

Dépendances syntaxiques de surface pour le français

Schéma d'annotation
pour un corpus en dépendances obtenu par conversion du FrenchTreebank

Marie Candito, Benoît Crabbé, Mathieu Falco

Version	date	Commentaire
V0	avril 2009	Version initiale
V1	oct 2009	Révisions mineures. Passage au format CONLL
V1.1	janvier 2010	Ajout documentation des catégories et traits
V1.2	mai 2011	ajout références guide d'annotation FTB en plus des références d'articles FTB
V1.3	dec 2013	Précisions dans l'introduction : ce document n'est pas un guide d'annotation mais une description de schéma

1	Introduction.....	3
1.1	<i>NB : Ce document n'est pas un guide d'annotation</i>	<i>3</i>
2	Contraintes Linguistiques et Formelles	4
2.1	<i>Noeuds des arbres de dépendances : les formes fléchies</i>	<i>4</i>
2.2	<i>Relations de dépendances utilisées pour le français</i>	<i>4</i>
2.2.1	<i>Distinctions supplémentaires</i>	<i>5</i>
2.2.2	<i>Récapitulatif</i>	<i>5</i>
2.3	<i>Format.....</i>	<i>6</i>
2.3.1	<i>Format CONLL</i>	<i>6</i>
2.3.2	<i>Format pivot</i>	<i>8</i>
3	Dépendants du verbe	9
3.1	<i>Auxiliaires</i>	<i>10</i>
3.2	<i>Clitiques et pronoms forts</i>	<i>10</i>
3.2.1	<i>Clitiques inversés</i>	<i>10</i>
3.2.2	<i>Clitiques et dislocations</i>	<i>10</i>
3.2.3	<i>Cas difficiles pour les clitiques</i>	<i>10</i>
3.2.4	<i>Pronoms forts</i>	<i>11</i>
3.3	<i>La fonction OBJ</i>	<i>12</i>
3.3.1	<i>Cas des pronoms objets dans le noyau verbal</i>	<i>13</i>
3.4	<i>La fonction ATS, ATO</i>	<i>13</i>
3.5	<i>La fonction A-OBJ</i>	<i>13</i>
3.6	<i>La fonction DE-OBJ</i>	<i>14</i>
3.7	<i>La fonction P-OBJ</i>	<i>14</i>
3.8	<i>La fonction MOD</i>	<i>15</i>
3.9	<i>La fonction SUJ.....</i>	<i>15</i>
4	Changements de diathèse	16
4.1	<i>Passif.....</i>	<i>16</i>

4.2	<i>Causatif</i>	16
4.3	<i>Impersonnel</i>	17
5	Dépendants du nom	17
5.1	<i>Incises au sein d'un NP (à faire)</i>	17
5.2	<i>Relatives</i>	17
5.2.1	Relatives avec antécédents	17
5.2.2	Relatives sans antécédent (à faire)	18
5.3	<i>Clivées</i>	18
6	Dépendants de la prep	18
7	Dépendants phrastiques et infinitivaux	19
7.1	<i>Dépendants infinitivaux</i>	19
7.1.1	objets directs sans complémenteur	19
7.1.2	objets directs avec complémenteur	20
7.1.3	indirects	20
7.2	<i>Dépendants phrastiques</i>	20
7.2.1	Sujets et Objets phrastiques	20
7.2.2	Ajouts phrastiques	20
7.2.3	Compléments indirects phrastiques.....	20
8	Coordination (incomplet)	20
8.1	<i>Enumérations</i>	21
8.2	<i>Coordinations de phrases</i>	21
8.3	<i>Head-Gapping (à faire)</i>	21
8.4	<i>Right Node Raising (à faire)</i>	21
8.5	<i>Left subj elision et RNR (à faire)</i>	21
8.6	<i>Zeugme (à faire)</i>	22
8.7	<i>Coordination de catégories différentes</i>	22
8.8	<i>Arg cluster (Coordination de séquences de constituants) (à faire)</i>	22
8.9	<i>Unbounded ellipsis + RNR (à faire)</i>	22
8.10	<i>Det coord</i>	22
9	Comparatives	22
10	Consécutives	24
11	Participiales et adjectivales	25
11.1	<i>participiales adnominales</i>	25
11.2	<i>participiales absolues, avec sujet réalisé</i>	25
11.3	<i>participiales absolues, sans sujet réalisé</i>	25
11.4	<i>adjectivales</i>	25
12	Dépendants adverbiaux	26
12.1	<i>Cas général</i>	26
12.2	<i>Superlatif</i>	26
13	Références	26

1 Introduction

Pour la représentation syntaxique, nous nous intéressons aux dépendances syntaxiques de surface, pour deux raisons principales. D'une part, elles constituent une annotation plus directement exploitable pour l'extraction de structures argumentales. D'autre part, elles constituent un format plus neutre pour évaluer et comparer différents analyseurs syntaxiques.

Nous choisissons de définir un schéma de dépendances de surface, syntaxiquement plus neutre que l'analyse en constituants, et facilitant la conversion vers différents schémas connus, comme le schéma Easy (Paroubek et al., 2005) ou des schémas internationaux tels le schéma GR (Carroll et al. 98), le schéma Stanford Dependencies (De Marneffe et al., 2006) ou le schéma PARC 700 (King et al., 2003). La multiplicité de ces schémas d'annotation tient pour partie à des choix linguistiques et pratiques différents, notamment sur le caractère surfacique ou pas des dépendances. Ainsi par exemple les relations de contrôle sont sensées être encodées pour les 4 formats cités, mais les dépendances sur mots sémantiquement vides n'apparaissent que pour GR et Stanford. Nous avons préféré définir un schéma d'annotation uniquement en dépendances, et en dépendances purement *de surface*, dont le schéma d'annotation soit un pivot convertible vers les différents standards cités, et un pivot enrichissable ou convertible en dépendances plus profondes.

L'objectif pratique de ce guide est la description d'un schéma de dépendances de surface obtenu par conversion des arbres syntagmatiques du French Treebank (Abeillé et al. 2003, ci-après FTB), et plus généralement les analyses fournies par les parsers statistiques appris sur le FTB.

Cette conversion a été réalisée sur l'ensemble du FTB, ainsi qu'une évaluation de cette conversion : nous avons constitué un corpus d'évaluation de 120 phrases, pour lesquelles la représentation en dépendances obtenue automatiquement a été manuellement corrigée.

Cela a amené à distinguer deux schémas d'annotation :

- un premier niveau, obtenable par conversion automatique des arbres en constituants du FTB. Les spécifications linguistiques pour ce niveau découlent largement d'une simple traduction « constituants vers dépendances » du schéma d'annotation en constituants fonctionnellement annotés du FTB (Abeillé et al. 2004 ; Abeillé, 2004) (ci-après le **schéma d'annotation FTB**). Il y a cependant certains cas où l'annotation doit être plus précise, et quelques cas de divergences.
- un second niveau, décrivant partiellement une annotation alternative pour des constructions que l'on sait erronées en sortie de la conversion automatique en dépendances => ce niveau est appelé infra « dépendances profondes », et dans toute la suite du document, sauf mention explicite « à inférer en dépendances profondes », nous décrivons ce qui est obtenu au premier niveau, par conversion automatique.

1.1 NB : Ce document n'est pas un guide d'annotation

Nous insistons sur le fait que ce document n'est pas un guide d'annotation en dépendances : il n'a ni été conçu ni utilisé pour une campagne d'annotation en dépendances de phrases. Ce document est la *description* d'un schéma d'annotation de dépendances de surface pour le français : le schéma d'annotation tel que reflété dans le résultat de la conversion automatique du French Treebank, des constituants vers les dépendances.

2 Contraintes Linguistiques et Formelles

La dépendance syntaxique représente le fait que la présence d'un mot et tous ses dépendants est légitimée par un autre mot, son gouverneur (cf. Kahane, 2001 pour un historique de la représentation en dépendances).

On entend ici annoter des dépendances dites “surfaciques”, c'est-à-dire des relations entre formes fléchies, où toute forme fléchie, même sémantiquement vide, est représentée. Par définition, une forme fléchie a un et un seul gouverneur, sauf pour la forme "racine" de la phrase. Pour récupérer facilement la racine de l'arbre de dépendances, on rajoute une dépendance root(null, racine). On obtient ainsi exactement une dépendance par mot de la phrase d'entrée.

Ainsi, formellement, la structure de dépendance associée à une phrase est un arbre orienté, dont les noeuds correspondent aux formes fléchies de la phrase, et les arcs qui correspondent aux dépendances, sont étiquetés par une relation de dépendance.

L'ordre linéaire de la phrase est encodé par un identifiant sur les noeuds, de type entier, local à chaque phrase (par convention le premier mot de chaque phrase reçoit l'identifiant 0).

Notre schéma d'annotation n'impose pas la projectivité, i.e. la projection d'un noeud peut correspondre à un segment discontinu de la phrase. Cependant, par construction, la conversion automatique des arbres de constituants FTB en dépendances produit des arbres de dépendances projectifs.

2.1 Noeuds des arbres de dépendances : les formes fléchies

Les noeuds sont étiquetés par l'identifiant d'ordre linéaire dans la phrase, une forme fléchie, un lemme, et par la catégorie syntaxique de ce lemme. Les catégories sont un jeu de 28 catégories intégrant les 13 catégories principales du French Treebank, plus des informations de mode du verbe, la distinction nom propre / nom commun (cf. (Crabbé et Candito, 2008)).

Les formes fléchies utilisées sont des mots au sens d'unités insécables et autonomes, elles peuvent donc correspondre graphiquement à une suite de tokens (un « mot composé »).

A noter le caractère formel de cette définition : les mots composés ont une syntaxe particulière (insertion bloquée, et/ou la suite de catégories des composants ne correspond pas à une syntaxe régulière).

Pour ce qui est de la sémantique, on peut ainsi avoir une forme fléchie asémantique, ou bien une forme fléchie faisant partie d'une unité sémantique formellement discontinu (comme dans une expression figée).

2.2 Relations de dépendances utilisées pour le français

Les relations de dépendances dans le cas de gouverneur **verbal** sont les fonctions définies dans le schéma d'annotation du FTB (Abeillé, 2004): *a_obj*, *ats*, *ato*, *de_obj*, *obj*, *mod*, *p_obj*, *subj*

Nous étendons l'utilisation de ces fonctions dans deux cas, non annotés dans le FTB :

- (i) les dépendants dans une participiale passée

- (ii) les dépendants adverbiaux réduits à un seul mot comme dans *ils sont là*. On ajoute par ailleurs

On ajoute par ailleurs les relations *aux_caus*, *aux_pass*, *aux_tps* pour les auxiliaires, et *aff* pour les clitiques figés.

Pour les autres catégories de gouverneurs, pour lesquels le FTB ne contient pas d'annotation fonctionnelle, on utilise les relations : *mod*, *mod_rel*, *coord*, *arg*, *arg_coord*, *arg_comp*, *arg_cons*, *det*, *ponct* et *dep*.

On ne gère pas la sous-catégorisation des têtes non verbales. En particulier, dans le cas de dépendants prépositionnels, pour un gouverneur non verbal, on ne distingue pas les arguments et les ajouts, et on utilise la relation générique *dep* par défaut.

2.2.1 Distinctions supplémentaires

Dans de nombreux cas, la structure formelle des dépendances définies ci-dessus ne permet pas de repérer certains traits cruciaux pour l'interprétation, notamment pour l'extraction des structures argumentales.

Aussi, lors de l'annotation manuelle, on choisit de préciser certaines relations par des informations de type sémantique.

On code ainsi le caractère locatif pour un *mod* ou un *p_obj* (*mod_loc* / *p_obj_loc*) : cette information est sémantique mais a aussi des effets formels : la préposition n'est pas figée, la cliticisation en y.

On spécifie *p_obj_agt* pour un complément d'agent.

On spécifie *subj_impers* pour un sujet explétif *il*, et on note *aff_moyen* pour le clitique *se* dans le cas d'une construction au moyen.

Dans les évaluations en dépendance ces relations plus spécifiques sont actuellement mappées sur leur relation générique.

2.2.2 Récapitulatif

<i>Relations pour gouverneurs verbaux (cf. schéma d'annotation FTB + ajouts)</i>	
<i>subj</i>	Sujet
<i>obj</i>	objet
<i>de_obj</i>	SP argumental en de, non locatif
<i>a_obj</i>	SP argumental en à, non locatif
<i>p_obj</i>	autre SP argumental
<i>ats</i>	Attribut du sujet
<i>ato</i>	Attribut de l'objet
<i>mod</i>	Modifieur
<i>aux_tps</i>	auxiliaires de temps
<i>aux_pass</i>	auxiliaires du passif
<i>aux_caus</i>	verbe causatif (en cas de complexe causatif + inf)
<i>aff</i>	clitiques figés
<i>Relations pour gouverneurs non verbaux</i>	
<i>mod</i>	Modifieurs repérés structurellement (par exemple adjectifs épithètes), autres que les relatives
<i>mod_rel</i>	Relatives adnominales
<i>coord</i>	Relation portée par un coordonnant, avec comme gouverneur le coordonné immédiatement précédent
<i>arg</i>	Utilisé dans le cas de prépositions « liées » : dans <i>de Charybde en Scylla</i> , parallèlement au traitement de la coordination, <i>en Scylla</i> est dépendant de type

	<i>arg</i> de la première préposition (<i>de</i>)
<i>dep_coord</i>	Relation portée par un coordonné (sauf le premier), avec comme gouverneur le coordonnant immédiatement précédent
<i>det</i>	Relation portée par les déterminants
<i>pnct</i>	Relation portée par tout dépendant typographique, sauf pour les virgules jouant le rôle de coordonnant (qui porte la relation <i>coord</i>)
<i>dep</i>	Relation sous-spécifiée, pour les dépendants prépositionnels (pas de gestion de la distinction argument / ajout pour les gouverneurs non verbaux)
Relations plus spécifiques, réservées actuellement à l'annotation manuelle	
<i>mod_loc</i>	Modifieurs sémantiquement locatifs (au propre ou au figuré). Pour faire le pendant avec les <i>p_obj_loc</i>
<i>mod_cleft</i>	La subordonnée dans le cas d'une clivée
<i>p_obj_agt</i>	Pour les compléments d'agent (en cas de passif, ou de causatif)
<i>p_obj_loc</i>	Pour les dépendants argumentaux locatifs (locatifs de source, destination, ou localisation)
<i>subj_impers</i>	Pour le sujet explétif <i>il</i>
<i>aff_moyen</i>	Pour le clitique <i>se</i> en cas de moyen
<i>arg_comp</i>	Utilisé pour relier une comparative (dépendant) et son gouverneur (section 9)
<i>arg_cons</i>	Utilisé pour relier une consécutive et son gouverneur adverbial (section 10)

2.3 Format

Le format d'écriture utilisé dans ce manuel (dit « format pivot ») est inspiré du format du schéma d'annotation PARC 700 : les dépendances *y* sont notées sous la forme «relation(gouverneur, dépendant)».

En dehors de ce manuel, le corpus en dépendances suit le format standard CoNLL (défini lors du *shared tasked* CoNLL de parsing en dépendances (<http://nextens.uvt.nl/~conll>)).

2.3.1 Format CoNLL

cf. <http://nextens.uvt.nl/~conll>

Seules les 8 premières colonnes sont renseignées (les colonnes 9 et 10 pour la tête projective et la relation projective sont non pertinentes).

Exemple

1	Lyonnaise_- Dumez	Lyonnaise- Dumez	N	NPP	g=f n=s s=p	2	subj	_	_
2	vient	venir	V	V	m=ind n=s p=3 t=pst	0	root	_	_
3	d'	de	P	P		2	de_obj	_	_
4	hispaniser	hispaniser	V	VINF	m=inf	3	obj	_	_
5	sa	son	D	DET	g=f n=s p=3 s=poss	6	det	_	_
6	filiale	filiale	N	NC	g=f n=s s=c	4	obj	_	_
7	espagnole	espagnol	A	ADJ	g=f n=s s=qual	6	mod	_	_

Jeu de catégories (tagset)

Le format CoNLL prévoit deux niveaux de granularité de catégories (colonnes 4 et 5).

Le jeu de catégories grossier est exactement constitué des 13 catégories de base du FTB.

Le jeu plus fin est construit à partir des 13 catégories de base du FTB, croisées avec le mode des verbes, et certaines distinctions apparaissant dans le trait subcat dans le FTB, soit :

- la caractère interrogatif ou pas des adverbes, adjectifs et déterminants
- une indication du cas des clitiques
- le trait commun/propre pour les noms
- la distinction conjonction de coord versus conjonction de subordination
- des tags particuliers pour les amalgames préposition+déterminant, préposition+pronom (*auquel*)

Catégorie	Catégorie de base du FTB	Trait supplémentaire du FTB	Description / exemple
V	V	mode	verbe indicatif
VS	V	mode	verbe subjontif
VINF	V	mode	verbe infinitif
VPP	V	mode	participe passé
VPR	V	mode	participe présent
VIMP	V	mode	verbe impératif
NC	N	subcat = C	nom commun
NPP	N	subcat = P	nom propre
CS	C	subcat =S	conjonction de subordination
CC	C	subcat =C	conjonction de coordination
CLS	CL	subcat =	clitique sujet
CLO	CL	subcat =	clitique objet
CLR	CL	subcat =	clitique réfléchi
P	P	subcat =	préposition non amalgamée
P+D			préposition+dét : <i>le lutin des alpages</i>
P+PRO			préposition+prorel : <i>le lieu auquel on pense</i>
I	I		interjection
PONCT	PONCT		ponctuation
ET	ET		mots étrangers
ADJ	A	subcat ≠ int	adjectifs non interrogatifs
ADJWH	A	subcat = int	adjectifs interrogatifs
ADV	ADV	subcat ≠ int	adverbes non interrogatifs
ADVWH	ADV	subcat = int	adverbes interrogatifs
PRO	PRO	subcat ≠ int ≠ rel	pronoms non interrogatifs ni relatifs
PROREL	PRO	subcat = rel	pronoms relatifs
PROWH	PRO	subcat = int	pronoms interrogatifs
DET	D	subcat ≠ int	déterminants non interrogatifs
DETH	D	subcat = int	déterminants interrogatifs

Traits (features)

FEATURE	MEANING	COMMENT
g=f	gender=feminine	
g=m	gender=m masculine	
m=imp	mood=imperative	redundant with cat=VIMP

m=ind	mood=indicative	redundant with cat=V
m=inf	mood=infinitive	redundant with cat=VINFINF
m=part	mood=participle	redundant with cat=VPP / cat=VPR
m=subj	mood=subjonctive	redundant with cat=VS
n=p	number=plural	
n=s	number=singular	
p=1	person=1	
p=2	person=2	
p=3	person=3	
s=c	common nouns	redundant with cat=NC
s=card	cardinal NC,A,D or PRO	
s=def	definite DET or P+D	
s=dem	demonstrative DET or PRO	
s=excl	exclamative DET ('quel'), ADV (e.g. 'combien')	
s=ind	indefinite DET, ADJ or PRO	
s=int	interrogative pronouns (PROWH:'qui'), adj (ADJWH:'quel'), det (DETHW:'quel'), adv (ADVWH:'comment') P+PRO ('auquel')	
s=neg	negative adverbs, determiners ('nul'), pronouns	
s=obj	object clitics	redundant with cat=CLO
s=ord	ordinal adjectives (eg. 'premier')	
s=p	proper nouns	redundant with cat=NPP
s=part	partitive determiners	
s=pers	personal pronouns (eg. 'eux')	
s=poss	possessive determiners and pronouns	
s=qual	qualificative adjectives	
s=refl	reflexive clitics and pronouns (eg:'eux-mêmes')	systematic for cat=CLR
s=rel	relative pronouns (PROREL or P+PRO)	systematic for cat=PROREL
s=s	conjunctions, or sentence final punctuation	systematic for cat=CS
s=suj	nominative clitic	redundant with cat=CLS
s=w	sentence internal punctuation	
t=cond	tense=conditional	
t=fut	tense=future	
t=impft	tense=imparfait	
t=past	tense=past for passé simple or past participles	
t=pst	tense=present	

2.3.2 Format pivot

Par rapport au format de Parc 700, on distingue cependant une zone pour les dépendances (ici de surface : zone « surf_deps ») et une zone pour des traits sur les mots (catégories, lemmes, traits sémantiques ...), facultatifs ou obligatoires (zone « features »).

Les formes sont notées en accolant ~ et un identifiant numérique, unique au sein d'une phrase. On choisit de numéroter dans l'ordre linéaire de la phrase, en partant de 0 (contrairement à Parc700 qui note 0 pour la tête de l'arbre de dépendances). Les caractères "(", ")", ",", et "-" sont respectivement encodés comme <LBR>, <RBR>, <C> et <D>.

Voici un exemple :

```

sentence(
  id(ftb_12,flmf7XXX.cat.xml)
  date(2008-07-18)
  validators(M.Falco)
  sentence_form(Comme prévu , M. Henri Krasucki , secrétaire_général depuis 1982 , n'en
fait pas partie .)
  surf_deps(
    mod(fait~13,Comme~0)
    OBJ(Comme~0,prévu~1)
    ponct(fait~13,<C>~2)
    SUJ(fait~13,M.~3)
    mod(M.-3,Henri~4)
    mod(M.~3,Krasucki~5)
    ponct(M.~3,<C>~6)
    mod_app(M.~3,secrétaire_général~7)
    mod(secrétaire_général~7,depuis~8)
    OBJ(depuis~8,1982~9)
    ponct(fait~13,<C>~10)
    mod(fait~13,n'~11)
    de-OBJ(fait~13,en~12)
    root(null~-1,fait~13)
    mod(fait~13,pas~14)
    OBJ(fait~13,partie~15)
    ponct(fait~13,~16)
  )
  features(
    pos(Comme~0,P)
    pos(prévu~1,V)
    ...
    lemma(fait~13,faire)
  ))

```

NB : dans toute la suite, on omet par simplicité les identifiants des formes, lorsqu'il n'y a pas d'ambiguïté.

3 Dépendants du verbe

Cette section spécifie les relations gouverneur verbal et dépendants du verbe. On suit d'une manière générale le guide d'annotations fonctionnelles du FTB (Abeillé, 2004). Des divergences apparaissent notamment pour les cas de dépendances à distance, pour lesquels on permet ici la non projectivité.

3.1 Auxiliaires

Dans le cas d'un complexe verbal avec auxiliaire (de temps, du passif, causatif) + verbe infinitif ou participe passé, on considère l'auxiliaire comme dépendant du verbe, y compris pour le sujet¹.

Les auxiliaires de temps sont annotés avec la relation AUX_TPS.

(1) *Les enfants ont vu le concert* => *AUX_TPS(vu-3,ont-2)*

Les auxiliaires de temps peuvent être eux même composés, et être gouverneur de leur AUX_TPS.

Dans les tours passifs, on annote l'auxiliaire passif avec la relation AUX_PASS.

(2) *Les enfants sont vus par les organisateurs* => *AUX_PASS(vus-3,sont-2)*

(3) *Les enfants ont été vus par les organisateurs*
=> *AUX_PASS(vus-4,été-3) AUX_TPS(été-3,ont-2)*

Dans les cas de causatif, on annote le causatif avec la relation AUX_CAUS.

(4) *Le surveillant fait dégager la foule* => *AUX_CAUS(dégager,fait)*

(5) *Le surveillant a fait dégager la foule* => *AUX_CAUS(dégager,fait),
AUX_TPS(fait, a)*

3.2 Clitiques et pronoms forts

3.2.1 Clitiques inversés

Les clitiques inversés redondants ou pas avec un argument sont notés comme dépendant de la tête verbale avec la relation correspondante

(6) *Donnez-les lui* => *OBJ(donnez, les), A-OBJ(donnez, lui)*

(7) *Quand Jean viendra-t-il ?* => *SUJ(viendra, Jean) SUJ(viendra, -t-il)*

3.2.2 Clitiques et dislocations

Le clitique redondant dans les dislocations est annoté également comme dépendant du verbe :

(8) *Ce filou, il a plus d'un tour dans son sac* => *SUJ(a, filou) [...] SUJ(a, il)*

3.2.3 Cas difficiles pour les clitiques

Les clitiques figés sont annotés avec la relation AFF (pour affixe).

(9) *Il se souvient de ses vacances* => *AFF(souvient, se) DE-OBJ(souvient, de),
OBJ(de, vacances)*

Les clitiques non figés sont annotés avec la relation donnée dans le FTB :

¹ Comme nous l'a fait remarquer Sylvain Kahane, il s'agit d'une entorse au principe d'annotation de surface puisque la présence du sujet est légitimée par le verbe fini (l'auxiliaire). Nous choisissons de violer ce principe, simplement de préparer le travail de dépendances plus profondes, dans un cas ici où il est trivial de le faire. A noter également que systématiquement choisir un gouverneur différent pour le sujet et les autres dépendants entraîne souvent de la non projectivité.

(10) *Il lui en parle* => *SUJ(parle,il) A-OBJ(parle,lui) DE-OBJ(parle,en)*

(11) *Il s'en souvient* => *SUJ(souvient,il) AFF(souvient,se) DE-OBJ(souvient,en)*

Les clitiques 'se' moyen sont notés avec la relation *AFF_MOYEN*

(12) *Le français se parle de moins en moins dans les conférences* =>
AFF_MOYEN(parle,se)

Quelques clitiques méritent plus d'attention :

- le est soit OBJ soit attribut du sujet (ATS)

(13) *Paul le voit* => *OBJ(voit, le)*

(14) *Paul l'est (en colère)* => *ATS(est, l')*

- en est soit DE-OBJ ou OBJ ou MOD

(15) *Paul en parlera (du film)* => *DE-OBJ(parlera,en)*

(16) *Paul en achète (des tomates)* => *OBJ(achète,en)*

Attention , très différent du FTB, qui d'ailleurs montre un codage très varié du en quantitatif ...

(17) *Paul en achète deux* => *OBJ(achète,en) DET(en,deux)*

(18) *Paul en achète deux autres*

(19) *La législation peut en amener une autre plus restrictive*

(20) *Paul en a vu la fin (du film)* => *DE-OBJ(fin, en) [ou DEP si on sous-spécifie la sous-catégorisation des noms]*

(21) *Le dynamisme en est célèbre (de cette personne)* => *MOD(dynamisme,en)*

- en peut également être figé dans ce cas on le note *AFF*

(22) *Paul en veut à Marie* => *AFF(veut, en)*

- y est soit A-OBJ, P-OBJ_LOC ou MOD

(23) *Paul y pense* => *A-OBJ(pense,y)*

(24) *Paul y va* => *P-OBJ_LOC(va,y)*

(25) *Le chômage y est endémique* => *MOD(est, y)*

- y peut également être figé :

(26) *Il y a trois ans* => *AFF(a, y)*

- se (réfléchi) est soit *a_obj* soit *obj* :

(27) *Les manifestants se donnent des fleurs* => *A-OBJ(donnent, se)*

(28) *Narcisse se regarde dans la glace* => *OBJ(regarde,se)*

3.2.4 Pronoms forts

Il y a quelques cas difficiles :

- Dont est soit DE-OBJ soit MOD selon le gouverneur :

(29) *le monsieur dont l'appartement est loué* => MOD(*appartement, dont*)

(30) *Le film dont il a vu la fin* => DE-OBJ(*fin, dont*)

(31) *La fille dont son frère lui a parlé* => DE-OBJ(*parlé, dont*)

- Tous (quantifieur flottant) est généralement un modifieur :

(32) *Paul les a tous convoqués* => MOD(*les, tous*)

3.3 La fonction OBJ

Les objets, notés OBJ, sont soit nominaux ou pronominaux (catégorie NP), soit clitiques (CLO), soit phrastiques (Ssub ou VPinf).

Le test pour identifier l'objet est la réalisation clitique en le/la/les (avec accord contrairement à l'attribut), le *en* quantitatif, le caractère direct (sans préposition dans le cas nominal).

Les objets sont typiquement des noms postverbaux sauf dans les cas de topicalisation, interrogation, relativisation, ou discours rapporté :

(33) *Il passe des heures à marcher* => OBJ(*passe-1, heures-3*)

(34) *le problème que vous avez résolu* => OBJ(*résolu-5, que-2*)

(35) *Le président a pris position* => OBJ(*pris-3, position-4*)

(36) *je crois que Paul viendra* => OBJ(*crois-1, que*), OBJ(*que, viendra-4*)

(37) *Il veut vous parler* => OBJ(*veut-1, parler-3*)

(38) *Quel désordre il a fait !* => OBJ(*fait-4, désordre-1*)

(39) *Deux heures de retard, annonce l'hôtesse* => OBJ(*annonce-4, heures-1*)

Pour les objets phrastiques ou infinitivaux, voir section 7.

Les interrogatives indirectes sont notées comme objets :

(40) *Je sais à qui tu penses* => OBJ(*sais-1, penses-5*)

(41) *Il demande où aller* => OBJ(*demande-1, aller-4*)

(42) *Ça explique pourquoi il est parti* => OBJ(*explique-1, parti-5*)

Certains adjectifs et compléments nominaux obligatoires mais non cliticisables sont objets Il s'agit de compléments indiquant une quantité (somme, poids,...) :

(43) *Cette robe coute cher* => OBJ(*coute-2, cher-3*)

(44) *Cette robe coute 50 kilos* => OBJ(*coute-2, kilos-4*)

(45) *La remorque pèse deux tonnes* => OBJ(*pèse-2, tonnes-4*)

(46) *La remorque pèse lourd* => OBJ(*pèse-2, lourd-3*)

Si l'adjectif peut être suivi d'un objet nominal alors il est noté MOD :

(47) *Paul crie fort sa joie* => MOD(*crie-1, fort-2*) OBJ(*crie-1, joie-3*)

(48) *Cette fleur sent bon la lavande* => MOD(*sent-2, bon-3*) OBJ(*sent-2, lavande-5*)

3.3.1 Cas des pronoms objets dans le noyau verbal

Ces pronoms sont notés comme objets, par exemple :

- (49) *Paul a tout compris* => OBJ(*compris-3,tout-2*)
- (50) *Ils n'ont rien vu* => OBJ(*vu-4,rien-3*)
- (51) *Il essaie de tout lire* => OBJ(*lire-4,de*), OBJ(*de, tout-3*)

3.4 La fonction ATS, ATO

Etant donné que l'on annote ici des dépendances de surface, on garde la copule comme gouverneur dans le cas d'un attribut du sujet ou de l'objet.

- (52) *Laurent est bizarre* => SUJ(*est, Laurent*), ATS(*est, magnifique*)
- (53) *Paul trouve Laurent bizarre* => SUJ(*trouve, Paul*), OBJ(*trouve, Laurent*), ATO(*trouve, bizarre*)

Sachant qu'en dépendances profondes, on voudrait plutôt :

- (54) *Laurent est bizarre* => SUJ(*bizarre, Laurent*)
- (55) *Paul trouve Laurent bizarre* => SUJ(*trouve, Paul*), OBJ(*trouve, bizarre*), SUJ(*bizarre, Laurent*)

=> **A inférer en dépendances profondes**

3.5 La fonction A-OBJ

Les objets indirects en à , notés A-OBJ, sont des compléments obligatoires soit nominaux ou pronominaux (catégorie PP), soit clitiques (CLO), soit des infinitives phrastiques (VPinf) introduites par à.

Le test pour identifier les A-OBJ est la cliticisation par lui, leur.

- (56) *Il ressemble à Martin* => A-OBJ(*ressemble-1,à*), OBJ(*à,Martin-3*)
- (57) *J'encourage Marie à venir* => A-OBJ(*encourage-1, à*), OBJ(*à,venir-4*)

La cliticisation en y indique généralement un locatif sauf dans certains cas où on notera A-OBJ :

- (58) *Jean pense à Marie* => A-OBJ(*pense-1,à*), OBJ(*à,Marie-3*)
- (59) *Jean va à Paris* => P-OBJ_LOC(*va-1,à*), OBJ(*à,Paris-3*)

Car on a pas Où pense Jean ? mais bien Où va Jean ?

3.6 La fonction DE-OBJ

Les objets indirects en *de*, notés DE-OBJ, sont des compléments obligatoires soit nominaux ou pronominaux (de catégorie PP), soit clitiques (CLO), soit des subordonnées (Ssub ou VPinf) introduites par *de*.

Le test pour identifier les compléments DE-OBJ est la cliticisation en *en* ou la pronominalisation en *dont*.

(60) *Il parle de ses vacances* => DE-OBJ(*parle-1, vacances-4*)

(61) *Il rêve de partir* => DE-OBJ(*rêve-1, de*), OBJ(*de, partir-3*)

(62) *Ça ne date pas d'hier* => DE-OBJ(*date-2, d'*), OBJ(*d', hier-4*)

Cas des locatifs introduits par *de* : Les locatifs obligatoires introduits par *de* ne sont pas des DE-OBJ. On les repère par une pronominalisation en *d'où* (et clitique en *en*)

(63) *Pierre vient de Paris* => P-OBJ_LOC(*vient-1, de*), OBJ(*de, Paris-3*)

(64) *Pierre est de Paris* => P-OBJ_LOC(*est-1, de*), OBJ(*de, Paris-3*)

Les infinitives introduites par *de* qui se pronominalisent par clitique accusatif sont notées OBJ :

(65) *Jean propose de venir* => OBJ(*propose-1, de*), OBJ(*de, venir-3*)

Les complétives introduites par *que* sont notées DE-OBJ quand elles se pronominalisent par *en*, ou si on a une infinitive en *de*, une complétive en *de ce que*, ou une nominalisation en *cela*

(66) *Paul s'étonne que Marie ne soit pas là* => DE-OBJ(*étonne, que*), OBJ(*que, soit-6*)

3.7 La fonction P-OBJ

Les objets indirects introduits par une préposition, notés P-OBJ, sont des compléments obligatoires nominaux (de catégorie PP), soit des subordonnées (Ssub ou VPinf) introduites par *de*.

Pour reconnaître les P-OBJ, il faut tester leur caractère obligatoire (non effaçable sans changement de sens), ou bien non mobile.

Il s'agit d'une fonction typiquement attribuée au complément d'agent au passif et aux locatifs obligatoires.

Dans le cas locatif (au propre et au figuré) on utilise un sous-type : P-OBJ_LOC

(67) *Pierre est attendu par Marie* => P-OBJ_AGT(*attendu, par*), OBJ(*par, Marie*),
AUX_PASS(*attendu, est*)

(68) *Je compte sur Pierre* => P-OBJ(*compte, sur*) OBJ(*sur, Pierre*)

(69) *Je vais à la pêche* => P-OBJ_LOC(*vais, à*), OBJ(*à, pêche*)

(70) *Il sort de la maison* => P-OBJ_LOC(*sort, de*), OBJ(*de, maison*)

- (71) *Il se comporte comme s'il était seul* ==> ici « comme » est préposition, et prend une complétive comme objet P-Obj(comporte, comme) Obj(comme, s'), Obj(s', était)

Les adverbes locatifs obligatoires et les NPs locatifs ont également la fonction P-Obj_LOC lorsqu'ils sont obligatoires :

- (72) *Paul va là-bas*
 (73) *Paul va très loin*
 (74) *Paul va rue Richard Lenoir*

Les adverbes de manière obligatoires et à place fixe sont également notés P-Obj

- (75) *Paul va bien* ==> P-Obj(va, bien)
 (76) *Paul se comporte mieux* ==> P-Obj(comporte, mieux)

3.8 La fonction MOD

La relation MOD couvre tous les dépendants optionnels du verbe (et des autres catégories ce qui n'est pas détaillé dans cette section). Les modificateurs sont optionnels et généralement mobiles. Leurs catégories sont variables : il peut s'agir de PP, d'adverbes, d'adjectifs, de NPs, d'infinitives, de relatives extraposées ou de subordonnées circonstancielles.

- (77) *Paul travaille le samedi* ==> MOD(travaille-1, samedi-3)
 (78) *A lire ce livre, on dirait un roman* ==> MOD(dirait-5, lire-1)
 (79) *Il y a des lions en Afrique* ==> MOD(a-2, en) Obj(en, Afrique-6)
 (80) *On progresse en se trompant* ==> MOD(progresse-1, en) Obj(en, trompant-4)
 (81) *Fatigué, Paul partit à la campagne* ==> MOD(partit-2, Fatigué-0)
 (82) *Paul est parti fatigué* ==> MOD(Paul-2, fatigué-3)

En particulier, les infinitives de but sont notées MOD :

- (83) *Je vais à Paris acheter des livres* ==> MOD(vais-1, acheter-4)

Les quantifieurs flottants sont également notés MOD mais ils modifient le nom qui est quantifié !

- (84) *Les délégués apporteront chacun leur contribution* ==> MOD(délégués-1, chacun-3)
 (85) *Les enfants viendront tous* ==> MOD(enfants-1, tous-3)

3.9 La fonction SUJ

Le sujet est obligatoire en phrase finie, sauf à l'impératif, il s'agit d'un nom, d'un pronom, d'un clitique ou d'une subordonnée (infinitive ou finie). Aucun sujet n'est un groupe prépositionnel.

Dans le cas de noms, de pronoms ou de clitiques il s'accorde en nombre et en personne avec le verbe. Il se cliticise en il, on, ce etc.

- (86) *la fille qui viendra* => *SUJ(viendra-3,qui-2)*
 (87) *Arriver à l'heure était impossible* => *SUJ(était-4,Arriver-0)*
 (88) *De 300 à 400 personnes étaient réunies* => *SUJ(étaient-4,personnes-3)*

Dans le dernier cas, la préposition introduit le quantificateur, on a également : deux personnes étaient réunies

Cas de sujets inversés on a des sujets inversés dans des phrases interrogatives, dans des relatives et dans certains cas de discours rapporté :

- (89) *Quelle est la température ?* => *SUJ(est-1,temperature-3)*
 (90) *le livre qu'a donné Paul à Marie* => *SUJ(donné-4,Paul-5)*

Cas de l'inversion complexe dans ce cas on note deux fois la fonction sujet :

- (91) *Quand Paul viendra-t-il ?* => *SUJ(viendra-2,Paul-1) SUJ(viendra-2,il-3)*

4 Changements de diathèse

On annote les fonctions de surface. Ainsi au passif par exemple, l'argument sémantique 2 porte la fonction sujet.

4.1 Passif

L'auxiliaire du passif est dépendant du participe (relation AUX_PASS) :

- (92) *Paul est apprécié de tous* => *P-OBJ_AGT(apprécie, de) SUJ(apprécie, Paul),*
OBJ(de, tous), AUX_PASS(apprécie, est)
 (93) *Paul est apprécié par tous* => *P-OBJ_AGT(apprécie, par), SUJ(apprécie, Paul),*
OBJ(par, tous), AUX_PASS(apprécie, est)

4.2 Causatif

On choisit de noter tous les dépendants sur l'infinitif (y compris les clitiques), et une relation AUX_CAUS entre faire et l'infinitif. Cela permettra de repérer les cadres de sous-catégorisation causatifs.

- (94) *Paul fait entrer Marie* => *SUJ(entrer, Paul), OBJ(entrer, Marie),*
AUX_CAUS(entrer, fait)
 (95) *Paul fait parler les enfants à leur mère* => *SUJ(parler, Paul), OBJ(parler, enfants),*
A-OBJ(parler, à), OBJ(à, mère),
AUX_CAUS(parler, fait)

4.3 Impersonnel

On note le sujet impersonnel comme dépendant de la tête verbale avec comme relation SUJ_IMPERS. Le « sujet logique » est OBJ du verbe:

- (96) *Il arrive un train* => SUJ_IMPERS(arrive, il), OBJ(arrive, train)
(97) *Il est impossible d'arriver à l'heure* => SUJ_IMPERS(est, il), OBJ(est, de),
OBJ(de, arriver), ATS(est, impossible)
(98) *Il en va de_même pour les étudiants* => SUJ_IMPERS(va, il), AFF(va,en),
MOD(va,de_même)
(99) *Il faut ajouter neuf candidats* => SUJ_IMPERS(faut, il), OBJ(faut,ajouter)
(100) *il n'est pas nécessaire de ...* => SUJ_IMPERS(est, il)

On code de la même manière les impersonnels à sujet *ça* ou *ce*

- (101) *Ça serait bien de finir* => SUJ_IMPERS(serait, Ça), OBJ(serait, de), OBJ(de,
finir), ATO(serait, bien)

(Rem pour l'infinitive directe *de finir*, le complémenteur *de* est considéré comme la tête, pour un traitement uniforme de tous les complémenteurs, cf. section 7)

Rem : le FTB ne note pas de fonction pour l'adverbe nu *bien*. Ici on uniformise, et dans ce cas il s'agit d'un attribut de l'objet (attribut du sujet dans le tour personnel *Finir serait bien*)

5 Dépendants du nom

Faire une distinction entre arguments et ajouts pour les dépendants d'un nom est souvent délicat.

On code une relation sous-spécifiée (relation générique DEP) dans le cas où on ne sait pas faire la distinction sans faire appel au lexique. Donc par exemple les adjectifs épithètes sont systématiquement MOD, en revanche les SP dépendants d'un nom portent seulement DEP.

5.1 Incises au sein d'un NP (à faire)

- (102) *Un voyage, j'espère, agréable* => ???

5.2 Relatives

5.2.1 Relatives avec antécédents

Relation MOD_REL, la tête de la relative étant le verbe (y compris pour relatives en qui et que)

- (103) *la série télévisée dans laquelle joue Karl* => MOD_REL(série, joue), MOD(joue,
dans), OBJ(dans,laquelle)
(104) *la série qui plaît* => MOD_REL(série, plaît), SUJ(plaît, qui)

5.2.1.1 Cas de dont

Quand dont correspond à un complément du verbe de la relative, il est annoté DE-OBJ.

(105) *la série dont Pierre parle* => *MOD_REL(série, parle), DE-OBJ(parle, dont)*

Quand il correspond à un complément d'un NP interne à la relative, on note la dépendance correcte (non projective), suffixée par # pour cas de dépendance non locale

(106) *la série dont Pierre connaît la fin* => *MOD_REL(série, connaît), DE-OBJ#(fin, dont)*

(107) *parmi les candidats, dont trente-deux sont des femmes, ...* => *MOD_REL(candidat, sont), MOD#(trente-deux, dont)*

(cf. on peut reconstruire "trente-deux des candidats sont des femmes")

Quand la relative en dont est sans verbe (tour elliptique), la tête de la relative est le NP, et dont en est un modifieur (cf. les spécifications du schéma Easy)

(108) *les organisations des pays membres, dont en France FO* => *MOD_REL(organisations, FO), MOD(FO, dont), MOD_LOC(FO, en), OBJ(en, France)*

5.2.2 Relatives sans antécédent (à faire)

(109) *Choisis qui tu veux*

(110) *Paul donne son livre à qui le veut*

5.3 Clivées

[A VALIDER]

Le foyer de clivée est attribut du sujet. La subordonnée (en que/qui) est notée comme un dépendant *mod_cleft* du verbe être.

La relation syntaxique entre le verbe de la subordonnée et le foyer de clivée n'est pas notée au niveau des dépendances de surface.

=> **à inférer en dépendances profondes**

Le schéma suit celui du French Treebank où la subordonnée en que/qui est codée comme un relative, non intégrée au foyer de clivée. Ainsi la subordonnée a comme tête le verbe.

(111) *C'est un difficile début d'année qui s'annonce* => *ats(est, début), mod_cleft(est, annonce), SUJ(annonce, qui)*

(112) *C'est à la fin mai que la BERD a adopté à son égard une nouvelle stratégie* => *ats(est, à), OBJ(à, fin), MOD(fin, mai), mod_cleft(est, adopté), comp(adopté, que)*

6 Dépendants de la prep

On note systématiquement la préposition tête (en surface) du SP, que la préposition soit sémantiquement pleine ou vide.

La prep sous-catégorise un OBJ nominal (pour la gloire), phrastique (pour que Paul chante), ou infinitival (pour chanter), ou un P-OBJ (jusque chez Paul).

(113) *jusque chez Pierre* => P-OBJ(*jusque, chez*), OBJ(*chez, Pierre*)

Pour les séquences de ... en ... ou de ... à ..., on note la deuxième préposition comme dépendant sous-catégorisé (ARG) de la première.

(114) *tomber de Charybde en Scylla* => MOD_LOC(*tomber, de*), OBJ(*de, Charybde*), ARG(*de, en*), OBJ(*en, Scylla*)

(115) *de 5 à 7 %* => OBJ(*de, 5*), ARG(*de, à*), OBJ(*à, %*)

(116) *depuis chez Paul jusqu'à chez Marcel* => P-OBJ_LOC(*depuis, chez*), OBJ(*chez, Paul*), ARG(*depuis, jusqu'à*), P-OBJ_LOC(*jusqu'à, chez*) OBJ(*chez, Marcel*)

(117) *Paul rêve de partir* => DE-OBJ(*rêve, de*), OBJ(*de, partir*)

(118) *Paul dort pour rêver* => MOD(*dort, pour*), OBJ(*pour, rêver*)

(119) *Paul parle tout en dormant* => MOD(*parle, en*), OBJ(*en, dormant*), MOD(*en, tout*)

(120) *Comme prévu, Paul chante* => MOD(*chante, Comme*), OBJ(*Comme, prévu*)

La préposition peut être modifiée par un adverbe :

(121) *peu avant le début* => MOD(*avant, peu*), OBJ(*avant, début*)

7 Dépendants phrastiques et infinitivaux

Attention : contrairement au schéma d'annotation du French Treebank, on choisit de normaliser par rapport au traitement de la préposition dans le cas d'un SP nominal : les complémenteurs (conjonctions) et prépositions introduisant les compléments ou ajouts phrastiques et infinitivaux servent d'intermédiaires entre le verbe principal et le verbe enchâssé.

La distinction complémenteur ou préposition sémantiquement vide / sémantiquement pleine n'est pas faite à ce niveau de représentation.

7.1 Dépendants infinitivaux

7.1.1 objets directs sans complémenteur

(122) *Paul pense partir.* => OBJ(*pense, partir*)

NB : Les équations de contrôle ne sont pas notées dans ce format pivot : on note SUJ(*pense, Paul*) mais pas SUJ(*partir, Paul*)

7.1.2 objets directs avec complémenteur

Repérés car alternant avec un complément nominal direct

Le Treebank étiquette le complémenteur comme une préposition. Mais la distinction par rapport au cas ci-dessus est la fonction OBJ et pas DE-OBJ ou A-OBJ.

- (123) *Paul promet de partir* => OBJ(promet, de), OBJ(de, partir)
(124) *Paul veille à fermer la porte* => A-OBJ(veille, à), OBJ(à, fermer)
(125) *Il arrive de faire des erreurs* => OBJ(arrive-1,de), OBJ(de, faire-4)

7.1.3 indirects

- (126) *Paul rêve de partir* => DE-OBJ(rêve, de), OBJ(de, partir)
(127) *Paul dort pour rêver* => MOD(dort, pour), OBJ(pour, rêver)

7.2 Dépendants phrastiques

7.2.1 Sujets et Objets phrastiques

- (128) *Paul promet qu'il partira* => OBJ(promet, que), OBJ(que, partira)
(129) *Que Paul parte surprend tout le monde.* => SUJ(surprend, que), OBJ(que, parte)

7.2.2 Ajouts phrastiques

- (130) *Paul rit parce qu'il a compris* => MOD(rit, parce_qu), OBJ(parce_qu, compris)
(131) *Quand Paul rit, Pierre sourit.* => MOD(sourit, quand), OBJ(quand, rit)

7.2.3 Compléments indirects phrastiques

Apparaissant soit avec un tour *de ce que*, soit sans préposition, mais repérables car alternant avec SP nominaux : Paul rêve qu'il part / Paul rêve de la lune)

- (132) *Paul rêve qu'il part* => DE-OBJ(rêve, qu), OBJ(qu, part)

8 Coordination (incomplet)

On choisit pour annoter la coordination un schéma où la tête est le premier conjoint (schéma « à la Melcuk »).

- (133) *Jean voit Paul et Marie* => OBJ(voir, Paul), COORD(Paul, et), DEP_COORD(et, Marie)

Problématique pour distinction : un étudiant et un prof intelligents / un étudiant et un prof intelligent => A envisager, introduire suffixe _DISTRIB ??

- (134) *Un étudiant et un professeur intelligents discutent* => SUJ(discutent, étudiant), COORD(étudiant, et), DEP_COORD(et, professeur), MOD_DISTRIB(professeur, intelligents)

(135) *Un étudiant intelligent et un professeur idiot discutent* => *SUJ(discutent, étudiant), MOD(étudiant, intelligent), COORD(étudiant, et), DEP_COORD(et, professeur), MOD(professeur, idiot)*

(136) *Ils vont à Paris et à Genève* => *P-OBJ_LOC(vont-1, à-2), COORD(à-2, et), DEP_COORD(et, à-5)*

8.1 Enumérations

La virgule peut servir de coordonnant (notée <C>)

(137) *Jean achète du pain le lundi, le mardi et le mercredi* => *MOD(achète, lundi), COORD(lundi, <C>), DEP_COORD(<C>, mardi), COORD(lundi, et), DEP_COORD(et, mercredi)*

8.2 Coordinations de phrases

(138) *Jean aime Marie et Paul aime Virginie* => *HEAD(null, aime-1), COORD(aime-1, et), DEP_COORD(et, aime-3)*

Cas d'un coordonnant sémantiquement plein : même annotation de dépendances de surface

(139) *Jean aime Marie mais Paul aime Virginie* => *HEAD(null, aime-1), COORD(aime-1, mais), DEP_COORD(mais, aime-3)*

8.3 Head-Gapping (à faire)

Tout se passe comme si un des conjoints a son élément tête qui est omis.

(140) *Jean aime Marie et Paul Virginie* => *si la tête n'était pas omise on aurait : HEAD(null, aime-1), COORD(aime-1, et), DEP_COORD(et, aime-3), SUJ(aime-3, Paul), OBJ(aime-3, Virginie) ... => avec omission, on n'a plus « DEP_COORD(et, aime-3), et on marque SUJ-ELL ??*

8.4 Right Node Raising (à faire)

(141) *Marie fabrique et Pierre vend des crêpes* => *en surface HEAD(null, fabrique), COORD(fabrique, et), DEP_COORD(et, vend)*

Et on ajouterait une dépendance profonde : une zone deep_dep, où l'on peut noter OBJ(fabrique, crêpes) en notant que c'est une ellipse ??, on va avoir ce pb pour toutes les ellipses

8.5 Left subj elision et RNR (à faire)

(142) *Marie cuit et vend des crêpes*

8.6 Zeugme (à faire)

(143) *Napoléon prit du poids et beaucoup de pays*

8.7 Coordination de catégories différentes

(144) *Jean est un républicain et fier de l'être* => *ATS(est, républicain),
COORD(républicain, et), DEP_COORD(et,
fier)*

8.8 Arg cluster (Coordination de séquences de constituants) (à faire)

Une possibilité : notation proche de celle du head gapping, les infos fonctionnelles permettent de repérer le scope.

(145) *Paul offre des baskets bleues à la petite fille et des rouges à la grande* =>
*SUJ(offre, Paul), OBJ(offre, baskets), A-
OBJ(offre, à-1), COORD(offre, et),
OBJ_ELL(et, rouges), A-OBJ_ELL(et, à-2)*

Rappel : Le treebank donne : (SENT (NP (N Paul)) (VN (V offre)) (NP des baskets bleues) (PP à la petite fille) (COORD (CC et) (NP des rouges) (PP à la grande)))

8.9 Unbounded ellipsis + RNR (à faire)

(146) *Jean aime et espère épouser Marie*

A représenter en déps. profondes ??

NP coord

(147) *Jean et Paul se marient*

8.10 Det coord

(148) *J'ai vu un ou des coureurs au départ* => *DET(coureurs, un), COORD(un, ou),
DEP_COORD(ou, des)*

Remarque : problématique pour le repérage cf. le treebank n'annote pas des constituants déterminant

9 Comparatives

On prend l'adverbe (plus, moins, aussi, davantage...) comme gouverneur de la comparative en que.

L'adverbe est dépendant du terme de la comparaison (adjectif, nom, verbe ou adverbe), par la relation MOD.

Le complémenteur introduisant la subordonnée est dépendant de l'adverbe et noté avec la relation ARG_COMP.

Marie : ce point à valider : La tête de la subordonnée (même dans le cas elliptique) est objet du complémenteur. Autre solution à envisager : accoler le nom de la fonction de cet élément dans ce qui a été élidé (OBJ_SUJ , OBJ_OBJ ...cf. exemples (151) et (152))

(149) *La voiture de Pierre est plus économe que celle de Jean* => => *ATS(est,grosse), MOD(économe,plus), ARG_COMP (plus,que), OBJ(_SUJ)(que, celle)*

(150) *L'emploi en France repart plus vite que les salaires* => *MOD(repart,vite), MOD(vite,plus), ARG_COMP (plus,que), OBJ(_SUJ)(que, salaires)*

(151) *Le Japon possède plus de main d'oeuvre qualifiée que la Chine* *OBJ(possède,main d'oeuvre), MOD(main d'oeuvre,plus), MOD(main d'oeuvre,qualifiée), ARG_COMP (plus,que), OBJ(_SUJ)(que, Chine)*

(152) *Le Japon possède plus de main d'oeuvre qualifiée que de paysans* => *OBJ(possède, main d'oeuvre), MOD(main d'oeuvre, plus), MOD(main d'oeuvre, qualifiée), ARG_COMP (plus, que), OBJ(_OBJ)(que, paysans)*

Pour gérer le cas des comparatives sans subordonnée, c'est-à-dire le cas où il y a une ellipse complète de la subordonnée :

(153) *Paul a acheté une maison plus grande* => *OBJ(acheté,maison), MOD(maison,grande) , MOD(grande,plus)*

(154) *Paul a acheté une plus grande maison* => *OBJ(acheté,maison), MOD(maison,grande) , MOD(grande,plus)*

Le cas le plus simple est celui des comparatives avec subordonnée réalisée :

(155) *Il est plus offensif qu'il (ne) l'était lors de la dernière réunion* => *ATS(est,offensif), MOD(offensif,plus), ARG_COMP(plus,que), OBJ(que,était) SUJ(était,il), ATS(était,l')...*

Toutefois, le cas le plus courant est celui des comparatives avec subordonnée à head-gapping :

(156) *Jean est plus petit que Paul* => *ATS(est,petit), MOD(petit,plus), ARG_COMP(plus,que), SUJ(que, Paul)*

(157) *La situation économique est moins encourageante que l'année dernière.* => *ATS(est,encourageante), MOD(petit,moins), ARG_COMP(moins,que), OBJ(que, année), MOD(année,dernière)*

(158) *Il mange plus rapidement que toi* => MOD(*rapidement, plus*), MOD(*mange, rapidement*)

On peut avoir une discontinuité entre l'élément modifié par l'adverbe quantifieur et la subordonnée.

(159) *L'élimination touchera davantage les sortants que les entrants.* => OBJ(*touchera, sortants*), MOD(*touchera, davantage*), ARG_COMP(*davantage, que*), OBJ(*que, entrants*)

(160) *Il finit plus rapidement son travail que toi* => OBJ(*finit, travail*), MOD(*finit, rapidement*), MOD(*rapidement, plus*), ARG_COMP(*plus, que*), OBJ(*que, toi*)

Cas de quantifieur flottant : on garde une représentation surfacique, où l'adverbe quantifieur modifie le verbe

(161) *Il a plus mangé de bonbons que toi* => OBJ(*mangé, bonbons*), MOD(*mangé, plus*), ARG_COMP(*plus, que*)

Alors que l'équivalent avec le quantifieur non flottant, on code le quantifieur modifieur du N :

(162) *Il a plus mangé de bonbons que toi* => OBJ(*mangé, bonbons*), MOD(*bonbons, plus*), ARG_COMP(*plus, que*)

On a également des cas de comparative sans terme de comparaison : la comparaison porte sur deux clauses. Dans ce cas, on note la construction subordonnée comme modifieur de la principale :

(163) *Il grossit plus qu'il ne grandit* => MOD(*grossit, plus*), ARG_COMP(*plus, que*)

(164) *Jean travaille plus que Luc* => MOD(*travaille, plus*), ARG_COMP(*plus, que*)

(165) *Jean rit autant qu'il pleure* => MOD(*rit, autant*), ARG_COMP(*autant, qu'*)

10 Consécutives

(166) *trop rapide pour être durable*

Le SP *pour être durable* étant syntaxiquement légitimé par *trop*, on le code comme dépendant de l'adverbe, avec une relation spécifique *arg_cons.* : MOD(*rapide, trop*), ARG_CONS(*trop, pour*), OBJ(*pour, être*)

(167) *suffisamment lent pour que cela passe inaperçu* => MOD(*lent, suffisamment*), arg_cons(*suffisamment, pour*), OBJ(*pour, que*), OBJ(*que, passe*) ...

(168) *d'une telle force que la capsule explosa* => MOD(*force, telle*), arg_cons(*telle, que*)

11 Participiales et adjectivales

11.1 participiales adnominales

On les considère comme des modificateurs du nom. La relation profonde sujet entre le participe et le nom modifié n'est pas codée en surface.

=> **A inférer en dépendances profondes**

(169) *Les personnes nées en 48 => MOD(personne, nées)*

(170) *Les personnes ayant un tel comportement => MOD(personnes, ayant)*

11.2 participiales absolues, avec sujet réalisé

(171) *La nuit venant, les chats sortirent => MOD(sortirent, venant), SUJ(venant, nuit)*

(172) *Les feuilles ramassées, les jardiniers partirent => MOD(partirent, ramassées),
SUJ(ramassées, feuilles)*

(173) *Une fois tous les participants arrivés, on commença => MOD(arrivés, une fois),
MOD(participants, tous), det(participants,
les), suj(arrivés, participants),
MOD(commença, arrivés)*

11.3 participiales absolues, sans sujet réalisé

(174) *Née le 15 janvier, Paulette fêtera son 15ème anniversaire. (= > d'après le treebank,
on a un VPpart, modifieur au niveau du VN, ce
qui donne MOD(fêtera, née))*

(175) *Tablant sur un cours de baril à 21 dollars, le président*

(176) *Marland procèdera à une augmentation de capital, permettant ainsi ...*

On reste sur un codage très surfacique : la participiale est modifieur du verbe principal. La relation de sujet entre le participe et le sujet de la principale n'est pas codée en surface.

=> **A inférer en dépendances profondes**

11.4 adjectivales

Même situation pour les "adjectivales" (??) (environ 50 cas dans tout le treebank)

(177) *Enjoué, volubile, XXX savait captiver son public. => MOD(savait, enjoué),
MOD(savait, volubile)*

(178) *Membre du CES, à Bruxelles, depuis 1986, ... François XXX a été élu président
=> MOD(élu, Membre), DEP(Membre, du),
OBJ(du, CES), DEP(Membre, à) ...*

Exemple construit où l'adjectif ne porte pas sur le sujet du verbe principal :

(179) *Mobiles, actifs et médiatiques, on les entend partout => MOD(entend, mobiles)*

12 Dépendants adverbiaux

12.1 Cas général

L'adverbe est en général noté mod. Sauf lorsqu'il est sous-catégorisé, auquel cas il reçoit la fonction p_obj, ou p_obj_loc, ou ats / ato.

- (180) *Paul dort beaucoup* => MOD(*dort, beaucoup*)
(181) *Paul va là-bas* => P_OBJ_LOC (*va, là-bas*)
(182) *Les photos sont bien* => ATS(*sont, bien*)
(183) *Paul se comporte bien* => P_OBJ(*comporte, bien*)

12.2 Superlatif

- (184) *Le plus grave cas* => det(*cas, le-1*), MOD(*cas, grave*), MOD(*grave, plus*)
(185) *Le cas le plus grave* => det(*cas, le-1*), MOD(*cas, grave*), det(*grave, le-3*), MOD(*grave, plus*)
(186) *Les cas les plus longuement débattus* => det(*cas, les-1*), MOD(*cas, débattus*), det(*débattus, les-3*), MOD(*débattus, longuement*), MOD(*longuement, plus*)

13 Références

ABEILLÉ A., TOUSSENEL F. & CHÉRADAME M., (2004). Corpus Le Monde, Annotation en constituants, guide pour les correcteurs. <http://www.llf.cnrs.fr/Gens/Abeille/guide-annot.pdf>

ABEILLÉ A. (2004). Guide des annotateurs : ANNOTATION FONCTIONNELLE <http://www.llf.cnrs.fr/Gens/Abeille/guide-fonctions.new.pdf>

ABEILLÉ A. & BARRIER N. (2004). Enriching a french treebank. In *LREC 2004*, Lisbon.

ABEILLÉ A., CLÉMENT L. & TOUSSENEL F. (2003). *Building a treebank for French*, In *Treebanks*. Kluwer : Dordrecht.

CARROLL J., BRISCOE T. & SANFILIPPO A. (1998). Parser evaluation : a survey and a new proposal. In *LREC 1998*, Granada.

CANDITO M., CRABBÉ B., DENIS P. & GUERIN F. (2009). Expériences d'analyse syntaxique statistique du français. In *TALN 2008*, p. 45–54, Avignon.

CRABBÉ B. & CANDITO M. (2008). Expériences d'analyse syntaxique statistique du français. In *TALN 2008*, p. 45–54, Avignon.

KAHANE S. (2001). Grammaires de dépendances formelles et théorie sens-texte. In *TALN 2001*, Tours, France.

KING T., CROUCH R., RIEZLER S., DALRYMPLE M. & KAPLAN R. (2003). The parc 700 dependency bank. In *EACL workshop on Linguistically Interpreted Corpora*, Budapest.

DE MARNEFFE M., MACCARTNEY B. & MANNING C. D. (2006). Generating typed

dependency parses from phrase structure parses. In *LREC 2006*.

PAROUBEK P., POUILLOT L.-G. , ROBBA I. & VILNAT A. (2005). Easy : campagne d'évaluation des analyseurs syntaxiques. In *Actes de TALN'05, Atelier EASy : campagne d'évaluation des analyseurs syntaxiques*, Dourdan.