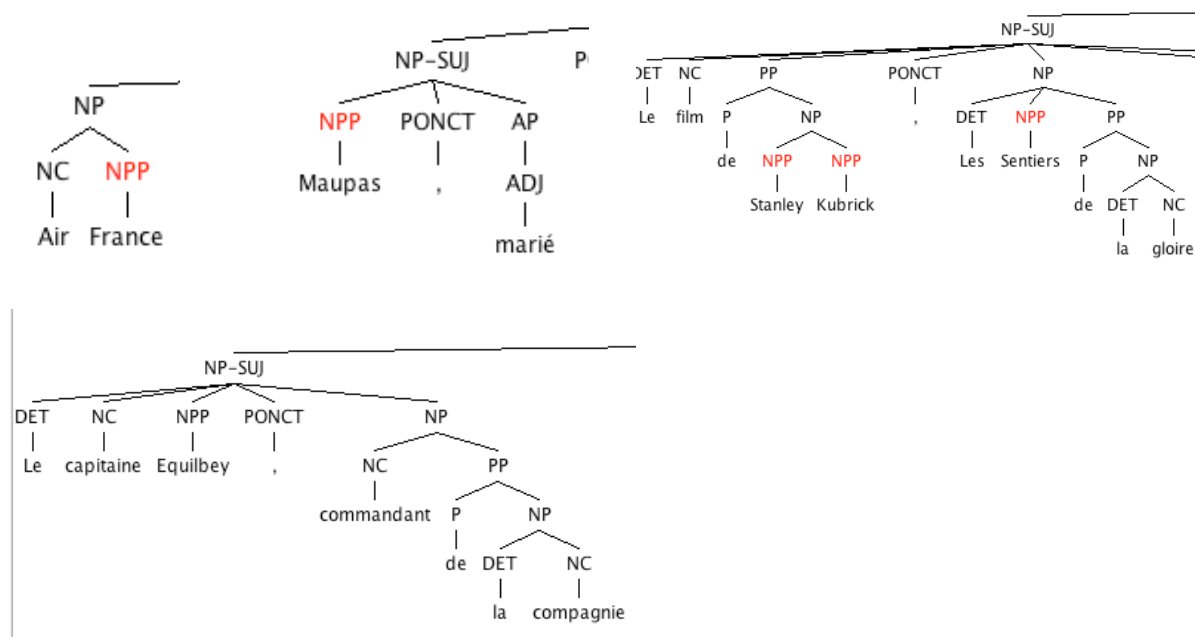
Les noms propres sont annotés NPP. Lorsque les NPP sont composés, seule la tête est NPP.

Les adjectifs épithètes du nom ne projettent pas de NP. Les appositions projettent un autre NP (une virgule ou ":" sont des indices)

<NP> mon ami:NC Pierre:NP </NP>

<NP> le général:NC Eyadéma:NPP </NP>

<NP> Pierre , <NP> mon ami :NC </NP> </NP> pour la plus grande gloire de <NP> ce monument : <NP> le foie gras </NP> </NP>



1.2 Annotation des NPP complexes

Dominique de Talancé (NP NPP Dominique (PP * P de * (NP NPP Talancé)))

Renaud van Ruymbeke (NP NPP Renaud * NPP Van * NPP Ruymbeke)

Guillaume 2 (NP NPP Renaud (A 2))

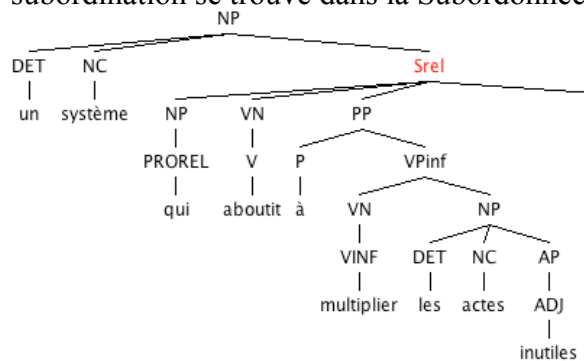
2.2 Annotation des NPP contenant des mots étrangers

très "raw ET food ET" ==> A N
 (exit VPP) les resto
 kitchen NC ou Juice NC bar NC
 de chez rice NC and CC fish NC
 (façon N) (american A) (x KK) (story N)

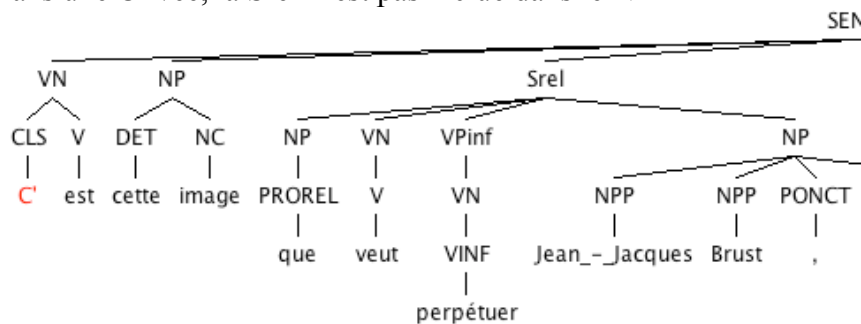
(Bob NPP) ('s ET) (juice ET) (bar ET) ==> NPP ET N N
 l'avant comptoir ==> D P NPP
 l'atelier ==> D NPP
 chez l'ami Jean ==> P D N NPP

2.5 Annotation des Subordonnées relatives (Srel)

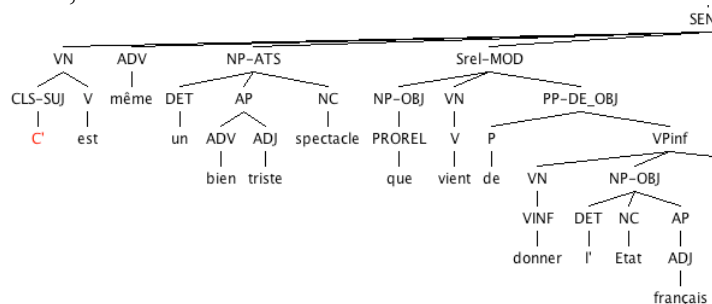
Les subordonnées relatives sont incluses dans le NP qu'elles reprennent et la conjonction de subordination se trouve dans la Subordonnée relative.



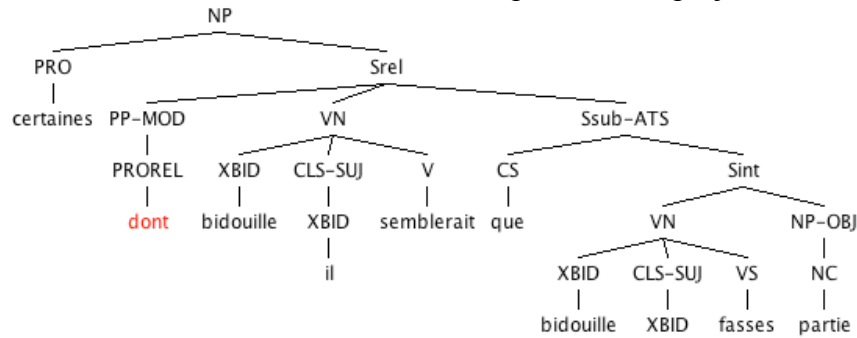
Dans une Clivée, la Srel n'est pas incluse dans le NP



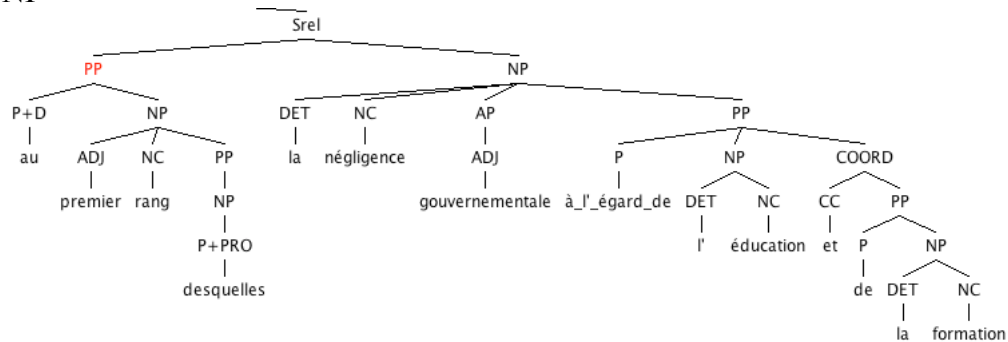
Dans les clivées, le syntagme entre « c'est » et « que »/ « qui » est ATS, la Srel a la fonction MOD, et est rattachée au VN



Les subordonnées relatives introduites par « dont » projettent un PP

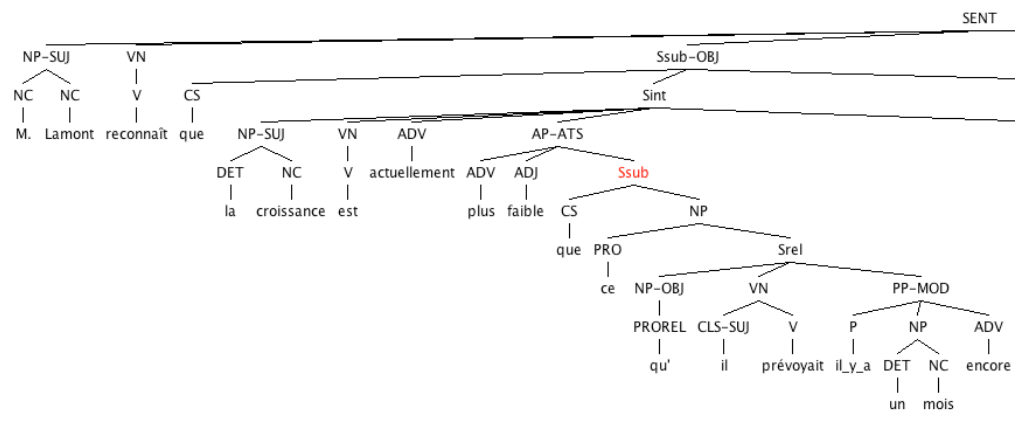


Dans les subordonnées relatives qui commencent par une préposition, on a un PP qui inclut un NP

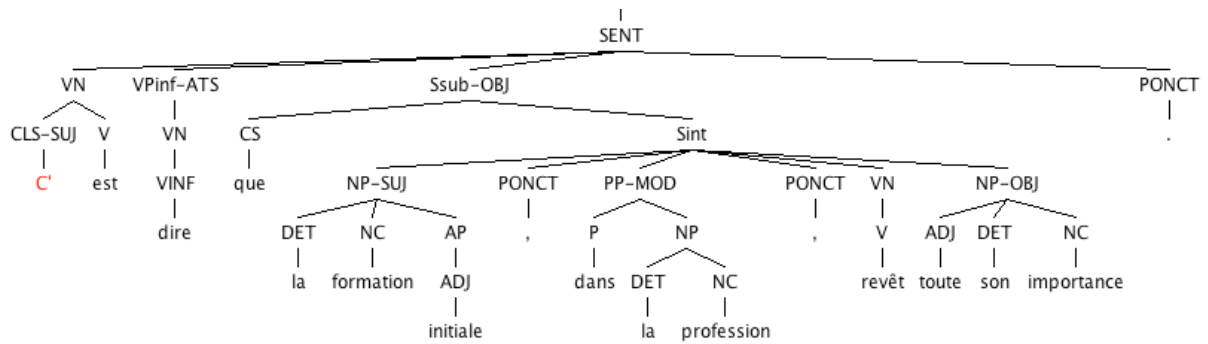


2.6 Annotation des syntagmes subordonnés (Ssub)

Le pronom introducteur est une conjonction de subordination (CS). Le syntagme subordonné **est dépendant de** la structure qui le précède.



Dans une tournure impersonnelle, la subordonnée **n'est pas incluse** dans le syntagme qui la précède. On la rattache à la racine.

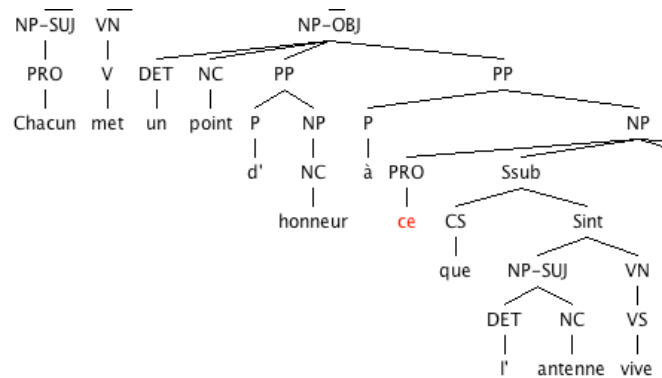


Ssub dans AdP ou AP (aussi souvent que, plus grand que)

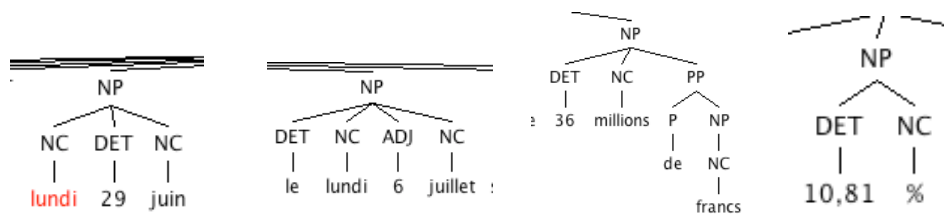
Une seule Ssub quand deux CS se suivent (comme quand, comme si)

Si une phrase commence par un CS (Si tu viens, je serai content), on fait une Ssub et si Sujet/ VN, alors on fait une Sint

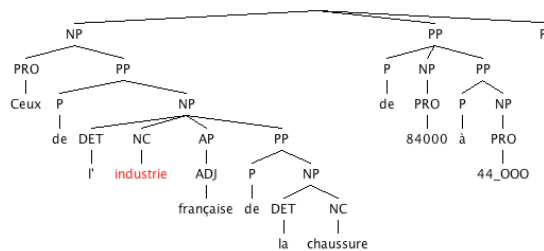
Dans une Ssub commençant par « à ce que » et « de ce que »



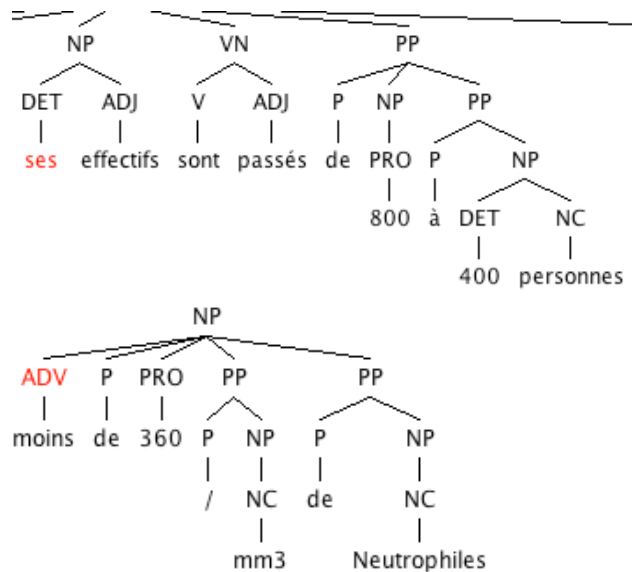
2.7 Annotation des dates et chiffres avec mise en facteur ou non



De X à Y de PRO à PRO



De X à Y (N) **de PRO à DET N**



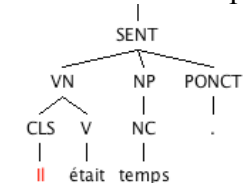
Le problème du cas rencontré dans le SMT « *Je passe de 250 à 95 mille* » n'a pas encore été résolu.

Autres cas :

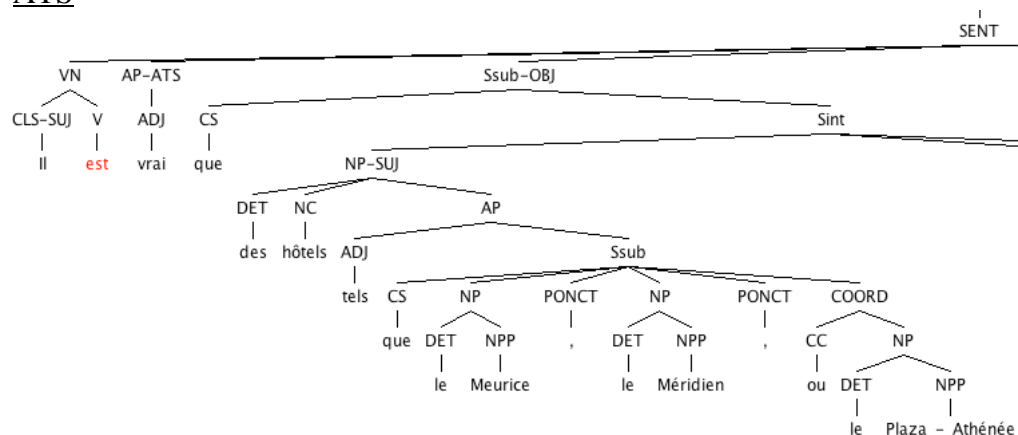
(VN (V passe)) (PP (P de) (NP (PRO 17)) (PP (P à) (NP (DET 9) (NC points))))
 (PP (P de) (NP (DET la) (ADJ seconde)) (PP (P à) (NP (DET la) (ADJ troisième) (NC place))))
 (VN (V passe)) (PP (P de) (NP (PRO 3)) (PP (P à) (NP (PRO 5))))

2.8 Formes impersonnelles

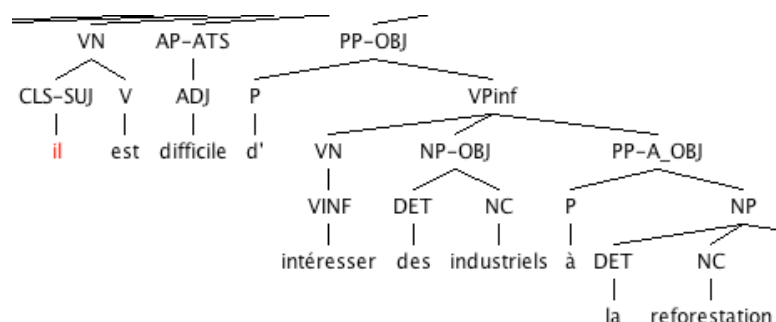
Les tournures impersonnelles ont pour sujet « il, ce ou ça ». Le sujet est dans le VN



Dans les tournures impersonnelles, le syntagme qui suit l'adjectif est Objet mais l'Adjectif est ATS

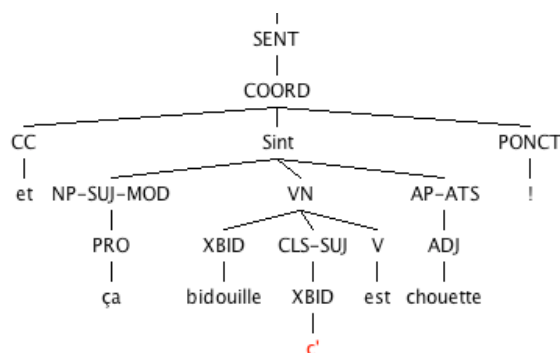
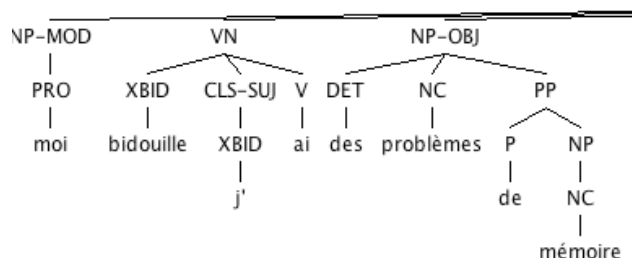
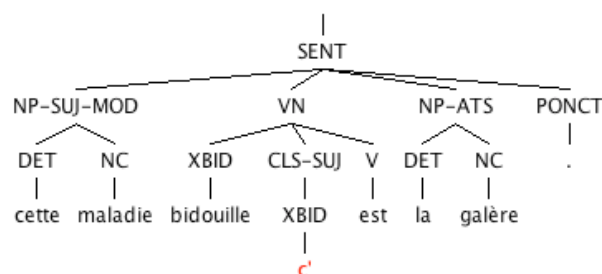


Dans une structure impersonnelle, l'infinitive extraposée introduite par une préposition doit être annotée sous la forme suivante :



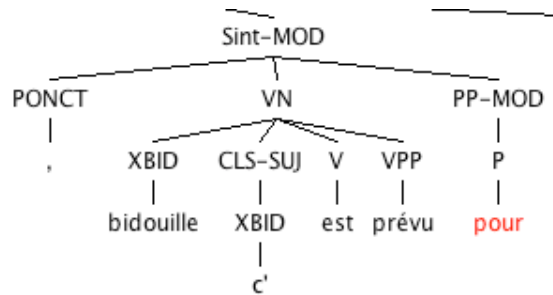
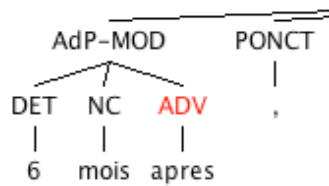
2.9 Annotation des phrases à deux sujets

Pour le moment, l'annotation est fluctuante...



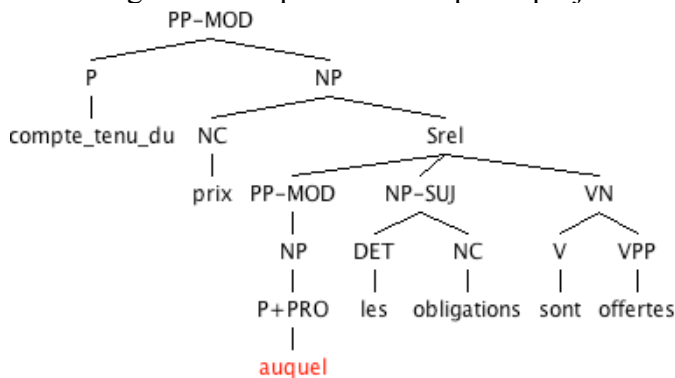
2.10 Annotation des prépositions isolées en début ou fin de phrase

Les prépositions employées sans dependants sont généralement recatégorisées en Adverbe et ne projettent pas de syntagme. (avant, après, ensuite, etc)



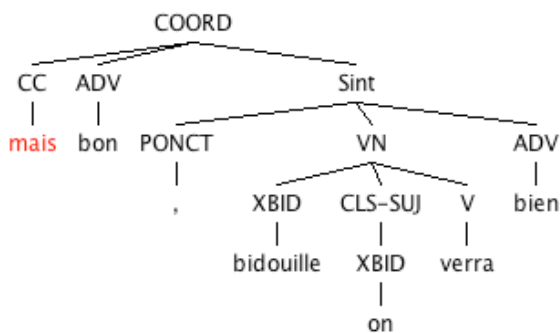
Si on rencontre une préposition seule comme attribut ou complément (je suis pour, je vote contre) alors il faut faire un PP unaire.

Les amalgames « auquel » et « duquel » projettent un NP et sont inclus dans un PP.

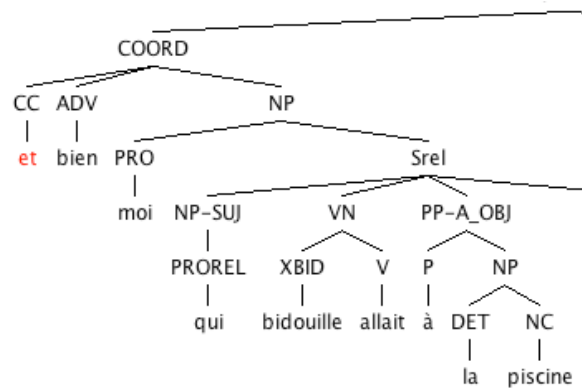


2.11 Annotation des phénomènes extra-prédicatifs dans les corpus bruités : interjection ou adverbe

« mais bon »



« et bien »

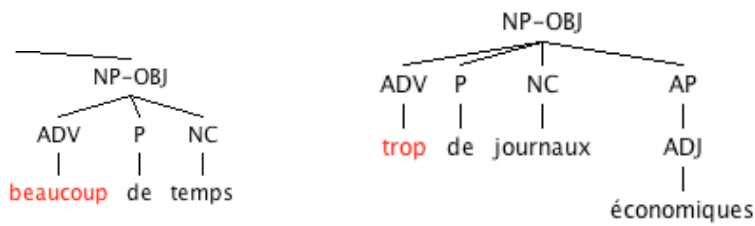


« blablabla », « basta », « bisous », « humm » sont annotés Interjection

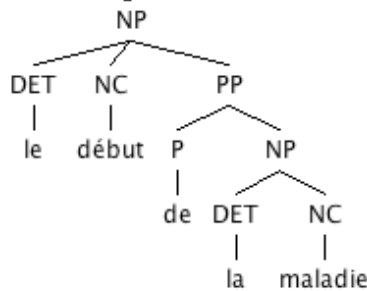
« merci » est annoté I si il est isolé (fin de post) mais il est NC dans « un grand merci à tous »

« bon, je vais » ???

2.12 Adverbe suivi de d'une préposition : pas de PP après ma Prep

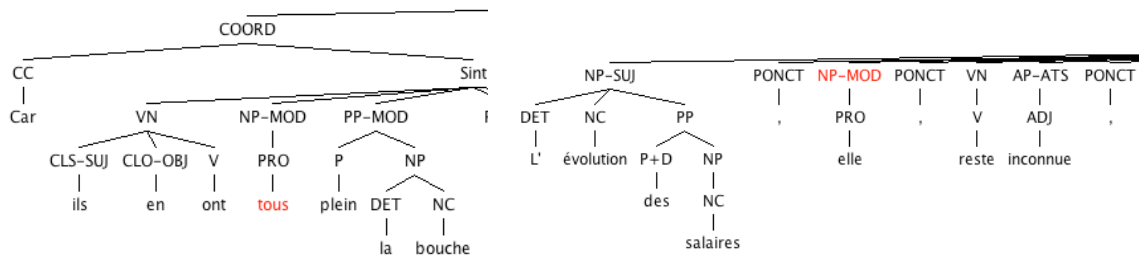


SAUF si la préposition « de » est suivie d'un DET. Dans ce cas, on fait un PP avec un NP interne et l'ADV initial est étiqueté PRO



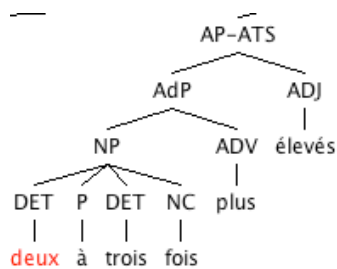
2.13 Les pronoms « flottants »

Les pronoms « flottants » sont généralement MOD

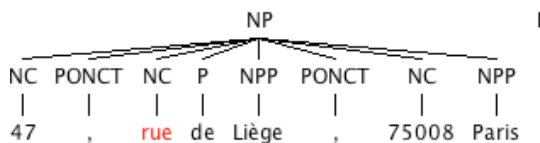


2.14 Les cas où on laisse les structures plates

Deux à trois N

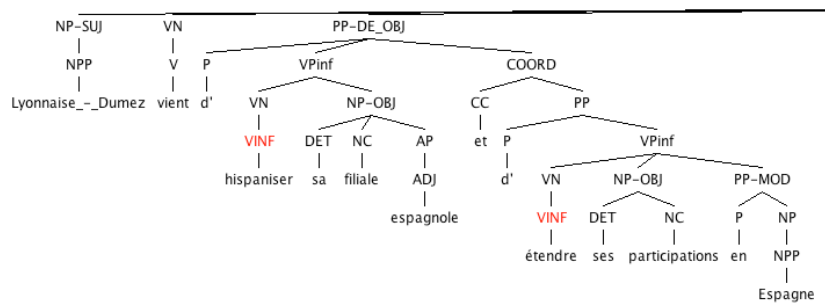
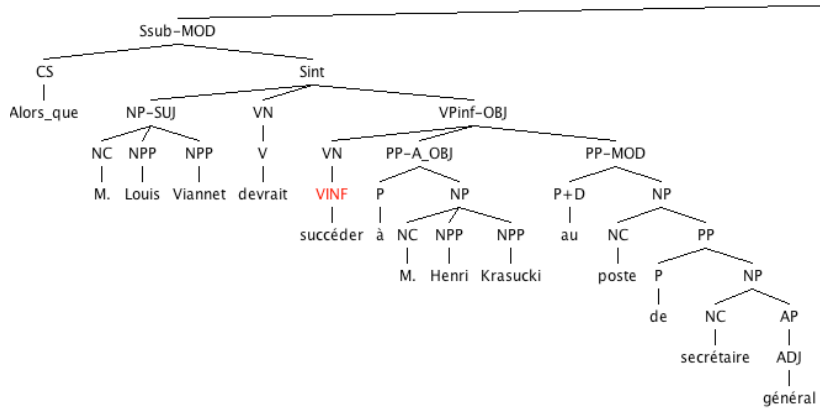


Titres de films, titres de livres, nom de rues



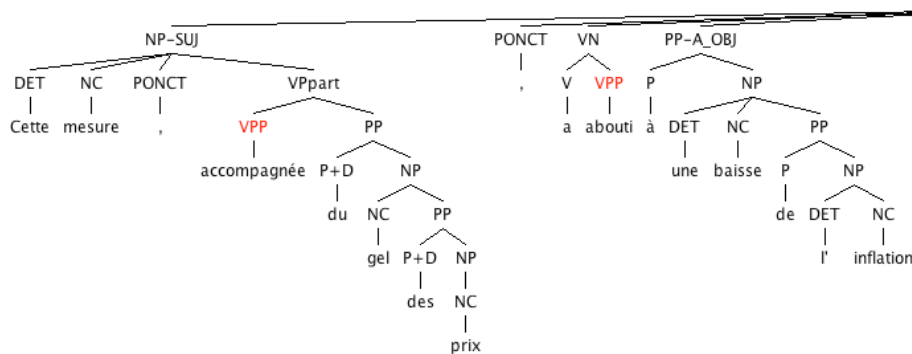
2.15 Les structures infinitives (VPinf)

Elles comportent un VN à l'infinitif et peuvent être ou non introduites par une préposition, un pronom quantifieur, un adverbe, un pronom interrogatif. Elles peuvent être indépendantes.

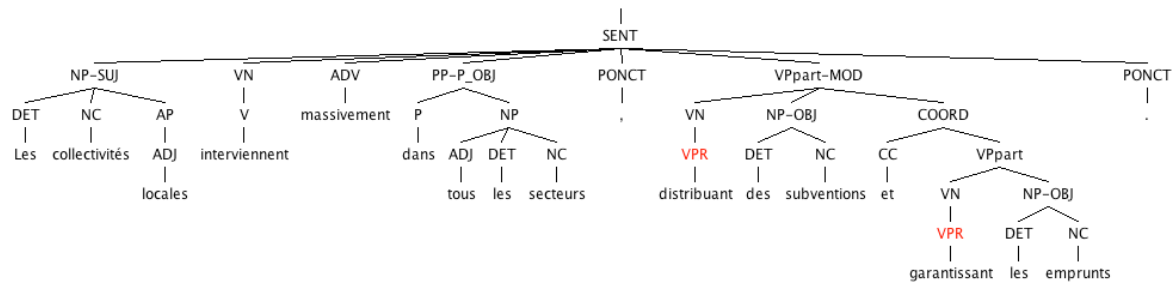


2.16 VPpart : annotation des VPP et VPR

VPP



VPR



Fonctions

ATO, ATS, A_OBJ, DE_OBJ, MOD, OBJ, SUJ,
P_OBJ ou (P_OBJ_LOC, P_OBJ_DLOC, P_OBJ_AGT)

Spécialisation de P_OBJ :

P_OBJ_LOC : commutation possible (au niveau du sens) avec "ici" ou "là"

P_OBJ_DLOC : commutation possible (au niveau du sens) avec "d'ici" ou "de là"

P_OBJ_AGT : les compléments d'agent du passif

P_OBJ_O : les autres cas

Disloquée

(NP-SUJ ça) (VN fait) (ADV longtemps) (Ssub-SUJ-MOD que Paul m'embête)

(NP-SUJ ça) (VN fait) (NP-OBJ 10 ans) (Ssub-SUJ-MOD que Paul m'embête)

(VN-SUJ il y a) (NP-MOD 3 ans) (Ssub-OBJ que Paul m'embête)

(VN voilà) (NP-OBJ des semaines) (Ssub-MOD que blabla)

disloquée repris par un clitique

Les syntagmes disloqués (en début ou en fin de phrase, repris par un clitique) n'ont pas de fonction: **(A DECIDER)**

<NP> Paul </NP>, <VN fct:SUJ/OBJ> je ne le connais </VN> pas

<VN fct:SUJ/A-OBJ> je ne lui parle </VN> plus, <PP> à Paul </PP>

<NP fct:SUJ> La récolte </NP>, <NP fct:MOD> elle </NP>, était fichue

Ce n'est pas le même cas : il y a une pause entre "la récolte" et "elle" et "était fichue"

Le « elle » n'est pas le clitique

OBJ

Le contrat interdit cette opération (cela / *de cela) => OBJ

Cette procédure risque d'avoir des répercussions (? cela / *de cela) => OBJ

il apparaît que Z (impersonnel) => (que Z) est OBJ

risquer / décider de VPinf => OBJ et ~~pas DE_OBJ~~

viser à Vinf => OBJ et ~~pas A_OBJ~~

A_OBJ

Ils vendent (NP-OBJ leurs produits) (PP-A_OBJ à des prix égaux)

Nous nous sommes engagés (...) (PP-A_OBJ à protéger le climat)

DE_OBJ

Marie rêve de voler dans le ciel (marie rêve (*cela/de cela)) => DE_OBJ
convaincre X que Z => le (que Z) est DE_OBJ
« Il a discuté des affaires nationales ... » => non polémique => DE_OBJ

*Certains verbes admettent les 2 sous catégorisations (OBJ et DE_OBJ). Dans ce cas, on met **DE_OBJ** (C'est le cas du verbe rêver : rêver cela / de cela et du verbe discuter : discuter cela / de cela / mais pour « discuter » il y a une distinction de sens)*

MOD

« il a déclaré (MOD au juge) (OBJ avoir blabla) »
« son directeur est venu (P_OBJ_LOC à l'Elysée) (MOD pour un diner) »
Une somme qui n'apparaît pas dans sa déclaration => MOD
Dans son ordonnance JChirac est évoqué => MOD

Je remercie (NP-OBJ le président en exercice) (PP-MOD pour sa réponse).
(NP-MOD Monsieur le Président) je tiens (PP A_OBJ remercier le rapporteur)

P_OBJ_O

Le PS arrive en deuxième position => P_OBJ_O (*en deuxième position, le PS arrive)
(...) la menace très grave qui pèse (P_OBJ_O sur le modèle européen)
Ils n'ont été impliqués (P_OBJ_O dans aucune action terroriste)

P_OBJ_LOC

une cinquantaine de personnes ont été renvoyée devant le tribunal => P_OBJ_LOC
envoie ses obus sur les tranchées => P_OBJ_LOC

ATS

Ils se trouvent directement pris => ATS
Sa mort a été classée sans suite => ATS

ATO

Je trouve (NP-OBJ cela) (AP-ATO dramatique)
Le verdict reconnaîtra (AP-ATO coupables) (NP-OBJ plusieurs dirigeants)

P_OBJ_DLOC

(...) les gendarmes sortaient (P_OBJ_DLOC de leur brigade)
Anita (...) est arrivée (...) très jeune, (P_OBJ_DLOC de sa chère Italie)
(...) je reviens (P_OBJ_DLOC du Koweït)

P_OBJ_AGT

L'expérience (...) a été affaiblie (PP-P_OBJ_AGT par la distance qui les sépare)
Ces systèmes (...) sont financés (PP-P_OBJ_AGT par les fonds du gouvernement)
Les frais supplémentaires (...) seront supportés (PP-P_OBJ_AGT par l'utilisateur)

En (DE_OBJ ou MOD)

(Ssub-MOD (CS comme) (Sint (VN (**CLO-DE_OBJ en**) (V témoignent)) (NP-SUJ (DET les) (NC appellations) (PP (P de) (NP blabla))))))

(VN (ADV n') (**CLO en**) (V porte)) (AdP-MOD (ADV pas) (ADV moins)) (NP-OBJ (DET un) (NC nom) (PP (P à) (NP (NC particule))))

(PP-MOD (P à) (VPinf (VN (**CLO en**) (VINF croire)) (NP-OBJ (NPP Alain) (NPP Fribourg))))

(NP (DET ces) (NC morts) (Srel (NP-SUJ (PROREL qui)) (VN (**CLO-OBJ en**) (V savaient)) (ADV trop)))

(NP-SUJ (DET Le) (NC juge) (PONCT ,) (VPpart (VN (VPR souhaitant)) (VPinf-OBJ (VN (**CLO en**) (VINF avoir)) (NP-OBJ (DET le) (NC coeur) (AP (ADJ net))))))

(NP-SUJ (DET le) (NPP Viet) (NPP Minh)) (VN (**CLO-DE_OBJ en**) (V profitait)) (ADV également)

"Un lion peut (VPinf-OBJ (**CLO-OBJ en** contrôler un autre)) »

dont (MOD ou De OBJ)

(NP (DET les) (NC effractions) (COORD (CC et) (NP (DET les) (NC pressions))))(Srel (**PP-MOD** (PROREL **dont**)) (VN (V auraient) (VPP eu)) (PP-OBJ (P à) (VPinf (VN (VINF subir)))) (NP-SUJ (DET les) (NC magistrats)))

Les producteurs (...) de mouton, (**PP-DE_OBJ** PROREL **dont** on attend aujourd'hui qu'ils vendent leurs produits)

2 Points linguistiques abordés précédemment

Amalgames

« **d'_où** » **amalgamé** correspond au cas introduisant un NP (~~et pas une relative~~). Il a le sens de « **de la vient NP** ». Il s'agit dans ce cas présent d'un **P**.

« **d'où** » **non amalgamé** correspond au cas de relative/interrogative

(Srel (PP (P d') (NP (**PROREL où**)) (VN (V provient))

(NP (NC port) (PP (P de) (NP (NPP Rotterdam) (PONCT -LRB-) (Srel (PP (P d') (NP (**PROREL où**)) (VN (CLS il) (V reçoit)) (NP (DET les) (NC tourteaux) (PP (P d') (NP (NPP Amérique)))))) (PONCT -RRB-))))

« **c'est_pourquoi** » **amalgamé** si CS, ADV
(SENT (Ssub (CS c'est_pourquoi)(Sint on ne délivrera pas....)))
(SENT (**ADV** c'est_pourquoi)(Sint j'estime que))

« **c'est pourquoi** » **non amalgamé** si ADVWH
(SENT (VN (c'est)) (ADVWH pourquoi) (Sint (Alonso se retrouve confiné)))

« **non seulement** » **amalgamé**
c'est important (PP (ADV non_seulement) (P sur) (NP le plan symbolique) (COORD (CC mais) (Ssub (ADV aussi) (CS parce que) (Sint (cela réaffirme le droit....)))))

« **qui que ce soit** » **amalgamé**
(COORD (CC et) (NP (PROREL qui_que_ce_soit) (PP (P d') (NP (PRO autre)))) (Srel qui vit au Kosovo)))

« **pas seulement** » **n'est pas amalgamé**
(VN Nous ne pouvons) (ADV pas) (VPinf (ADV seulement) (VN exiger) (Ssub que les Serbes se comportent correctement)) (Sint (VN nous devons) (VPinf (ADV également) (VN obliger) (NP)))

VPR, le cas de « concernant » (dans le sens de « en ce qui concerne »)
(VPpart (VN (VN (VPR concernant) ...

« pessimiste **concernant** la reprise » => VPR
« **concernant** la reprise, les ministres ont annoncé un rapport ... » => P qui projette un PP

Clitiques réflexifs

nous/CLS nous/CLR penchons/V sur le problème ...

"tout ce qui" et « tout ce que »

Tout:PRO ce:PRO qui: PROREL évoque la mer
(NP (ADV tout) (PRO ce) (Ssub (CS que)))

"l'un de/des"

SREL rattachée à « l'un » « C'est **l'un** de ces traités **qui** a attiré notre attention »
SREL rattachée à « députés » « C'est l'un des **députés que** nous avons convoqués »
SREL rattachée à « exemple » « l'un des quelques **exemples** SREL **où** les trois institutions... »

2 QUESTIONS NON RESOLUES

omission de catégories syntaxique
répétition de catégories syntaxique
omission du "ne" quasi systématique "je sais pas"
omission du "ne" et du "pas" "t'inquiète" "tinkiete" (le dernier est inventé mais je crois l'avoir vu)
constructions elliptiques du fait du support : forum, twitter, facebook. Ce qui n'est pas le cas de Yelp -> omission du sujet
"je suis lvl 56 " -> omission de la Prep (jv.com)

“Le Samson C01U, branchement USB sans installation, marche superbement bien et (est) livré avec le pied” -> omission du V
 “pour ma part, patch depuis 7 mois et j’en suis super satisfaite” -> ellipse du VN ?? (docti aj)
 “un_peu déçue” -> idem (docti aj)
 ? “cela mettai du temps” -> ? beaucoup_de ?
 (Facebook) afida toujours oci belle c est une star pour moi -> omission du VN. Nous aurions dû avoir deux phrases. et une analyse de “oci belle” AP-ATS

Constructions syntaxiques anarchiques

Ssub « il y a des jours que nous n’avons pas envie de nous battre (docti) »
 COORD « SALUT tu es très belle mais j'adore ta crinière changer de temps en temps c'est sympa mais il vaut mieux rester soi-même »
 Srel « Les 1D font partis des gens que si je les croisent dans la rue , j'en aurais rien a faire ...tout comme Lady Gaga , elle jla tape krément »

annoter ce qui aurait du être écrit ou ce qui est écrit ?

(Docti)

“il a juste manqué le réalisme”

il en a manqué DE_OBJ

il l’a manqué ? OBJ

(Comment traiter les :

@Bidulle (phrase))

(Doctiab)

“en plus des symptômes que je vous ai parler plus haut”

-> dont je vous ai parlé plus haut

DE_OBJ avec une Srel en que ? ou OBJ (annotatrice 1) et DE_OBJ (annotatrice 2, adju)

Liste de mots considérés comme composés pour Sequoia:

. en_tête_à_tête
 . il_y_a
 . hors-d’_oeuvre
 . en_fait
 . tout_le_monde
 . peut_être
 . d’_ailleurs
 . chou-fleur
 . new_yorkais
 . l’_on
 . à_la_va_vite
 . là_bas
 . près_de

. préfixe « ex » et « non » => décollés et taggés PREF

Choix pour le projet déjà tranché :

on concatène les prix qui ne le sont pas 10euros