Université Paris 7

UFR de Linguistique
TaLaNa

Organisation modulaire et paramétrable de grammaires électroniques lexicalisées

Application au français et à l'italien

Marie-Hélène Candito

Thèse de Doctorat
de Linguistique Théorique, Formelle et Automatique

soutenue le 26 janvier 1999

Directrice : Anne Abeillé

Jury : Roberto Basili
Laurence Danlos
Aravind Joshi
Philip Miller (Rapporteur)
Eric Wehrli (Rapporteur)

Remerciements

J'ai passé plusieurs années à préparer cette thèse (!). Ces années ont été riches, en rencontres, en discussions, sur la linguistique, l'informatique, et bien sûr, sur le lien entre les deux disciplines...

Je remercie énormément Anne Abeillé, ma directrice, pour son enthousiasme, ses compétences, sa rigueur. Son aide constante et son énergie ont été essentielles à mon travail. Merci beaucoup.

Je voudrais remercier chaleureusement Sylvain Kahane, pour avoir bien voulu collaborer avec moi. Une partie de cette thèse est le résultat d'un travail commun, sur la jonction entre les grammaires d'arbres adjoints et les grammaires de dépendances, et en particulier la Théorie Sens-Texte.

Je remercie sincèrement Laurence Danlos, pour la création et la direction de TaLaNa, laboratoire vivant, plein d'énergies diverses et variées : les membres de TaLaNa, que je suis heureuse de connaître.

Merci beaucoup également à Benoît Robichaud, pour avoir mis en place la procédure d'évaluation de la grammaire sur le corpus TSNLP, et à Bangalore Srinivas pour ses heures non comptées de travail, à mettre en place le système XTAG à TaLaNa.

Je remercie enfin les membres du jury, et tout particulièrement mes rapporteurs Philip Miller et Eric Wehrli, d'avoir bien voulu relire attentivement mon travail, dans des délais assez courts! Merci de votre patience.

Merci à Benjamin pour son soutien sans faille...

Sommaire

INTRODUCTION **1**

PARTIE 1 MOTIVATIONS PRATIQUES ET AMBITION THEORIQUE POUR LA METAGRAMMAIRE **5**

CHAPITRE 1 LES GRAMMAIRES LEXICALISEES D'ARBRES ADJOINTS : UNE INTRODUCTION **7**

1. RAPPELS SUR LE FORMALISME **7**

1.1 PROCEDURE DE DERIVATION ASSOCIEE AUX TAG 8
1.2 LES TRAITS ASSOCIES AUX ARBRES ELEMENTAIRES 12

2. TAG ET LINGUISTIQUE **15**

2.1 TAG EST UN FORMALISME MATHEMATIQUE, UTILE EN LINGUISTIQUE. 15
2.1.1 TAG est un formalisme mathématique contraint *15*
2.1.2 L'opération d'adjonction permet un domaine de localité étendu, et en particulier la lexicalisation *16*
2.2 LES PRINCIPES LINGUISTIQUES ASSOCIES AUX TAG 17
2.3 TAG ET QUELQUES THEORIES SYNTAXIQUES 20
2.4 TAG ET INTERPRETATION SEMANTIQUE 22
2.4.1 Une version sémantique du Principe de Cooccurrence Prédicat-Arguments *23*

3. L'ORGANISATION INTERNE D'UNE LTAG : MOTIVATIONS POUR UN NIVEAU SUPPLEMENTAIRE DE REPRESENTATION **27**

3.1 ORGANISATION PRATIQUE D'UNE LTAG : EXEMPLE DU SYSTEME XTAG 28
3.2 MOTIVATIONS POUR UN NIVEAU SUPPLEMENTAIRE DE DESCRIPTION 30
3.2.1 Solutions existantes pour la représentation compacte d'une LTAG *31*

CHAPITRE 2 LA METAGRAMMAIRE **33**

1. LA STRUCTURATION DES CONNAISSANCES LINGUISTIQUES DANS LA MG **34**

1.1 LA SOUS-CATEGORISATION EN TERMES DE FONCTIONS GRAMMATICALES 35
1.2 LA DISTINCTION DE FONCTIONS GRAMMATICALES INITIALES 38
1.2.1 Les redistributions des FG *39*
1.3 LES REALISATIONS DES FONCTIONS GRAMMATICALES 41
1.4 LES DIFFERENTS DOMAINES D'INFORMATION DE LA MG 41
1.5 JUSTIFICATION DES DOMAINES D'INFORMATION RETENUS 42
1.5.1 Pour la notion de sous-catégorisation initiale *45*
1.6 QUESTIONS LIEES A L'INTERFACE SYNTAXE/SEMANTIQUE 46
1.6.1 La correspondance entre arguments sémantiques et actants grammaticaux *46*
1.6.2 Le lemme principal n'est pas toujours la tête syntaxique *48*
1.6.3 Les actants grammaticaux non référentiels *49*
1.6.4 Un argument du lexème peut être réalisé comme un actant d'un autre lexème *50*
1.7 UNE FORMALISATION DU PCPA, EN TERMES FONCTIONNELS 52
1.8 DES LEXIQUES POSSIBLES POUR LA MG 52

2. L'APPAREIL FORMEL UTILISE **54**

2.1 UN GRAPHE D'HERITAGE MULTIPLE ET (PRESQUE) MONOTONE 54
2.1.1 Héritage : définition formelle et exemples d'utilisation en linguistique *54*
2.1.2 Héritage dans la MG *59*
2.2 DESCRIPTIONS PARTIELLES D'ARBRES 61

2.3 STRUCTURES DE TRAITS ASSOCIEES AUX CONSTANTES ... 62

3. ORGANISATION GENERALE DE LA MG ... 63

3.1 LA HIERARCHIE NOYAU ... 65
3.1.1 Dimension 1 : sous-catégorisation initiale ... 65
3.1.2 Dimension 2 : redistributions des fonctions ... 66
3.1.3 Dimension 3 : réalisations grammaticales des fonctions ... 67
3.2 DES CLASSES CROISEES POUR REPRESENTER DES SCHEMES ... 68
3.2.1 Procédure de génération de l'ensemble des classes croisées ... 69
3.2.2 Un exemple détaillé ... 69
3.3 LES CONTRAINTES DE COMPATIBILITE ... 72
3.4 COMPARAISON AVEC LES AUTRES TRAVAUX DE REPRESENTATION COMPACTE DE TAG ... 73

4. GENERATION DE SCHEMES ... 75

5. IMPLEMENTATION ... 76

5.1 FORMAT D'ENTREE DU COMPILATEUR ... 77
5.2 ALGORITHME DE GENERATION DE SCHEMES ... 77
5.3 RESTRICTION DE LA COUVERTURE DE LA TAG COMPILEE ... 79
5.4 LACUNES ET AMELIORATIONS ENVISAGEES ... 79
5.4.1 Ordre respectif des fonctions grammaticales ... 79
5.4.2 Composition automatique de redistributions ... 80
5.4.3 Permettre un plus grand partage entre fonctions grammaticales : définir une hiérarchie de fonctions grammaticales ... 81

PARTIE II METAGRAMMAIRES POUR LE FRANÇAIS ET L'ITALIEN ... 83

CHAPITRE 3 LA SOUS-CATEGORISATION INITIALE POUR LES VERBES ... 86

1. LES VERBES "DEFICIENTS" ... 86

1.1 LES AUXILIAIRES DE TEMPS ... 87
1.2 LES VERBES CAUSATIFS ... 89
1.3 LA COPULE ... 91
1.4 LES VERBES RESTRUCTURANTS EN ITALIEN ... 91
1.4.1 Les données ... 91
1.4.2 La sémantique de la restructuration ... 94
1.4.3 L'analyse en TAG ... 94
1.5 LES VERBES A MONTEE ... 96
1.5.1 L'analyse pour le français ... 96
1.5.2 L'analyse pour l'italien ... 98
1.5.3 Les verbes à montée de l'objet ... 98

2. LES VERBES PLEINS ... 100

2.1 LA PROJECTION D'UN VERBE : "L'ARTERE" DANS UN ARBRE A ANCRE VERBALE ... 100
2.1.1 La représentation de l'artère verbale dans la MGf et la MGi ... 103
2.1.2 La gestion du mode ... 104
2.2 L'HYPOTHESE INACCUSATIVE ... 106
2.2.1 L'inaccusativité, syntaxe ou sémantique? ... 107
2.2.2 Caractérisation des formes inaccusatives en italien ... 111
2.2.3 Le français n'a pas de propriétés d'inaccusativité de surface ... 119
2.3 LES SOUS-CATEGORISATIONS INITIALES PREVUES POUR LES VERBES PLEINS ... 122
2.3.1 Les fonctions grammaticales distinguées ... 123
2.3.2 Familles de schèmes et sous-catégorisations initiales ... 125
2.3.3 Liste des familles pour le français ... 126
2.3.4 Liste des familles pour l'italien ... 132

CHAPITRE 4 LES REDISTRIBUTIONS DES FONCTIONS, POUR LES VERBES....139

1. LES REDISTRIBUTIONS POUR LES VERBES EN FRANÇAIS....141

1.1 L'IMPERSONNEL....142
1.1.1 Les verbes et locutions impersonnels....143
1.1.2 La redistribution "impersonnel"....144
1.2 LE PASSIF....147
1.2.1 Interprétations du passif et morphologie passive....147
1.2.2 Prise en compte du passif impersonnel....148
1.2.3 Passif en de....150
1.2.4 Passif et actants phrastiques....150
1.2.5 Interactions avec les autres redistributions....151
1.2.6 Description des classes utilisées....151
1.3 LE MOYEN....154
1.3.1 Interactions avec les autres redistributions....156
1.3.2 Description des classes utilisées....156
1.4 LE CAUSATIF....157
1.4.1 Analyse des données....157
1.4.2 Traitement en TAG....164
1.4.3 Description des classes utilisées....165
1.4.4 Remarques et points faibles....166
1.5 LA "REFLEXIVATION"....167
1.5.1 Analyse des données....167
1.5.2 Description des classes utilisées....176
1.6 RECAPITULATIF....177

2. LES REDISTRIBUTIONS POUR LES VERBES EN ITALIEN....178

2.1 PROMOTION SPONTANEE DE L'OBJET D'UN INACCUSATIF....179
2.2 L'IMPERSONNEL....179
2.2.1 Ce qui s'appelle impersonnel en italien....179
2.2.2 Ce qui s'appelle impersonnel en français....182
2.2.3 Représentation dans la MG....183
2.3 LE PASSIF....184
2.3.1 L'équivalent du passif en de français....186
2.3.2 Interactions avec les autres redistributions....186
2.4 LE MOYEN....187
2.4.1 Contraintes et interactions avec les autres redistributions....188
2.5 LE CAUSATIF....191
2.5.1 Contraintes....193
2.5.2 Interaction avec les autres redistributions....194
2.6 LA REFLEXIVATION....195
2.6.1 Propriétés communes aux verbes réfléchis....195
2.6.2 Le vrai réfléchi est-il intransitivant en italien?....196
2.6.3 Contraintes et interaction avec les autres redistributions....197
2.6.4 Représentation dans la MG....198
2.7 RECAPITULATIF....198
2.7.1 Implémentation simplifiée....200

CHAPITRE 5 LA REALISATION DES FONCTIONS DEPENDANT D'UN VERBE...203

1. LES REALISATIONS DU SUJET....203

1.1 DIVERGENCES FRANÇAIS / ITALIEN....204
1.1.1 L'inversion du sujet....204
1.1.2 Différentes propriétés d'extraction en français et en italien....205
1.2 LE SUJET DANS LA MGF....206
1.2.1 Les participiales....207

1.2.2 Les sujets non réalisés ... *209*
1.2.3 Les sujets réalisés ... *210*
1.3 LE SUJET DANS LA MGI ... 211
1.3.1 Représentation des participiales adnominales ... *212*
1.3.2 Les propriétés du "sujet inacc+" ... *214*
1.4 L'ACCORD DE LA FORME VERBALE ... 214
1.4.1 Si le sujet est réalisé ... *215*
1.4.2 Si le sujet n'est pas réalisé (localement) ... *215*

2. LES REALISATIONS DE L'OBJET DIRECT ... 217

2.1 L'ACCORD DU PARTICIPE PASSE ... 219
2.1.1 Pour le français ... *220*
2.1.2 Pour l'italien ... *222*
2.2 AJOUTS ENVISAGES ... 223

3. L'ATTRIBUT ... 223

4. LES COMPLEMENTS PREPOSITIONNELS ... 224

4.1 A1-OBJET ET A2-OBJET ... 224
4.1.1 Représentation dans la MGf et la MGi ... *225*
4.2 GEN-OBJET ... 226
4.3 LES AUTRES COMPLEMENTS PREPOSITIONNELS : OBL-OBJET ET AGT-OBJET ... 228
4.4 LE LOCATIF ET LE LOCATIF DE PROVENANCE ... 229

5. LES REALISATIONS COMME CLITIQUES PRONOMINAUX ... 230

5.1 LES FAITS LINGUISTIQUES ... 231
5.1.1 Caractérisation des clitiques ... *231*
5.1.2 Les clitiques en syntaxe ou morphologie? ... *233*
5.2 L'ANALYSE DES CLITIQUES PRONOMINAUX COMME AFFIXES ... 234
5.2.1 Le système utilisé pour le français ... *235*
5.2.2 Le système utilisé pour l'italien ... *240*
5.2.3 Traitement des clitiques ajouts ... *245*
5.2.4 Analyse des clitiques et interprétation sémantique ... *245*
5.3 LES CLITIQUES PRONOMINAUX DANS LA MGF ET LA MGI ... 247
5.4 LES DIFFERENTS TRAITEMENTS DES CLITIQUES PROPOSES POUR LA TAGF ... 248
5.4.1 Correspondance systématique entre analyse en morphologie et analyse en syntaxe ... *250*
5.5 LA MONTEE DES CLITIQUES ET LE POUVOIR GENERATIF DE TAG ... 251
5.6 CAS PROBLEMATIQUES POUR UN TRAITEMENT EN SYNTAXE ... 253

6. LES REALISATIONS PHRASTIQUES ... 255

6.1 GENERALITES SUR LA REPRESENTATION DES ACTANTS PHRASTIQUES ... 255
6.1.1 Complémenteurs pleins et complémenteurs régis ... *255*
6.1.2 La représentation des extractions à distance, hors des complétives et infinitives ... *256*
6.1.3 Structures des complétives et des infinitives ... *259*
6.2 LES FONCTIONS A REALISATION SEULEMENT PHRASTIQUE ... 259
6.2.1 Infinitive de perception et infinitive de but ... *259*
6.3 LES FONCTIONS ALTERNANT ENTRE REALISATION NOMINALE ET PHRASTIQUE ... 262
6.3.1 Sujet phrastique ... *262*
6.3.2 Objet phrastique ... *265*
6.3.3 Gén-objet phrastique ... *266*
6.3.4 A2-objet phrastique ... *267*
6.3.5 Interrogatives indirectes ... *268*

7. LES POSITIONS EXTRAITES : RELATIVES, CLIVEES, QUESTIONS ... 270

7.1 LES RELATIVES ... 270
7.1.1 Les relatives en français ... *270*
7.1.2 Les relatives en italien ... *276*

7.2 LES CLIVEES 277
7.2.1 Les clivées en français 278
7.2.2 Les clivées en italien 279
7.3 LES ACTANTS GRAMMATICAUX "INTERROGES" 280
7.3.1 Pronoms, déterminants, et adverbes interrogatifs 280
7.3.2 Les arbres élémentaires pour les actants syntaxiques interrogés en français 282
7.3.3 Les arbres élémentaires pour les actants grammaticaux interrogés en italien 283
7.3.4 Représentation dans la MGf et la MGi 283

CHAPITRE 6 LA REPRESENTATION DES MODIFIEURS 285

1. LES DIFFERENTES SOUS-CATEGORISATIONS INITIALES DISTINGUEES 286

1.1 SOUS-CATEGORISATIONS INITIALES POUR LES PREPOSITIONS 287
1.2 SOUS-CATEGORISATIONS INITIALES POUR LES COMPLEMENTEURS (PLEINS) 288
1.3 SOUS-CATEGORISATIONS INITIALES POUR LES ADVERBES 289
1.3.1 Placement des adverbes 290
1.4 SOUS-CATEGORISATIONS INITIALES POUR LES ADJECTIFS 291
1.5 SOUS-CATEGORISATIONS INITIALES POUR LES NOMS (EMPLOIS DE MODIFIEUR UNIQUEMENT) 192

2. LES REDISTRIBUTIONS POUR LES MODIFIEURS 292

3. LES REALISATIONS DES FONCTIONS MODIFIEUR ET DEPENDANTS DE MODIFIEURS 295

3.1 LE MODIFIEUR 295
3.2 LES FONCTIONS DEPENDANT D'UN MODIFIEUR 297

PARTIE III PROLONGEMENTS 299

CHAPITRE 7 TAG ET L'INTERFACE SYNTAXE/SEMANTIQUE 301

1. DEPENDANCES ET TETES 302

1.1 DEPENDANCES SEMANTIQUES 303
1.2 DEPENDANCES SYNTAXIQUES ET TETE SYNTAXIQUE 303
1.3 DIVERGENCES ENTRE DEPENDANCES SYNTAXIQUES ET SEMANTIQUES 305
1.3.1 Modifieurs syntaxiques et modifieurs sémantiques 306
1.4 CARACTERISATION DES ARBRES ELEMENTAIRES SOUS L'ANGLE DE L'INTERFACE SEMANTIQUE/SYNTAXE 309
1.5 CRITERES DE TYPAGE DES NOEUDS ARGUMENTAUX EN NOEUDS PIED ET NOEUDS A SUBSTITUER 310

2. ARBRE DE DERIVATION ET GRAPHE DE DEPENDANCES SEMANTIQUES 313

2.1 PROCEDURE D'INTERPRETATION DE L'ARBRE DE DERIVATION 314
2.2 PROCEDURE DE TRADUCTION D'UN GRAPHE DE DEPENDANCES SEMANTIQUES EN UN ARBRE DE DERIVATION 317

3. CAS PROBLEMATIQUES 321

3.1 DIFFERENTES SEMANTIQUES POUR UNE SYNTAXE UNIFORME 321
3.2 QUELQUES CAS OU LE PCPA N'EST PAS RESPECTE 323
3.3 DIVERGENCES DANS L'INTERFACE SYNTAXE/SEMANTIQUE 324
3.3.1 Le PCPA ne garantit pas à lui seul une interprétation correcte de l'arbre de dérivation 327

4. APPLICATION A L'ANALYSE OU LA GENERATION 329

CHAPITRE 8 APPLICATIONS ET PERSPECTIVES 333

1. GRAMMAIRES DE REFERENCE DU FRANÇAIS ET DE L'ITALIEN 333

1.1 EVALUATION DE LA TAG DU FRANÇAIS, SUR LE CORPUS DE TEST TSNLP 333

2. APPLICATION EN GENERATION DE TEXTES : COMPILATION DE TAG POUR LE SYSTEME G-TAG **338**

3. SUPERTAGGING **339**

4. TRADUCTION AUTOMATIQUE **340**

CONCLUSION **345**

REFERENCES BIBLIOGRAPHIQUES **349**

ANNEXE A : GUIDE D'UTILISATION DU COMPILATEUR MG→TAG **363**

ANNEXE B : MG POUR UN FRAGMENT DE GRAMMAIRE ANGLAISE **371**

ANNEXE C : EXEMPLES D'ANALYSES POUR LE FRANÇAIS ET L'ITALIEN **381**

Conventions et notations utilisées

Notations linguistiques

*	Phrase inacceptable
?	Phrase à acceptabilité douteuse
Jean parle (E + Chiara)	Notation abrégée représentant les deux phrases *Jean parle* et *Jean parle à Chiara*
E	Element vide
forme	Forme fléchie
LEMME	Lemme
/LEXEME/	Lexème
'sémantème'	Sémantème

Lorsque nous étudions des phénomènes italiens, nous donnons immédiatement en dessous, en caractères plus petits la glose en français, et éventuellement si cela nous paraît nécessaire, la traduction en français, alors entre parenthèses.

(1) Ne sono stati mandati tanti al fronte.
NE-part ETRE-passé.composé.3mpl envoyés au front.
(Il en a été envoyé tant au front.)

Lorsque l'on étudie en parallèle phénomènes français et italiens, on adopte la convention que les exemples numérotés dans une langue, sont repris dans l'autre langue avec la même numérotation plus une apostrophe. On ne donne pas dans ce cas de glose ou de traduction en italien.

(2) a. Paul court à la boulangerie acheter du pain.
a'. Paul corre al panificio **a** comprare pane.
b. Paul entend Pierre chanter.
b'. Paul sente Pierre cantare.

Les abbréviations pour les gloses sont les suivantes: m masculin, f féminin, p pluriel, s singulier, 1 (resp. 2, 3) 1ère (resp. 2ème, 3ème) personne. Les clitiques sont classés en nom (nominatif), ACC (accusatif), DAT1 (datif en fr. *lui*, it. *gli*), DAT2 (datif en fr. *y*, it. *ci*), gén (génitif, fr. *en*, it. *ne*), part (partitif, fr. *en*, it. *ne*), loc (locatif, fr. *y*, it. *ci*), dloc (locatif de provenance, fr. *en*, it. *ne*).

Notations pour la métagrammaire

$A \Rightarrow B$	Contrainte de cooccurrence positive (A ne peut être héritée par une classe croisée que si B est également héritée)
$A \Rightarrow \neg B$	Contrainte de cooccurrence négative (A et B sont imcompatibles)
$\rightarrow$	Révision de la valeur d'un champ ou attribut

Notations pour les grammaires d'arbres adjoints

$\downarrow$	Signale un noeud à substituer
*	Signale un noeud pied
$\lozenge$	Signale un noeud dominant une ancre (lexicale)
ε	Chaîne vide

Introduction

Une grammaire électronique est un objet qui se situe à la charnière entre la complexité et la diversité linguistiques et les besoins de régularité et d'efficacité pour le traitement informatique.
Au départ indissociables des programmes qui les utilisaient, les grammaires électroniques ont acquis un statut à part entière, à la fois comme représentation garante de la cohérence d'analyses linguistiques et comme ensemble de données indépendant, permettant un traitement automatique de la langue. Il reste cependant que ces données organisées ne sont en général compatibles qu'avec un système de traitement précis. La réutilisation de grammaires par divers systèmes est quasiment inexistante à l'heure actuelle.
Dans le même temps, il existe désormais des modèles linguistiques relativement stabilisés, qui chacun réutilisent des notions comparables. L'utilisation de la notion de fonction ajoutée à la notion de constituant est un schéma que l'on retrouve dans plusieurs théories linguistiques importantes, telles LFG ou HPSG, qui ont recueilli les fruits d'un effort considérable de formalisation de phénomènes linguistiques.
Un fossé sépare cependant souvent ces formalisations des grammaires électroniques effectivement implémentées. Cela provient principalement de la diversité des architectures pour les programmes de traitement qui utilisent ces grammaires, ainsi que du fait que les formalisations linguistiques restent peu orientées vers une efficacité de traitement. Il reste fréquent qu'une grammaire électronique soit construite à partir de zéro, sans pouvoir emprunter à des implémentations déjà existantes.

Cette thèse a pour ambition de fournir un niveau de représentation, la métagrammaire, intermédiaire entre d'un côté une théorie syntaxique dans toute sa complexité et de l'autre une grammaire électronique utilisable par un système d'analyse ou de génération syntaxique. L'utilisation de ce niveau de description pour représenter la syntaxe (ou un fragment de syntaxe) française et italienne constitue un test et fournit également deux grammaires électroniques parallèles.

La métagrammaire que nous proposons, parce qu'elle s'appuie sur une représentation modulaire et hiérarchique des informations syntaxiques, peut être abordée à différents niveaux de généralité. Avec les métagrammaires instanciées pour le français et pour l'italien que nous décrivons dans cette thèse, nous espérons proposer une base de données syntaxiques pour les deux langues, à laquelle on peut accéder en choisissant le niveau de précision voulu. Les deux métagrammaires peuvent fournir un panorama des syntaxes des deux langues, ainsi que de leur contraste. Mais elles permettent également d'appréhender des phénomènes plus précis, en descendant à un point d'observation plus spécifique.

En outre, d'un point de vue pratique, la métagrammaire a été conçue pour pouvoir être compilée en une grammaire d'arbres adjoints lexicalisée (TAG), directement utilisable par un système opérationnel. Les métagrammaires du français et de l'italien sont compilables en deux TAG[1] dont le parallélisme peut être mis à profit pour des applications multilingues. Le système dans son ensemble allie les avantages d'une grammaire qui laisse la place à l'expression de généralisations, indispensables pour limiter la taille d'une grammaire et faciliter la maintenance, ainsi que les avantages d'une grammaire TAG.

Formalisme mathématique au départ, les TAG ont trouvé une application dans la représentation de connaissances linguistiques, en particulier en syntaxe, mais également en sémantique. Après les critiques formulées à l'encontre de la grammaire générative des années 70, ont émergé des théories linguistiques d'une nouvelle génération (dont LFG, GPSG, HPSG). Les principaux objectifs étaient de limiter la capacité générative du modèle, tant par principe théorique, qu'en vue d'un traitement informatique; de s'attacher à la représentation de phénomènes de surface, directement observables; de fournir une bonne articulation entre informations lexicales, syntaxiques et sémantique. En particulier, avec l'utilisation généralisée de structures de traits combinées par unification, les informations lexicales ont pu prendre une place prépondérante, après des décennies d'hégémonie de la syntaxe. La diversité des comportements des différents éléments lexicaux, mise en évidence par Gross (1975), impose l'enrichissement des informations venant du lexique, et impose que ces informations puissent se propager aux niveaux de description plus abstraits, ce aussi bien dans le cadre d'applications pratiques, que dans le cadre de la définition théorique d'un modèle de langue.
Le formalisme TAG, dans son application en linguistique computationnelle, permet de répondre à ces objectifs. Formellement, les théories citées ci-dessus s'appuient sur le formalisme mathématique de règles de réécriture, et l'augmentent par l'utilisation de structures de traits, combinées par unification. TAG constitue un appareil formel différent, ce qui lui confère certains avantages. Du point de vue formel, les TAG ont une capacité générative contrainte, entre les grammaires contextuelles et les grammaires hors-contextes (Joshi et al., 1991), ce qui permet la définition d'algorithmes d'analyse en temps polynomial. Du point de vue linguistique, elles permettent la définition d'unités de description larges, il offre la possibilité d'un rapport différent entre les composantes généralement considérées pour une grammaire computationnelle. En particulier, les TAG sont bien connues pour avoir la possibilité d'être intégralement lexicalisées (Schabes, 1990; Abeillé, 1991). TAG répond de fait de manière idéale à l'exigence de disposer d'un lexique riche en informations. La lexicalisation confère également à TAG des propriétés computationnelles évidentes, puisque la grammaire pour une phrase à analyser peut n'être constituée que de règles faisant intervenir un élément lexical effectivement présent dans la phrase (Schabes et al., 1988).

Cependant la propriété de lexicalisation intégrale de la grammaire ne s'accompagne pas d'une organisation des informations linguistiques. L'ajout d'un niveau supplémentaire de description, la métagrammaire, comble cette lacune.

[1] Plus précisément en deux ensembles de structures "prélexicalisées" qu'il reste à associer à un lexique pour fournir des structures lexicalisées.

Une autre lacune du modèle TAG est explorée dans cette thèse. Il s'agit d'une étude précise sur le statut linguistique exact de la structure représentant une dérivation en TAG. Cette structure est en effet souvent présentée comme ayant des avantages particuliers, et comme pouvant constituer une bonne base pour l'interprétation sémantique. Pourquoi? Dans quelle mesure cela est-il exact? Nous avons tenté de répondre à cette question, en collaboration avec Sylvain Kahane, et proposons une lecture précise de cette structure de dérivation, comme graphe de dépendances sémantiques.

La thèse est composée de trois parties. Dans la première, nous développons les motivations pratiques et l'ambition théorique de la métagrammaire : au Chapitre 1, nous introduisons le formalisme des TAG et son utilisation en linguistique computationnelle, ainsi que les motivations pour un niveau supplémentaire de description, et pour une étude plus approfondie de l'interface avec la sémantique. Au Chapitre 2, nous décrivons la métagrammaire, dans ses aspects théoriques et formels, ainsi que l'implémentation réalisée pour la compilation d'une métagrammaire en une TAG.
La seconde partie est consacrée à la description des métagrammaires instanciées pour le français et l'italien. La métagrammaire du français a été construite en partant d'une TAG du français existante (Abeillé, 1991), et la métagrammaire de l'italien a été écrite par contraste systématique avec le français. Les Chapitres 3, 4 et 5 sont consacrés à la représentation des structures projetées par un verbe, le Chapitre 6 traite de la représentation des modifieurs.
Dans une troisième partie, nous étudions des questions d'interface entre sémantique et syntaxe, et l'interprétation sémantique possible du résultat de la dérivation en TAG (Chapitre 7). Nous fournissons au Chapitre 8 une description des applications réalisées et de celles envisagées.

Partie I

La métagrammaire : Motivations pratiques et ambition théorique

Chapitre 1
Les grammaires lexicalisées d'arbres adjoints : état de l'art

Dans ce chapitre, nous commençons par rappeler les caractéristiques formelles du formalisme des TAG, puis nous développons pour quelles raisons et de quelle manière il est utilisé en linguistique et en linguistique computationnelle, en fournissant les principes linguistiques associés aux structures élémentaires d'une grammaire TAG. Enfin nous développons les motivations pour une organisation plus modulaire d'une grammaire TAG, ce que fournit la métagrammaire que nous proposons dans cette thèse.

1 Rappels sur le formalisme

A partir de la proposition formelle de Joshi et al. (1975), les grammaires d'arbres adjoints, dans leur application en linguistique computationnelle, ont évolué, avec la définition de diverses extensions du formalisme. On se référera aux "TAG standards" pour désigner le formalisme des TAG augmentées de structures de traits (Vijay-Shanker, 1987) et lexicalisées (Schabes et al., 1988; Schabes, 1990; Abeillé, 1991).

Etant donnés un vocabulaire N non terminal et un vocabulaire T terminal, une TAG[2] est un ensemble d'**arbres élémentaires**, qui sont des arbres tels que les noeuds non feuilles sont étiquetés par des non terminaux et les noeuds feuilles sont soit étiquetés par un terminal, soit étiquetés par un non terminal, et sont alors soit un **noeud à substituer** (repéré dans les figures par ↓) soit un **noeud pied** (repéré dans les figures par *).
Les arbres **initiaux** sont les arbres qui ne comportent pas de noeud pied. Les arbres **auxiliaires** sont les arbres qui en comportent exactement un. Les TAG standards ne permettent pas qu'un arbre élémentaire ait plus d'un noeud pied.
Les TAG utilisées en syntaxe reposent sur un vocabulaire non terminal correspondant à des catégories syntagmatiques, et un vocabulaire terminal correspondant à des éléments lexicaux. On écrira dorénavant qu'un noeud "est de catégorie X" pour un noeud étiqueté par un non terminal X[3].

[2] Pour désigner le formalisme des TAG, on emploiera parfois l'expression abrégée "les TAG". L'expression "une TAG" se réfère à un objet grammaire, qui suit la définition formelle fixée par le formalisme des TAG.
[3] Mais on peut imaginer une utilisation du formalisme des TAG autre. Par exemple une des applications des TAG synchrones (Shieber et Schabes, 1990, cf. infra Section 2.4) couple une TAG syntaxique avec une TAG sémantique, dont le vocabulaire non terminal est constitué de symboles pour un terme, une relation et une formule.

Une TAG est **lexicalisée**, si et seulement si tout arbre élémentaire comporte au moins un noeud feuille terminal.

1.1 Procédure de dérivation associée aux TAG

Le formalisme des TAG comprend une procédure de dérivation. Une **dérivation** lie une séquence de terminaux (dorénavant une phrase) à un **arbre dérivé**, dont les noeuds feuilles forment la phrase, et qui est obtenu par la combinaison d'arbres élémentaires, par les opérations d'**adjonction** et de **substitution**. Un arbre dérivé est auxiliaire s'il a un noeud pied.

L'opération de substitution, schématisée Figure 1[4], consiste à substituer à un noeud à substituer (dit **site** de la substitution) d'un arbre γ, un arbre α. Les arbres γ et α peuvent être élémentaires ou dérivés, mais α ne peut pas être auxiliaire (Schabes, 1990). Le noeud site de substitution et la racine de α doivent être de même catégorie.

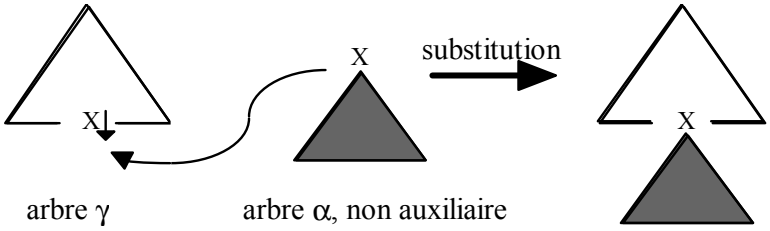


Figure 1. L'opération de substitution

L'opération d'adjonction consiste à insérer un arbre auxiliaire β, élémentaire ou dérivé, à un noeud *non feuille* d'un arbre γ, dit alors noeud **site** de l'adjonction. L'arbre β prend dans l'arbre dérivé la position du sous-arbre de γ constitué du noeud site et des ses fils directs ou indirects. Ce sous-arbre est dans l'arbre dérivé rattaché au noeud pied de β. Le noeud site, la racine de β et le noeud pied de β doivent être de même catégorie.

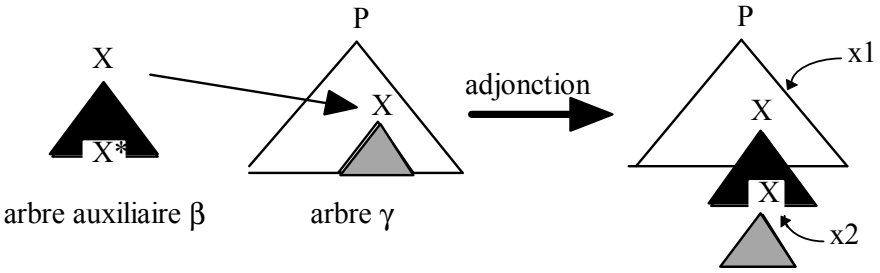


Figure 2. L'opération d'adjonction

Le noeud site d'adjonction est non feuille, donc ni un noeud à substituer, ni un noeud pied ne peuvent recevoir d'adjonction. Ces noeuds restent non adjoignables une fois

[4] Par convention, on utilise la lettre α pour désigner un arbre initial (élémentaire ou dérivé), la lettre β pour un arbre auxiliaire (élémentaire ou dérivé) et la lettre γ ou autre lorsque la distinction n'est pas pertinente. Dans les figures, les noeuds pertinents sont étiquetés par leur catégorie. Par la suite on pourra également utiliser des noms distincts pour différents noeuds de même catégorie, pour pouvoir les distinguer.

qu'ils sont "consommés" par substitution ou adjonction. Ainsi par exemple dans l'arbre de la Figure 1, le noeud de catégorie X de l'arbre dérivé est non adjoignable. Et dans l'arbre de la Figure 2, dans l'arbre dérivé, le noeud de catégorie X le plus bas (nommé x2), est lui aussi non adjoignable. Quant au noeud nommé x1 dans cette figure, on verra infra, que le fait d'y permettre ou pas l'adjonction correspond à deux définitions différentes de la dérivation.

Une dérivation est **complète** si tous les noeuds feuilles de l'arbre dérivé sont des terminaux.

Contrairement à ce qui est le cas pour la procédure de dérivation associée à des règles de réécriture hors-contexte, avec la dérivation TAG, l'arbre dérivé ne définit pas de manière unique une dérivation. C'est-à-dire que plusieurs dérivations distinctes peuvent aboutir au même arbre dérivé. Cela est illustré Figure 3 où un même arbre dérivé, représentant *Jean brise la glace*, peut être obtenu par deux dérivations, une pour l'interprétation littérale, où chaque mot de la phrase ancre un arbre différent, et une pour l'interprétation idiomatique, où un arbre élémentaire est ancré par *brise la glace* (cf. le traitement des idiomes par Abeillé (1991)).

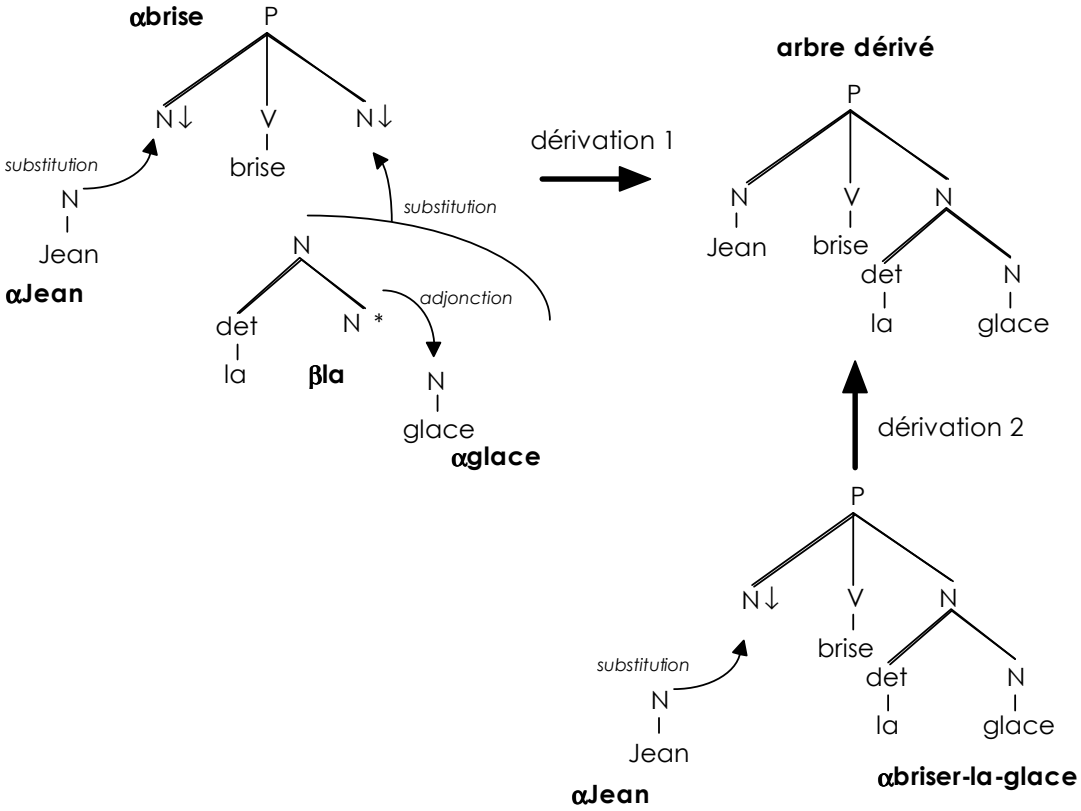


Figure 3. Un même arbre dérivé pour deux dérivations différentes

TAG comporte donc une structure supplémentaire pour représenter une dérivation, l'**arbre de dérivation**, dont les arcs représentent les opérations de combinaison définissant la dérivation représentée. Les noeuds de l'arbre de dérivation sont étiquetés par des arbres élémentaires. L'arbre de dérivation comporte deux types d'arcs, les arcs de

substitution (représentés graphiquement par un trait pointillé) et les arcs d'adjonction (représentés graphiquement par un trait plein). Pour représenter une opération de substitution (resp. d'adjonction) où un arbre γ1 est substitué (resp. adjoint) à un noeud site X d'un arbre γ2, on note un arc de substitution (resp. d'adjonction) entre γ2 et γ1, tel que γ2 domine γ1. L'arc est étiqueté de telle sorte que le noeud site X soit caractérisé. On a l'habitude de noter comme étiquette de l'arc, l'adresse, selon la convention de Gorn[5], du noeud site. Cela permet de s'affranchir de l'éventuel nommage des noeuds, qui peut varier lors d'une opération de combinaison.
On donne Figure 4 les arbres de dérivation pour les dérivations présentées Figure 3.

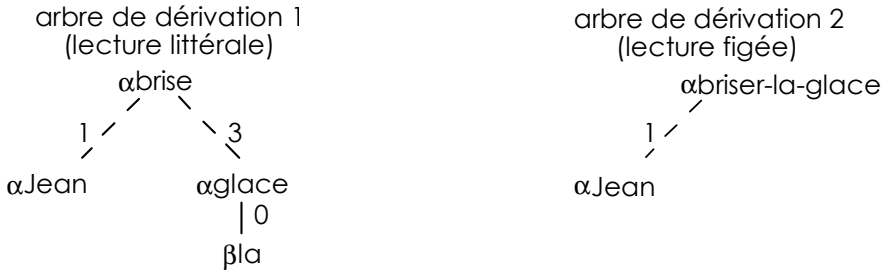


Figure 4. Les deux dérivations pour l'arbre dérivé de la Figure 3

Plusieurs définitions de la dérivation ont été données dans la littérature. Dans la définition de Vijay-Shanker (1987), les branches de l'arbre de dérivation ne sont pas ordonnées. Mais alors, pour qu'un arbre de dérivation définisse un seul arbre dérivé (l'inverse n'étant pas vrai), il faut ajouter une contrainte supplémentaire, qui est que deux adjonctions ne peuvent pas avoir lieu au même noeud d'un arbre élémentaire. Si l'on considère la Figure 2, cela signifie que dans l'arbre dérivé, le noeud de catégorie X nommé x1 (qui a l'adresse qu'avait le noeud site d'adjonction avant l'adjonction), ne peut pas recevoir une adjonction.

Shieber et Schabes (1994) ont proposé une définition alternative, mieux adapté à l'interprétation sémantique de l'arbre de dérivation (nous reprenons leurs motivations au Chapitre 7). Ils permettent que deux adjonctions aient lieu au même noeud[6], et pour que l'arbre de dérivation ne définisse qu'un seul arbre dérivé, ils imposent que les adjonctions ayant lieu au même noeud soit ordonnées.
On donne Figure 5 un exemple de dérivation selon cette définition. Dans l'arbre de dérivation les deux adjonctions, qui se font à la même adresse, sont ordonnées dans le temps (l'adjonction exprimée par la branche la plus à gauche se faisant en premier).

[5] Selon la convention de Gorn, les adresses des noeuds d'un arbre (à branches ordonnées) sont définies par récurrence. L'adresse de la racine est 0. Si l'adresse d'un noeud N est a, et si N' est un noeud fils de N, tel que N a y noeuds fils qui précédent N', alors l'adresse de N' vaut (10 x a) + y.
[6] Plus précisément les auteurs le permettent pour deux adjonctions d'arbres auxiliaires *modifieurs*, mais cette distinction est d'ordre linguistique (cf. infra Section 2.4), nous nous concentrons pour l'instant sur les aspects formels.

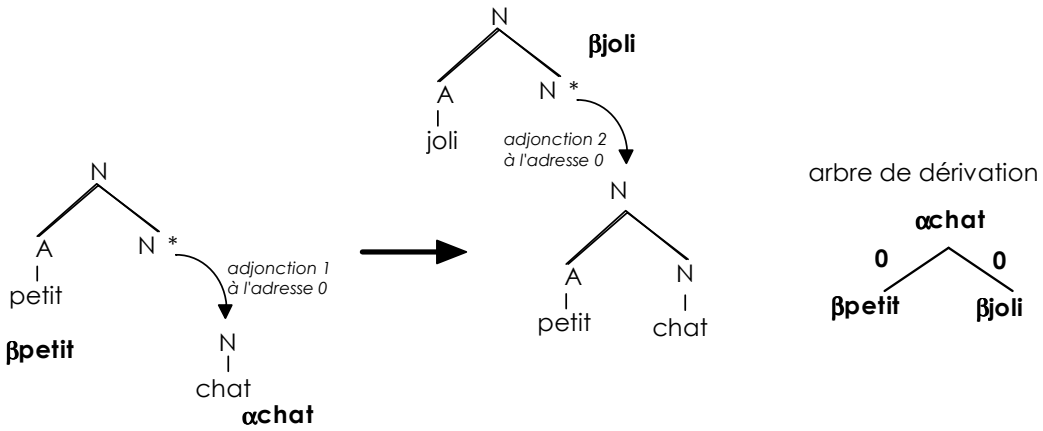


Figure 5. Dérivation selon Shieber et Schabes (1994)

La Figure 6 montre la dérivation selon la définition de Vijay-Shanker, pour obtenir le même arbre dérivé. Une des adjonctions (βjoli) se fait sur un arbre auxiliaire (βpetit) et l'arbre auxiliaire dérivé est adjoint sur l'arbre initial (αchat).

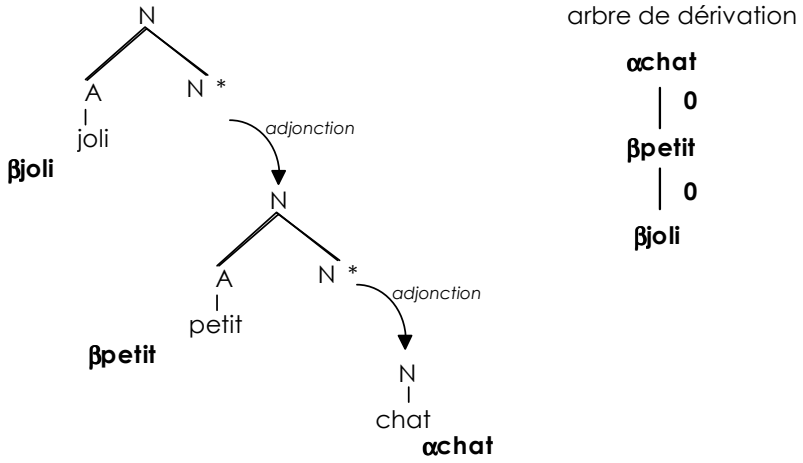


Figure 6. Dérivation selon Vijay-Shanker (1987)

En pratique, la définition de Vijay-Shanker est celle couramment utilisée pour les implémentations du formalisme des TAG (par exemple par l'analyseur intégré au système XTAG (Paroubek et al., 1992)). La définition de Shieber et Schabes n'a été implémentée[7] à notre connaissance que dans le système de génération G-TAG (Danlos et Meunier, 1996), basé sur les TAG[8].

[7] Shieber et Schabes fournissent cependant un algorithme de compilation des TAG avec leur définition de la dérivation en des grammaires linéaires indexées, et un algorithme de parse.

[8] Dans G-TAG, l'arbre de dérivation est une représentation qui a un double statut. Du point de vue linguistique, elle représente des relations concept-arguments, et du point de vue procédural, elle encode une dérivation. Il est donc normal que les auteurs aient pris la définition de la dérivation de Shieber et Schabes, avec laquelle l'arbre de dérivation est plus facilement interprétable comme représentation sémantique.

1.2 Les traits associés aux arbres élémentaires

Suite au développement de l'utilisation de structures de traits pour la linguistique computationnelle, Vijay-Shanker (1987) a défini les **TAG augmentées de structures de traits**. Tout noeud d'un arbre élémentaire est augmenté d'une structure de traits bipartite, les traits **amont** et les traits **aval**[9]. Les structures de traits sont une liste de traits de la forme :

- attribut = valeur, où valeur peut être soit atomique soit être elle-même une structure de traits
- chemin = variable, où chemin est une séquence d'attributs $:a_1:a_2:...:a_n$, tels que la valeur de l'attribut a_j est une structure de traits comportant l'attribut $:a_{j+1}$, et la variable doit apparaître dans une autre structure de traits du même arbre élémentaire. Il s'agit d'une équation de coindexation entre valeurs de deux chemins.

Les structures ne sont pas cycliques, c'est-à-dire qu'un attribut :a ne peut pas avoir pour valeur une structure de traits qui comprend à un certain niveau le même attribut :a.
Même si les traits à valeur non atomique sont utilisés en TAG, ils le sont de manière très réduite par rapport à d'autres théories utilisant l'unification (comme LFG ou HPSG) (cf. Vijay-Shanker, 1987; Abeillé, 1991).
Lors de la combinaison par adjonction ou substitution de deux arbres, la mise à jour des traits se fait de la façon suivante :

- En cas de substitution, les traits amont du noeud racine de l'arbre substitué sont unifiés avec les traits amont du noeud site de substitution.

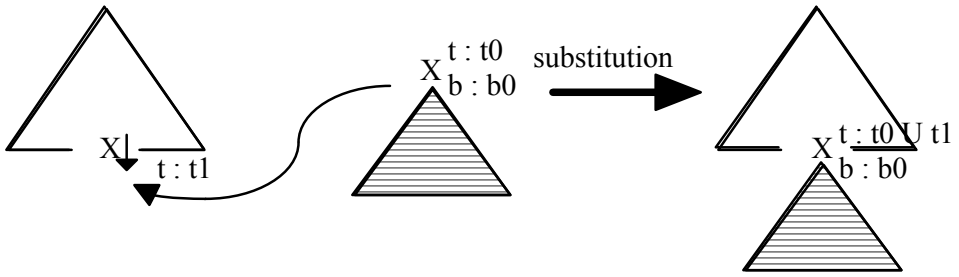


Figure 7. Mise à jour des traits en cas de substitution

- En cas d'adjonction, les traits amont du noeud racine de l'arbre auxiliaire sont unifiés avec les traits amont du noeud site d'adjonction, et les traits aval du noeud pied de l'arbre auxiliaire sont unifiés avec les traits aval du noeud site d'adjonction.

[9] ou traits *top* et *bottom* en anglais. Sur les figures, les traits amont sont précédés de t:, et les traits aval de b:.

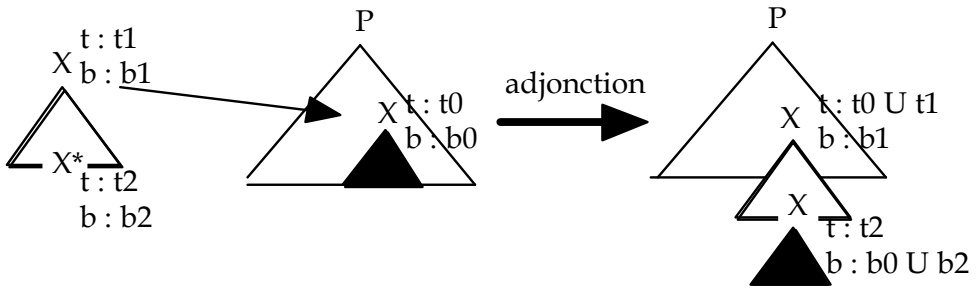


Figure 8. Mise à jour des traits dans le cas d'une adjonction

Si les unifications en jeu lors d'une substitution (respectivement d'une adjonction) échouent alors la substitution (respectivement l'adjonction) ne peut avoir lieu.
A la fin d'une dérivation complète, les structures de traits aval et amont de tous les noeuds sont unifiées, et chaque noeud reste associé à une seule structure de traits. Une dérivation complète n'est donc **valide** que si, à tous les noeuds, les structures de traits amont et aval sont unifiables.

Les structures de traits bipartites à un noeud tiennent compte du fait qu'une adjonction peut avoir lieu en ce noeud. Les traits amont concernent les relations du noeud avec les noeuds le dominant, et les traits aval concernent les relations du noeud avec les noeuds qu'il domine. En cas d'adjonction, ces deux types de relation sont séparées par l'arbre adjoint. Les noeuds à substituer, qui sont non adjoignables, n'ont qu'une structure de traits amont. Les noeuds pied, qui sont également non adjoignables, ont eux une structure de traits aval uniquement.

Vijay-Shanker (1992) propose de voir les arbres élémentaires augmentées de structures de traits bipartites comme des quasi-arbres, c'est-à-dire des structures qui ont les propriétés des arbres, sauf que certains de leurs noeuds — appelés quasi-noeuds pour marquer clairement qu'ils composent un quasi-arbre — ne sont pas liés par précédence linéaire ou dominance immédiate, mais par une relation de dominance au sens large. Cette relation de dominance signifie qu'il peut exister un chemin de longueur positive ou nulle entre les deux quasi-noeuds. La Figure 9 représente de façons schématique un quasi-arbre, avec deux quasi-noeuds, reliés par une relation de dominance (notée en pointillés). Le quasi-noeud dominant (resp. dominé) porte la structure de traits amont (resp. aval).

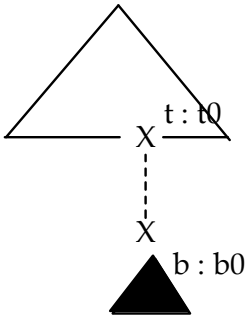


Figure 9. Un quasi-arbre

S'il n'y a pas adjonction, le chemin est de longueur nulle. Les structures de traits des deux quasi-noeuds doivent s'unifier. S'il y a adjonction, le chemin entre les deux quasi-noeuds est spécifié par le chemin entre racine et noeud pied de l'arbre adjoint. L'arbre adjoint peut lui-même comporter des liens de dominance large. Il s'agit d'une réinterprétation de l'opération d'adjonction comme spécification du chemin entre deux quasi-noeuds (Figure 10), qui a l'avantage de faire de l'adjonction une opération monotone : toutes les relations vraies de l'arbre recevant l'adjonction le sont aussi de l'arbre dérivé.

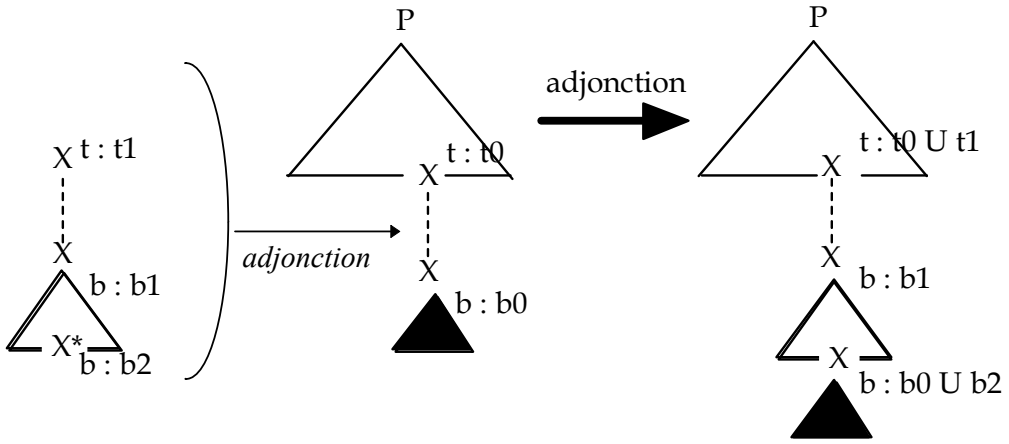


Figure 10. L'adjonction réinterprétée comme spécification d'un chemin

Dans la suite, nous gardons les notations classiques des arbres élémentaires (i.e. sans liens de dominance large), mais il est utile d'avoir cette interprétation de l'adjonction à l'esprit.

Contraintes d'adjonction

Dès l'écriture des premières grammaires TAG, il est apparu nécessaire de pouvoir contrôler quelles adjonctions peuvent se faire à un noeud donné.

L'adjonction d'un arbre auxiliaire β à un noeud A peut être impossible (i) parce que la racine et le noeud pied de β n'ont pas la même catégorie, (ii) parce que les unifications accompagnant l'adjonction échouent (iii) parce que l'adjonction rend incompatible les traits aval et amont à un noeud, sans qu'aucun arbre auxiliaire de la grammaire ne permette de résoudre le conflit en s'adjoignant à ce noeud.

On peut en outre vouloir spécifier qu'un noeud donné A ne doit jamais recevoir d'adjonction. Il est difficile d'exprimer ce genre de contrainte en jouant sur la valeur de traits amont et aval. On utilise alors explicitement une **contrainte de non-adjonction** (notée graphiquement par un "NA" à côté du noeud en question).

Inversement, une adjonction peut être obligatoire à un noeud du fait que ses structures de traits amont et aval ne s'unifient pas. On parle de **contrainte d'adjonction obligatoire**. Il faut pour rendre la dérivation valide que s'adjoigne à ce noeud un arbre auxiliaire, élémentaire ou dérivé, de telle sorte qu'après les unifications ayant lieu lors de l'adjonction, le conflit soit résolu.

Une contrainte d'adjonction obligatoire peut ne pas être localisée sur un noeud précis, mais sur un ensemble de noeuds dont la valeur d'un attribut est coïndicée par une même

variable, mais dont plusieurs valeurs distinctes sont spécifiées pour cet attribut. Par exemple Figure 11, tel quel l'arbre ne peut pas constituer un arbre dérivé final, puisqu'à l'unification des traits aval et amont à chaque noeud, un conflit apparaît entre val1 et val2. Une adjonction en X2, X1 ou X0 peut résoudre le conflit sur l'attribut :attr.

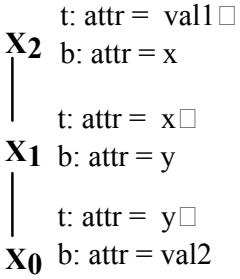


Figure 11. Contrainte d'adjonction sur un ensemble de noeuds

Enfin, il faut remarquer que les contraintes d'adjonction obligatoire peuvent apparaître dynamiquement lors d'une dérivation.

2 TAG et linguistique

2.1 TAG est un formalisme mathématique, utile en linguistique.

Le formalisme mathématique des TAG n'est pas associé à une théorie linguistique unifiée. Mais TAG a été reconnu utile en linguistique, du fait de différentes propriétés formelles (cf. entre autres Kroch et Joshi, 1985; Abeillé, 1991; Rambow, 1994), qui sont en particulier le domaine de localité étendu, la localisation des dépendances, la lexicalisation. Nous détaillons ces notions ci-dessous.

2.1.1 TAG est un formalisme mathématique contraint

Les grammaires à règles de réécriture hors-contexte (ci-après CFG, pour *context-free grammars*) sont telles quelles un formalisme mathématique trop contraint, dans la mesure où certains phénomènes de langue naturelle dépassent les langages hors contextes. D'autre part, sans l'augmentation des CFG par des traits, le pouvoir descriptif est faible et ne permet pas l'expression des généralisations adéquates. Les théories linguistiques utilisant les CFG, les ont augmentées de diverses manières (avec l'ajout de transformations pour la grammaire transformationnelle, d'une structure fonctionnelle pour LFG, de structures de traits complexes pour HPSG, plus dans ces deux derniers cas, par des systèmes de règles lexicales).
Ces augmentations font sortir ces théories du cadre strict des langages hors contextes. La classe des langages générés par LFG est incluse dans celle des langages contextuels (Kaplan et Bresnan, 1982). Pour HPSG, Carpenter (1991) montre que les règles

lexicales pourraient si elles sont non contraintes faire que HPSG simule une machine de Turing.

Utiliser le formalisme des TAG constitue une démarche différente, qui est non pas d'augmenter les CFG, mais de remplacer les CFG (par les TAG justement). Le formalisme des TAG permet d'engendrer des langages de la classe des langages légèrement contextuels (en anglais *mildly context-sensitive*, (Joshi et al. 1991)). Il s'agit d'une classe de langages strictement incluse dans la classe des langages contextuels, et qui inclut strictement la classe des langages hors contexte. On peut donc traiter d'une part certaines constructions qui sortent du cadre des langages contextuels, sans inclure pour autant tous les langages contextuels. L'avantage des langages légèrement contextuels est de pouvoir être reconnus en un temps polynomial et non pas exponentiel.

2.1.2 L'opération d'adjonction permet un domaine de localité étendu, et en particulier la lexicalisation

Les TAG offrent la possibilité de disposer de règles (les arbres élémentaires) qui sont des unités linguistiques plus larges que des règles de réécriture. Une règle de réécriture est en effet un arbre de profondeur 1, alors qu'un arbre élémentaire en TAG est de profondeur quelconque, supérieure ou égale à 1.
On dit que les TAG définissent un **domaine de localité étendu**. C'est-à-dire que certaines informations mettant en relation deux éléments linguistiques (lexicaux ou pas) qui dans des formalismes basés sur les CFG sont non locales, c'est-à-dire concernent des éléments qui ne sont pas forcément exprimés ensemble dans une même règle, seront locales en TAG. Par exemple, l'accord du participe avec l'objet extrait en français ne peut pas être local en CFG, c'est-à-dire qu'on ne peut pas localiser dans la même règle de réécriture l'objet extrait et le participe (à moins de perdre toute généralisation, en écrivant des règles très plates, et en fixant a priori la distance possible entre l'élément extrait et le participe). En TAG, on peut au contraire définir une analyse où le participe et le noeud syntagmatique représentant l'objet extrait sont localisés dans le même arbre élémentaire, l'accord est alors spécifié localement (cf. Chapitre 5, la section sur l'objet et sur le traitement de l'accord du participe en français et en italien).
Ce domaine de localité étendu est permis par l'opération d'adjonction[10]. En effet, pour pouvoir localiser dans la même règle l'objet extrait et son verbe, étant donné qu'il peut exister une distance "non bornée" entre les deux, il faut disposer d'une opération de type adjonction qui puisse éventuellement venir rompre la localité existante au niveau des arbres élémentaires.

En particulier, l'adjonction permet la propriété de lexicalisation intégrale de la grammaire, en gardant une liberté pour le choix des éléments lexicaux ancrant les arbres élémentaires (cf. Schabes, 1990; Abeillé, 1991). Pour que des structures élémentaires qui représentent par exemple une phrase soient lexicalisées, sans avoir à prévoir les positions pour des modifieurs, il faut disposer d'une opération pour ajouter les modifieurs à cette structure.

[10] Alternativement, avec l'interprétation par Vijay-Shanker des arbres comme des quasi-arbres, le domaine de localité étendu est permis pas la sous-spécification de chemins entre noeuds.

Si TAG n'est pas en soi une théorie linguistique, des éléments de théorie linguistique sont cependant classiquement associés aux TAG, qui mettent à profit les propriétés formelles originales de TAG. Nous rappelons ces principes ci-dessous.

2.2 Les principes linguistiques associés aux TAG

Par principes linguistiques associés aux TAG, on entend les principes qui ont été énoncés par différents auteurs (principalement Abeillé, Frank, Joshi, Kroch, Schabes), principalement dans le cadre de la construction de TAG pour l'anglais à l'Université de Pennsylvanie, par le groupe XTAG (XTAG group, 1998), construction initiée il y a une dizaine d'année; et pour le français, initialement au sein du groupe XTAG, puis à l'université Paris 7 (Abeillé, 1991; Abeillé et al. 1996c; Abeillé et Candito, à paraître). Cette thèse s'inscrit en partie dans le cadre de l'augmentation de la TAG du français.
Ces principes linguistiques sont :

- **Lexicalisation** : tout arbre élémentaire comporte au moins un noeud feuille lexical (Schabes, 1990; Abeillé, 1991).
  On appelle **ancre** tout noeud feuille lexical (c'est-à-dire, dans la définition formelle des TAG, tout noeud feuille dont la catégorie appartenant au vocabulaire terminal).
  Avec l'importance croissante de la place réservée au lexique dans les théories linguistiques depuis les années 80, TAG s'est révélé être un formalisme particulièrement intéressant, en cela qu'il permet la lexicalisation intégrale de la grammaire, ce qui n'est possible qu'avec une opération de type adjonction (c'est-à-dire qui permette l'insertion de matériel à un noeud non feuille d'un arbre syntagmatique).
  Schabes et al. (1988) mettent en évidence les avantages linguistiques et computationnels de la lexicalisation. La grammaire d'une phrase, au moment de l'analyse, peut être réduite aux seuls arbres élémentaires qui ne comportent comme ancres que des éléments lexicaux qui apparaissent dans la phrase. Sur le plan linguistique, la lexicalisation permet une souplesse dans la description, qui s'affranchit (en partie) de la trop grande généralisation apportée par la catégorisation des éléments lexicaux. Ainsi par exemple, *matin* et *maison* sont des noms, qui n'ont pourtant pas la même distribution. Pour représenter l'emploi de *matin* comme modifieur (comme dans *Ce matin, je me suis levée tôt*), en TAG il faut définir un arbre auxiliaire, ancré par *matin*. Les noms qui n'entrent pas dans cette construction, comme *maison*, n'ancreront tout simplement pas d'arbres de ce type.

- **Principe de cooccurrence prédicat-arguments (PCPA)** (formulation classique): tout arbre élémentaire ancré par un prédicat comporte au moins un noeud pour chacun des arguments que le prédicat sous-catégorise (Kroch et Joshi, 1985; Abeillé, 1991).
  Frank (1992) exprime un équivalent, dans le cadre de l'utilisation des TAG comme instanciation d'une grammaire GB, sous la forme de la *Condition on Elementary Tree Minimality*, qui prévoit qu'un arbre élémentaire corresponde à la projection étendue d'une tête lexicale. La formulation de Frank s'attache cependant à fixer une limite supérieure à un arbre élémentaire.

L'interprétation exacte de ce principe repose sur la définition de sous-catégorisation. Nous reviendrons sur ce principe infra Section 2.4, où nous proposons une reformulation, plus précise, qui permet de rendre compte de manière uniforme de tous les arbres élémentaires.
Mais on peut déjà citer le cas "emblématique" en TAG d'application de ce principe, qui est sans doute celui des arbres élémentaires utilisés pour représenter l'extraction. En effet, un élément extrait sous-catégorisé doit, d'après ce principe, être localisé sous la forme d'une position à étendre dans l'arbre de son prédicat. Etant donnée que la distance entre l'élément extrait et son prédicat n'est pas bornée (dans une grammaire de compétence), cette localisation n'est possible que du fait de l'existence de l'adjonction. Par exemple pour représenter *le livre que Pierre pense que Paul a écrit*, un arbre auxiliaire pour *Pierre pense que* est adjoint dans celui pour *le livre que Paul a écrit*.

- **Principe de minimalité sémantique (PMS)** (Abeillé, 1991)[11] : l'ensemble des ancres de tout arbre élémentaire correspond à exactement une unité sémantique non vide.
  Afin d'être plus précis, nous introduisons certains termes.
  On appelle **lemme** une abstraction sur un ensemble de formes fléchies, qui différent par leur flexion. On appelle **lexème** un ensemble de lemmes, tel que la combinaison de ces lemmes forme une unité sémantique non vide, telle que l'unité sémantique n'est pas décomposable en sous-parties dont chacune est exprimée par une sous-partie de l'ensemble des lemmes du lexème.
  Un **lexème fléchi** est constitué d'une forme fléchie pour chaque lemme du lexème.
  On appelle **sémantème**, toute unité sémantique qui est le signifié d'un lexème. C'est le noyau de sens commun à tous les lexèmes fléchis associés à ce lexème. La notion de sémantème étant associée à celle de lexème, les sémantèmes sont dépendants de la langue. De plus, un lexème étant composé de lemmes et non pas de morphèmes, la notion de sémantème est dépendante du découpage en mots adopté (c'est-à-dire des limites de la morphologie fixées).[12]

  Le principe de minimalité sémantique peut être reformulé de la manière suivante :

  L'ensemble des ancres d'un arbre élémentaire forme un lexème fléchi.

  L'application de ce principe a plusieurs conséquences. D'une part, les expressions à sens non compositionnel (les expressions figées) sont représentées par des arbres élémentaires à ancres multiples (Abeillé et Schabes, 1989; Abeillé, 1991).
  D'autre part, les mots fonctionnels (les prépositions régies, les complémenteurs régis), qui sémantiquement ne sont pas autonomes[13], n'ancrent pas à eux seuls un arbre élémentaire, une ancre sémantiquement pleine doit les accompagner.

[11] Ce principe est le résultat direct de la proposition de Abeillé (1991) de traiter les constructions à verbes support et les expressions figées par des arbres élémentaires à ancres multiples. Cependant la formulation du principe a varié. Celle que nous donnons correspond aux principes de "consistance sémantique" et de "non compositionnalité" de Abeillé (1993). Nous préférons en donner une formulation unique.
[12] On adopte les notations suivantes : *forme fléchie*, LEMME, /LEXEME/, 'sémantème'.
[13] Ces mots sont souvent considérés comme sémantiquement vides, mais on peut tout de même remarquer que deux lexèmes (à sens différents) peuvent ne différer que par ces mots-là (/PARLER-A/, /PARLER-DE/). Sans les considérer comme totalement vides de sens, on peut se limiter à les déclarer non autonomes, et à sémantique non régulière (Abeillé, 1991).

Les combinaisons verbe support + nom prédicatif sont également représentées par des arbres élémentaires comprenant le verbe support et le nom.
Un lexème pour une expression figée peut comprendre des lemmes qui n'ont pas toute la variation morphologique possible, suivant le degré de figement de l'expression. Des traits peuvent être ajoutés aux lemmes, de manière à contraindre leur réalisation, en nombre par exemple. De plus, l'expression figée suit les règles d'accord générales (entre sujet-verbe, déterminant-nom...), ce qui est représenté de manière régulière dans les arbres élémentaires associés. Ainsi par exemple pour *briser la glace*, le lexème est /BRISER-LE-GLACE(sing)/. L'accord du déterminant est géré dans les arbres élémentaires associés. On donne Figure 12 l'arbre élémentaire pour cette expression, dans le cas actif, avec sujet en position canonique. Par simplification, lorsqu'un la forme d'un lemme est complètement déterminée, on donne la forme fléchie directement, plutôt que lemme + traits. On donne également un exemple d'arbre pour le lexème avec verbe support /POSER-QUESTION/. Le déterminant pour le nom prédicatif étant libre (cf. *poser une/la/cette/des/quelle question(s)*), il n'apparaît pas dans l'arbre élémentaire, mais est traité comme les autres déterminants[14] par adjonction.

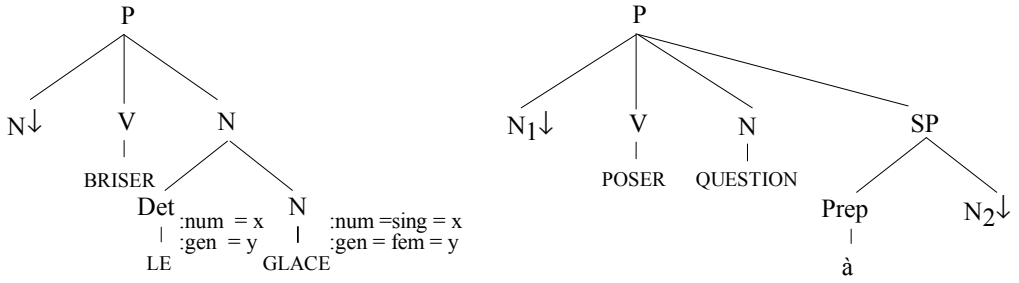


Figure 12. Arbres élémentaires pour une expression figée et pour un nom et son verbe support

Parmi les différentes ancres, on distingue l'**ancre principale**, qui est celle dominante en syntaxe (mais pas forcément en sémantique), c'est-à-dire le verbe dans le cas d'expressions figées en comprenant un, encore le verbe dans le cas de combinaisons verbe support + nom prédicatif. Les autres ancres sont appelées **co-têtes**.
Parmi les co-têtes, il faut cependant faire la distinction entre les mots fonctionnels et les autres (noms ou adjectifs dans une expression figée, ou une construction à verbe support). L'apparition des mots fonctionnels dépend en effet de la construction syntaxique, par exemple la préposition régie *à* n'est pas présente en cas de cliticisation.

Comme derniers principes linguistiques énoncés dans le cadre de TAG, on peut citer ceux concernant les traits. Peu de travaux linguistiques en TAG concernent les traits, mais on peut trouver une tentative d'organisation systématique de la propagation des traits dans (Abeillé, 1993). Abeillé reprend les différents principes de propagation de GPSG et adapte leur propagation au cas de TAG, spécifique du fait de l'adjonction

[14] On verra cependant infra que si l'on adopte un principe strict d'interface avec la sémantique, les déterminants qui correspondent à un argument (les possessifs dans certains emplois) doivent être substitués.

pouvant intervenir dans la propagation d'un trait le long d'une séquence de noeuds en relation de dominance. L'idée générale est que la structure de traits amont d'un noeud concerne les relations qu'il entretient avec les noeuds qu'il ne domine pas et que la structure de traits aval concerne les relations que le noeud entretient avec les noeuds qu'il domine.
Les traits décorant les arbres élémentaires sont qualifiés de traits de tête[15], traits d'accord et traits de pied (un trait de tête pouvant être un trait d'accord également).

- Les équations sur les traits d'accord : elles coindicent les traits d'accord des parties amont des noeuds concernés.
- Les équations sur les traits de tête : le long de la chaîne de noeuds que projette la tête syntaxique, les équations de tête coindicent les traits de tête amont d'un noeud avec les traits de tête aval du noeud immédiatement dominant.
- Les équations sur les traits de pied : elles coindicent un trait de pied amont du noeud pertinent pour le trait de pied avec ce même trait de pied aval du noeud immédiatement dominant.

En pratique, les équations sur les traits de tête et d'accord rentrent assez bien dans la définition donnée ci-dessus, mais pour les traits de pied, il y a plus d'idiosyncrasies.

2.3 TAG et quelques théories syntaxiques

En dehors des principes linguistiques énoncés, plusieurs travaux, implémentés ou pas, portent sur l'instanciation en TAG de principes linguistiques énoncés dans d'autres théories linguistiques.
Certaines des contraintes d'îlots distinguées en grammaire générative ont trouvé une représentation dite "naturelle" en TAG, c'est-à-dire qui fait appel à des contraintes du formalisme lui-même, plutôt qu'à des principes stipulés en plus par la théorie linguistique (Kroch, 1987, 1989). Rambow et Vijay-Shanker (1998) étudient quelles sont les caractéristiques de TAG sur lesquelles reposent la représentation de contraintes d'îlots. Pour la plupart, c'est le domaine de localité étendu qui permet de bloquer certaines extractions, simplement par la non existence de certaines topologies d'arbres élémentaires. Par exemple pour bloquer (1) (on note les traces uniquement pour faciliter la lecture), il suffit qu'il n'existe pas d'arbres élémentaires pour une relative qui extraient un autre complément que le complément relativisé (cf. l'arbre de la Figure 13, qui pourrait s'adjoindre sur *homme*, le tout se substituant dans l'arbre de *vu*, en position objet).
(1) * J'ai vu à qui$_i$ l'homme qui parlait e$_i$.

[15] Il s'agit da la tête "morpho-syntaxique", c'est-à-dire l'élément qui transmet les propriétés morphosyntaxiques de cétagorie, genre, nombre, temps, personne, etc... au constituant qui la contient (cf. Zwicky, 1985, pour une étude des différentes notions qui ont été qualifiées de tête en syntaxe). Nous prenons cette notion-là comme notion de tête syntaxique (cf. infra).

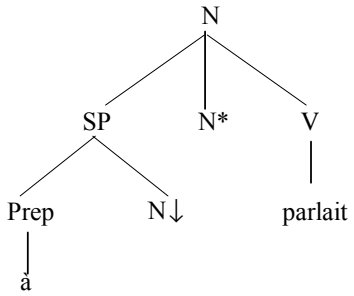


Figure 13. Un arbre élémentaire interdit, qui permettrait de dériver (1)

Un autre moyen "classique" en TAG pour bloquer l'extraction est de jouer sur la distribution des noeuds pied. Par exemple, la contrainte de l'îlot adverbial est représentée par le fait que le complément de l'adverbe (au sens large d'ajout) est toujours à substituer, l'adverbe comportant déjà un noeud pied pour s'adjoindre sur l'élément qu'il modifie. On ne peut donc avoir extraction hors du complément de l'adverbe.
Par exemple, en (2a), l'arbre ancré par *quand* s'adjoint sur la principale *Paul est parti*, son complément phrastique (*le baron demandait Isabelle en mariage*) est donc substitué. On voit que l'on ne peut pas obtenir (2b) : en substituant *qui le baron demandait en mariage*, le *qui* ne peut pas être antéposé à la principale (Abeillé, 1991).

(2) a. Paul est parti quand le baron demandait Isabelle en mariage.
b. *Qui Paul est-il parti quand le baron demandait en mariage?

Mais on peut remarquer que formellement, rien n'empêche d'inverser les rôles de la principale et du complément de *quand*. On pourrait formellement, même en gardant le PCPA, inverser les rôles de la principale et de la subordonnée pour représenter *quand* : il s'agirait d'un arbre pour *quand* avec la principale à substituer (*Paul est parti*), et qui s'adjoigne sur *qui le baron demandait en mariage*. L'ordre des mots de (2b) peut alors être obtenu. On voit donc, que même si le formalisme peut aider à représenter des contraintes (par exemple, par la distribution des noeuds pied et à substituer), c'est la théorie linguistique qui doit contraindre de toutes façons la topologie des arbres élémentaires, et le choix des noeuds pied et à substituer. La différence en TAG est qu'en disposant d'un domaine de localité étendu, on a un contrôle plus aisé sur la forme des structures que l'on veut interdire. Mais par exemple Abeillé (1991) indique également que certaines contraintes d'îlots doivent de toutes façons être représentées en TAG par des traits, ce qu'elle considère être extérieur au formalisme des TAG proprement dit.

TAG a été également considéré comme une implémentation possible d'une grammaire HPSG. Plus exactement Kasper et al. (1995) ont proposé une procédure de compilation de HPSG en TAG. A partir d'une grammaire HPSG, avec les principes propres à HPSG, une TAG est compilée, qui est utilisable par les analyseurs TAG existants. La compilation construit des arbres élémentaires[16] à partir de la description d'un élément lexical en HPSG. L'idée principale sous-jacente à la compilation est de construire des arbres élémentaires par instanciation de plusieurs schémas de dominance immédiate de HPSG. Par exemple, pour un verbe, les schémas de DI qui consomment les traits de

[16] Plus précisément, la compilation n'aboutit pas à des TAG standards, mais à des quasi-arbres (cf. supra, la redéfinition d'une TAG comme un ensemble de qusi-arbres, par Vijay-Shanker, 1987).

valence peuvent être appliqués, pour construire un arbre élémentaire de projection phrastique.

On peut également citer les travaux de Vijay-Shanker et Joshi (1989) qui proposent de définir une grammaire LFG, mais dont la composante syntagmatique (la structure-c) ne serait pas basée sur des règles de réécriture mais sur une grammaire TAG. L'avantage d'utiliser TAG est d'une part que l'on limite la propagation de traits (on peut se passer des équations $\uparrow = \downarrow$, étant donné le domaine de localité étendu). D'autre part le principe de complétude/unicité/cohérence est représentable en TAG par la topologie des arbres élémentaires (ils doivent comprendre les positions à étendre appropriées) et par le fait que les noeuds à substituer ou pied sont consommés une seule fois. Donc la bonne formation des dépendances fonctionnelles peut être séparée de la procédure de dérivation, ce qui n'est pas le cas en LFG.
Les auteurs étudient plus particulièrement le principe d'incertitude fonctionnelle de LFG, et arguent qu'il s'agit d'un corollaire si une TAG est utilisée pour le composant syntagmatique. C'est-à-dire que le formalisme des TAG permet de contrôler le chemin entre un élément extrait et son prédicat. Mais certains emplois de l'incertitude fonctionnelle pour l'allemand notamment ne peuvent pas être modélisées en TAG (O. Rambow, communication personnelle). D'autre part certaines analyses en LFG ne peuvent pas être transposées telles quelles en TAG. Par exemple le traitement du contrôle en LFG, ou un actant grammatical peut consommer un élément de plusieurs listes de sous-catégorisation n'est pas transposable directement en TAG.

Enfin les TAG et les grammaires de dépendances ont été également comparées. Il s'agit d'une comparaison entre les représentations linguistiques en grammaire de dépendances (les arbres de dépendances) et l'arbre de dérivation TAG. On aborde cette question à la section suivante, ainsi qu'au Chapitre 7.

2.4 TAG et interprétation sémantique

Deux démarches sont possibles pour ce qui est de faire dépasser les limites de la seule syntaxe aux TAG.
La première consiste à définir un module d'analyse sémantique, intervenant sur le résultat de l'analyse syntaxique. Dans le cas de TAG, on dispose à la fois de l'arbre dérivé et de l'arbre de dérivation. Or on s'accorde généralement à dire que l'arbre de dérivation peut être vu comme une représentation linguistique relativement proche de la sémantique, du fait que les noeuds de l'arbre de dérivation sont des arbres élémentaires qui correspondent à une unité sémantique (d'après le PMS), et que les arbres élémentaires localisent les relations prédicat-arguments (d'après le PCPA), relations qui se traduisent par autant d'arcs dans l'arbre de dérivation. La représentation linguistique que constitue un arbre de dérivation s'affranchit de la syntaxe de surface, et fournit une bonne base pour une analyse sémantique.
Mais, l'arbre de dérivation tel quel ne peut pas fournir les informations qui concernent la portée des quantificateurs et des modifieurs. Une deuxième démarche pour l'analyse (ou la génération) sémantique a été proposée avec les TAG synchrones (Shieber et Schabes,

1990; Shieber, 1994; Rambow et Satta, 1996[17]). Les TAG synchrones permettent de définir des correspondances entre les structures engendrées par plusieurs TAG. Une application possible proposée par les auteurs est le couplage d'une TAG syntaxique et d'une TAG sémantique[18]. La TAG sémantique est définie avec comme "catégories" non terminales, les symboles pour une logique des prédicats. Les structures élémentaires d'une TAG synchrone sont des paires d'arbres élémentaires, avec un arbre élémentaire sémantique et un arbre élémentaire syntaxique. Les deux éléments d'une paire ne sont pas forcément tous les deux initiaux ou tous les deux auxiliaires. Pour une paire d'arbres élémentaires, des liens sont définis entre un noeud de l'arbre syntagmatique et un noeud de l'arbre sémantique. Le même noeud en sémantique peut être lié à différents noeuds en syntaxe, qui sont dans une relation de projection. Lors de la dérivation, ces liens sont consommés, et deux arbres de dérivation sont construits de manière synchrone.

Si l'idée des TAG synchrones est attractive, il nous semble qu'il est dommage de ne pas tirer profit des propriétés particulières de l'arbre de dérivation en TAG. Or un travail de fond sur le statut exact de l'arbre de dérivation manque. En particulier, l'apparente hétérogénéité de l'utilisation de l'adjonction (pour la complémentation phrastique et pour les modifieurs) est considérée comme problématique (cf. Rambow et Joshi, 1992; Shieber et Schabes, 1994). Dans le cadre d'une collaboration avec Sylvain Kahane (Candito et Kahane, 1998a; 1998b), nous avons essayé de combler cette lacune, en étudiant à quelles conditions l'arbre de dérivation peut être considéré comme une représentation de dépendances sémantiques, à la lumière des notions de dépendance sémantique et dépendance syntaxique de la Théorie Sens-Texte (Mel'cuk, 1988).
Nous étudions en détail cette interprétation de l'arbre de dérivation au Chapitre 7. Mais nous introduisons d'ores et déjà une version sémantique du principe de cooccurrence prédicat-arguments, qui permet une caractérisation uniforme des arbres élémentaires d'une grammaire TAG, telle qu'elle est généralement conçue. Nous fournissons ensuite une caractérisation des différents arbres élémentaires d'une TAG, en fonction de l'interface syntaxe/sémantique qu'ils mettent en jeu.

2.4.1 Une version sémantique du Principe de Cooccurrence Prédicat-Arguments

TAG est souvent présenté comme un formalisme qui permet de localiser dans une même structure les relations entre un prédicat et ses arguments sémantiques, ce que nous appellerons dorénavant les dépendances sémantiques.
C'est dans ce cadre que s'inscrivent tous les travaux sur la dépendance non bornée en TAG ou dans des extensions de TAG comme les TAG ensemblistes (voir Kroch et Joshi, (1985) pour l'extraction hors de compléments phrastiques; Kroch (1989) pour un

[17] La proposition initiale de Shieber et Schabes (1990) permet des arbres de dérivation non isomorphes pour les deux TAG, ce qui s'est révélé trop puissant d'un point de vue computationnel (Shieber, 1994). Rambow et Satta (1996) proposent une solution intermédiaire, formellement plus puissante que TAG, mais tout de même contrainte, et qui permet de traiter les exemples initiaux de portée des quantificateurs de Shieber et Schabes.
[18] Outre leur application pour le couplage d'une analyse syntaxique et d'une analyse sémantique, il a également été proposé de les utiliser en traduction automatique, en reliant deux TAG syntaxiques pour deux langues différentes (Abeillé et al., 1990).

traitement en TAG ensembliste de l'extraction hors de groupes nominaux; Bleam (1994) pour un traitement de la montée des clitiques en espagnol). Ces travaux s'attachent à trouver des analyses en TAG, ou au besoin dans une extension de TAG, telles que l'élément extrait soit localisé dans l'arbre de son gouverneur sémantique.
Cependant, la démarche n'est pas poussée jusqu'au bout, en particulier l'extension aux dépendances sémantiques mettant en jeu un modifieur syntaxique n'est pas prise en compte. De plus une étude de la faisabilité de la localisation systématique des dépendances sémantiques au sein d'un arbre élémentaire manque.

Abeillé, avec ses travaux sur les idiomes et les constructions à verbes support (Abeillé, 1991), résumés par la formulation du principe de minimalité sémantique, fait le pas décisif de considérer que les TAG manipulent des unités sémantiques. Dans la continuation de cela, il est logique de donner une version sémantique au principe de cooccurrence prédicat-arguments, plutôt qu'une formulation qui fasse appel à la notion de sous-catégorisation. Nous allons voir en outre, que cela permet de donner une caractérisation uniforme des arbres élémentaires d'une TAG, qu'il s'agisse d'arbres pour modifieurs syntaxiques ou pas. Soit donc le principe suivant :

Principe de cooccurrence prédicat-arguments (reformulation 1): un arbre élémentaire est la structure syntagmatique minimale qui comprend un ensemble d'ancres lexicales correspondant à une unité sémantique, et qui comporte une position à étendre pour la réalisation de tous et seulement les arguments, *exprimés en syntaxe*, de cette unité sémantique (Candito et Kahane, 1998a).

Les positions à étendre sont soit des noeuds à substituer (et sont donc étendues par la substitution), soit des noeuds pied (auquel cas la position est étendue par l'adjonction de l'arbre qui la contient sur un autre arbre).
Le principe précise qu'il s'agit des positions pour les arguments *exprimés en syntaxe*, ce pour représenter d'une part le fait que certains arguments peuvent être non exprimés, soit par omission (par exemple le bénéficiaire pour le sémantème 'donner'), soit que la diathèse utilisée réduit le nombre d'arguments exprimés (par exemple le moyen ne permet pas l'expression de l'argument agent, bien qu'un agent soit interprétable, cf. Chapitre 4). D'autre part, les arguments peuvent également être exprimés mais pas en syntaxe. En effet si l'on considère les pronoms personnels comme des affixes flexionnels (Miller, 1991), ils constituent des cas de réalisation d'arguments en morphologie. Comme nous utilisons TAG avec pour éléments lexicaux des mots (et pas des morphèmes), on ne peut pas considérer de positions à étendre pour ces réalisations d'argument.

2.4.1.1 Caractérisation uniforme des arbres élémentaires d'une TAG

On peut remarquer que la formulation de PCPA ci-dessus ne fait pas de distinction entre noeud pied et noeud à substituer. Il s'en suit que cette formulation donne une caractérisation uniforme des arbres élémentaires d'une LTAG, qu'il s'agisse d'arbres pour des modifieurs ou pas, d'arbres auxiliaires ou pas.
On donne quelques exemples d'arbres élémentaires ci-dessous, dont la topologie suit la description de la TAG du français de Abeillé (1991). Dorénavant, les positions argumentales à étendre seront numérotées en partant de 0 (on suit un ordre d'oblicité

syntaxique, dans le cas non marqué; ainsi pour un verbe l'argument 0 est le sujet à l'actif)[19].

Pour un verbe transitif strict à arguments nominaux (Figure 14), les positions à étendre N0 et N1 sont à substituer. Pour un nom sans argument (Figure 15), l'arbre élémentaire ne comprend aucun noeud pied ou à substituer. Pour un adjectif, dans son emploi épithète (Figure 16), sans complément, l'arbre élémentaire comporte une position à étendre sous la forme d'un noeud pied pour le nom modifié, qui est correspond à un argument sémantique de l'adjectif. Enfin, on donne l'arbre élémentaire pour un verbe à complétive, avec sujet et complétive en position canonique (Figure 17). Le verbe a deux arguments sémantiques, et l'arbre comprend bien une position pour chacun (N0 et P1). Mais les verbes à complétive sont traités en TAG par adjonction sur la complétive (cf. infra), c'est pourquoi P1 est un noeud pied.

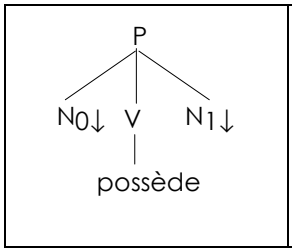


Figure 14

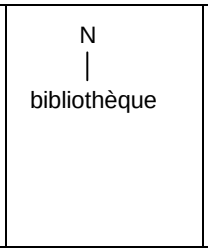


Figure 15

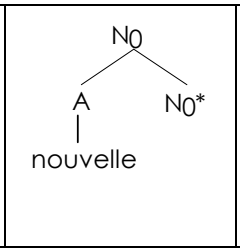


Figure 16

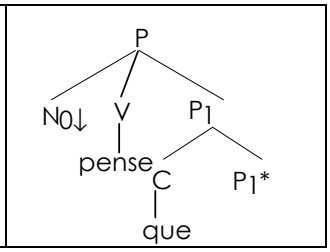


Figure 17

Or ces arbres doivent être distingués sur le plan syntaxique, selon le statut syntaxique de l'ancre par rapport au tout (l'ancre est-elle la tête syntaxique ou pas) et sur le plan formel (l'arbre est-il auxiliaire ou pas), bien que les deux distinctions ne coïncident que partiellement.

Pour les arbres auxiliaires, on peut trouver chez plusieurs auteurs (Kroch, 1989:70; Shieber et Schabes, 1994:96) une distinction, selon la localisation de la tête syntaxique du constituant représenté par l'arbre.

Nous précisons que nous employons la notion de **tête syntaxique** définie comme l'élément qui transmet les propriétés morpho-syntaxiques au constituant qui le contient, ce constituant pouvant le transmettre à sont tour ou pas[20]. En particulier, la tête transmet la catégorie syntaxique. Ainsi la tête d'une combinaison det + N est-elle le nom, la tête d'une phrase est-elle le verbe (puisque le statut de phrase n'est possible, sauf cas très particuliers qu'en présence d'un verbe, éventuellement seul, comme un intransitif à l'impératif).

La tête syntaxique peut être l'ancre de l'arbre auxiliaire, l'exemple typique étant celui d'un arbre auxiliaire ancré par un verbe à complétive (permettant l'extraction); ou bien ce peut être le noeud pied de l'arbre auxiliaire, il s'agit alors d'un modifieur syntaxique, c'est-à-dire que le noeud pied représente un argument sémantique de l'ancre, qui se

[19] Les exemples suivent certains choix linguistiques de Abeillé (1991) : les arbres de projection phrastique ont une structure plate, sans noeud SV. Les différentes projections du nom sont représentées par des traits, et pas par des catégories distinctes. Dans les arbres, on étiquette les noeuds par leur catégorie, et leur indice argumental, s'ils représentent un argument. Plusieurs noeuds peuvent donc avoir la même étiquette.

[20] Il s'agit donc pour certains auteurs de la tête morpho-syntaxique (ou pour Zwicky, 1985 du *locus morphologique*). C'est la notion de tête employée en GPSG ou en HPSG, qui l'utilisent pour généraliser la percolation des traits dits "de tête".

réalise en syntaxe comme le gouverneur syntaxique de l'ancre. C'est le cas par exemple pour un arbre élémentaire ancré par un adjectif dans son emploi épithète.
On voit que l'utilisation de l'adjonction dans les deux cas est très différente. Dans le premier cas, pour les arbres que nous désignerons comme **enchâssants** (on redonne Figure 19, l'exemple d'un arbre élémentaire pour un verbe à complétive), dans l'arbre dérivé la tête syntaxique est l'ancre de l'arbre adjoint. Dans le second cas, pour les arbres que nous désignerons comme **modifieurs** (au sens où un des arguments sémantique de l'ancre, représenté par le noeud pied, se réalise en syntaxe comme le gouverneur syntaxique de l'ancre)[21], la tête syntaxique est inchangée après adjonction (on redonne ci-dessous Figure 18 l'exemple d'un arbre élémentaire pour un adjectif dans son emploi épithète).

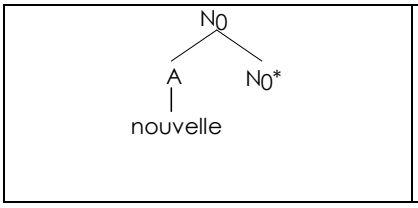


Figure 18. Arbre auxiliaire modifieur

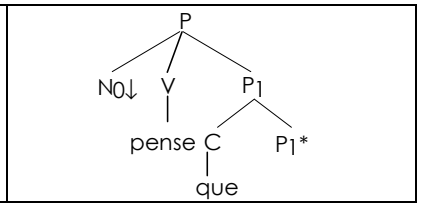


Figure 19. Arbre auxiliaire enchâssant

On comprend alors la dissymétrie posée par Shieber et Schabes (1994), dans leur définition de la dérivation (cf. supra note 6). Seuls les arbres modifieurs peuvent s'adjoindre plusieurs fois sur le même noeud. Il n'aurait pas de sens en effet d'adjoindre plusieurs arbres enchâssants à un même noeud, puisque cela reviendrait à fournir plusieurs têtes syntaxiques pour un même élément.

Avec une version sémantique du PCPA, le noeud pied a le même statut que les noeuds à substituer. Ces noeuds sont censés représenter un argument sémantique de l'ancre, que l'arbre soit initial ou auxiliaire, enchâssant ou modifieur. On peut vérifier que c'est en général un principe implicite dans les LTAG existantes, c'est-à-dire que, en dehors des cas de divergences fortes entre syntaxe et sémantique, le principe est suivi.

Avec le PMS et le PCPA tel que nous le formulons ci-dessus, une grammaire TAG peut être vue comme une grammaire lexicalisée qui réalise l'interface entre sémantique et syntaxe, en ce qui concerne les dépendances (les questions de portée de la quantification doivent être traitées par un mécanisme supplémentaire).

Nous étudions au Chapitre 7 la possibilité d'interpréter l'arbre de dérivation comme représentant des dépendances sémantiques, et quelles sont les configurations qui

[21] Nous ne prenons pas tout à fait la terminologie de Shieber et Schabes (arbres prédicatifs/arbres modifieurs). Nous ne voulons pas employer le terme "prédicatif" pour une notion syntaxique, d'où le terme "enchâssant". L'adjonction d'un arbre auxiliaire enchâssant a pour effet d'amener une tête syntaxique, mais pas forcément sémantique (cf. Chapitre 7). Kroch (1989) a également proposé une terminologie, il appelle les arbres auxiliaires modifieurs des *athematic* trees et les arbres enchâssants des *complement* trees. "Un arbre complément sert à introduire un prédicat qui sous-catégorise et qui assigne un théta-rôle à un syntagme de la catégorie de son noeud pied. [...] Un arbre athématique [est un arbre auxiliaire tel que] son noeud pied est la tête syntaxique du syntagme dominé par le noeud racine." (Kroch, 1989:70, traduit par moi).

peuvent empêcher cette interprétation. Nous étudierons également plus précisément les différents cas d'interface syntaxe/sémantique instanciés par les arbres élémentaires, ainsi que les cas de divergences syntaxe-sémantique problématiques, c'est-à-dire qui font qu'il paraît impossible de respecter à la fois les contraintes de représentation de la syntaxe, et les contraintes imposées par les principes sémantiques (PMS et PCPA).

2.4.1.2 Prise en compte des phénomènes de contrôle

Nous avons fourni une formulation "sémantique" du PCPA, où les positions à étendre que comporte un arbre élémentaire sont les arguments exprimés de l'ancre, celle-ci pouvant ne pas être la tête syntaxique.
Il faut cependant introduire la notion de domaine de réalisation d'un argument, étant donnés les phénomènes de contrôle. Si l'on considère en effet l'exemple (3), cas de contrôle anaphorique, l'argument 0 de 'honorer' est effectivement exprimé en syntaxe, par *les villageois* (les villageois honorent le cortège), mais il est clair qu'au niveau syntaxique, il s'agit d'une réalisation "hors domaine".

(3) Pour honorer le cortège, les drapeaux ont été déployés par les villageois.

Cependant la notion de domaine en TAG est difficile à formaliser, autrement que par la notion d'arbre élémentaire. On ne peut pas en effet simplement considérer la ou les projections phrastiques de l'ancre comme domaine : dans le cas par exemple d'une clivée, illustré Figure 20, le foyer de clivée, argument sémantique du verbe ancre, est réalisé hors de la stricte projection phrastique de cette ancre (la tête syntaxique de tout l'arbre est en effet la copule, et l'arbre dans son ensemble en représente une projection).

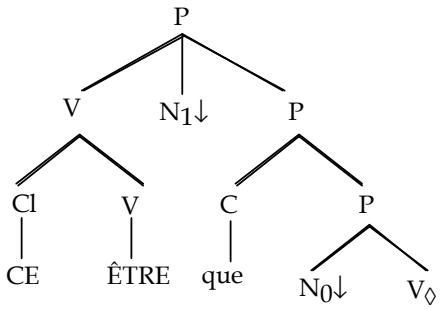


Figure 20. Schème pour une clivée (*c'est N1 que N0 V*)

Cela conduirait pour le PCPA à une définition circulaire de la notion d'arbre élémentaire : un arbre élémentaire contiendrait les positions pour les arguments sémantiques à l'intérieur d'un domaine qui serait l'arbre élémentaire.
Nous préférons considérer les phénomènes de contrôle comme un cas exceptionnel (bien défini), où un des arguments est partagé par deux lexèmes, et est en TAG représenté dans l'arbre du lexème dominant en syntaxe (dans notre exemple, dans l'arbre de *déployés*).
Nous revenons sur les cas de contrôle au Chapitre 2, pour illustrer des cas de divergences dans la correspondance entre arguments sémantiques et actants grammaticaux.

3 L'organisation interne d'une LTAG : motivations pour un niveau supplémentaire de représentation

Après avoir précisé les principes linguistiques retenus pour la bonne formation des arbres élémentaires, nous revenons sur la question de l'organisation d'une TAG.
Une caractéristique des TAG lexicalisées est que les éléments lexicaux ont un accès direct aux structures syntaxiques dans lesquelles ils peuvent apparaître. Cela permet un codage fin des phénomènes linguistiques, limite la propagation de traits et permet de disposer à l'exécution — au moment de l'utilisation effective de la grammaire, en analyse ou en génération — de structures qui sont toutes lexicalisées (cf. Schabes et al. (1988) pour les avantages computationnels de la lexicalisation).
Cependant, la lexicalisation complète lors de l'utilisation de la grammaire — *à l'exécution* — n'empêche pas de formuler des généralisations et d'utiliser, pour la *représentation* de la grammaire, le partage d'informations entre éléments lexicaux. Or une forme verbale peut ancrer des dizaines voire des centaines d'arbres élémentaires, pour peu qu'elle ait plusieurs arguments sémantiques. En effet, étant donné le domaine de localité étendu, un même lexème peut ancrer des arbres élémentaires qui diffèrent selon la diathèse de l'ancre (la correspondance entre arguments sémantiques et actants grammaticaux, cf. infra), et pour une diathèse donnée, selon la réalisation grammaticale des actants grammaticaux (en position canonique, extraite, clitique, inversée ...). Donc le partage d'informations entre arbres élémentaires est indispensable pour limiter la physique de la grammaire.

3.1 Organisation pratique d'une LTAG : exemple du système XTAG

Un premier niveau de partage a été proposé aux débuts de l'écriture de LTAG à couverture réelle, dans le système XTAG (Paroubek et al., 1992). Les arbres élémentaires d'une LTAG y sont compilés à partir de 3 sources d'information :

- un ensemble de **schèmes**[22] (des arbres élémentaires dont l'ancre lexicale n'est pas instanciée)
- un **lexique syntaxique** associant chaque lexème[23] à un ensemble de schèmes, plus des traits morpho-syntaxiques
- un **lexique morphologique** où les formes fléchies pointent vers un lemme plus des traits morphologiques

[22] Nous avons utilisé dans d'autres publications le terme, trop lourd, de "schéma d'arbre". Dans les travaux en anglais, on trouve le terme "tree sketch".
[23] Le système XTAG ne contraint bien sûr pas que les unités qui apparaissent dans le lexique syntaxique correspondent effectivement à une unité sémantique (c'est dans cette acception que nous employons le terme de lexème). Mais le système XTAG prévoit qu'un lexème peut être une unité polylexicale, et on peut donc ainsi représenter une expression figée par une entrée dans le lexique syntaxique.

Un schème est une abstraction sur un ensemble d'arbres élémentaires, en cela qu'une ou plusieurs ancres sont non spécifiées.

La Figure 21 montre l'arbre élémentaire canonique associé au lexème fléchi *parlait*[24].

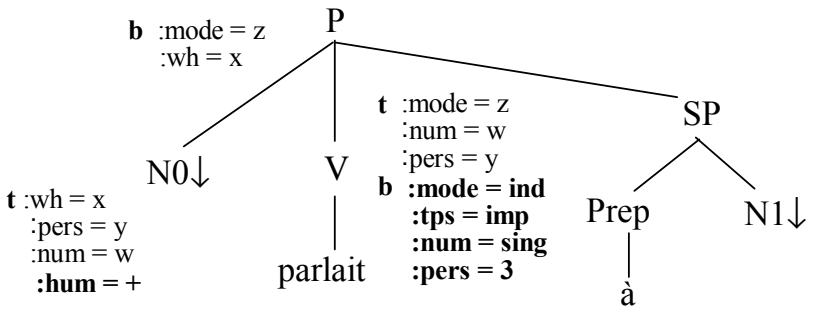


Figure 21. Un exemple d'arbre élémentaire lexicalisé

La Figure 22 montre les 3 sources d'informations qui le représentent en interne. La forme *parlait* pointe vers le lemme PARLER, le lexème /PARLER/, comprenant un lemme unique PARLER, sélectionne à son tour la famille n0Van1 (un ensemble de schèmes, cf. infra), où la préposition *à* apparaît comme co-tête (sauf dans le cas du clitique datif).

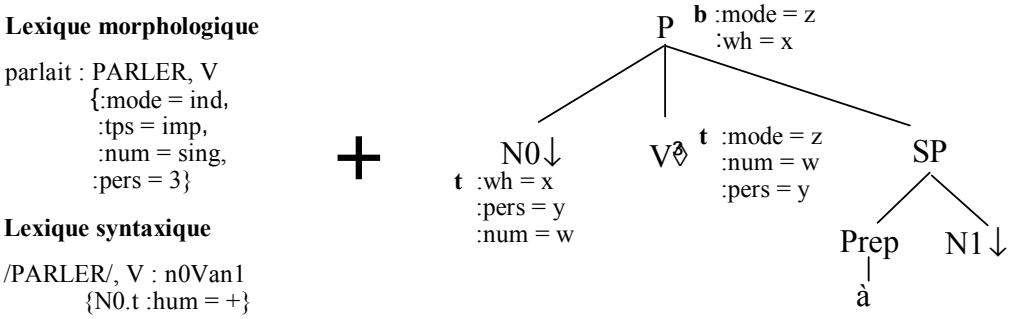


Figure 22. Trois sources d'informations pour un arbre élémentaire, dans le système XTAG

Dans les schèmes, les préterminaux qui sont destinés à dominer une ancre sous-spécifiée sont notés d'un ◊.

Les traits spécifiés dans le lexique morphologique concernent une des ancres, et sont ajoutés comme traits aval sur le noeud dominant immédiatement cette ancre. Les traits spécifiés dans le lexique syntaxique précisent à quel noeud de l'arbre élémentaire ils s'appliquent, et s'il s'agit de traits amont ou aval[25].

La sélection lexicale des schèmes est contrôlée par les traits des lexiques morphologique et syntaxique. Elle utilise en outre le regroupement des schèmes en **familles**, qui rassemblent les schémas de même structure argumentale (dans la littérature on trouve

[24] Dans l'arbre lexicalisé, on a noté en gras les traits provenant des lexiques syntaxique et morphologique.

[25] Les traits ajoutés par le lexique syntaxique peuvent être utilisés pour ajouter des informations sémantiques sur les arguments, ou bien pour des informations morpho-syntaxiques, comme spécifier le mode d'une complétive, qui est également un argument sémantique. Dans le système XTAG, il n'est pas très satisfaisant que l'on ait recours au nom des noeuds pour ajouter les traits. Il vaudrait mieux utiliser une caractérisation des positions argumentales (i.e. les noeuds pied ou à substituer qui représentent les arguments sémantiques).

"de même cadre de sous-catégorisation (XTAG group, 1998)", mais les familles comprennent les schèmes avec changement de diathèses, comme le passif). L'idée est de grouper des schèmes susceptibles d'être sélectionnés par un même élément lexical.
Un lexème peut dans le lexique syntaxique spécifier simplement une liste de schèmes, à tout noeud desquels il peut ajouter des traits. Ou bien il peut spécifier une ou plusieurs familles de schèmes et par un système de traits peut en sélectionner éventuellement un sous-ensemble. Par exemple, le lexème /COMPORTER/ sélectionne la famille des transitifs moins les arbres passifs et moyens, ce qui est représenté dans le lexique syntaxique, et qui vaut donc pour toutes les formes de /COMPORTER/. La "non sélection" peut également provenir d'une incompatibilité de traits pour une forme donnée d'un lexème. Ainsi par exemple, bien que le lexème /MANGER/ sélectionne toute la famille des transitifs, la forme *mangeait* ne sélectionne pas les arbres du passif, qui imposent une ancre au participe passé.

On a ainsi 3 niveaux de généralisation par rapport aux structures lexicalisées utilisées par une LTAG, le lexème, le schème et la famille de schèmes. L'ensemble des schèmes forme un composant purement syntaxique, mais lors de l'*utilisation* d'une LTAG, ces schèmes n'interviennent jamais tels quels, mais toujours avec leur(s) ancre(s), c'est pourquoi on peut dire qu'une LTAG est complètement lexicalisée.

3.2 Motivations pour un niveau supplémentaire de description

Les schèmes restent cependant des unités larges, dont chacune rassemble des informations qui dans d'autres formalismes sont codées dans plusieurs structures[26], de même type ou de types différents. Comparée à d'autres formalismes ou théories linguistiques, on peut dire qu'une LTAG se présente sous une forme déjà "compilée". Cette compilation est en partie automatique, avec la lexicalisation des schèmes, mais la construction des schèmes est réalisée "à la main" : le formalisme des LTAG ne fournit pas tel quel de mécanismes pour l'expression de généralisations syntaxiques au sein de l'ensemble des schèmes.
Cette lacune tient au fait que les travaux en TAG ont surtout porté sur la lexicalisation et sur le pouvoir expressif permis par l'opération d'adjonction. Mais cela n'est pas justifié du point de vue théorique. Un formalisme linguistique se juge notamment par sa capacité de capturer les régularités existantes entre langues et au sein d'une langue. Représenter l'ensemble des schèmes sans aucune organisation revient à nier à peu près toute régularité syntaxique à l'intérieur de la phrase simple.
Et du point de vue pratique, comme cela est déjà souligné par Vijay-Shanker et Schabes (1992), la redondance dans les schèmes rend difficile l'écriture et la mise à jour de la grammaire. Le concepteur d'une LTAG doit veiller à l'uniformité du traitement d'un phénomène sur toute la grammaire et à l'exhaustivité des croisements de phénomènes

[26] Un arbre élémentaire spécifie à la fois les relations de dominance et de précédence, là où d'autres formalismes tels HPSG ou la Théorie Sens-Texte, permettent de dissocier les deux. Et surtout un arbre élémentaire en LTAG est en général de profondeur supérieure à 1 et correspond donc à un morceau de dérivation dans un formalisme basé sur des règles de réécriture (cf. Kasper et al. 1995 pour une présentation d'une grammaire LTAG comme grammaire HPSG précompilée).

linguistiques (par exemple si la grammaire traite le passif et les relatives, pour être cohérente elle doit aussi traiter les relatives au passif). On peut ajouter que cela rend difficile la comparaison entre différentes grammaires et la vision globale d'une grammaire pour une langue donnée.

3.2.1 Solutions existantes pour la représentation compacte d'une LTAG

La première tentative d'organisation des grammaires TAG est celle citée du système XTAG, avec en particulier une organisation en familles de schèmes. Cependant, les familles sont de simples listes de schèmes, sans aucune organisation interne, et sans aucun partage entre différentes familles.
Plusieurs solutions ont été proposées et/ou implémentées pour représenter de manière compacte une TAG lexicalisée. Ces solutions utilisent les deux outils de représentation du lexique, les graphes d'héritage et les règles lexicales. Dans le cas de la représentation d'une TAG, ces règles s'appliquent à des structures syntagmatiques lexicalisées (ou pré-lexicalisées, i.e. les schèmes) et il s'agit plutôt de règles lexico-syntaxiques.

Vijay-Shanker et Schabes (1992) ont proposé de représenter une LTAG par un graphe d'héritage monotone, augmenté de règles "lexico-syntaxiques". Une implémentation pour la LTAG anglaise du XTAG group a été réalisée[27], (Xia et al., 1998). Becker (1993) a proposé et implémenté un système surtout basé sur des règles lexico-syntaxiques (qu'il appelle "métarègles"). Enfin, Evans et al. (1995) ont proposé de représenter totalement une LTAG dans le formalisme DATR (Evans et Gazdar, 1989), donc sous la forme d'un graphe d'héritage, non monotone, avec des règles lexicales prises en charge par DATR. Actuellement le système est utilisé au sein du projet LEXSYS, en cours à l'université du Sussex, pour la représentation compacte d'une D-Tree Grammar[28] de l'anglais (Smets et Evans, 1998). Ces solutions reposent sur la construction de schèmes canoniques — i.e. représentant des constructions non marquées — par combinaison d'unités plus petites, et de dériver les constructions marquées par règles lexico-syntaxiques.

Nous proposons de représenter les TAG lexicalisées comme la compilation d'un cadre de description linguistique plus abstrait et plus modulaire : la métagrammaire (MG), qui s'appuie en partie sur les propositions formelles de Vijay-Shanker et Schabes (1992). MG permet l'expression de généralisations syntaxiques par l'utilisation du partage de propriétés par héritage. Mais la MG a également l'ambition de fournir des éléments de théorie linguistique, qui mettent à profit l'utilisation du formalisme des TAG, ce qui la distingue des solutions ci-dessus, qui ont un caractère plutôt "technique". La MG formalise les principes rendant compte de la bonne formation des structures lexicalisées,

[27] Cette implémentation reprend en outre notre proposition (Candito, 1996) de dissocier règles de changement de diathèse et règles de réalisations syntagmatiques, bien que les fonctions grammaticales soient considérées comme des positions structurales.
[28] Les D-Tree Grammars (Rambow et al., 1995) sont un formalisme "de la famille des" TAG. Elles utilisent une opération de substitution généralisée, qui permet à des parties de l'arbre substitué de s'intercaler dans l'arbre recevant la substitution.

que nous avons précisé supra. Cette formalisation utilise une représentation de la sous-catégorisation en termes fonctionnels.
En définissant une organisation générale des informations linguistiques, ainsi que des principes indépendants de la langue à respecter, la MG a également une vocation multilingue.

Un outil implémenté permet la compilation d'une MG en une TAG (plus précisément les schèmes d'une TAG), en opérant automatiquement les croisements valides de phénomènes.

Nous avons écrit une MG pour le français (MGf), à partir d'une TAG existante (Abeillé 1991), ce qui a permis d'en augmenter la couverture et d'en assurer la cohérence. En procédant par contraste systématique avec le français, nous avons également écrit une MG pour l'italien (MGi). La compilation automatique fournit deux TAG parallèles, compatibles pour des applications multilingues.
Au delà de l'application de compilation en TAG, les deux MG fournissent une représentation hiérarchisée de la syntaxe des deux langues, cela permet un contraste entre les deux langues, plus ou moins précis, selon le niveau de généralité considéré.

Le plan de la thèse est le suivant. Chapitre 2 nous décrivons la métagrammaire, en détaillant ses caractéristiques théoriques et formelles. Les Chapitre 3, 4, 5 décrivent les parties de la MGf et la MGi concernant le comportement syntaxique des lexèmes de lemme principal verbal, et de leurs actants grammaticaux. Chapitre 6, nous décrivons la représentation des autres catégories de lexèmes, mais de manière moins détaillée. Chapitre 7, nous revenons sur l'interprétation de l'arbre de dérivation comme graphe de dépendances sémantiques. Enfin, nous présentons Chapitre 8 les applications et perspectives ouvertes par notre travail.

Chapitre 2
La métagrammaire

Nous avons précisé des principes de bonne formation des arbres élémentaires d'une TAG au Chapitre 1. Une TAG apparaît comme une grammaire dont les unités décrivent le comportement syntaxique possible d'un lexème fléchi, qui par définition correspond à une unité sémantique.

La métagrammaire a été conçue pour (i) représenter de manière modulaire et hiérarchisée et (ii) générer, les schèmes d'une TAG, en tant qu'unités respectant les principes de minimalité sémantique et de cooccurrence prédicat-arguments. Au-delà de la réalisation pratique de la compilation en une TAG, la MG pose des principes à vocation universelle pour l'organisation des connaissances syntaxiques.

On a choisi le nom "métagrammaire" (dorénavant MG), du fait qu'il ne s'agit pas, comme pour une grammaire computationnelle, d'un ensemble de règles pour décrire des énoncés de la langue, mais d'un ensemble de règles pour décrire des règles (en l'occurrence des schèmes), elles-mêmes utilisables par des règles combinatoires (l'adjonction et la substitution) pour décrire des énoncés de la langue. Mais bien sûr, par transitivité, la MG[1] constitue un modèle (partiel) de compétence, puisqu'elle décrit des règles formant elles-mêmes un modèle partiel de compétence.
Avec ce système, l'analyse d'un énoncé (ou inversement la génération) est le résultat de deux dérivations. Une première dérivation qui engendre les schèmes par combinaison des règles de la MG, et une dérivation qui engendre l'énoncé par combinaison des schèmes (lexicalisés). L'originalité de cette approche en deux temps, tient au fait que la première dérivation peut se faire (et est actuellement faite) par compilation, avant l'analyse (ou la génération) proprement dite. Elle est en cela moins contrainte par des exigences d'efficacité, et on peut utiliser au sein de la MG un appareil formel plus coûteux en termes computationnels que ceux utilisés en analyse syntaxique classique (i.e. par exemple les TAG, ou les règles de réécriture).

Dans une première section, nous décrivons les concepts linguistiques utilisés et leur répartition au sein de la MG, puis nous décrirons l'appareil formel que nous avons retenu, l'organisation générale d'une MG et la procédure de génération d'arbres élémentaires. Enfin nous terminons ce chapitre par la description de l'implémentation du générateur de schèmes que nous avons réalisée.

[1] On parlera de "la MG" pour désigner le formalisme et de "une MG" pour désigner une métagrammaire instanciée pour une langue particulière. On utilisera également la MGf et la MGi pour désigner les métagrammaires écrites pour le français et l'italien.

1 La structuration des connaissances linguistiques dans la MG

La MG se définit par rapport aux TAG comme un niveau supplémentaire de description, qui utilise autant que faire se peut la généralisation, afin de faire apparaître les régularités existant au sein d'une langue et entre langues.
Son but est de décrire (et permettre de générer) les structures syntagmatiques que peuvent ancrer les lexèmes fléchis de la langue considérée, telles que ces structures comprennent une position pour chaque argument sémantique exprimé du léxème. C'est-à-dire que ces structures décrites sont des schèmes tels qu'ils ont été définis au Chapitre 1.
La MG doit également faciliter la sélection par un lexème de l'ensemble des schèmes qu'il peut ancrer, c'est-à-dire que la MG doit disposer de la notion de famille de schèmes.

La MG est telle quelle une grammaire non lexicalisée. Mais on a vu que l'existence de composantes non lexicalisées est une condition nécessaire à l'expression de généralisations. Couplée à un ou plusieurs lexiques adéquats, la MG permet cependant de générer des unités qui sont elles toutes lexicalisées : des arbres élémentaires au sens de TAG. Avec l'architecture comprenant la représentation d'une TAG lexicalisée en MG + lexiques, on obtient ainsi les avantages pratiques de la lexicalisation, et les avantages pratiques et théoriques fournis par l'expression de généralisations syntaxiques.

La clé de l'expression de généralisations pour la représentation d'un ensemble de schèmes est de décomposer les informations contenues dans un schème en unités plus petites. Plus précisément il s'agit de définir quelles sont les types d'information qui doivent être fournies pour définir totalement un schème.
Nous pouvons annoncer dès maintenant, pour faciliter la lecture, les trois domaines d'information que nous considérons :

- la sous-catégorisation initiale (en termes de fonctions grammaticales)
- les éventuelles redistributions (changements de diathèse)
- les réalisations des fonctions grammaticales

Un schème est défini dans la MG comme une combinaison (contrôlée) de ces trois domaines d'information (que nous appellerons des "dimensions"). Chaque domaine d'information est organisé de manière hiérarchique, et comprend un éventail de possibilités, mutuellement exclusives. Deux schèmes peuvent être différents mais partager certaines sous-parties. Ces sous-parties communes peuvent être la sous-catégorisation initiale (dans les deux schèmes, elle peut être complètement identique, ou bien simplement partager une ou plusieurs fonctions grammaticales); les redistributions (les deux schèmes peuvent représenter un passif par exemple, mais différer sur les réalisations des fonctions); une ou plusieurs réalisations de fonction en commun.

Nous décrivons ci-dessous plus précisément les notions linguistiques employées pour décomposer la syntaxe sous-jacente à un ensemble de schèmes en ces trois domaines d'information.

Dans le cadre de la compilation en une TAG, nous intégrons à la MG les principes linguistiques énoncés pour les schèmes au Chapitre 1, qui prévoient l'interface avec la sémantique. Ainsi la MG décrit-elle des schèmes tels que chacune comprend une position (syntaxique) pour tout argument sémantique exprimé d'un lexème pouvant ancrer ce schème. La MG est donc concernée par l'interface sémantique / syntaxe, dans la mesure où les arguments sémantiques doivent y recevoir une représentation syntaxique. Nous faisons dans un premier temps l'hypothèse, simpliste, que les arguments sémantiques d'un lexème se réalisent comme des dépendants syntaxiques de ce lexème. Cela est la plupart du temps le cas pour un lexème verbal par exemple. Nous reviendrons sur cette hypothèse, de manière à intégrer les cas où elle n'est pas vérifiée.

1.1 La sous-catégorisation en termes de fonctions grammaticales

Un schème représente un des comportements syntaxiques possibles pour un ensemble de lexèmes.
Il s'agit donc d'organiser les différentes sources d'information qui définissent complètement ces "comportements syntaxiques possibles" pour les lexèmes.

Lemme principal

Pour cela, on définit le **lemme principal** d'un lexème, comme (i) le lemme lui-même dans le cas d'un lexème réduit à un seul lemme, ou (ii) dans le cas d'un lexème polylexical, le lemme qui gouverne syntaxiquement les autres éléments du lexème, c'est-à-dire le verbe dans le cas d'une expression figée, ou d'une combinaison verbe support plus nom prédicatif.

Les différents schèmes représentés par une MG sont classés selon la catégorie syntaxique du lemme principal des lexèmes qui peuvent les ancrer.

Actants grammaticaux

Ensuite, le lemme principal d'un lexème porte des contraintes quant aux expressions qui doivent ou peuvent l'accompagner, c'est ce qui est couramment appelé la sous-catégorisation. On appelle **actants grammaticaux** d'un lemme, les syntagmes qui sont soumis à des contraintes d'existence et de réalisation morpho-syntaxique par le lemme. Nous réservons le terme d'argument pour un argument sémantique d'un sémantème.

La **sous-catégorisation** d'un lexème *dans un emploi donné*, est la liste d'actants grammaticaux du lemme principal du lexème qui accompagnent le lemme dans cet emploi donné. Cette liste d'actants est exprimée de telle sorte que les contraintes morphologiques et syntaxiques sur l'expression de ces actants soient explicitées.

Un schème représente un emploi possible pour les lexèmes qui peuvent l'ancrer, et ce schème est en partie caractérisé par la sous-catégorisation qu'ont ces lexèmes dans cet emploi.

Fonctions grammaticales

Nous posons que chaque langue comporte des types prédéfinis de contraintes morphologiques et syntaxiques imposées aux actants grammaticaux, c'est-à-dire des **fonctions grammaticales**, et que l'on peut exprimer les contraintes morpho-syntaxiques imposées aux actants grammaticaux d'un lemme en assignant à ces actants la fonction grammaticale appropriée. C'est-à-dire qu'on considère que la sous-catégorisation d'un lexème peut être exprimée en termes fonctionnels.

Nous suivons en cela toute une tradition grammaticale, dont on peut citer en linguistique moderne les travaux de Tesnière pour les grammaires de dépendances (Tesnière, 1959), la Théorie Sens-Texte (Mel'cuk, 1988) et pour une intégration entre grammaire de constituants et grammaire de dépendances, la Grammaire Relationnelle (RG) (Perlmutter et Postal, 1983), et la Grammaire Lexicale Fonctionnelle (LFG) (Bresnan, 1982).

Les fonctions grammaticales permettent de formuler des généralisations tant internes à une langue qu'entre différentes langues (cf. les travaux en RG ou en LFG). En effet, au sein d'une langue, spécifier qu'un lexème doit être accompagné d'un constituant remplissant telle fonction grammaticale permet de s'abstraire des différentes réalisations de surface pour cette fonction. D'autre part, entre différentes langues, les fonctions permettent d'établir des parallèles entre des constructions différentes en surface (soit par l'ordre linéaire, soit par le marquage morphologique), mais qui partagent une représentation en termes de dépendances fonctionnelles. Nous rejoignons donc les auteurs cités dans leur critique de la grammaire transformationnelle et ses développements, qui réduisent les fonctions grammaticales à des positions structurales.

Nous posons que l'on peut pour chaque langue définir des fonctions grammaticales (dorénavant **FG**), caractérisées par un ensemble de propriétés morpho-syntaxiques propres à la langue, mais que l'on peut établir des correspondances entre fonctions pour différentes langues.

La sous-catégorisation illustrée par un schème (i.e. la sous-catégorisation des lexèmes pouvant ancrer ce schème) est exprimée dans la MG en termes de fonctions grammaticales.

Nous considérons que les contraintes qu'impose un lexème sur la ou les catégories syntaxiques possibles pour ses actants grammaticaux s'ajoutent au typage des actants grammaticaux en fonctions grammaticales. Nous pensons en effet qu'il est dommage de distinguer plusieurs fonctions dès que les catégories diffèrent. Premièrement parce qu'on ne représenterait pas le caractère mutuellement exclusif de ces fonctions, deuxièmement parce qu'en français par exemple, il est général qu'un verbe qui sélectionne un sujet phrastique permet également un sujet nominal. De plus un verbe à objet phrastique peut pour certains verbes apparaître au passif (4).

(4) a. Les experts n'avait pas prévu que l'irruption puisse être si violente.
    b. Que l'irruption puisse être si violente n'avait pas été prévu par les experts.

Ainsi, si comme c'est ce qui est généralement fait en LFG par exemple, on donne des noms différents pour les FG de catégorie nominale ou phrastique, on doit représenter par

un autre moyen l'exclusion mutuelle, et on doit exprimer des règles différentes pour le passif par exemple, selon que l'objet est phrastique ou nominal.

De plus, pour certaines langues (comme l'allemand ou le russe), certaines FG peuvent être réalisées par des actants portant des cas différents, et cela est déterminé lexicalement (par l'élément sous-catégorisant ces FG). Il faut donc prévoir de pouvoir en outre spécifier le cas lexical associé aux FG, dans la sous-catégorisation initiale.

Les schèmes représentés dans la MG, qui partagent la même sous-catégorisation sont cependant distingués selon la catégorie syntaxique des actants grammaticaux apparaissant dans la sous-catégorisation, ainsi que le cas lexical (s'il existe) des actants grammaticaux.

"Exceptions"

La généralisation que constitue le typage des actants en FG est bien sûr soumise à exceptions. C'est-à-dire que dans le détail, les actants de différents lexèmes, dont on considère qu'ils portent la même FG, peuvent différer par certaines propriétés (cf. les travaux de Gross (1975) et la classification très isolante à laquelle il aboutit, c'est-à-dire que si l'on prend en compte de nombreuses propriétés, les classes d'éléments les partageant toutes sont réduites à 1,5 éléments en moyenne).
Pour la métagrammaire, il s'agit de fournir les moyens expressifs pour gérer une sélection lexicale fine des schèmes par les éléments lexicaux. En pratique, dans l'implémentation, c'est le système de sélection lexicale par traits du système XTAG qui est utilisé.

Ordre d'oblicité des fonctions grammaticales

Nous suivons Keenan et Comrie (1977) en considérant qu'il existe un ordre universel d'accessibilité des fonctions grammaticales. Ces auteurs ont tiré cette conclusion après une étude de la formation des relatives restrictives pour une cinquantaine de langues, et ont proposé la hiérarchie universelle suivante :

sujet > objet direct > objet indirect > objet oblique > génitif > ocomp[2]

Pour une langue particulière, certaines de ces fonctions peuvent ne pas être distinguées.
HPSG définit la notion comparable d'ordre d'oblicité des actants grammaticaux d'un lexème, représenté par l'ordre d'apparition dans la liste valeur de l'attribut ARG-ST. Dans le cadre d'HPSG, cet ordre d'oblicité est utilisé à diverses fins : représenter les contraintes de liage; représenter des contraintes d'ordre linéaire par défaut exprimées en fonction de l'oblicité relative des actants grammaticaux (Sag, 1987). Par exemple pour l'anglais, un principe général d'ordre des mots, indépendant des schémas de dominance immédiate, peut être que les compléments sont ordonnés du moins oblique au plus oblique, dans le cas non marqué. Enfin Miller (1991)[3] utilise l'ordre d'oblicité pour définir les règles de calcul sémantique associées aux règles syntaxiques, pour le français.

[2] Ocomp est utilisé pour "objet de comparaison", comme *la nuit* dans *Mon coeur est plus noir que la nuit*.
[3] Miller (1991) utilise en fait le formalisme GPSG (Gazdar et al., 1985) mais y a intégré des modifications en s'inspirant de HPSG, et notamment l'utilisation d'une liste ordonnée pour représenter la sous-catégorisation, ainsi que l'utilisation de règles lexicales.

Nous nous référons à l'ordre d'oblicité pour la représentation des redistributions des fonctions grammaticales (cf. infra Section 1.2). En ce qui concerne l'ordre des compléments, la MG n'utilise pas actuellement l'ordre d'oblicité, mais cela pourrait être intégré. Pour le français et l'italien, l'ordre respectif des compléments est relativement libre, même si l'ordre d'oblicité fournit un ordre linéaire préférentiel. Pour une langue comme l'anglais, la spécification d'un ordre fixe des compléments selon l'ordre d'oblicité nous paraît plus justifiée.

1.2 La distinction de fonctions grammaticales initiales

Nous avons justifié en partie l'utilisation de fonctions grammaticales. Une autre justification classique peut être avancée, qui prend en compte le fait qu'un même lexème peut être employé avec différentes diathèses, i.e. différentes correspondances entre arguments sémantiques et actants grammaticaux.
Si l'on considère par exemple une phrase à l'actif pour un transitif en français, et son correspondant passif (personnel), dans les deux phrases le verbe comporte un sujet , qui dans les deux cas a tout un ensemble de propriétés communes bien connues (l'accord, la position préverbale possible, la non réalisation à l'infinitif, l'alternance avec un clitique nominatif, avec le relatif *qui*, ..., cf. Chapitre 5, pour la description du sujet dans la MG du français). Pourtant le sujet dans les deux cas ne correspond pas au même argument sémantique.

(5) a. La crue a inondé le village.
    b. Le village a été inondé par la crue.

Le fait que des arguments sémantiques différents puissent montrer le même comportement grammatical nous semble être une justification évidente de la pertinence de la notion de FG.
Jusqu'à présent, en s'en tenant à la sous-catégorisation fonctionnelle, on retient la même sous-catégorisation pour (5b) et pour un verbe intransitif à l'actif (*Paul dort*). Pourtant la différence est que *Paul dort* n'est pas le correspondant passif d'un phrase active. Dans la MG, les schèmes doivent être représentés non seulement de manière à faire apparaître des généralisations, mais également de manière à pouvoir représenter quel est l'ensemble des schèmes qu'un même lexème peut ancrer.

Pour cela, nous introduisons la notion de fonction grammaticale initiale. Nous distinguons pour un lexème donné, un cas d'emploi non-marqué, pour ce qui est de la correspondance entre arguments sémantiques et actants grammaticaux.
Le cas **non-marqué** est défini pour chaque catégorie de lemme principal, comme le cas d'emploi du lemme le moins marqué morphologiquement et le plus productif dans la langue. On peut remarquer que cette définition n'est pas toujours opératoire. Il est clair en français de définir l'actif comme moins marqué que le passif, car ce dernier n'apparaît pratiquement pas sans son correspondant actif. Mais dans d'autres cas, le choix peut être arbitraire, comme par exemple entre *Paul charge le camion d'oranges* et *Paul charge les oranges dans le camion*.

Une caractéristique du cas non-marqué est de mettre en jeu le nombre maximal d'arguments stricts[4] du lexème.
On peut alors définit la **sous-catégorisation initiale** d'un lexème comme la sous-catégorisation qu'il a dans le cas d'emploi non marqué, c'est-à-dire la liste d′actants grammaticaux, qui doivent être réalisés localement dans le cas non-marqué, et qui portent chacun, dans ce cas non-marqué, une fonction grammaticale, dite **initiale**.

Les FG effectivement portées par les actants grammaticaux d'un lexème peuvent différer des FG initiales, du fait des différentes diathèses qu'un lexème peut montrer. Ce que l'on a appelé jusqu'à présent sous-catégorisation, qui est défini pour un emploi donné d'un lexème, est maintenant appelé **sous-catégorisation finale**. Il s'agit de la sous-catégorisation effectivement portée par un lexème dans un emploi donné.

Les schèmes représentés dans la MG sont classés à la fois selon la sous-catégorisation finale qu'ils illustrent, et selon la sous-catégorisation initiale des lexèmes qui peuvent les ancrer.

Cette distinction entre sous-catégorisation initiale et sous-catégorisation finale est inspirée de la représentation en différentes strates de la sous-catégorisation en Grammaire Relationnelle. Dans cette théorie, les actants ont une relation grammaticale initiale qui leur est assignée, et peut différer de leur relation finale. Il faut cependant remarquer que la Grammaire Relationnelle ne fait pas référence à un cas non marqué pour assigner les relations grammaticales initiales. L'assignation est faite de telle sorte que des généralisations sur les sous-catégorisations trouvées sur les différentes strates associées à une phrase puissent être exprimées. C'est par exemple typique pour ce qui est de l'hypothèse inaccusative, où un intransitif inaccusatif sélectionne un initial 2 (en gros un objet initial), qui n'est jamais final 2. Pour nous, même si nous retenons l'hypothèse inaccusative pour l'italien (cf. Chapitre 3), de telle sorte qu'un intransitif inaccusatif sélectionne un objet initial, cet objet peut être l'objet final. Et on peut considérer que le cas non marqué des inaccusatifs est d'avoir leur argument réalisé comme un objet[5].

La notion de famille de schèmes (à la XTAG), c'st-à-dire le groupement de schèmes en ensembles susceptibles d'être sélectionnés par un même lexème est traduite par la notion de sous-catégorisation initiale. Appartiendront à la même famille, les schèmes qui partagent la même sous-catégorisation initiale, ainsi que la même catégorie pour le lemme principal.

[4] Par strict on entend exclure le cas de l'argument supplémentaire causateur à considérer pour un prédicat complexe fait d'un lexème plus un verbe causatif.
[5] Sornicola (1994) après l'étude d'un corpus italien, indique que 35% des formes strictement intransitives (inaccusatives et inergatives) à l'écrit et 45% à l'oral, sont employées avec leur argument placé après le verbe. Il n'est donc pas impossible de considérer que le cas non marqué est l'emploi postverbal pour les inaccusatifs, c'est-à-dire l'emploi où l'objet initial de l'inaccusatif est objet final (i.e. de surface).

1.2.1 Les redistributions des FG

Nous venons d'introduire la distinction entre sous-catégorisation initiale et sous-catégorisation finale pour un lexème dans un emploi donné. Les deux sous-catégorisations ne sont identiques que dans l'emploi non marqué.
En s'inspirant de la Grammaire Relationnelle, nous considérons que la sous-catégorisation finale d'un lexème est dérivable de sa sous-catégorisation initiale par une suite de **redistributions** des FG associées aux actants grammaticaux (ou changements de diathèse)[6].

La grammaire d'une langue comprend les redistributions possibles à partir d'une sous-catégorisation, redistribution au sens large, comprenant :

- les redistributions pures qui ne modifient pas le nombre d'arguments exprimés (comme le passif avec agent)
- les redistributions qui réduisent le nombre d'arguments exprimés (comme le passif sans agent)
- les redistributions qui augmentent le nombre d'arguments (comme le causatif).

Les redistributions ont une productivité variable et un lexème donné peut bloquer certaines redistributions a priori applicables, a priori possible étant donnée la catégorie du lemme principal du lexème. Ces "exceptions" peuvent elles-mêmes se ramener à des régularités exprimées en termes sémantiques, par exemple le moyen en français est bloqué pour les lemmes prédicatifs à sujet initial non agentif (Ruwet, 1972).

Dans la MG, nous verrons qu'une redistribution peut être décomposée en une série de changement sur la FG associée à un actant grammatical. Par exemple on peut décomposer le passif personnel en français en la "démotion"[7] de l'actant sujet vers la fonction agt-objet (pour complément d'agent), et "promotion" de l'objet vers la fonction sujet. Nous réservons le terme redistribution à un ensemble de changements de FG, *tel que la sous-catégorisation résultante est valable pour la langue considérée*. On parlera de **redistribution partielle** pour le changement de la FG d'un seul actant, dans le cas où celui-ci doit s'accompagner d'une autre redistribution partielle. Par exemple en français la démotion de l'actant sujet vers une autre FG, par exemple dans le cas du passif, vers la fonction agt-objet, doit s'accompagner de la promotion d'un actant vers la FG sujet (l'objet ou un *il* impersonnel, cf. Chapitre 4).

Une redistribution peut s'appliquer à une sous-catégorisation non initiale. C'est-à-dire qu'une sous-catégorisation finale peut être dérivée d'une sous-catégorisation initiale par l'application de plusieurs redistributions en séquence. Par exemple, en italien, un lexème employé au causatif peut également être employé au causatif passif (cf. l'exemple (6a),

[6] Plusieurs termes équivalents existent dans la littérature, *revaluation* pour Relational Grammar (Blake, 1990), *changement de diathèse* (Melcuk,93) ou *syntactic alternation*, ce dernier terme étant plutôt utilisé pour des redistributions à productivité assez faible, sans marque morphologique.
[7] Nous prenons le terme anglais de Comrie (1977) *demotion*, qui fait référence à l'ordre d'oblicité des FG. Un changement de FG qui fait descendre un actant dans cette hiérarchie de FG est une démotion, l'inverse étant une promotion.

qui sera détaillé au Chapitre 4). C'est-à-dire que la sous-catégorisation résultant de l'application du causatif peut subir l'application du passif. On peut également citer l'exemple classique d'application du passif après le dative-shift en anglais (6b).

(6) a. Il libro è stato fatto leggere agli studenti da Maria
Le livre est été fait lire aux étudiants par Marie

b. Paul was given a book by Mary.
Paul a été donné un livre par Mary.

Nous parlerons de **composition de redistributions** pour l'application séquentielle de plusieurs redistributions (non partielles).

1.3 Les réalisations des fonctions grammaticales

Nous avons déjà vu qu'un lexème est associé à une sous-catégorisation initiale; et que *dans un emploi donné*, il est associé à une sous-catégorisation finale. Pour caractériser totalement l'emploi d'un lexème, qui correspond à un schème dans la MG, il reste un dernier paramètre : le type de réalisation grammaticale pour chaque actant grammatical de la sous-catégorisation finale.

Nous avons en effet défini les FG par des ensembles de propriétés morphologiques et syntaxiques. Parmi ces propriétés, une bonne partie concerne les alternances possibles d'un constituant avec par exemple un pronom personnel, un relatif ...
C'est qu'il faut considérer que lorsqu'un lexème comprend dans sa sous-catégorisation finale un actant grammatical portant une FG donnée, le lexème va pouvoir être accompagné de cet actant *sous différentes formes et dans différentes positions mutuellement exclusives*.

Les différentes réalisations possibles pour un actant grammatical portant une FG finale donnée sont représentées dans la MG *indépendamment de la FG initiale pour l'actant grammatical*. C'est-à-dire que la fonction sujet est décrite indépendamment du fait qu'il s'agit d'un sujet à l'actif ou au passif. C'est une des motivations théoriques justifiant le typage du comportement grammatical des actants grammaticaux en fonctions grammaticales.
En pratique, on réalise grâce à cela un partage d'informations important.

1.4 Les différents domaines d'information de la MG

Nous sommes maintenant en mesure de fournir la liste des paramètres qui caractérisent un schème parmi l'ensemble des schèmes représentés par une MG, et de là les domaines d'information qui sont distingués au sein de la MG, que nous avions déjà annoncés supra.

Un schème est caractérisé au sein de la MG par :

- La catégorie de l'ancre principale (la catégorie du lemme principal des lexèmes pouvant ancrer ce schème)

- Une sous-catégorisation initiale (tout lexème ayant cette sous-catégorisation initiale, sauf traits incompatibles, peut ancrer ce schème)
- Une séquence ordonnée de redistributions
- Pour chaque FG apparaissant dans la sous-catégorisation finale, un type de réalisation grammaticale.

En représentant au sein de la MG des schèmes, et non pas des unités lexicalisées, on garde le principe de lexicalisation à l'exécution, puisqu'il n'a pas de sens dans le cas de lexiques conséquents de précompiler des arbres lexicalisés. On peut remarquer que cela implique en particulier que les différentes sous-catégorisations possibles pour un lexème (ou du moins celles couvertes par la MG considérée) soient prévues, et précompilées[8].

Nous pouvons redonner les différents domaines d'information considérés au sein de la MG, que nous appelons **dimensions**, pour représenter la syntaxe d'une langue :
- Dimension 1 : sous-catégorisation initiale
- Dimension 2 : séquences possibles de redistributions
- Dimension 3 : réalisations des fonctions grammaticales

Nous pensons que cette répartition est valable indépendamment de l'application de compilation en une TAG. TAG a la particularité de comprendre des unités de description larges, qui combinent les trois types d'informations.
Nous justifions dans la section suivante cette organisation en trois domaines d'informations, et la comparons avec celles que l'on peut trouver dans quelques autres théories linguistiques.

1.5 Justification des domaines d'information retenus

Nous utilisons dans la MG une distinction nette entre les phénomènes qui relèvent de sous-catégorisations alternatives pour un lexème, dérivables l'une de l'autre par redistribution, et ce qui relève de réalisations alternatives pour une même fonction grammaticale.

Classiquement dans une théorie linguistique on distingue (i) des "règles" syntaxiques qui modifient le nombre d'arguments, qui peuvent modifier la morphologie du lexème, et qui modifient également le sens du lexème, en y apportant des nuances aspectuelles par exemple; et (ii) des règles syntaxiques sans changement de sens, sans changement de morphologie du lexème, et surtout sans changement dans le nombre d'arguments exprimés. Les règles (i), qui sont représentées dans la MG comme des redistributions,

[8] On pourrait cependant imaginer un système avec une construction des schèmes à l'exécution, notamment dans le cas de sous-catégorisations initiales non prévues. C'est une solution qui est envisageable dans le cadre de la génération par exemple. A partir d'une entrée lexicale fournissant une sous-catégorisation (en termes fonctionnels) non prévues, on peut construire les schèmes possibles, en considérant toutes les séquences de redistributions applicables (y compris le cas sans redistributions), et en considérant pour chaque sous-catégorisation finale, toutes les réalisations de fonctions alternatives pour chaque fonction apparaissant dans la sous-catégorisation finale.

sont plutôt sujettes à idiosyncrasies, les règles (ii), sui sont représentées dans la MG comme des réalisations alternatives de FG, sont plutôt régulières.

En GB, les règles de type (i) sont représentées par des transformations locales vers des positions argumentales, ou bien sans transformations (elles sont générées dans la base); les règles (ii) sont représentées par des transformations cycliques vers des positions non argumentales. Les grammaires des années 80, qui utilisent l'unification, ont d'abord critiqué l'utilisation de transformations pour les règles (i), du fait du caractère moins régulier de celles-ci, et du fait que les sous-catégorisations résultantes existent par ailleurs dans la base.
LFG, à ses débuts (Bresnan, 1982) a introduit la notion de règles lexicales pour traiter les règles de type (i), celles de type (ii) y sont représentés comme des règles de réécriture augmentées d'équations pour construire la structure fonctionnelle. Bresnan et Kanerva (1989) ont ensuite proposé une évolution du modèle qui remplace les règles lexicales (pour représenter les règles de type (i)) par une correspondance directe entre le niveau sémantique et le niveau de représentation de la sous-catégorisation. Les règles sur la modification du nombre d'arguments sont déplacées vers le niveau sémantique.
HPSG utilise pour les règles de type (i) des règles lexicales. On peut cependant citer Dini (1995), qui a proposé une correspondance directe entre le niveau sémantique et le niveau de représentation de la sous-catégorisation, i.e. les traits de valence. Ces traits de valence acquièrent une valeur en fonction des rôles des arguments sémantiques.
Pour les règles de type (ii), leur traitement a varié. Par exemple pour l'extraction, Pollard et Sag (1987) ont d'abord proposé un traitement par schémas de DI, avec traces; puis par règles lexicales (Pollard et Sag, 1994, chapitre 9). Cette solution par règles lexicales a cependant été critiquée (voir par exemple Robichaud, 1995), du fait de l'utilisation de règles lexicales pour un processus très régulier, qui n'est du coup plus distingué de règles de type (i). Nous pensons également qu'il n'est pas justifié de traiter sur le même plan les deux types de règles, ne serait-ce que parce que les différentes réalisations de fonctions possibles (en utilisant nos termes) dépendent de la sous-catégorisation de surface du lexème, et donc des règles de type (i) qu'il a pu subir.
Plus récemment, Abeillé et al. (1998) proposent de construire des unités lexicales à partir de trois sources d'informations : 1) une représentation de la flexion et 2) le type de réalisation des actants du mot (le mot a ses actants en position canonique, ou bien certains sont cliticisés, ou bien certains sont extraits), 3) une représentation abstraite de la sous-catégorisation, qui compte tous les actants (qu'ils soient extraits, cliticisisés ou pas).

On peut donc conclure à un consensus sur la pertinence de la distinction entre redistributions et réalisations alternatives de fonctions (en utilisant nos termes). En pratique cependant, le classement des phénomènes selon les deux types de règles peut être discuté dans certains cas, et peut dépendre de la langue considérée. Nous avons retenu dans la MG principalement le critère de changement/non changement du nombre d'actants. Mais l'interprétation du critère peut être également sujette à discussion dans certains cas.
Prenons le cas des clitiques pronominaux. Nous suivons Miller (1991) pour le français et (Monachesi, 1995) pour l'italien en considérant qu'il s'agit d'affixes flexionnels. La question de savoir si le nombre d'actants est modifiée ou pas dépend de ce que l'on

appelle exactement actant. Nous considérons que les actants grammaticaux peuvent être réalisés en syntaxe ou en morphologie (comme clitiques) mais que dans les deux cas ils constituent des actants. Pour nous donc, les clitiques pronominaux constituent des réalisations alternatives de fonctions grammaticales, réalisations particulières en cela qu'elles s'opèrent au niveau morphologique. Cela est cohérent avec le fait que les verbes ne changent pas de sens selon que leurs actants sont pronominalisés ou pas, que la pronominalisation est une propriété plutôt régulière (elle fait partie des propriétés définitoires de bon nombre de FG), que la pronominalisation de l'objet direct ne modifie pas la transitivité d'un verbe. donnent des indications claires qu'il faut considérer les pronoms clitiques comme des réalisations alternatives de FG.

Un autre cas typique d'hésitation possible entre règles de type (i) et règles de types (ii) est celui où un actant peut ne pas être exprimé, toutes choses égales par ailleurs. C'est le cas par exemple pour un complément d'agent au passif. S'agit-il alors d'un cas particulier de redistribution, qui "efface" un actant de la sous-catégorisation, ou faut-il considérer qu'une des propriétés de la FG portée par l'actant est d'être optionnelle? Dans ce deuxième cas, cela revient à considérer qu'une réalisation particulière de fonctions grammaticale est la "non-réalisation". Dans le cas du complément d'agent, nous avons opté dans la MG du français et de l'italien pour la deuxième solution, sur la base des observations suivantes :

- la non-réalisation du complément d'agent n'est pas associée à un changement de sens du lexème. Dans le cas avec et sans agent, un agent est interprétable.
- elle n'est pas associée non plus à un changement de morphologie du lexème (contrairement par exemple au moyen, qui fait intervenir un clitique réfléchi fr. *se*, it. *si*).
- elle est en outre très régulière, c'est-à-dire que si un complément d'agent peut être réalisé, il peut très généralement être également omis[9].

Un autre cas litigieux est celui des vrais réfléchis dans les langues romanes (au sens constructions à sens réfléchi ou réciproque). Dans le cas d'un vrai réfléchi alternant avec un objet, nous rappellerons au Chapitre 4 que le réfléchi n'est pas à considérer comme un objet, car d'une part les verbes le portant sont intransitifs au regard des tests du causatif et de l'impersonnel, et d'autre part, le réfléchi peut être cooccurrent (pour certains locuteurs) à un clitique accusatif indéfini *en* comme dans *Il s'en lavait trois dans la fontaine*. On considère donc la réfléxivation dans ce cas comme un phénomène représentable par réduction de la sous-catégorisation (un cas particulier de redistribution, qui assigne une FG vide à un actant), comme le faisait par exemple Grimshaw (1982), dans le cadre de LFG. Un autre argument en faveur d'un traitement par redistribution est le fait que le réfléchi modifie la morphologie du lexème, en imposant l'auxiliaire de temps fr. ETRE, it. ESSERE. C'est d'ailleurs le seul argument restant, bien maigre, pour représenter également comme une redistribution le cas d'un réfléchi alternant avec un a-objet (objet indirect) (cf. Chapitre 4), s'il n'était également le souci de donner une représentation homogène des deux cas de réfléxivation.

[9] L'inverse n'est cependant pas vraie. On peut citer Zribi-Hertz (1982), qui note que l'expression du complément d'agent est exclue dans le cas d'un passif impersonnel pour un intransitif strict (cf. *Il a été dansé et ri (E + *par toute l'assistance*).

Au contraire, pour les réfléchis en anglais, le réfléchi portant clairement la fonction objet (ou objet indirect), il n'y a pas de raison de ne pas traiter les réfléchis par des règles de type (ii) (cf. Sells et al., 1987, pour un recensement de divers statuts du réfléchi selon les langues).
Enfin on peut également citer le cas du *si* impersonnel en italien. Il partage avec le réfléchi le fait d'imposer l'auxiliaire ESSERE, ce qui est plutôt une propriété de type "redistribution". Cependant, nous verrons Chapitre 4, qu'il a les propriétés d'un sujet, en termes distributionnels, d'où sa représentation dans la MG de l'italien, comme une réalisation alternative de sujet.

1.5.1 Pour la notion de sous-catégorisation initiale

Si la distinction entre redistributions et réalisations de fonctions paraît consensuelle (modulo les variations théoriques sur l'utilisation explicite/implicite des fonctions syntaxiques), il n'en va pas de même pour la distinction d'une sous-catégorisation initiale pour un lexème, qui peut ne pas être sa sous-catégorisation dans un emploi donné, i.e. sa sous-catégorisation finale.
Nous rejoignons en cela la Grammaire Relationnelle, voire GB, sauf que les fonctions grammaticales que nous considérons ne sont pas ramenées à des positions structurales, ce qui permet à notre sens de prétendre à plus d'universalité, étant donné l'existence de langues non configurationnelles.
Cependant la plupart des théories linguistiques nées des critiques de la grammaire transformationnelle s'attachent à ne donner qu'une description mono-stratale de la syntaxe (LFG, HPSG par exemple).
Mais d'une part on peut remarquer que les versions antérieures de LFG ou HPSG s'appuient sur des règles lexicales, qui peuvent modifier la sous-catégorisation (l'équivalent de nos redistributions).
Il nous semble exagéré de parler de modèles déclaratifs en utilisant l'astuce qui consiste à considérer que les règles lexicales (ou équivalents) peuvent avoir une interprétation purement déclarative. Les règles lexicales sont par définition ordonnées (avec une partie gauche et une partie droite). Les concepteurs de ces règles ont à notre avis à l'esprit une sous-catégorisation initiale des lexèmes, qui met en jeu le nombre maximal d'arguments sémantiques stricts du lexème (hors causatif donc), sur laquelle ils s'appuient pour définir les règles.
Le mérite de modèles comme LFG et HPSG est plutôt de circonscrire la procéduralité à un domaine précis de la grammaire.
Par exemple en HPSG, Abeillé et al. (1998) précisent :

> "We assume further that inflectional information always combines [...] monotonically. By contrast, any phenomenon involving a change in the number of arguments in the ARG-ST list will be treated as derivational, rather than inflectional, i.e. in terms of a relation between distinct lexemes."

Ainsi la spécification d'un lexème est effectivement déclarative (à quelques exceptions près[10]), mais au prix d'un éclatement plus grand du nombre de lexèmes. La procéduralité

[10] Nous avons relevé par exemple dans les travaux de Abeillé et al. (1998) le cas de la réfléxivation, pour laquelle les auteurs considèrent qu'il n'y a pas variation du nombre d'actants sous-catégorisés (dans la liste abstraite ARG-ST), et qui devrait donc recevoir un traitement monotone. On peut cependant remarquer que

a été circonscrite à un domaine précis de la grammaire, les règles lexicales. Et par un jeu de traits, on peut bloquer ou permettre certaines compositions de règles (c'est-à-dire séquences ordonnées d'applications de règles). Par exemple, Abeillé et al. (1998) donnent une règle lexicale pour le moyen (que les auteurs appellent medio-passif) où l'entrée porte le trait TRANS+ et la sortie le trait TRANS-. Seule la sortie est compatible avec la règle lexicale pour l'impersonnel. Cela revient à dire, de manière organisée, que pour un transitif, la règle du moyen doit s'appliquer avant la règle de l'impersonnel.

Dans le cas de la MG, la procéduralité est circonscrite au domaine d'expression des redistributions.[11]

Plus récemment, les règles lexicales en LFG ou HPSG ont tendance à être supplantées par un autre mécanisme pour gérer les différentes diathèses qu'un lexème peut montrer, comme on l'a cité plus haut (cf. Bresnan et Kanerva, 1989; Bresnan et Zaenen, 1990 pour LFG; ou Dini, 1995 pour HPSG) de calculer la sous-catégorisation à partir de la représentation de la sémantique lexicale. Les différentes diathèses découlent de structures argumentales différentes, la structure argumentale d'un lexème pouvant être modifiée. Le passif est par exemple représenté comme une démotion mais au niveau des rôles sémantiques : une démotion de l'argument agent. La notion de "fonction profonde" n'est alors pas pertinente.
Mais il nous semble que le fait que des redistributions s'appliquent à certaines expressions figées prouvent une certaine autonomie de la syntaxe, et la pertinence de la notion de fonction initiale. Comment assigner les fonctions syntaxiques aux actants d'une expression figée, qui sont non-référentiels[12], si cela est fait sur la base de rôles thématiques associés aux arguments? Il est certes vrai que les expressions figées n'admettent pas toutes des redistributions, mais il demeure que cela existe.

1.6 Questions liées à l'interface syntaxe/sémantique

Nous avons fait l'hypothèse simpliste que tous les arguments d'un lexème se réalisent comme des dépendants syntaxiques du lexème. Or nous avons déjà cité au Chapitre 1 des cas où cela n'est pas vrai. En particulier, typiquement un modifieur est un foncteur en sémantique (il gouverne sémantiquement au moins un argument), mais un de ses arguments se réalise comme le gouverneur syntaxique du modifieur.
Nous revenons donc sur les questions liées à la correspondance entre arguments sémantiques et actants syntaxiques, et nous verrons comment les cas "non simples" (i.e. où il est faux que tous les arguments se réalisent comme des dépendants syntaxiques) sont intégrés à la MG.

---

les auteurs ont besoin pour représenter les réfléchis d'introduire une valeur par défaut (ce que l'on peut considérer comme procédural) pour le trait représentant la transitivité.

[11] Pour la MG du français et de l'italien, dans quelques cas cependant nous avons dû utiliser le masquage de propriétés pour des phénomènes ne relevant pas de redistributions, comme par exemple la sélection de l'auxiliaire ESSERE dans le cas d'un SI-nominatif en italien (cf. Chapitre 4).

[12] Nous abordons infra, Section 1.6.3, la question des actants grammaticaux non référentiels.

1.6.1 La correspondance entre arguments sémantiques et actants grammaticaux

De nombreux travaux visent à expliciter des règles générales et sémantiques de correspondance[13] entre les arguments sémantiques et les actants grammaticaux qui les représentent. Le mécanisme pour représenter cette correspondance (qu'il intègre des règles générales de correspondance ou qu'il s'appuie entièrement sur une correspondance explicitée dans un lexique) ne nous concerne pas ici. Nous ne considérons pas à notre niveau de typage des arguments sémantiques (en rôles thématiques par exemple) pour exprimer les contraintes sur leur réalisation syntaxique[14]. Notre propos est d'explorer les généralisations exprimables en termes syntaxiques, même si certaines se ramènent à des généralisations faisant intervenir un classement sémantique des arguments.

Comme les actants grammaticaux sont caractérisés par une FG initiale, la correspondance entre arguments sémantiques et actants grammaticaux se ramène à une correspondance entre arguments sémantiques et FG initiales. Nous nous inspirons directement des travaux des débuts de LFG pour la définition de cette interface (Bresnan, 1982a)[15].
Il existe des contraintes de bonne formation d'un lexème, qui limitent les possibilités de correspondance entre les arguments sémantiques et les actants grammaticaux. La fonction d'association des FG initiales aux arguments sémantiques est (Bresnan (1982b:163) donne ces généralisations mais pour les fonctions de surface) :

- une fonction totale : chaque argument sémantique a exactement une FG initiale
- une injection : une même FG est la FG initiale d'au plus un argument sémantique

La première propriété n'est pas forcément vraie d'une sous-catégorisation finale, puisque certains arguments peuvent être démis (le sujet initial d'une forme verbale au moyen par exemple). LFG considère une fonction spéciale ϕ pour ces arguments, comme une fonction qui n'a pas de réalisation grammaticale, et énonce les propriétés ci-dessus pour la sous-catégorisation de surface. Cela suppose que ϕ puisse être portée par au plus un argument. Nous ne considérons pas que cela doive être forcément le cas, donc pour nous, la correspondance entre arguments sémantiques et FG finales n'est pas une fonction totale, certains arguments pouvant ne pas porter de FG finale. Mais c'est toujours une injection, c'est à dire qu'une sous-catégorisation initiale ou finale ne comporte pas de doublons.

Mais ni la fonction d'association des FG initiales, ni la fonction d'association des FG finales ne sont des bijections. En effet :

---

[13] Nous utilisons le terme de Jackendoff (1990). On trouve aussi souvent en anglais "linking" (par exemple (Levin et al., 1995)).

[14] Cependant, un tel typage de même que les moyens expressifs pour relier différents sémantèmes, au sein d'une langue ou entre langues, en permettant leur décomposition sont des composants nécessaires du module sémantique d'une théorie linguistique adéquate. Mais cela dépasse le cadre de notre travail.

[15] Plus récemment Bresnan et Kanerva (1989) ajoutent à LFG une théorie de la correspondance (*lexical mapping theory*), qui permet d'apparier les rôles sémantiques aux fonctions grammaticales (de surface) de manière déclarative, par sous-spécification de la fonction associée au rôle *patient/theme* qui, selon le contexte, peut être réalisé comme un sujet ou un objet.

- un argument sémantique d'un lexème peut se réaliser comme le gouverneur syntaxique du lexème, et ne pas porter alors de FG par rapport au lexème
- un actant grammatical peut ne correspondre à aucun argument sémantique (c'est un explétif, comme le *il* impersonnel en français, ou c'est une composante d'une expression figée), on parlera alors d'actant grammatical **non-référentiel**
- un argument du lexème peut être réalisé comme un actant d'un autre lexème

Nous détaillons ces trois cas de divergences dans la correspondance arguments/actants.

1.6.2 Le lemme principal n'est pas toujours la tête syntaxique

Nous avons défini la sous-catégorisation initiale d'un lemme principal comme une liste d'actants grammaticaux portant une FG initiale. Mais nous avons vu au Chapitre 1 la possibilité pour un argument sémantique de se réaliser en syntaxe comme un gouverneur syntaxique du lexème. Donc il faut prévoir cette possibilité dans la définition de la sous-catégorisation initiale. On distingue deux cas, selon que :

1. dans le cas non marqué, le lemme principal est la tête syntaxique de la structure comprenant tous les actants grammaticaux (c'est le cas d'un verbe par exemple)

2. dans le cas non marqué, un des arguments sémantiques se réalise par un actant grammatical qui est la tête syntaxique de la structure comportant le lemme principal et ses actants grammaticaux (c'est-à-dire si le lexème est un modifieur syntaxique, cf. Chapitre 1)

Dans le cas 1, on numérote les arguments par leur ordre d'oblicité dans le cas non marqué, en partant de 0 (par exemple pour un verbe, le sujet initial est l'argument 0).
Dans le cas 2, par convention, l'argument sémantique qui se réalise comme tête syntaxique est l'argument 0. Cet argument ne porte pas de fonction grammaticale par rapport à l'ancre, mais au contraire, l'ancre porte une fonction (de type modifieur) par rapport à lui. Les autres arguments sémantiques éventuels se réalisent comme dépendants syntaxiques de l'ancre, et portent une fonction initiale par rapport à l'ancre. Par exemple, on peut considérer que le lexème /FIDELE/ a deux arguments sémantiques (le fidèle, argument 0, et l'entité à qui on est fidèle, argument 1). Le cas non marqué peut être illustré en (7) (c'est l'emploi comme épithète), où la réalisation syntaxique de l'argument 0 gouverne syntaxiquement l'adjectif.

(7) les personnes fidèles à leurs convictions

L'argument 1 de l'adjectif porte une fonction complément par rapport à l'adjectif et le constituant fait de l'adjectif et son complément portent la fonction "modifieur" par rapport à l'argument 0.

Par commodité, on se référera à la sous-catégorisation initiale d'un lexème comme la liste de ses arguments avec fonction initiale, plus la fonction modifieur, **sachant qu'il s'agit en fait de la fonction inverse** (c'est le lexème qui est modifieur de l'argument 0).

Une solution alternative serait de considérer pour notre exemple d'un adjectif épithète que c'est le nom qu'il modifie qui est un sujet par rapport à l'adjectif (et se ramener au

cas général où tous les arguments portent une fonction par rapport à la réalisation du gouverneur sémantique. Cependant, cela constituerait un cas où un élément (le nom) porte une fonction par rapport à un autre (l'adjectif), et est cependant la tête syntaxique du tout (cf. le nom est la tête syntaxique du tout, qui est un SN, dont en outre les marques de genre et nombre sont ceux du nom).
De plus, considérer que le nom est le sujet de l'adjectif épithète oblige à considérer qu'un même élément peut porter plusieurs fonctions grammaticales : le nom pourrait être le sujet de plusieurs adjectifs épithètes, et il peut porter en plus une fonction par rapport à un verbe.
A notre avis, cette solution définit la fonction grammaticale comme correspondance directe avec la structure argumentale. Nous préférons garder la définition traditionnelle de fonction grammaticale, telle qu'un élément peut porter au plus une fonction, et ne peut constituer la tête syntaxique du constituant qui le contient[16]. La définition traditionnelle a l'avantage de rendre compte du fait qu'un élément en syntaxe a au plus un gouverneur, ce qui n'est pas le cas en sémantique.
Aussi, nous faut-il considérer qu'un sémantème peut se réaliser comme un dépendant syntaxique portant une fonction par rapport à un de ses arguments sémantiques.

1.6.3 Les actants grammaticaux non référentiels

Pour les actants grammaticaux non référentiels, la sous-catégorisation initiale doit prévoir la forme de la réalisation de ces actants. On révise donc la définition de la sous-catégorisation initiale de la manière suivante :

> La **sous-catégorisation initiale** est une liste d'actants grammaticaux caractérisés par leur fonction initiale, leurs catégories possibles, éventuellement leur cas lexical, *plus pour chaque actant grammatical non référentiel, la forme de sa réalisation*.

Copiant la notation de LFG, on note entre parenthèses la liste des actants grammaticaux référentiels, avec entre crochets leurs catégories possibles) et hors parenthèses les non référentiels, dont la forme est donnée entre crochets. On donne également les catégories possibles pour les actants. Ainsi par exemple pour *avoir du pain sur la planche*, la sous-catégorisation initiale est notée:

/AVOIR-DU-PAIN-SUR-LA-PLANCHE/ (sujet[N]) objet[*du pain*] locatif [*sur la planche*]

Il faut cependant introduire l'appareil adéquat dans le cas où les actants grammaticaux non référentiels peuvent varier selon les traits morphologiques des autres actants, où dans le cas où certains actants ont une des traits morphologiques fixés. On peut donc ajouter dans les entrées lexicales, des traits sur les actants grammaticaux (spécifiant une valeur ou une variable pour coindexation).

[16] On laisse cependant la possibilité qu'un autre élément Z, extérieur à la relation entre un élément X portant une fonction par rapport à Y, soit la tête syntaxique. Par exemple, dans le cas d'un verbe à montée du sujet, comme dans *Paul semble rêver*, on considère que *Paul* est le sujet de *rêver*, la tête syntaxique du tout étant cependant le verbe à montée *sembler* (cf. Chapitre 3). Les verbes à montée sont considérés comme déficients, en cela qu'ils ne sous-catégorisent pas d'éléments.

Nous pouvons répéter un argument en faveur de la représentation des redistributions en syntaxe (et pas par une correspondance sémantique/syntaxe). Il est à noter que des actants grammaticaux non référentiels peuvent être affectés par une redistribution. Il faut donc considérer que celles-ci s'appliquent sur les actants grammaticaux et pas sur les arguments sémantiques.

Dans l'implémentation actuelle de la MG, et de la compilation MG→TAG, nous nous appuyons sur les lexiques TAG. Les MG du français et de l'italien implémentées ne comprennent aucun arbre spécifiquement pour une expression figées. Les seules expressions figées représentables actuellement sont celles qui ne requièrent que la lexicalisation d'une position à substituer, d'un schème existant par ailleurs (en particulier, on ne dispose pas des arbres qui comportent directement les noeuds pour le déterminant, en cas de déterminant figé). Cette lexicalisation est réalisée dans le lexique syntaxique (au format XTAG).

1.6.4 Un argument du lexème peut être réalisé comme un actant d'un autre lexème

Nous revenons maintenant sur le fait qu'un argument d'un lexème peut être réalisé comme actant d'un autre lexème. C'est-à-dire que l'hypothèse d'une stricte localité dans l'interface sémantique lexicale / syntaxe est fausse.
On a vu déjà au Chapitre 1 ce genre de disparités (en étudiant les différentes configurations pour les dépendances sémantiques et syntaxiques entre deux lexèmes).
Nous étudions plus en détail des cas de divergences entre dépendances sémantiques et dépendances syntaxiques au Chapitre 7. Mais il est intéressant de citer d'ores et déjà, outre des cas relativement marginaux, dépendant des lexèmes considérés, on peut relever des cas productifs en syntaxe, qui sont la montée du sujet ou de l'objet, et le contrôle[17].

Pour le cas des verbes à montée du sujet et de l'objet, le principe de cooccurrence prédicat-arguments, dans sa version sémantique, ne permet pas que les actants grammaticaux non gouvernés sémantiquement par le verbe à montée apparaissent comme position à étendre dans l'arbre élémentaire de ce verbe à montée. D'où le traitement des verbes à montée du sujet par adjonction sur une projection verbale non phrastique, étendu pour l'anglais (XTAG group, 1998) aux verbes à montée de l'objet. Cette analyse, qui privilégie l'interface avec la sémantique, le fait au détriment de régularités syntaxiques, c'est-à-dire que l'objet, même non gouverné sémantiquement localement, a des propriétés d'objet (bien que souvent il n'ait pas toutes les propriétés d'objet) (cf. Chapitre 3, et Chapitre 7).
Pour le français et l'italien, suivant Abeillé (1991), nous gardons l'analyse des verbes à montée du sujet par adjonction. Ainsi, le "sujet" du verbe à montée du sujet n'est pas considéré comme faisant partie de sa sous-catégorisation. En revanche pour les verbes à montée de l'objet (rares) on permet exceptionnellement une sous-catégorisation avec un actant grammatical référentiel (donc ouvert, et représenté par une position à étendre en

[17] On les considère comme productifs étant donné que la montée et le contrôle peuvent s'appliquer a priori pour tout verbe enchâssé sous certains verbes particuliers (à montée ou à contrôle).

TAG) ne corresponde pas à un argument sémantique. D'où la sous-catégorisation initiale suivante pour /EMPECHER/ :

/EMPECHER/ (sujet[N], de-objet[P]) objet[N])

Nous discutons au Chapitre 3, de la possibilité de considérer que l'objet peut avoir tendance dans certains cas à prendre le statut d'argument sémantique du verbe enchâssant, on se ramène alors à une configuration de type "contrôle".

Pour les phénomènes de contrôle en effet, la différence est qu'un actant grammatical réalisé (phonologiquement) est l'argument sémantique à la fois du lexème enchâssant et de celui enchâssé. En TAG, on ne peut pas localiser cet actant dans l'arbre de ses deux gouverneurs sémantiques. On choisit de prévoir une position pour l'argument partagé au niveau du verbe enchâssant. Pour l'arbre du verbe enchâssé, on considère que l'argument non réalisé localement est représenté par une trace (i.e. que l'infinitive est représentée comme une phrase à sujet vide). Une justification technique de cette représentation est fournie au Chapitre 5, dans la section sur la représentation des actants phrastiques.
Outre le cas de verbes à contrôle, comme en (8a), où les deux verbes partagent un argument sémantique ('Jean'), on peut citer également d'autres situations que l'on peut qualifier de contrôle, comme par exemple en (8b), où un verbe locatif (ici de mouvement) et une préposition locative (ici *à*) partagent un argument. On peut encore combiner les deux, en (8c) où l'argument *Paul* est partagé par le verbe principal, la préposition locative et le verbe infinitif de l'infinitive de but.

(8) a. Jean veut partir.
b. Paul va à la campagne.
c. Paul court à la boulangerie chercher du pain.

Dans le cas de contrôle de l'argument sémantique d'un lexème, on ne peut donner qu'une seule position à étendre pour l'actant grammatical correspondant à l'argument partagé. Dans les analyses TAG, cette position est prévue dans l'arbre élémentaire pour le lexème gouvernant en syntaxe (le verbe enchâssant). Les arbres élémentaires ancré pas l'infinitif en (8a), la préposition locative en (8b) (et les deux en (8c)) ne respectent pas le PCPA à la lettre, puisqu'un des arguments sémantiques n'a pas de position locale.
On choisit dans les TAG française et italienne de représenter cet argument non réalisé localement par une trace (nous revenons sur ce point au Chapitre 5, dans la section sur la représentation des actants phrastiques).

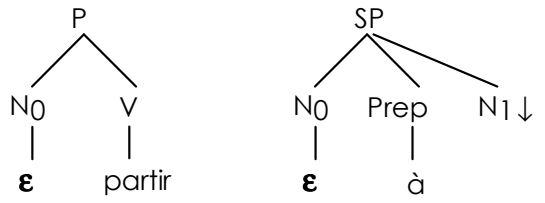


Figure 23. Arbre élémentaire pour infinitive et arbre pour préposition locative sans arg0

Dans la MG, cela est représenté par des réalisations grammaticales particulières pour les fonctions grammaticales concernées. Ainsi, pour représenter les infinitives, on considère qu'elles illustrent un cas particulier de réalisation du sujet (la non-réalisation). On

distingue différents cas de "non-réalisation", pour distinguer les infinitives des impératives, et en italien le cas du pro-drop. On verra au Chapitre 4, dans la description des réalisations grammaticales du sujet ce qu'il en est pour les participiales. Pour les prépositions locatives nous renvoyons au Chapitre 6.

Nous venons de détailler des cas où un argument est réalisé par un actant qui porte une GF à un niveau syntaxique enchâssant. Inversement, dans le cas d'un lexème pour une expression figée, un argument du lexème peut être réalisé par un actant grammatical qui porte une GF non pas par rapport au lemme principal (le verbe), mais par rapport à un actant grammatical enchâssé, non référentiel. Par exemple pour l'expression *N0 se payer la tête de N1*, l'argument 1 est complément de nom de l'actant non référentiel *tête*. Il a les réalisations régulières d'un complément, y compris le possessif (*Jean s'est payé ma tête*).
TAG offre pour ces expressions figées une grande souplesse, en permettant de construire directement les arbres élémentaires appropriés (Abeillé, 1991). Dans la MG, il nous faut introduire un paramètre supplémentaire pour l'expression de la sous-catégorisation d'un lexème : on permet d'indiquer explicitement le gouverneur syntaxique d'un actant, lorsqu'il ne s'agit pas du lemme principal. D'où la sous-catégorisation initiale suivante pour *N0 se payer la tête de N1* :
/SE-PAYER-LA-TETE(sing)/
(sujet[N], complt-de-nom[N/Det]{gouverneur = objet}**)** objet[tête (dét+, déf+, sing)]

Dans l'implémentation actuelle des MG du français et de l'italien, les schèmes avec actants figés enchâssés ne sont pas représentés.

1.7 Une formalisation du PCPA, en termes fonctionnels

Nous avons donné au Chapitre 1 une version sémantique du principe de cooccurrence prédicat-arguments, qui stipule qu'un arbre élémentaire comprend des positions à étendre pour tous les arguments exprimés du lexème ancrant l'arbre.
Avec les considérations sur la correspondance arguments sémantiques / actants grammaticaux, et la représentation de la sous-catégorisation (syntaxique) en termes fonctionnels, nous sommes en mesure de préciser ce principe, en y intégrant des contraintes syntaxiques.
Nous pouvons maintenant préciser que les arguments exprimés sont les actants grammaticaux non référentiels du lexème. Les cas où un actant référentiel n'est pas un argument sémantique du lexème sont traités par adjonction (cf. l'adjonction d'un verbe à montée du sujet). Le cas particulier de l'argument sujet d'une infinitive est représenté par uns position trace (ce qui n'est effectivement pas une position à subsituer).

Le PCPA est formalisé dans la MG par le principe (parallèlement à LFG) suivant :

Principe de cohérence/complétude/unicité : Un arbre élémentaire, ancré par un lexème, illustre un emploi de ce lexème avec une sous-catégorisation finale, qui est dérivable par une suite de 0, 1 ou plusieurs redistributions à partir de la sous-catégorisation initiale du lexème. En outre, l'arbre élémentaire doit contenir

exactement une position à étendre pour chaque fonction de cette sous-catégorisation finale qui est portée par un actant référentiel et qui est réalisée en syntaxe.

1.8 Des lexiques possibles pour la MG

La MG ne comprend pas à l'heure actuelle de lexiques propres. L'implémentation réalisée génère des schèmes, et repose sur le format des lexiques à la XTAG, avec un lexique morphologique associant forme fléchie à lemme, et un lexique syntaxique, associant lexème (unité éventuellement polylexicale) à une famille de schèmes. La notion de famille dans la MG a été formalisée comme tous les schèmes qui partagent la même sous-catégorisation initiale (avec parfois précision des catégories possibles pour les actants grammaticaux, et également de certains traits morpho-syntaxiques[18]).

On peut cependant imaginer différents types de lexiques pouvant être utilisés. La condition est de pouvoir récupérer à partir des informations du lexique une représentation fonctionnelle de la sous-catégorisation, de manière à pouvoir associer les schèmes appropriés.

Nous pouvons donner un exemple d'organisation possible pour le lexique à "brancher" à une MG.
Lexique morphologique + **lexique syntaxique** comme un ensemble d'**entrées lexicales**, tel que chaque entrée lexicale comprend les informations suivantes:

Entrée lexicale :

lexème +
- sémantème et liste d'arguments sémantiques
- sous-catégorisation initiale
- correspondance entre arguments sémantiques et actants grammaticaux

Exemples:

Lexème	/RIRE/	/ACCROCHER/
Catégories	V (aux avoir)	V
Structure prédicat-arguments	'rire' (0)	'accrocher' (0, 1)
Association	|	| |
Sous-catégorisation initiale	(sujet[N])	(sujet[N], objet[N])

Il peut être également intéressant de permettre à l'exécution qu'un lexème prenne une sous-catégorisation initiale non prévue pour ce lexème-là. Cela permettrait le traitement d'emplois "créatifs" comme l'exemple anglais (9) (Mandelblit, 1997) (ces emplois créatifs qui semblent plus fréquents en anglais que dans des langues comme le français

[18] Ainsi par exemple pour le français, pour les sous-catégorisations intransitives (sans objet direct), on a systématiquement considéré deux familles, selon l'auxiliaire de temps considéré (cf. Chapitre 3).

ou l'italien). La difficulté demeure bien sûr de contrôler ces sélections "créatives" de sous-catégorisations.

(9) She trotted the horse into the stable. (Mandelblit, 1997).
Elle trotta le cheval dans l'écurie.
(Elle mena le cheval qui trottait dans l'écurie)

2 L'appareil formel utilisé

La MG reprend la proposition de Vijay-Shanker et Schabes (1992) d'utiliser, pour représenter une TAG, un graphe d'héritage où pour chaque classe peut être spécifiée une **description partielle d'arbres** (Rogers et Vijay-Shanker, 1994)[19]. En revanche, nous n'utilisons pas explicitement de règles lexico-syntaxiques (cf. Section 3.4). Nous détaillons ces notions ci-dessous.

2.1 Un graphe d'héritage multiple et (presque) monotone

Les structures élémentaires de la MG sont donc des classes organisées en graphe d'héritage.

Empruntés aux travaux en informatique sur la programmation orientée objet, les graphes d'héritage ont été utilisés en linguistique computationnelle, principalement pour la représentation du lexique (cf. par exemple les travaux en Construction Grammar (voir par exemple Koenig, 1994[20]), HPSG (Pollard et Sag, 1994), ou le lexique génératif de Pustejowsky (1995)).

Nous rappelons les définitions formelles liées à la notion de graphe d'héritage, ainsi que l'utilisation de l'héritage en linguistique computationnelle, avant de décrire celui qui est utilisé dans la MG.

2.1.1 Héritage : définition formelle et exemples d'utilisation en linguistique

Si une classe[21] désigne un ensemble de propriétés, on peut considérer l'ordre partiel défini par la relation de subsomption entre classes:
Une classe A subsume une classe B si l'ensemble des propriétés de A est inclus dans l'ensemble des propriétés de B. La classe A est alors une **superclasse** de B. C'est une

[19] Vijay-Shanker et Schabes (1992) proposaient d'utiliser le langage des descriptions partielles décrites par Rogers et Vijay-Shanker (1992). Nous avons utilisé la version plus récente de (Rogers et Vijay-Shanker, 1994). La différence réside principalement dans la définition des quasi-arbres, initialement vus comme des modèles partiels d'arbres, puis comme des cas particuliers de descriptions partielles.
[20] Les travaux en Construction Grammar ont pour ambition de représenter toute la grammaire dans une structure hiérarchisée. Cela reste cependant principalement appliqué dans le cadre de descriptions lexicales.
[21] Nous utilisons le terme classique en programmation objet, car il nous semble prêter moins à confusion que le terme "type", également employé en représentation des connaissances.

superclasse indirecte s'il existe une classe C, distincte de A et B, telle que A subsume C et C subsume B. Dans le cas contraire c'est une superclasse directe. Plus généralement, on dit que B est plus spécifique que A et que A est plus générale que B.
Pour représenter un domaine de connaissances, on peut classifier ses éléments comme autant d'instances des classes définies dans la hiérarchie. Un élément est une **instance** d'une classe si il en a toutes les propriétés. On parle de "la" classe d'un élément, comme la classe la plus spécifique dont l'élément est une instance. Il faut distinguer la relation d'instanciation et la relation de subsomption. Les classes modèlent un domaine, et sont organisées selon une hiérarchie. Les instances sont les objets du domaine considéré, elles sont associées à une classe.

Il s'agit d'une formalisation d'un principe de classification généralisé en représentation des connaissances. On peut par exemple citer la classification du vivant (Figure 24).

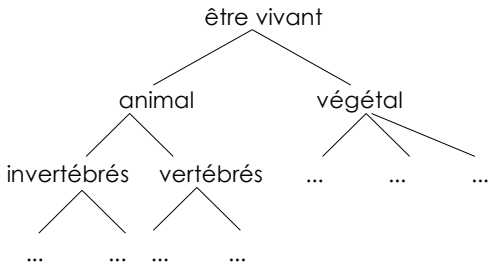


Figure 24. La classification hiérarchique (partielle...) des êtres vivants

A partir de ce principe déclaratif de classification, l'héritage peut être défini comme un processus dynamique, de transmission des propriétés à partir de relations de subsomption assertées[22]. Pour chaque classe, en disposant d'un principe d'héritage, on peut ne spécifier que les superclasses directes et les **propriétés locales**, c'est-à-dire les propriétés de la classe qui ne sont pas des propriétés des superclasses. Une instance *a*, dont la classe est A, a par définition des propriétés locales de A, ainsi que les propriétés locales de toute superclasse de A, directe ou indirecte, on parle alors de propriétés locales et de **propriétés héritées**. On rappelle ci-dessous les critères qui définissent différents types d'héritage, et différents types de graphes d'héritage.

L'interprétation dynamique d'une classification hiérarchique (i.e. le processus d'héritage), est utilisé pour faire physiquement partager des propriétés à un ensemble d'objet. Par exemple, pour la représentation de la sous-catégorisation d'un verbe, on peut faire partager la propriété "a un objet direct" à tous les cadres de sous-catégorisation comprenant un objet.

Héritage strict / héritage par défaut

L'héritage strict correspond à la définition donnée plus haut : toutes les propriétés d'une super-classe sont héritées. Une conséquence est que si une propriété locale d'une classe est incompatible avec une propriété héritée, ou si plusieurs propriétés héritées de

[22] Il existe des systèmes de représentation des connaissances (dérivant de KL-ONE) dans lesquels les liens de subsomption sont inférés dynamiquement à partir des propriétés.

différentes superclasses sont incompatibles, on a un paradoxe et la hiérarchie n'est pas cohérente.

Avec l'héritage par défaut, les propriétés d'une classe sont révisables. Elles ne seront héritées que si elles ne sont pas incompatibles avec les propriétés amenées par des classes plus spécifiques. La révision de propriétés est généralement utilisée pour le traitement des exceptions. En haut de la hiérarchie, on a les propriétés générales, qui sont révisables au fur et à mesure que l'on descend dans la hiérarchie pour arriver aux cas particuliers. On donne ci-dessous l'exemple simpliste d'une représentation des verbes en anglais, avec révision de la propriété "forme du participe" : au niveau général est spécifié que le participe est en /ed/, mais certains verbes (comme *to beat*) peuvent réviser cette propriété.

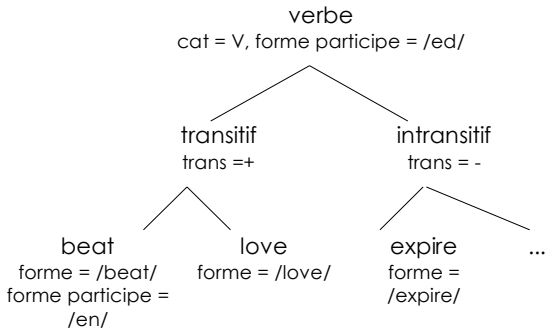


Figure 25. Exemple d'héritage non monotone

La possibilité de réviser des propriétés pour ce que l'on considère être des exceptions, s'accompagne d'un inconvénient : on n'a plus forcément inclusion des propriétés d'une super-classe dans les propriétés d'une sous-classe[23], la hiérarchie est donc moins directement compréhensible.

Héritage simple / héritage multiple

Si une classe ne peut avoir qu'une seule superclasse directe, on parle d'héritage simple. Le graphe représentant les relations de subsomption est un arbre (ou plusieurs arbres non connexes). Cependant, certains cas nécessitent qu'une classe puisse hériter de plusieurs superclasses directes. On parle alors d'héritage multiple.
Formellement la hiérarchie est plus complexe. Le graphe d'héritage est un graphe acyclique orienté (il n'y a aucune raison de permettre les cycles, qui intuitivement correspondent à des chemins dans le graphe d'héritage qui n'apportent aucune information, et en pratique font exploser inutilement la complexité).

L'héritage multiple est utile si les objets à représenter ont des propriétés indépendantes les unes des autres. On peut représenter ces propriétés par des hiérarchies indépendantes, appelées **dimensions**, et définir des objets tels qu'ils doivent pour être bien formés, hériter de propriétés appartenant à des dimensions bien définies. On parle alors d'héritage multiple **orthogonal**.

[23] On ne peut donc plus parler de relations de subsomption; la hiérarchie n'est faite que de liens d'héritage.

Par exemple, les propriétés qui caractérisent un verbe peuvent être distinguées en différents domaines d'information, comme la classe flexionnelle (le groupe en français) et la sous-catégorisation par exemple. Une classification croisée (i.e. par héritage multiple orthogonal) est naturelle pour représenter ces données.

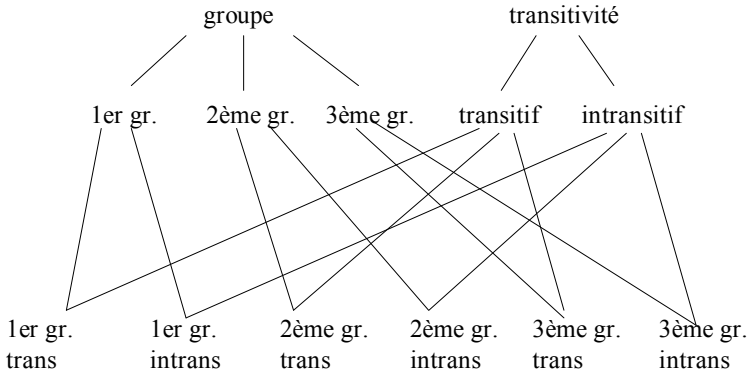


Figure 26. Exemple d'héritage multiple

Il faut alors distinguer des classes **disjonctives**, telles que leurs sous-classes sont mutuellement exclusives (i.e. une classe ne peut pas hériter de deux sous-classes d'une classe disjonctive), des classes **conjonctives** (Koenig et Jurafsky, 1994), qui dominent la racine des différentes dimensions. Une classe conjonctive détermine les types de propriétés (partitionnées en dimensions différentes) dont doit hériter un objet pour être bien formé[24]. Ainsi si l'on reprend l'exemple ci-dessus de classification croisée selon groupe morphologique et transitivité, on peut considérer une classe "verbe", qui stipule qu'une description de verbe, pour être bien formée, doit hériter d'une classe de groupe et d'une classe de transitivité. Les classes groupe et transitivité sont elles disjonctives (un verbe appartient à seul groupe à la fois).

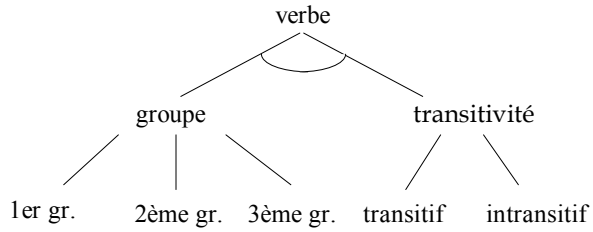


Figure 27. Exemple d'une classe conjonctive (verbe), et de deux classes disjonctives (groupe, transitivité)

Il existe aussi l'héritage multiple non restreint, où une classe peut hériter directement de plusieurs classes correspondant à des propriétés non-disjointes. Il faut alors gérer les "conflits" entre propriétés : une classe peut être atteinte par différents chemins dans le graphe d'héritage, ce qui peut entraîner l'incohérence de la hiérarchie si ces différents chemins véhiculent des propriétés incompatibles. Un moyen de gérer ces conflits est l'**héritage prioritaire** (aussi appelé linéarisation). Avec l'héritage par défaut on a vu

[24] On représente les classes conjonctives en ajoutant un arc de cercle reliant les différentes dimensions. Flickinger (1987) parle de *subset link* pour les liens entre une classe et une sous-classe qui en représente un sous-ensemble (dans l'optique de l'héritage strict) et de *perspective link* pour un lien entre une classe A et une sous-classe qui en représente une dimension.

comment on peut éviter un conflit en masquant une propriété générale par une propriété plus spécifique. Cette solution se base sur le fait qu'en héritage simple, les superclasses d'une classe sont totalement ordonnées, de la plus spécifique à la plus générale. Or on n'a plus cet ordre total avec l'héritage multiple. L'héritage prioritaire consiste à définir un ordre total de spécificité sur toutes les superclasses d'une classe (suivant un algorithme d'ordonnancement), pour pouvoir ensuite utiliser le masquage de propriétés, et donc éviter les conflits.
Un exemple d'algorithme d'ordonnancement peut être celui de Common Lisp Object System (CLOS, voir Steele, 1990)[25]. Sans détailler tout l'algorithme, nous pouvons donner les trois règles suivies pour ordonner les super-classes d'une classe de la hiérarchie :

1 Une classe est plus spécifique que sa ou ses superclasses directes.

2 Les superclasses directes d'une classe A sont ordonnées de la plus spécifique à la plus générale (ordre inclus dans la définition de la classe A)

3 On isole les sous-hiérarchies si elles sont indépendantes.

Considérons par exemple la hiérarchie illustrée Figure 28, pour une classe (C1 par exemple) on note de gauche à droite les superclasses directes de la plus spécifique à la plus générale.

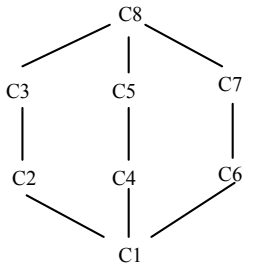


Figure 28. Un exemple de hiérarchie à héritage multiple

L'algorithme de CLOS donne comme liste ordonnée des super-classes de C1 : C1 C2 C3 C4 C5 C6 C7 C8. En effet, d'après la règle 1 :

C1 < C2 C2 < C3 C3 < C8
C1 < C4; C4 < C5 C5 < C8
C1 < C6 C6 < C7 C7 < C8

d'après la règle 2

C2 < C4 < C6

et d'après la règle 3 :

C3 < C4 <C7

Si on veut que la hiérarchie conserve une certaine lisibilité, il faut bien sûr que le calcul de l'ordre des superclasses soit bien clair. D'autre part la construction de cet ordre total exclut certains types de configurations, comme par exemple celles de la Figure 29 (tiré de Russel et al., 1993) où l'on doit avoir à la fois C < B < A et A < B.

[25] Le lexique d'ELU (Environnement Linguistique d'Unification) déjà cité utilise l'héritage prioritaire avec l'algorithme d'ordonnancement des classes de CLOS, (Russel et al., 1993).

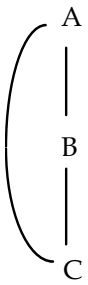


Figure 29

Structure de traits typées

L'utilisation de structures de traits organisées en graphe d'héritage (comme cela est fait par exemple en HPSG) se fait en définissant des types (ce que nous appelons des classes) de structures de traits (définis formellement par Carpenter, 1992). Les types expriment des contraintes sur la présence de certains attributs et/ou sur la valeur de certains attributs. Ces contraintes sur la valeur de certains attributs peuvent être exprimées en fournissant une valeur ou en fournissant le type dont doit être une instance la valeur de l'attribut. Une structure de traits typée, comme instance d'un type, doit satisfaire les contraintes du type locales ou héritées.

2.1.2 Héritage dans la MG

Si l'application de l'héritage pour la représentation du lexique est désormais courante, la représentation hiérarchisée de structures syntaxiques l'est moins. L'idée originale de l'application des outils de représentation du lexique à la représentation d'une TAG (et donc de structures syntaxiques) est de Vijay-Shanker et Schabes (1992). La MG s'appuie en partie sur leurs propositions (nous avons également cité les autres travaux de représentation compacte de TAG, au Chapitre 1; nous faisons une comparaison plus poussée, Section 3.4).

La MG utilise l'héritage multiple, puisque nous considérons qu'un schème pour être bien formé doit hériter d'informations sur la sous-catégorisation initiale, la séquence de redistributions qui dérivent la sous-catégorisation finale illustrée par le schème (y compris le cas sans aucune redistribution), et les réalisations de fonctions. On peut représenter cela en considérant une classe abstraite "schème", d'où partent trois liens d'héritage conjonctifs, vers les 3 dimensions (Figure 30). Cependant, la bonne formation des schèmes est contrôlée de manière plus fine, puisque ce dont un schème hérite en dimension 3 dépend de la sous-catégorisation finale, qui dépend de la sous-catégorisation initiale et des redistributions appliquées.

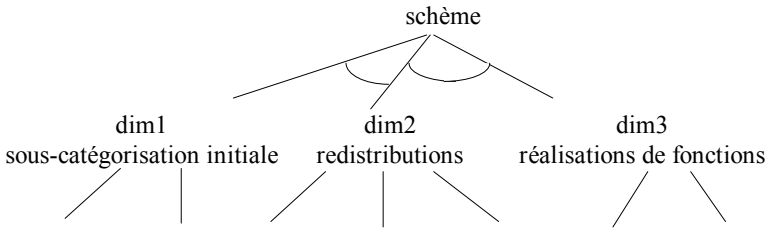


Figure 30. Les trois dimensions composant un schème

Les dimensions que nous avons retenues pour représenter l'ensemble des schèmes, sont représentés comme trois graphes d'héritage, en général disjoints, bien que l'on permettent que certaines classes servent comme superclasses pour représenter des phénomènes dans plusieurs dimensions[26].

L'utilisation d'un graphe d'héritage est classique pour représenter la sous-catégorisation. La sous-catégorisation étant en effet un objet de type liste, il est naturel de faire partager les éléments communs (i.e. la spécification des actants grammaticaux) à différentes listes. Ainsi toutes les sous-catégorisations comportant un objet, peuvent partager l'information "sous-catégorise un objet".

La MG utilise également un graphe d'héritage pour représenter la dimension 2, les redistributions, ce qui est assez original. Cela nous permet de décomposer les redistributions permises par une langue en redistributions partielles, qui mettent en jeu le changement de fonction d'un seul actant, et de faire partager ces redistributions partielles par plusieurs redistributions. Par exemple, pour la MG du français, la promotion de l'actant portant la fonction objet vers la fonction sujet est utilisée pour le passif personnel ainsi que pour le moyen personnel.

Enfin, la représentation en graphe d'héritage des réalisations de fonctions est elle aussi originale (mais là l'idée est de Vijay-Shanker et Schabes, 1992), en cela qu'il s'agit de représenter hiérarchiquement des structures syntaxiques.

Au sein de la MG, l'héritage des informations sur la topologie des arbres est monotone, mais on permet la révision de propriétés pour certains traits linguistiques décorant les arbres, notamment le trait codant la fonction grammaticale, ce pour représenter la redistribution des fonctions.
Les éventuels conflits entre propriétés héritées ne sont pas résolus, au contraire ils sont un des moyens utilisés pour représenter des incompatibilités entre classes (cf. Section 3.3).
On utilise cependant l'héritage prioritaire non pas pour la gestion des conflits, mais pour garantir l'ordre dans lequel la spécification des informations se fait. Cela est nécessaire pour représenter les redistributions, qui sont par définition procédurales et dépendantes d'un ordre d'application (voir Section 3.1).

[26] C'est le cas par exemple des classes qui représentent le clitique *si* en italien. Une classe générale (qui introduit la propriété de sélection de l'auxiliaire de temps ESSERE) sert à la fois comme superclasse à la morphologie moyenne (entrant dans la définition de la redistribution représentant le moyen) et comme superclasse au SI-nominatif (une réalisation particulière de la fonction sujet).

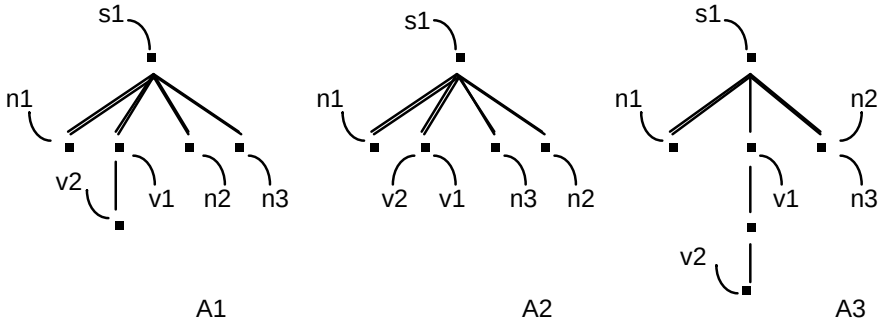
2.2 Descriptions partielles d'arbres

Formellement, Rogers et Vijay-Shanker (1994) définissent une description partielle comme un ensemble fini de formules, exprimées dans un sous-langage de la logique du premier ordre. Ce sous-langage n'utilise ni variables, ni fonctions, ni quantificateurs, mais un ensemble fini de constantes (qui sont donc les seuls termes), liées par 4 prédicats binaires : précédence linéaire (notée <), parenté ($\rightarrow$), dominance ($\Rightarrow$) et égalité ($\approx$).

Une description partielle D caractérise un ensemble d'arbres, c'est-à-dire les modèles qui satisfont à la fois les formules de D et les axiomes des arbres. Pour un modèle donné — un arbre — est définie une fonction d'interprétation qui associe à une constante un noeud dans l'arbre. On dira qu'une constante nomme ou désigne un noeud du modèle. Les relations de parenté et précédence linéaire ont l'interprétation classique sur les arbres. Il est en outre possible de spécifier dans une description une relation de dominance (notée $\Rightarrow$) entre deux constantes sans spécifier une chaîne de relations de parenté entre elles, par exemple entre les constantes v1 et v2 dans la description D1 ci-dessous.

$$\text{D1:} \quad (s1 \rightarrow n1) \wedge (s1 \rightarrow v1) \wedge (s1 \rightarrow n2) \wedge (s1 \rightarrow n3) \wedge (n1 < v1) \wedge (v1 \Rightarrow v2) \wedge (v2 < n2) \wedge (v2 < n3)$$

Dans un arbre satisfaisant D1, un chemin de longueur supérieure ou égale à zéro doit être défini entre les deux noeuds nommés par v1 et v2 (le chemin peut être de longueur nulle, v1 et v2 nomment alors le même noeud). Deux constantes nomment le même noeud si et seulement si elles sont égales, c'est-à-dire liées par la relation $\approx$. Cette égalité peut être spécifiée dans la description, ou bien peut être inférée de par les axiomes des arbres. Ainsi par exemple, si deux constantes dominent directement une même constante alors les deux premières sont égales (la conjonction $(s1 \rightarrow n1) \wedge (s2 \rightarrow n1)$ implique $s1 \approx s2$).
On donne Figure 31 des exemples d'arbres satisfaisant D1[27]. On peut remarquer qu'un noeud peut être nommé par 0, 1 ou plusieurs constantes. On ne considère dorénavant que des modèles ou tout noeud est nommé par au moins une constante (c'est le cas de A1 et A2 ci-dessous), il s'agit de modèles **référents**. Comme la description partielle est finie, l'ensemble des référents est finie.



[27] Les noeuds sont représentés par des carrés. Les constantes nommant un noeud y sont reliées par des petits arcs.

Figure 31. Exemples d'arbres satisfaisant la description D1

Ajouter des formules à une description réduit de manière monotone l'ensemble des arbres satisfaisants. Par exemple si on ajoute (v1 $\rightarrow$ v2) alors parmi les arbres de la Figure 31 seul A1 reste satisfaisant.

La MG représente les structures syntagmatiques comme des conjonctions de descriptions partielles : la description partielle associée à une classe est la conjonction de toutes les descriptions partielles héritées et de la description spécifiée localement (au niveau de la classe).
Grâce à la relation de dominance, on peut laisser sous-spécifié le chemin entre deux noeuds à un niveau général du graphe d'héritage, chemin qui sera éventuellement spécifié dans une classe soit plus spécifique, soit orthogonale. On peut également laisser sous-spécifié l'ordre relatif de deux noeuds (par exemple les constantes n2 et n3 de la description D1 ne sont pas ordonnées). On dispose ainsi de l'appareil formel pour dissocier dominance et précédence, ce qui rejoint d'autres théories linguistiques comme HPSG, ou GPSG.

La démarche est de représenter au final des schèmes comme combinaison d'unités plus petites exprimées sous la forme de descriptions partielles. On peut remarquer que c'est ce qui est fait dans un formalisme basé sur des règles de réécriture : un arbre dérivé est la combinaison d'arbres de profondeur 1 (les règles de réécriture) qui toutes décrivent partiellement l'arbre dérivé. Cependant la différence ici est que cette combinaison n'est pas faite lors de l'utilisation effective de la grammaire, mais avant. De ce fait, on n'est pas contraint par des exigences d'efficacité. Les unités combinées (les descriptions partielles d'arbres) offrent un pouvoir expressif plus grand que les règles de réécriture qui sont limitées à une profondeur 1 et qui spécifie totalement les relations des noeuds qu'elles introduisent.

2.3 Structures de traits associées aux constantes

Dans le cadre de la MG, les constantes des descriptions partielles sont augmentées de structures de traits : les traits déjà présents dans le formalisme des TAG, plus des traits propres à la MG, que nous nommons méta-traits (car ils ne vont pas apparaître sous la forme de traits[28] après la compilation en une TAG). Ces méta-traits sont entre autres :

- `pos` liste des parties du discours possibles
- `ind` indice
- `fonction` fonction dans la sous-catégorisation finale
- `fonction-i` fonction dans la sous-catégorisation initiale
- `qualité` type d'un noeud frontière
  (noeud à substituer, noeud pied, ancre, co-tête)

[28] La catégorie et l'indice apparaissent dans le système XTAG dans le nom des noeuds des arbres élémentaires. La qualité apparaît comme un trait. Les traits concernant les fonctions syntaxiques ne sont pas utilisés en TAG classique; ils sont utilisés dans G-TAG (Danlos & Meunier 1996), un système de génération basé sur les TAG.

Le metatrait :ind est utilisé pour ajouter le numéro d'argument sémantique aux positions argumentales et leurs projections éventuelles.[29]
Les constantes (au sens logique) utilisées dans les descriptions partielles sont des variables globales (au sens informatique) sur tout le graphe d'héritage : le même identificateur de constante utilisé dans deux classes différentes réfère à la même constante. La structure de traits associée à une constante est l'unification des structures de traits héritées pour cette constante, et pour les constantes qui lui sont égales (relation ≈), à ceci près que :

- La valeur du méta-trait `pos` est l'intersection de toutes les valeurs héritées. Par exemple on peut spécifier à un certain niveau de la hiérarchie qu'une constante ?i est nominale ou phrastique (?i:pos = N/P) et raffiner plus bas (i.e. dans une classe plus spécifique) que la catégorie est nominale (?i:pos = N). L'unification de N et N/P est N.
- Le méta-trait `fonction` est révisable. Les redistributions sont en effet représentées comme des modifications du métatrait `fonction` associé aux constantes représentant les actants grammaticaux.

Pour pouvoir nommer le même noeud, deux constantes doivent avoir des structures de traits unifiables. Si au sein d'une classe, on infère d'un côté l'égalité de deux constantes, et que leurs structures de traits ne s'unifient pas, il y a échec de la définition de la classe (plus exactement de toute instance de cette classe).

3 Organisation générale de la MG

Une fois posées ces notions formelles, nous pouvons donner la composition d'une MG pour une langue donnée (en dehors des lexiques morphologique et syntaxique).
Nous rappelons que nous reprenons les lexiques tels qu'ils sont organisés pour une TAG, et que l'on doit pour cela fixer les sous-catégorisations initiales possibles (qui vont correspondre à la notion de famille en TAG). En dehors des lexiques morphologique et syntaxique, une MG est alors constituée des composants suivants :

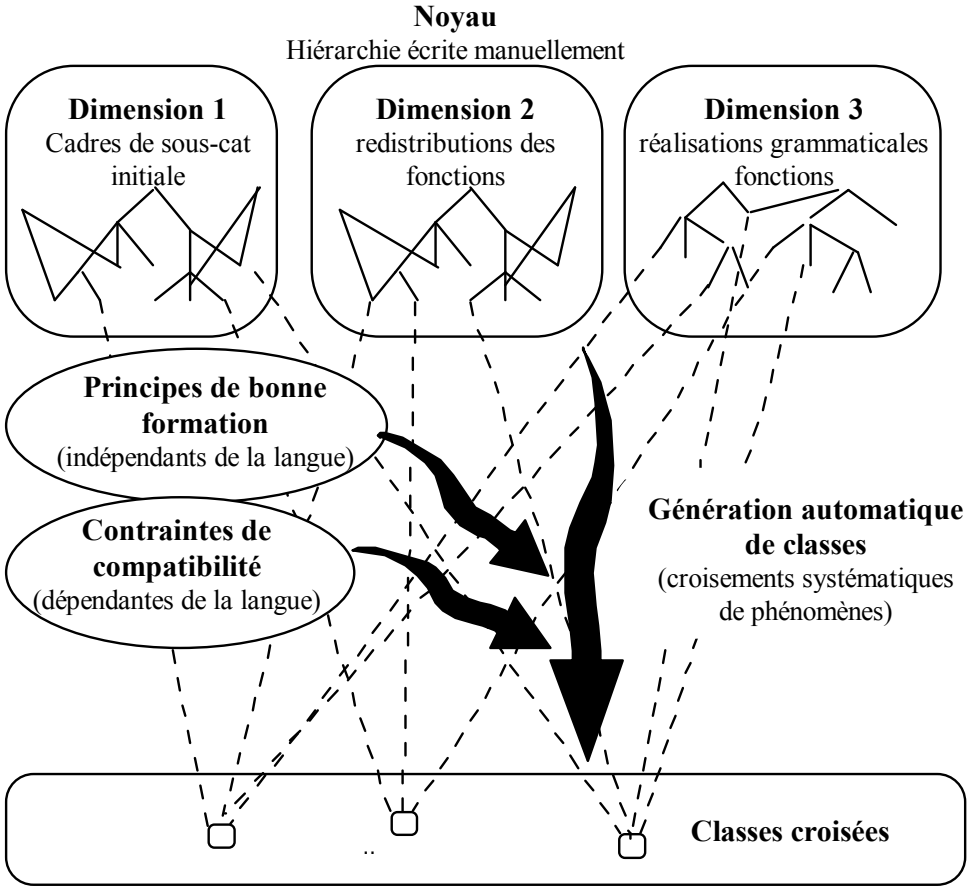
- la description des sous-catégorisations initiales prévues
- la description des redistributions possibles
- la description des réalisations des fonctions grammaticales
- un ensemble de contraintes de compatibilité dépendantes de la langue

Les 3 premiers types d'information (sous-catégorisations initiales, redistributions, réalisations des fonctions grammaticales) sont organisées en un graphe d'héritage, la **hiérarchie noyau**. Les **contraintes de compatibilité** sont fournies à part, dans un langage de contraintes fourni par le formalisme (Section 3.3).

[29] Dans l'implémentation, on construit des schèmes dans le format XTAG. Or celui-ci requiert que différents noeuds de même catégorie soient distingués par des indices alphabétiques. On utilise le métatrait :ind pour cela également.

La Figure 32 schématise l'organisation générale pour une MG et la compilation en TAG. On nomme **classes noyau** les classes de la hiérarchie noyau, et **classe terminale** une classe noyau dont aucune autre classe noyau n'hérite. Les classes terminales sont typées selon la dimension (1, 2, 3) à laquelle elles appartiennent. Les classes terminales de la dimension 3 sont en outre typées selon la fonction grammaticale dont elles représentent une réalisation[30].

Lors de la compilation en TAG, la hiérarchie noyau est augmentée de classes, dites **classes croisées**, générées automatiquement par croisements systématiques de phénomènes linguistiques. Une classe croisée est entièrement définie par sa liste de super-classes, qui sont des classes terminales. L'ensemble des classes croisées est caractérisé par des principes de bonne formation, dont une partie est imposée par le formalisme et est indépendante de la langue (il s'agit du principe de complétude/cohérence/unicité défini à la Section 1.7). Une autre partie est représentée par les contraintes de compatibilité, dépendantes de la langue.



[30] Nous utilisons le mot classe pour type. Donc les classes constituent en elles-mêmes un typage des objets qui en sont des instances. En disant que les classes terminales sont typées selon la dimension, et pour les classes terminales de la dimension 3, selon la fonction, nous exprimons que ces classes terminales doivent comprendre dans leur définition (via les classes dont elles héritent) l'information sur la dimension à laquelle elles appartiennent, et pour les classes de la dimension 3, la FG qu'elles représentent.

Figure 32. Organisation de la métagrammaire

3.1 La hiérarchie noyau

On peut considérer une MG comme une grammaire formelle qui engendre (via la procédure de compilation en TAG) un ensemble de schèmes. Les "règles" de la MG sont les classes noyau.

Chaque classe peut spécifier des informations du type suivant :
- Liste des superclasses directes (classes dont elle hérite directement), ordonnée par ordre de spécificité
- Description partielle d'arbres
- Structures de traits associées aux constantes (méta-traits et traits)
- Actions de révision de fonction
- Tête

Cette dernière information, le champ "tête", a pour valeur la constante qui représente la tête syntaxique de la structure. Nous verrons Chapitre 3, que dans le cas de schèmes qui représentent un prédicat complexe (*faire* + *Vinf* par exemple), la tête syntaxique n'est pas l'ancre principale. Nous détaillons au Chapitre 3 l'utilisation de ce champ.
Les actions de révision de fonction sont utilisées pour représenter les redistributions, elles opèrent en révisant la valeur du métatrait `fonction` des constantes concernées.

Seule la liste de superclasses directes doit être obligatoirement spécifiée (elle peut être vide, il s'agit alors d'une racine du graphe).

La MG impose une partition des classes terminales de la hiérarchie noyau en 3 domaines d'information, que nous appelons dimensions (il s'agit de trois dimensions nécessaires à la caractérisation d'un schème) :

- **Dimension 1** : sous-catégorisation initiale
- **Dimension 2** : redistributions des fonctions
- **Dimension 3** : réalisations grammaticales des fonctions

3.1.1 Dimension 1 : sous-catégorisation initiale

Les classes de cette dimension spécifient les informations pour la sous-catégorisation initiale d'un lexème, et fixent également la topologie commune à tous les schèmes dont le lemme principal est de même catégorie.
On donne par exemple la classe a-une-ancre-verbale, dont hérite toute classe terminale qui spécifie une sous-catégorisation initiale pour un verbe. Le contenu de la classe (dont la description partielle)[31] est expliqué au Chapitre 3.

a-une-ancre-verbale superclasses: nil

[31] Par convention, les constantes des descriptions partielles ont un identificateur commençant par ?.

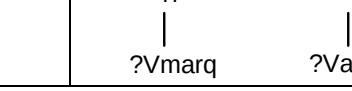
metatraits associés aux constantes :	description:
?Vanc:cat = V ?Vmarq:cat = V ?P:cat = P/N	

Figure 33. La classe a-une-ancre-verbale

Les classes terminales de cette dimension doivent spécifier localement ou hériter de la catégorie des lemmes du lexème pouvant ancrer le schème en construction, la liste des arguments avec leurs catégories possibles, leur indice argumental et leur fonction initiale, qui est également leur fonction par défaut. A titre d'exemple, la classe n0Vn1, qui sert pour les transitifs stricts à arguments nominaux (ou clitiques) en français est donnée Figure 34 (elle est entièrement définie par ses superclasses directes). Outre la classe a-une-ancre-verbale ci-dessus, elle hérite de sujet-n-initial et objet-n-initial, et n'ajoute aucune information[32]. Ces deux superclasses sont données Figure 35.

n0Vn1
superclasses: a-une-ancre-verbale sujet-n-initial objet-n-initial

Figure 34. Classe pour les transitifs à arguments nominaux

sujet-n-initial
superclasses: nil
metatraits :
?arg0:fonction-i = sujet ?arg0:fonction = sujet ?arg0:cat = N/Cl[33] ?arg0:ind = 0

objet-n-initial
superclasses: nil
metatraits :
?arg1:fonction-i = objet ?arg1:fonction = objet ?arg1:cat = N/Cl ?arg1:ind = 1

Figure 35. Les classes sujet-n-initial et objet-n-initial

3.1.2 Dimension 2 : redistributions des fonctions

Les classes de cette dimension spécifient les redistributions possibles, classées selon la catégorie de lexème auquel elles s'appliquent.
Les redistributions sont principalement pertinentes dans le cas d'un lexème verbal. Elles peuvent alors être accompagnées d'une morphologie verbale particulière.

Les classes de cette dimension représentent les redistributions partielles, qui stipule le changement de FG pour un actant grammatical. On donne un exemple valant pour le

[32] On simplifie légèrement. En effet les classes peuvent comprendre un champ "nom" supplémentaire, utilisé pour construire un nom pour chaque schème, par concaténation des valeurs locale et héritée du champ "nom".
[33] Nous spécifions comme catégorie possible N et Cl pour clitique, bien que les clitiques soient traités comme des affixes (et donc en morphologie). Le métatrait ?arg:cat = Cl est utilisé pour marquer que le type de réalisation de arg0 est clitique, mais une position syntaxique de clitique n'est pas effectivement placée dans les schèmes en construction. On peut ainsi dans certains cas imposer une réalisation clitique d'un actant (pour le *il* impersonnel par exemple), en spécifiant :cat= Cl.

français, de la redistribution partielle qui promeut l'actant grammatical objet vers la fonction sujet.

objet→sujet
superclasses: promotion-vers-sujet
action: ce qui est objet devient sujet

Les redistributions (complètes) sont représentées par des classes qui héritent éventuellement de plusieurs redistributions partielles, plus éventuellement d'une morphologie particulière. Par exemple, on donne ci-dessous la redistribution (complète) pour le moyen (comme dans *ces livres se vendent bien*) (cf. Chapitre 4).

moyen
superclasses: sujet→vide objet→sujet morpho-moyenne

L'ordre des superclasses est pertinent, puisque les différentes redistributions partielles sont ordonnées (la démotion du sujet doit avoir lieu avant la promotion de l'objet).
Enfin, on a vu qu'une sous-catégorisation finale peut être dérivable par une séquence de plusieurs redistributions.
Dans l'implémentation actuelle, on définit dans la hiérarchie noyau les séquences de redistributions possibles (qui sont peu nombreuses).
Ainsi par exemple le moyen impersonnel en français (*Il s'est vendu 300000 exemplaires du Goncourt*), est explicitement représenté par la classe suivante :

moyen-impersonnel
superclasses: moyen impersonnel

L'ordre de la séquence de redistributions est représenté par l'ordre des superclasses spécifiées.

3.1.3 Dimension 3 : réalisations grammaticales des fonctions

Les classes terminales de cette dimension représentent les différentes réalisations grammaticales pour chaque fonction pouvant apparaître dans un cadre de sous-catégorisation finale.

Ces classes héritent d'informations concernant une fonction et concernant une construction syntaxique. Par exemple, la classe pour le sujet relativisé hérite des propriétés générales du sujet et de celles des relatives.
L'idée est de représenter le plus séparément possible la réalisation d'une fonction de celle d'éventuelles autres fonctions de la sous-catégorisation finale. Cependant, lorsque les réalisations de surface de plusieurs fonctions sont interdépendantes (par exemple pour les clitiques), une même classe terminale peut fixer la réalisation de plusieurs fonctions.
On donne Figure 36 la classe agt-obj-interrogé, qui spécifie la réalisation du complément d'agent en position extraite (interrogée) (on donne les propriétés locales et

les propriétés héritées). Elle hérite d'une part de position-agt-obj, qui donne les propriétés générales des agt-objet, et de sp-interrogé, pour les propriétés des compléments interrogés, de catégorie SP.

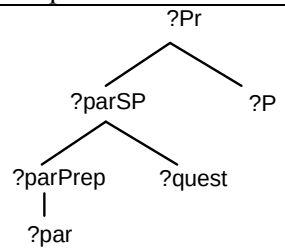
agt-obj-interrogé superclasses: position-agt-obj sp-interrogé	
traits et metatraits associés aux constantes :	description:
?parPrep:cat = Prep ?parSP:cat = SP ?quest:cat = N **?quest:fonction = agt-objet** ?quest:t:wh=+ ?parPrep:ind = ?parSP:ind = ?quest:ind	

Figure 36. Classe pour le complément d'agent interrogé (propriétés locales et héritées)

On peut remarquer que la description partielle associée à cette classe fait appel à la constante ?P, qui est introduite dans la classe a-une-ancre-verbale. C'est une illustration du fait que les constantes sont accessibles globalement dans toute la hiérarchie (ce sont des variables globales au sens informatique). D'autre part, cette classe étant une réalisation de la fonction agt-objet, elle utilise une constante dont la fonction est agt-objet (la constante ?quest). Il faut remarquer que la classe ne fait pas référence à la fonction initiale de cette constante (métatrait `fonction-i`) mais seulement à la fonction finale (métatrait `fonction`). En effet, les réalisations de fonctions sont spécifiées indépendamment de la fonction initiale des actants.

3.2 Des classes croisées pour représenter des schèmes

Une classe croisée se définit de manière à ce qu'elle représente un ou plusieurs schèmes (qui diffèrent éventuellement par l'ordre des branches, cf. infra), et que ce(s) schème(s) vérifient le principe de complétude/cohérence/unicité (Section 1.7).

Une classe croisée est une classe entièrement définie par sa liste de superclasses, et dont les caractéristiques sont :

- Une classe croisée hérite d'exactement une classe terminale de la dimension 1
- Une classe croisée hérite d'exactement une classe terminale de la dimension 2
- Une classe croisée hérite de classes terminales de la dimension 3 telles qu'une réalisation grammaticale unique est spécifiée pour chaque fonction de la sous-catégorisation finale, définie par la combinaison des classes terminales de la dimension 1 et 2.
- Une classe croisée vérifie les contraintes de compatibilités de la MG pour la langue considérée (définies Section 3.3).

On a annoncé en outre l'utilisation de l'héritage prioritaire, pour gérer l'ordre d'application du processus d'héritage. Cela implique que les superclasses d'une classe en générale et d'une classe croisée en particulier soient ordonnées par ordre de spécificité.

> Les superclasses d'une classe croisée sont ordonnées de la plus générale à la plus spécifique dans l'ordre dimension 1 < dimension 2 < dimension 3

L'ordre respectif des classes héritées de la dimension 3 n'est pas pertinent.

3.2.1 Procédure de génération de l'ensemble des classes croisées

Pour générer l'ensemble des classes croisées, on doit disposer de l'ensemble des classes terminales de la dimension 1, l'ensemble des classes terminales de la dimension 2, et pour chaque FG, l'ensemble des réalisations grammaticales.
L'ensemble des classes croisées est généré en épuisant toutes les possibilités de classe terminale de la dimension 1, puis 2, puis pour chaque FG apparaissant dans la sous-catégorisation finale, toutes les possibilités de réalisations sont utilisées.
Parmi les classes croisées ainsi définies ne sont retenues que celles qui vérifient les contraintes de compatibilité.

On a vu que les réalisations de fonction en dimension 3 ne dépendent que de la fonction finale des actants. Ainsi, en dimension 3, lorsqu'une constante représente un sujet, on ne sait pas a priori si cette constante sera unifiée avec le sujet initial ou avec un objet initial (au passif par exemple). Il faut donc établir un lien entre les constantes argumentales (i.e. pour représenter les arguments) introduites en dimension 1, qui reçoivent leur fonction finale (métatrait `fonction`) soit directement en dimension 1, soit par révision en dimension 2, et les constantes introduites en dimension 3, qui portent une fonction. Ce lien est réalisée en unifiant systématiquement les constantes portant la même fonction, une fois que la sous-catégorisation finale est définie :

> Après que les redistributions ont eu lieu, toute constante C1 est déclarée égale à toute autre constante C2, telle que le métatrait `fonction` de C1 a même valeur que celui de C2.

3.2.2 Un exemple détaillé

Soit le schème de la Figure 37 (sans traits), pouvant être ancré par exemple par *accompagnée*, pour un participe, pour dériver *Par qui sera accompagnée Marie?*[34].

[34] Le même schème est utilisé pour l'emploi dans une interrogative indirecte et l'emploi dans une interrogative directe. Il y a dans ce deuxième cas adjonction obligatoire d'un clitique inversé.

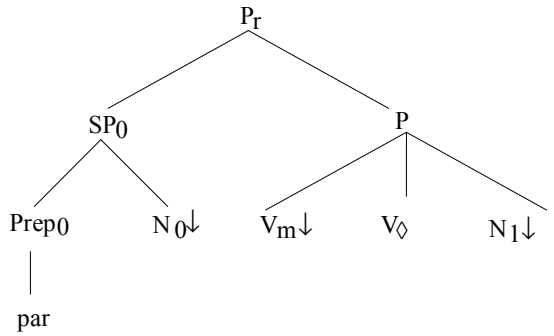


Figure 37. Schème pour *Par N0 Vétat Vpart N1*

Pour représenter ce schème, une classe croisée est définie, comme cela est illustré Figure 38.

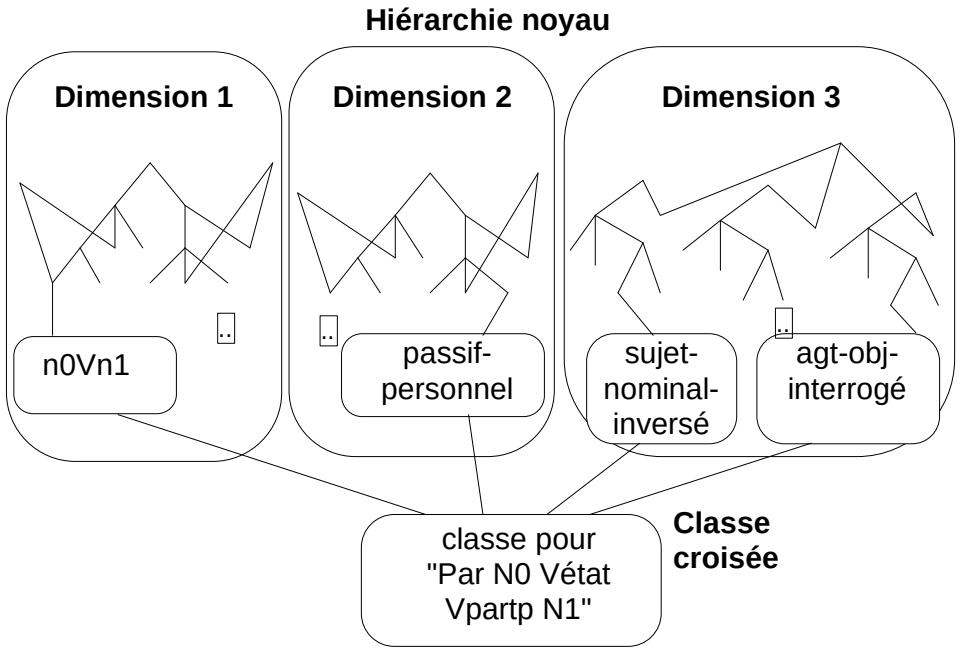


Figure 38. Un exemple de création de classe croisée

La combinaison des classes n0Vn1 (cf. supra Figure 34) et passif-personnel définit la sous-catégorisation finale : sujet (portée par arg1) et agt-objet (portée par arg0). Pour chacune des fonctions apparaissant dans la sous-catégorisation finale, on choisit une réalisation de surface, ici sujet-nominal-inversé et agt-obj-interrogé (cf. supra Figure 36).

Afin de détailler précisément le processus d'héritage mis en oeuvre pour cette classe croisée, nous détaillons les classes manquantes.
La classe passif-personnel est donnée ci-dessous[35], on donne à la fois les propriétés locales et les propriétés héritées. On simplifie par rapport à ce qui sera détaillé au Chapitre 4, pour se concentrer sur la description partielle. La classe hérite des

[35] Elle est légèrement simplifiée par rapport à ce qui sera détaillé Chapitre 4.

redistributions partielles sujet→agt-objet et objet→sujet, qui modifient le métatrait fonction du sujet et de l'objet. La classe comprend une constante ?Vpass, pour le verbe qui domine la copule ETRE, qui est unifiée avec la constante ?Vmarq introduite dans la classe a-une-ancre-verbale (cela est réalisé par l'intermédiaire du champ "tête", cela est détaillé Chapitre 4). L'ancre est placé à droite de l'artère allant de la copule (constante ?copule) au premier noeud phrastique (constante ?P).

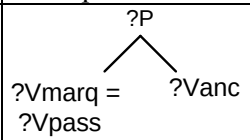
passif-personnel	
superclasses: sujet→agt-objet objet→sujet morpho-passive-avec-cop	
action: ce qui est sujet (et tel que :fonction-initiale=sujet, :lie-réfl=-) devient agt-objet ce qui est objet devient sujet	
tête→Vpass	
métatraits et traits associés aux constantes:	description:
?Vpass:cat = V ?Vanc:b:mode = partp ?Vanc:b:num = ?const-sujet:b:num ?Vanc:b:gen = ?const-sujet:b:gen	

Figure 39. Classe passif-personnel (simplifiée)

La classe sujet-nominal-inversé (simplifiée pour les traits) est donnée ci-dessous. Elle place simplement la constante ?const-sujet, dont la fonction est sujet, à droite de l'ancre (?Vanc). Les contraintes sont détaillées dans la section suivante[36].

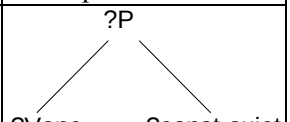
Sujet-nominal-inversé	
superclasses: sujet-nominal	
contraintes: ⇒⌉ objet-position-canonique, ⇒ extraction	
métatraits et traits associés aux constantes:	description:
?const-sujet:fonction = sujet ?const-sujet:cat = N ?supVmarq:t:num = ?const-sujet:t:num ?supVmarq:t:gen = ?const-sujet:t:gen	

Figure 40. Classe sujet-nominal-inversé (simplifiée)

A la création de la classe croisée de la Figure 38, ?arg0 porte le métatrait `fonction` = agt-objet, après l'héritage de la redistribution partielle sujet→agt-objet. ?arg1 porte le métatrait `fonction` = sujet, après l'héritage de la redistribution partielle objet→sujet. Donc ?arg0 est unifié à ?quest, et ?arg1 est unifiée à ?const-sujet. La conjonction de toutes les descriptions partielles héritées peuvent être représentée graphiquement comme dans la Figure 41.

[36] On représente les contraintes de compatibilité au niveau de la définition de la classe, mais toutes les contraintes apparaissant dans une MG, sont déclarées à part, de manière extérieure à la hiérarchie noyau.

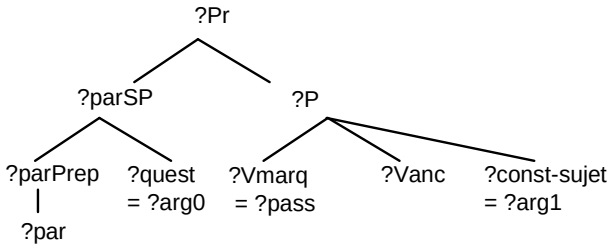


Figure 41. Conjonction des descriptions partielles

3.3 Les contraintes de compatibilité

Les principes de bonne formation cités sont insuffisants tels quels. Reste à représenter le fait que les classes terminales ne sont pas toutes compatibles entre elles. Certaines redistributions ne sont pas compatibles avec certaines sous-catégorisations initiales, certaines réalisations de fonction ne sont pas compatibles entre elles. Par exemple on ne peut pas en français inverser le sujet si l'objet est en position canonique. Ces compatibilités/incompatibilités sont exprimées soit par le contenu des classes elles-mêmes soit par des contraintes externes, qui peuvent être plus lisibles.
Plus précisément, on impose les principes suivants :

- La description partielle d'une classe croisée (conjonction des descriptions héritées) doit être satisfiable et les structures de traits associées aux constantes nommant le même noeud doivent être unifiables (cf. supra).
- Les super-classes d'une classe croisée doivent satisfaire des **contraintes de compatibilité** (dépendantes de la langue et de la couverture de la grammaire). On peut exprimer ces contraintes:
  - En spécifiant qu'une classe noyau est **disjonctive**, ce qui signifie que ses sous-classes (directes ou indirectes) sont mutuellement exclusives et donc ne peuvent pas apparaître ensemble comme super-classes d'une classe croisée. Par exemple en français, une seule extraction est possible pour un même prédicat. Dans la métagrammaire du français, on a déclaré la classe EXTRACTION (dont hérite les questions, les relatives et les clivées) comme disjonctive.
  - Par des **contraintes de cooccurrence** respectivement **négatives** ou **positives**, qui imposent soit que deux classes noyau sont incompatibles comme super-classes d'une classe croisée (contrainte négative, par définition symétrique), soit au contraire qu'une superclasse impose la présence d'une autre superclasse (contrainte positive, non symétrique). Par exemple, en français, un sujet nominal inversé est incompatible avec un objet (direct) postverbal et n'est productif que dans le cas d'une extraction d'un complément. Cela peut être représenté par les contraintes (que l'on a déjà données supra, Figure 40) :

    (sujet-nominal-inversé $\Rightarrow$ extraction)

    (sujet-nominal-inversé $\Rightarrow$ $\neg$ objet-position-canonique)

On peut utiliser ces contraintes pour de "vraies" contraintes linguistiques, mais également pour une application donnée, pour réduire la couverture de la grammaire générée.

3.4 Comparaison avec les autres travaux de représentation compacte de TAG

Plusieurs travaux existent désormais, qui fournissent une représentation compacte d'une TAG, ou d'un formalisme approchant.
Les trois propositions de base sont celles de Vijay-Shanker et Schabes (1992); Becker (1993) et Evans et al. (1995).
Elles s'appuient toutes trois sur une hiérarchie de descriptions syntaxiques, qui représente les schèmes "canoniques", c'est-à-dire, illustrant la sous-catégorisation initiale, et les positions canoniques pour tous les actants (en utilisant nos termes). Les schèmes pour les constructions marquées sont dérivés des premiers par application de règles "lexico-syntaxiques" (l'équivalent de règles lexicales, mais appliquées à des descriptions syntaxiques).
Les différentes propositions diffèrent selon l'emphase mise sur l'utilisation de règles lexico-syntaxiques, selon l'implémentation réalisée ou pas, selon le langage utilisé pour référer aux structures syntaxiques.
L'article de Vijay-Shanker et Schabes (1992), est une simple proposition, dont le coeur est d'utiliser des descriptions partielles d'arbres pour représenter la topologie des structures syntaxiques. Les descriptions partielles permettent une combinaison monotone des informations sur la topologie des arbres, avantage appréciable, lorsque l'on considère la complexité que peut montrer une représentation de la syntaxe d'une langue. Outre notre implémentation, l'implémentation de Xia et al. (1998) pour l'anglais peut être citée.

Becker (1993) a proposé un système, implémenté, où une emphase particulière est mise sur les règles lexico-syntaxiques permettant de dériver de nouveaux schèmes à partir de schèmes de base ou dérivés (l'auteur les appelle des métarègles). Becker a développé son propre langage de description des topologies d'arbres. Les règles lexico-syntaxiques comportent une partie gauche (l'entrée) où une description topologique est fournie, et une partie droite, comportant des modifications par rapport à l'entrée. Le contrôle sur l'application des règles est en théorie fait par subsomption, c'est-à-dire que toute règle dont la partie gauche subsume un schème de base ou dérivé est applicable à ce schème. En pratique, pour l'implémentation qui a été réalisée et intégrée au système XTAG, les règles sont activées manuellement, ou bien une séquence d'application de règles peut être spécifiée.

Evans et al. (1995) ont proposé d'utiliser DATR (Evans et Gazdar, 1989) pour représenter une TAG. Cette proposition a l'avantage de reposer sur un système de représentation hiérarchisée, DATR, implémenté, qui a fait ses preuves pour la représentation des connaissances lexicales. Toutes les informations sont représentées par des structures de traits, y compris la topologie des arbres. L'héritage est non monotone,

en particulier la topologie des arbres représentés peut être révisée. Les règles lexico-syntaxiques sont prises en charge par DATR.
Le contrôle sur l'application des règles lexico-syntaxiques est réalisé en spécifiant à un noeud du graphe d'héritage, quelles sont les règles applicables.
Actuellement le système est utilisé au sein du projet LEXSYS, en cours à l'université du Sussex, pour la représentation compacte d'une D-Tree Grammar de l'anglais (Smets et Evans, 1998).

Formellement, nous avons préféré suivre la proposition de Vijay-Shanker et Schabes (1992) d'utiliser un graphe d'héritage où la combinaison des informations sur la topologie des structures syntaxiques est monotone. Nous pensons que cela permet une plus grande transparence.

Nous avons en outre choisi de ne pas utiliser de règles lexico-syntaxiques explicites. Nous avons dissocié formellement deux types de "cas marqué" pour l'emploi d'un lexème, les cas marqués pour lesquels la sous-catégorisation finale n'est pas la sous-catégorisation initiale du lexème, et ceux pour lesquels les actants grammaticaux ne sont pas dans leur position canonique.
Mais nous utilisons le même appareil formel, le graphe d'héritage, pour gérer tous les cas de constructions marquées, qu'ils mettent en jeu ou pas des redistributions de fonction.

- Pour les cas qui ne mettent pas en jeu de redistribution de fonction (règles pour l'extraction, la cliticisation, l'inversion...), nous avons choisi de ne pas privilégier une position canonique pour la réalisation des fonctions grammaticales. La définition d'une fonction comporte au sein de la MG l'ensemble des réalisations possibles pour cette fonction (que ce soit en position canonique, extraite, inversée ...). Le processus dynamique de génération par règle lexico-syntaxique des "positions marquées" est équivalent dans la MG au processus dynamique de génération systématique de classes croisées avec toutes les réalisations grammaticales possibles pour une fonction apparaissant dans la sous-catégorisation finale.

- Les cas qui mettent en jeu une redistribution des fonctions (passif, impersonnel, se-moyen, dative-shift en anglais ...), sont traités en dimension 2. Les classes de cette dimension modifient la valeur du trait `fonction` pour certains arguments, ce qui pourrait être pris en charge par un mécanisme du type règle lexicale, comme cela est fait en LFG. L'avantage d'intégrer les redistributions au sein du graphe d'héritage est de pouvoir utiliser le partage d'informations entre différentes redistributions. Ainsi pour le français, nous avons pu décomposer par exemple le passif personnel (cf. supra) en trois phénomènes (démotion du sujet, promotion de l'objet vers la fonction sujet et morphologie passive) qui sont tous trois utilisés séparément pour d'autres redistributions.

L'avantage général de ne pas utiliser de règles lexico-syntaxiques est celui de ne pas avoir à gérer leur ordre d'application. L'implémentation de Evans et al. par exemple (1995) repose sur la définition d'un ordre d'application des règles.

Plus généralement, l'avantage de notre système par rapport à ceux cités est que la métagrammaire a l'ambition de fournir des éléments de théories linguistiques, en imposant une organisation des informations syntaxiques. Nous avons discuté Section 1.5 de la distinction faite dans de nombreuses théories entre règles de type "redistributions" et règles de types "réalisations de fonctions". Ces deux types de règles se retrouvent confondues dans les systèmes cités[37]. Le concepteur de la représentation compacte doit veiller à les séparer (par exemple en définissant explicitement qu'une règle pour le passif doit s'appliquer avant une règle pour l'extraction).
Au sein de la MG au contraire, le principe de cooccurrence prédicat-arguments a été formalisé, et est suivi par le compilateur MG→TAG, pour générer des schèmes qui respectent ce principe. Ainsi le formalisme est-il plus riche, puisqu'il intègre des principes linguistiques de bonne formation.
Sur le plan pratique, la formalisation du PCPA en principe de complétude/cohérence/unicité nous a permis de concevoir une procédure entièrement automatisée pour la compilation en TAG. Il suffit de typer les classes terminales de ls MG en classes de la dimension 1, ou 2, ou 3, avec pour les classes de la dimension 3 un typage supplémentaire selon la fonction grammaticale représentée.

4 Génération de schèmes

La compilation en une TAG comporte deux phases, la première étant la génération des classes croisées, en suivant la procédure donnée ci-dessus (Section 3.2). La deuxième phase est simplement la "traduction" d'une classe croisée en schème(s), la différence principale étant qu'une classe croisée hérite d'une description partielle d'arbres (augmentée de structures de traits associées aux constantes), comme conjonction de toutes les descriptions héritées, alors que les TAG manipulent des arbres.
Pour passer des descriptions partielles aux arbres, nous utilisons la notion de *référent minimal* définie par Rogers et Vijay-Shanker (1994). Une description partielle caractérise un ensemble d'arbres (ceux qui la satisfont) et les référents minimaux sont des arbres "représentatifs" de cet ensemble.
Intuitivement, pour passer d'une description partielle à un arbre qui la satisfait, il faut spécifier totalement les relations de dominance et de précédence linéaire laissées sous-spécifiées dans la description. En effet dans un arbre, chaque couple de noeud doit être lié soit par une relation de précédence (dans un sens ou dans l'autre) soit par une relation de dominance (également dans un sens ou dans l'autre). Les référents minimaux sont les arbres qui mettent en jeu le moins de noeuds possibles, une fois fixés le type (précédence ou dominance) et le sens de la relation existant entre chaque couple de noeuds.
Dans la description D1 donnée Section 2.2, chaque couple de constantes est dans une relation de dominance ou de précédence sauf le couple (n2,n3). Comme n2 et n3 ont le même père (s1), il y a 3 possibilités pour lier ces deux constantes :

$$(n2 < n3) \vee (n3 < n2) \vee (n2 \approx n3)$$

[37] L'implémentation de Xia et al. (1998) a repris cependant notre proposition (Candito, 1996) de dissocier règles de changement de diathèse et règles de réalisations syntagmatiques, bien que les fonctions grammaticales y soient considérées comme des positions structurales.

Il faut en outre prendre pour chaque couple dans une relation de dominance, le chemin le plus court, pour mettre en jeu le moins de noeuds possibles. C'est-à-dire que les liens de dominance qui sont sous-spécifiés dans une description sont de longueur zéro dans les référents minimaux, si les noeuds aux extrémités sont compatibles (si leurs structures de traits sont compatibles). Cela implique pour les référents minimaux de la description D1, que v1 et v2 nomment le même noeud. La Figure 42 donne les 3 référents minimaux obtenus.

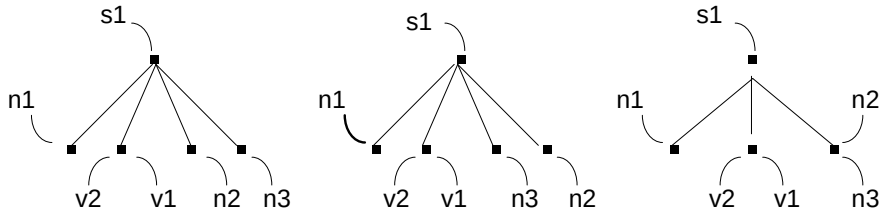


Figure 42. Les référents minimaux de la description D1

On peut donc remarquer, qu'une classe croisée peut a priori laisser sous-spécifiée la relation entre deux noeuds. D'un point de vue linguistique, cette sous-spécification est pertinente pour représenter la variation d'ordre de certains constituants. C'est ce qui est utilisé pour le français et l'italien pour représenter l'ordre libre des compléments post-verbaux.
Ainsi pour représenter les phrases *Paul donne un livre à Marie / Paul donne à Marie un livre*, on utilise la même classe croisée, qui place l'objet direct et l'objet indirect à droite du verbe. C'est la compilation en TAG qui génère les schèmes correspondants (comme ils portent des fonctions différentes et des indices différents, ils ne peuvent pas être unifiés et on n'a pas la troisième solution de la Figure 42).

Si l'on reprend l'exemple de création de classe croisée donné supra, la Figure 41 donne la description partielle héritée par la classe croisée. A la traduction en schèmes, le lien sous-spécifié entre ?supVmarq et ?Vmarq est spécifié en un chemin de longueur nulle. On peut vérifier que l'on obtient ainsi le schème de la Figure 37. Ce schème est unique, car tous la description Figure 41 n'a qu'un seul référent minimal.
Si l'on crée une classe croisée parallèle, mais spécifiant un complément supplémentaire, par exemple un a-objet, elle définirait les schèmes pour *Par N0 ETRE Vpartp N1 à N2* et pour *Par N0 ETRE Vpartp à N2 N1*, par sous-spécification de l'ordre respectif de l'a-objet et du sujet inversé.

5 Implémentation

L'implémentation du compilateur MG → TAG a été réalisée, en utilisant pour gérer le processus d'héritage un module objet de lisp (Common Lisp Objet System), version Lucid 4.1 (compatible 4.0).
Nous donnons en annexe A, un manuel pour utilisateur (en anglais) pour le système, ainsi qu'en annexe B, une hiérarchie de démonstration, qui représente une partie des

schèmes à ancre verbale de la grammaire anglaise du XTAG group (XTAG group, 1998).

Le compilateur fonctionne en prenant en entrée un ou plusieurs fichiers décrivant une MG pour une langue donnée (et éventuellement un domaine ou une application donnée) et fournit en sortie un ensemble de fichiers, qui représentent les schèmes de la grammaire TAG générée, dans le format requis par l'analyseur intégré au système XTAG.

5.1 Format d'entrée du compilateur

Le format d'entrée du compilateur est précisément défini à l'annexe A. Une MG dans ce format est constituée de :

- liste des noms de classes terminales de la dimension 1 (stockée dans ***families***)
- liste des noms de classes terminales de la dimension 2 (stockées dans ***redistributions***)
- liste des noms de classes terminales de la dimension 3, partitionnées selon la fonction dont elles représentent une réalisation (stockées dans ***realizations***)

- liste des contraintes de cooccurrence positives
- liste des contraintes de cooccurrence négatives
- liste des contraintes de croisement

- liste de définition de classes

Table 1. Eléments constitutifs d'une MG, au format d'entrée requis par le compilateur

La syntaxe de définition d'une classe peut être représentée par la grammaire CF suivante (partielle) :

définition-de-classe → **(**nom-de-classe
(liste-ordonnée-de-superclasses-directes**)**
liste-de-spécifications
)

liste-ordonnée-de-superclasses-directes → nom-de-classe*
liste-de-spécifications-de-propriétés → spécification-de-propriétés*
spécification-de-propriétés → spécification-de-description |
spécification-de-traits |
spécification-de-métatraits |
spécification-de-nom-de-schème

Table 2. Syntaxe de la définition d'une classe

5.2 Algorithme de génération de schèmes

L'algorithme de génération de schèmes est donné Table 3.

```
1  tant que il reste des classes non choisies dans *families*, choisir une classe C1 de
   *families* et  faire
   {
       ouvrir un fichier F1
2      tant que il reste des classes non choisies dans *redistributions*, choisir une
       classe C2 de *redistributions* et  faire
       {
3          si C1 et C2 sont incompatibles[38]
           alors retourner en 2
           sinon
           {
4              listeC reçoit (C2 C1)[39]
               calculer la sous-catégorisation finale SF, définie par C1 + C2
5              tant que il reste des sous-ensembles non choisis de *realizations* tels
               qu'ils contiennent une réalisation pour chaque fonction de SF, choisir
               un tel sous-ensemble listeC3 et faire
               {
                   listeC reçoit concaténation de listeC3 et listeC
                   si listeC ne respecte pas les contraintes de compatibilité
                   alors retourner en 5
                   sinon faire
6                  {
                       retourner le résultat R de l'unification des propriétés héritées
                       de toutes les superclasses de listeC, de la plus générale à la
                       plus spécifique
                       si l'unification a échoué (R vaut 0)
                       alors retourner en 5
                       sinon traduire R en un ou plusieurs schèmes, et les écrire
                       dans le fichier F1
                   fin sinon}
               fin tant que}
           fin sinon}
       fin tant que}
       fermer le fichier F1
   fin tant que}
```

[38] Les redistributions en jeu dans C2 doivent être applicables à partir de la sous-catégorisation initiale définie dans C1.

[39] L'ordre est pertinent, puisque l'on utilise l'héritage prioritaire, pour ordonner totalement les superclasses d'une classe. On construit à ce niveau une liste de superclasses directes (stockée dans listeC) qui définit totalement une classe croisée. Dans la listeC, les superclasses sont ordonnées de la plus spécifique à la plus générale, ainsi, C1, la classe de la dimension 1, est plus générale que C2, la classes de la dimension 2, qui sera à son tour plus générale que les classes de la dimension 3 choisies au niveau 5. Le processus d'héritage se fait de la classe la plus générale à la classe la plus spécifique.

Table 3. Algorithme de génération des schèmes représentés par une MG

On peut remarquer que l'algorithme rassemble dans un même fichier tous les schèmes définis à partir de classes croisées qui partagent la même classe C1, c'est-à-dire la même classe de la dimension 1, i.e. qui appartiennent à la même famille.
Nous ne détaillons pas l'algorithme d'unification des propriétés héritées, en particulier l'algorithme d'unification des description partielles. Les descriptions partielles au sein de la MG sont exprimées comme des conjonctions de quasi-arbres (intuitivement des arbres, sauf que certains chemins entre noeuds peuvent être sous-spécifiés, en l'occurrence représentés par un lien de dominance). Il est à noter que la conjonction de plusieurs quasi-arbres n'est pas forcément un quasi-arbre.
Plutôt qu'un algorithme par propagations de contraintes, nous avons implémenté un algorithme qui calcule l'unification de paires de quasi-arbres, et fournit un ensemble de quasi-arbres, interprété comme une disjonction de quasi-arbres possibles. L'unification avec des quasi-arbres supplémentaires ajoute de l'information de manière monotone, et peut réduire éliminer certains quasi-arbres de l'ensemble précédemment solution.

5.3 Restriction de la couverture de la TAG compilée

A la compilation MG→TAG, des coupes peuvent être opérées dans la hiérarchie noyau. On peut en effet spécifier un ensemble réduit de familles à générer, et/ou un ensemble réduit de redistributions, et/ou un ensemble réduit de réalisations de fonctions (en commentant respectivement certains noms de classes dans les variables *families*, *redistributions*, *realizations*).
De plus, on peut bloquer la génération de certains schèmes non pas en commentant certains noms de classe (ce qui revient à faire comme si elles n'avaient pas été définies), mais en ajoutant certaines contraintes de cooccurrence positives ou négatives. Par exemple, pour la MGi (la métagrammaire écrite pour l'italien, décrite dans la partie II, Chapitres 3, 4, 5, 6), on peut restreindre l'occurrence du sujet nominal inversé, aux cas où le verbe n'a pas d'objet en position canonique (postverbale). D'un point de vue théorique cela n'est pas formellement exclu, mais en pratique, ces cas sont extrêmement minoritaires par rapport aux cas de sujet inversé pour une forme sans objet en position canonique (cf. Sornicola, 1994). Pour bloquer ces cas, il suffit d'ajouter une contrainte de cooccurrence négative (sujet-N-inversé $\Rightarrow \neg$ objet-position-canonique). Il faut donc distinguer parmi les contraintes, celles qui constituent des vrais contraintes linguistiques, de celles qui servent à réduire la couverture de la grammaire.

5.4 Lacunes et améliorations envisagées

Nous détaillons maintenant certaines lacunes de l'implémentation actuelle, ainsi que les solutions envisagées pour les résorber.

5.4.1 Ordre respectif des fonctions grammaticales

La MG permet de définir les propriétés morpho-syntaxiques associées à une fonction grammaticale. La factorisation permise par la MG repose en grande partie sur le découpage des informations contenues dans un schème en une sous-catégorisation

initiale + une séquence de redistributions + une réalisation grammaticale pour chaque fonction de la sous-catégorisation finale. La description des réalisations grammaticales pour chaque fonction est faite indépendamment de la présence dans la sous-catégorisation finale d'autre fonction. Ainsi par exemple, un a-objet est décrit indépendamment du fait qu'un objet existe dans la sous-catégorisation finale ou pas.
Pour le français et l'italien, on a considéré que l'ordre respectif des compléments est libre. Mais, pour décrire par exemple la syntaxe anglaise dans la MG, on a dû ajouter un mécanisme pour spécifier l'ordre relatif des compléments (mécanisme qui peut être utilisé pour le français et l'italien si l'on désire contraindre plus l'ordre des compléments).

Le mécanisme permet d'ajouter des contraintes, dès lors que deux classes sont héritées par une classe croisée donnée. Plus précisément, on définit des triplets (X Y Z), appelés contraintes de croisement, tel que si une classe croisée C hérite de X et Y, alors on ajoute à la liste de superclasses de C, la classe Z.

Cependant nous sommes consciente que ce mécanisme n'est pas très satisfaisant, tant sur le plan théorique que sur le plan pratique. En pratique, il oblige à s'occuper de l'ordre respectif de chaque couple de fonctions grammaticales. Sur le plan théorique, il est dommage de ne pas utiliser la notion d'ordre d'oblicité pour représenter le placement respectif des réalisations des fonctions grammaticales. Nous envisageons donc de définir un placement par défaut, selon un ordre prédéfini de fonctions grammaticales. Plusieurs fonctions peuvent être placées au même rang, de manière à laisser la possibilité de ne pas fixer l'ordre respectif de deux fonctions (et donc générer tous les ordres possibles).
Techniquement, on peut envisager d'augmenter le pouvoir expressif des descriptions partielles, en permettant de placer dans des descriptions des constantes "optionnelles", i.e. qui ne seront effectivement placées que si elles sont indépendamment placées de manière non optionnelle dans une autre description partielle.

5.4.2 Composition automatique de redistributions

Actuellement, si une redistribution R1 peut être composée avec une redistribution R2 (c'est-à-dire si la sortie de R1 peut servir d'entrée à R2), le concepteur de la MG doit définir une classe R12, définie par la liste de superclasses directes (R2 R1)[40].
D'un point de vue pratique, étant donné que les compositions de redistributions ne sont pas absolument régulières mais sont sujettes à des contraintes diverses, il n'est pas si fâcheux que ça de devoir fournir explicitement la liste des compositions, mais cela n'est pas très satisfaisant d'un point de vue théorique, et risque de devenir fastidieux si le nombre de redistributions augmente.

Nous envisageons donc de modifier l'algorithme de génération de classes croisées, de manière à permettre l'application de 0, 1, ou n redistributions. Pour améliorer l'efficacité,

[40] Par exemple, dans la hiérarchie de démonstration fournie à l'annexe B, pour représenter le fait que le passif peut en anglais s'appliquer après le dative-shift (pour dériver *Mary was given a book by Trevor*), on définit une classe dative-shift-passive, qui hérite de passive et dative-shift (appliqués dans le sens dative-shift puis passive).

on peut précompiler dans un premier temps les séquences de redistributions possibles, et se ramener à l'algorithme tel qu'il est implémenté actuellement.

Pour les MG écrites pour le français et l'italien (cf. Partie II), bien que les compositions de redistributions doivent être explicitement déclarées, nous avons écrit les classes pour les redistributions en y prévoyant les contraintes sur la composition de redistributions

5.4.3 Permettre un plus grand partage entre fonctions grammaticales : définir une hiérarchie de fonctions grammaticales

Une autre amélioration qui paraît intéressante serait de pouvoir exprimer des généralisations valant pour plusieurs fonctions syntaxiques. Cela est déjà en partie possible, mais le système est limité par la contrainte que les constantes utilisées pour exprimer les descriptions partielles sont des variables globales sur toute la MG[41]. Cela constitue souvent un avantage, car on n'a pas à gérer la transmission de constantes d'une classe à l'autre. Mais il faut reconnaître que cela limite parfois le partage d'informations.
Ainsi si deux réalisations de fonctions sont susceptibles d'être héritées par une même classe croisée, on doit déclarer des constantes distinctes pour nommer des noeuds dont on veut qu'ils soient distincts en cas d'héritage par une même classe croisée.
Nous envisageons donc de définir deux types de comportement pour des constantes, un comportement de type "variable globale", et un comportement de type "variable locale" à une classe.
Avec un tel mécanisme, on pourrait envisager de définir une hiérarchie de fonctions syntaxiques, en laissant la possibilité de définir des "super fonctions" au sens de fonctions sous-spécifiée qui subsume un ensemble de fonctions. Il peut être intéressant de pouvoir se référer par exemple à toute fonction grammaticale de type "complément prépositionnel". Ou bien pour l'italien même, le sujet et l'objet partagent la même représentation dans le cas d'extraction (le pronom relatif est le même pour les deux fonctions, les pronoms interrogatifs sont également les mêmes, la possibilité d'extraction à distance est la même, cf. Chapitre 5). Il peut être intéressant de pouvoir définir une fonction sous-spécifiée "sujet-ou-objet", ou bien une fonction sous-spécifiée "complément-oblique".
Une telle hiérarchie de fonctions grammaticales a été proposée par exemple par Barbero et al. (1998), qui définissent une représentation hiérarchisée de la sous-catégorisation des mots, pour une grammaire de dépendances.

[41] Si un même identificateur de constante est utilisé dans plusieurs classes de la hiérarchie, et si ces différentes classes sont héritées par une même classe, un même identificateur réfère à une constante unique.

Partie II :
Métagrammaires
pour le français et l'italien

Nous décrivons dans cette partie, l'instanciation de la MG pour le français et l'italien, c'est-à-dire deux métagrammaires que nous désignerons dorénavant par MGf et MGi. La MGf a été écrite à partir d'une LTAG existante pour le français (Abeillé 1991), ce qui a permis de vérifier sa cohérence ainsi que d'étendre sa couverture de manière substantielle. L'ajout de nouveaux phénomènes est un travail réalisé conjointement avec Anne Abeillé (cf. Abeillé et Candito, à paraître). La formalisation dans la MG est de notre fait. La MGi a été écrite par contraste systématique avec la MGf[1].

Nous rappelons que la hiérarchie noyau sert à représenter la syntaxe d'une langue, via l'ensemble de schèmes qu'elle permet d'engendrer. Ces schèmes ne sont jamais utilisés tels quels par la TAG compilée à partir d'une MG, puisque la TAG n'utilise que des structures lexicalisées. Les formes fléchies ne sélectionnent que des schèmes compatibles avec leurs traits propres et les traits de leur lemme (cf. Chapitre 1). Donc, le nombre de schèmes d'une TAG n'a pas une influence directe sur son efficacité en analyse automatique. Ce qui compte pour l'efficacité, c'est le nombre de structures lexicalisées associées aux formes fléchies de la phrase à analyser.
Lorsqu'une TAG est écrite entièrement manuellement, un nombre de schèmes important rend difficile la mise à jour de la grammaire. Mais, si comme nous le faisons avec la MG, un outil aide à l'écriture de ces schèmes, leur nombre n'est pas un problème. On peut ainsi représenter à peu de frais des constructions peu productives, qui ne seront sélectionnées que par un nombre restreint de formes fléchies.

[1] Je précise que bien que connaissant bien l'italien, ce n'est pas ma langue maternelle. Le travail linguistique a été énormément facilité par l'existence de la *Grande Grammatica di consultazione* en trois volumes, dont j'ai utilisé principalement les deux premiers (Renzi, 1988; Renzi et Salvi, 1991). Il s'agit d'un ouvrage d'une grande précision et complétude, dont l'équivalent n'existe pas pour le français. J'ai en outre pour certains points interrogés des locuteurs italiens.
J'ai également esquissé une métagrammaire pour l'anglais, permettant de générer un sous-ensemble des schèmes de la grammaire du groupe XTAG. Elle n'est pas détaillée ici, mais est fournie en annexe, car elle est réduite et est plus accessible que la MGf et la MGi. Pouvoir représenter ce fragment de grammaire anglaise au sein de la MG permet de montrer que la MG est adaptable à des types d'analyse très différents : la grammaire XTAG comporte des noeuds VP, utilise des traces en cas d'extraction, ...

Actuellement la MGf comporte 240 classes et permet de générer environ 2400 schèmes (sans compter les arbres pour le causatif)[2]. La quasi-totalité des classes de la MGf a été reprise pour l'italien[3], avec des changements minimes, dont bien sûr ceux sur les éléments lexicaux co-têtes (prépositions, complémenteurs).
Les contraintes de compatibilité sont légèrement différentes en italien (et entraînent donc une combinatoire différente des classes croisées), ce qui est généralement dû au fait que l'ordre des mots est plus libre en italien. Dans la suite nous insisterons sur les points de divergence.

Le gros de notre travail concerne la représentation des schèmes à ancre verbale. Nous fournissons Chapitre 3, la description de la sous-catégorisation initiale pour les verbes (dimension 1), Chapitre 4, les redistributions de fonctions (dimension 2) et Chapitre 5, la réalisation grammaticale des fonctions (dimension 3). Nous avons cependant également travaillé sur la représentation des schèmes dont l'ancre peut constituer un modifieur, ce qui est détaillé au Chapitre 6.
La description des dimensions 1 et 2 reste relativement indépendante des structures syntagmatiques retenues pour les différentes constructions. C'est en dimension 3 que nous fournissons le détail de ces structures.

Pour l'italien, en ce qui concerne les catégories de constituants utilisées, nous avons repris les choix de (Abeillé 1991). En particulier, une structure plate de la phrase (sans SV) a été retenue également pour l'italien.
Les catégories syntaxiques utilisées sont, pour le français et l'italien :
V (verbe), P (phrase), N (nom, syntagme nominal), A (adjectif, syntagme adjectival), Ad (adverbe, syntagme adverbial), Prep (préposition), SP (syntagme prépositionnel, ou adverbial locatif/de manière), Dét (déterminant), C (complémenteur régi, conjonction de subordination), Conj (conjonction de coordination).

Ainsi, d'une manière générale, la même catégorie est utilisée pour représenter un nom et sa ou ses projections, de même que pour un adjectif, la distinction étant réalisée par des traits. Pour la projection du verbe, on ne distingue pas de SV rassemblant le verbe et ses compléments, mais on distingue tout de même la phrase par une catégorie distincte[4]. La préposition et le SP sont distingués. En revanche, un SP peut constituer la projection d'un adverbe locatif ou de manière, ces derniers ayant effectivement une distribution voisine de celle des "vrais SP" (comportant une préposition) locatifs.

Les traits morpho-syntaxiques utilisés pour les TAG sous-jacentes à la MGf et la MGi, sont les mêmes, sauf le trait :inv, qui en français code l'inversion éventuelle du sujet

[2] A titre de comparaison, la grammaire dont nous sommes partie comportait 850 schèmes. L'augmentation du nombre de schèmes provient de la systématisation des croisements de phénomènes linguistiques, de l'ajout de certains phénomènes, ainsi que de la duplication de certains schèmes, ne différant que par des traits (par exemple les schèmes pour les intransitifs sont systématiquement dupliqués selon l'auxiliaire de temps sélectionnés).
[3] Actuellement dans l'implémentation, il n'y a pas partage effectif des classes communes, mais copie.
[4] En pratique, par rapport à une distinction basée sur des traits, cela a l'avantage de pouvoir séparer les modifieurs de phrase des modifieurs de verbe, rien que par la catégorie de la racine et noeud pied des arbres auxiliaires. Mais on verra au Chapitre 5 que cela combiné avec le traitement de l'extraction par adjonction, oblige à considérer que les infinitives sont phrastiques.

(trait à 3 valeurs : "-" si pas d'inversion; "n" pour l'inversion du sujet nominal et "cl" pour l'inversion du sujet clitique). Ce trait n'est pas pertinent en italien (où il existe bien une inversion du sujet nominal, mais qui n'est pas sujette aux contraintes du français, cf. Chapitre 5).

Chapitre 3 : La sous-catégorisation initiale pour les verbes italiens et français

Pour décrire la sous-catégorisation initiale des verbes, nous commençons par les verbes déficients, qui ne sont utilisés que par combinaison avec un autre verbe, dont ils "héritent" d'un ou plusieurs arguments. Nous étudions ensuite l'hypothèse inaccusative, afin de déterminer la sous-catégorisation initiale retenue pour les verbes intransitifs. Enfin nous fournissons les différents cadres de sous-catégorisation prévus pour les verbes pleins italiens et français. Les cadres de sous-catégorisation sont exprimés en termes fonctionnels, et les différentes réalisations de fonctions sont détaillées au Chapitre 5, qui traite de la dimension 3 pour les verbes.

1 Les verbes "déficients"

Nous développons ici les cas de verbes déficients, tant sur le plan sémantique que sur le plan syntaxique.
Nous considérons des cas de verbes prenant un complément phrastique, infinitif ou participe, et qui se combinent avec ce verbe pour former par certaines propriétés un prédicat complexe. On parle de "déficience", dans la mesure où la sous-catégorisation de ces verbes est déterminée par le verbe enchâssé. Ces verbes ne sont pas représentés par des structures phrastiques, ils ne suivent donc pas la représentation en 3 dimensions. Nous allons considérer les cas suivants :

- auxiliaire de temps
- verbe causatif
- verbe restructurant (en italien seulement)

On traite également dans cette section des verbes à montée, qui sont également déficients par certains aspects, étant donné qu'ils constituent des prédicats unaires en sémantique, et qu'ils ne sélectionnent pas leur sujet.
Parmi ces différents cas, une barrière de phrase entre le "verbe enchâssant" (le verbe déficient) et le verbe "enchâssé" peut être plus ou moins présente. Dans le cas où cette barrière n'existe pas c'est-à-dire dans le cas où les deux verbes forment par certains aspects une seule unité lexicale, on parle généralement de prédicat complexe. Entre les deux situations on a un continuum, certaines constructions pouvant avoir un comportement hybride, suivant les propriétés étudiées. Par commodité on parle de verbe enchâssant dans tous les cas, mêmes pour les cas clairs de prédicat complexe où il n'y a pas d'enchâssement à proprement parler.

Les indices principaux de formation de prédicats complexes sont (i) la position des clitiques pronominaux et de la négation sur le verbe enchâssant, (ii) le fait que le verbe

enchâssé et ses compléments ne forme pas un constituant, et (ii) la possibilité pour le prédicat complexe de subir des redistributions comme le passif et le moyen.

Le traitement général en TAG pour ces verbes varie selon ces propriétés. On passe en revue les différents cas de verbes enchâssants, en donnant, de façon schématique le traitement qu'ils reçoivent en TAG généralement, ou bien spécifiquement dans la FTAG de (Abeillé, 1991) dont nous sommes partie. L'analyse précise sera donnée en dimension 3, sauf pour les constructions causatives, traitées par redistribution, donc en dimension 2.

On va voir que tous les cas de verbes enchâssants cités sauf le verbe causatif, sont analysés par adjonction du verbe enchâssant sur l'artère verbale de l'arbre pour le verbe enchâssé.

1.1 Les auxiliaires de temps

Les auxiliaires de temps constituent le cas le plus flagrant de formation d'un prédicat complexe entre l'auxiliaire et le participe pour un temps composé. Les clitiques se placent sur l'auxiliaire. Les tests classiques de constituance montrent que pour une phrase avec une forme verbale composée, le participe et les compléments ne forment pas un constituant (Abeillé et Godard, 1996a).

(10) Marie te l'a donné.

(11) a. *Ce qu'il a, c'est bu du vin.
b. *Bu du vin, Paul ne l'a pas.
c. *Paul voulait boire du vin mais il ne l'a pas.

Ces tests s'appliquent aussi aux équivalents en italien (voir par exemple Napoli, 1981).

L'analyse de Abeillé (1991) des auxiliaires de temps consiste à les adjoindre sur le participe, ce qui est en accord avec le fait que le participe et les compléments ne forment pas un constituant.
Un participe et une forme simple peuvent ancrer les mêmes schèmes, avec dans le cas d'un participe adjonction obligatoire d'un auxiliaire de temps par incompatibilité sur le trait mode. On donne Figure 43 l'adjonction d'un auxiliaire sur un arbre ancré par *dormi* (avec traits :mode et :aux uniquement). Dorénavant, on adopte la convention que les traits provenant des lexiques sont notés en gras. L'arbre pour le verbe intransitif doit recevoir une adjonction soit sur le noeud P, soit sur le noeud V, pour résoudre le conflit sur l'attribut mode.

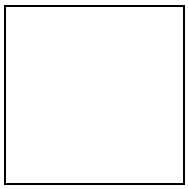


Figure 43. Adjonction d'un auxiliaire (forme *a*) sur un arbre pour une phrase simple

L'adjonction de l'auxiliaire résout le conflit. La sélection de l'auxiliaire est gérée par le trait :aux. Dans la MGf et la MGi, la valeur du trait :aux est spécifiée non pas au niveau des lexiques, mais au niveau des schèmes. En effet, l'auxiliaire effectif de temps ne dépend pas seulement de l'auxiliaire habituel, étant donné que la présence de réfléchis déclenchent l'auxiliaire /ETRE/ en français et /ESSERE/ en italien.

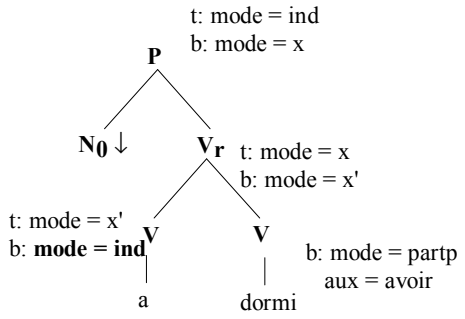


Figure 44. Résultat de l'adjonction de la Figure 43 (*N0 a dormi*).

La sélection de l'auxiliaire (ETRE/AVOIR en français, ESSERE/AVERE en italien) est gérée par le trait :aux, qui peut valoir être ou avoir, ou bien -. L'auxiliaire *effectif* de temps dépend de l'auxiliaire *habituellement* sélectionné par le verbe, mais également de la présence éventuelle d'un réfléchi en français, et en italien de la présence éventuelle d'un *si*, qu'il s'agisse du SI-nominatif ou du réfléchi). Donc dans la MGf et la MGi, le trait pour la sélection de l'auxiliaire n'est jamais spécifié dans le lexique, mais est directement codé dans les schèmes. Un transitif ne spécifie pas dans le lexique qu'il sélectionne l'auxiliaire AVOIR (AVERE en italien), cela est représenté dans les schèmes de la famille des transitifs, sauf les schèmes portant un réfléchi (un *si* en italien). Les intransitifs qui sélectionnent des auxiliaires différents sélectionnent des schèmes distincts, c'est-à-dire que l'on a systématiquement défini des familles différentes pour les intransitifs, selon leur auxiliaire[5].

L'analyse des auxiliaires par adjonction respecte le principe de cohérence sémantique et le PCPA. On peut en effet considérer qu'une forme verbale composée est constituée d'un sémantème (exprimé par le participe) dont les arguments se réalisent comme actants grammaticaux de la forme verbale composée; et que l'auxiliaire apporte une contribution sémantique de temps ou d'aspect, fonctionnant comme un foncteur sur un événement, ce qui est en accord avec le fait que l'arbre pour les auxiliaires comportent une seule position à étendre (le noeud pied verbal).
On peut cependant noter une légère disparité, étant donné que pour une forme composée, on considère les marques de temps comme autonomes et ancrant un arbre élémentaire (l'auxiliaire), alors que pour les formes simples, ces marques n'ancrent pas à elles seules un arbre élémentaire. Cela vient du parti pris de lexicalisme, qui ne dépasse pas la frontière du mot.

[5] C'est une propriété corrélée en français à la possibilité d'avoir une participiale adnominale ou absolue. En italien, étant donné que l'on a retenu l'hypothèse inaccusative, on a déjà vu que l'on distinguait les intransitifs sélectionnant ESSERE ou AVERE (i.e. inaccusatifs ou inergatifs) par des sous-catégorisations initiales différentes.

Pour l'italien, on peut considérer comme un auxiliaire le verbe STARE, qui marque le progressif. Il impose cependant le mode gérondif (12). Il est incompatible avec l'impératif (13a) et le passif (13b).

(12) Il treno sta partendo.
Le train STARE.3ind.prés PARTIRE.gérund
(Le train est en train de partir.)

(13) a. * Sta lavorando, fannullone!
STARE.3.impér.prés. TRAVAILLER.gérund, fainéant!
(Sois en train de travailler, fainéant)

b. * Il pranzo sta essendo servito al piano di sotto.
Le repas STARE.3ind.prés ETRE.gérund servi à l'étage au dessus.
(Le repas est en train d'être servi à l'étage au-dessus)

Nous le classons comme auxiliaire, car il porte les clitiques s'ils sont présents (14a). Il est représenté par un arbre similaire à celui des auxiliaires, sauf que le mode imposé est le gérondif. On bloque cette possibilité de mode pour les schèmes représentant un passif ou un impératif.
Cependant, le placement des clitiques sur le gérondif n'est pas formellement exclu (14b) (Cordin et Calabrese, 1988:551). Nous ne traitons cependant que le cas de type auxiliaire (pour traiter l'autre cas, il faudrait comme pour les verbes restructurants considérer deux structures différentes).

(14) a. **Gli** sto dando troppa fiducia.
DAT1.3ms je-STARE.3ind.prés DONNER.gérund trop de confiance
(Je lui fais petit à petit trop confiance)

b. Sto dando**gli** troppa fiducia.
je-STARE.3ind.prés DONNER.gérund-DAT1.3ms trop de confiance

1.2 Les verbes causatifs

Les verbes causatifs dans les langues romanes sont bien connus pour mettre en jeu un type particulier de complémentation phrastique. On donne l'analyse complète des données et la représentation dans la MGf et la MGi au Chapitre 4, sur les redistributions. Nous ne donnons ici qu'un résumé des arguments en faveur de la formation d'un prédicat complexe, et le traitement que cela nécessite en TAG.
Dans le cas d'une construction causative *faire + Vinf*, le sujet de l'infinitif se réalise comme un objet si l'infinitif est intransitif (15a), comme un a1-objet ou un agt-objet si l'infinitif est transitif (15b).

(15) a. Marie fait rire Jean
b. Marie fait visionner les bandes (à Jean + par Jean).

L'infinitif ne peut pas porter de clitiques pronominaux ni une négation (en fait pour le français, d'une part il faut distinguer une construction concurrente où il n'y a pas formation d'un prédicat complexe (cf. Abeillé et al., 1997, et les références citées), d'autre part les clitiques idiomatiques ou réfléchis peuvent se placer sur l'infinitif, même dans le cadre de la structure avec prédicat complexe, cf. Chapitre 4). Les compléments pronominalisés de l'infinitif se placent sur le verbe causatif en français (16) et en italien.

(16) Marie le fait résoudre par ses élèves / à ses élèves.

Pour ce qui est de la possibilité de mettre au passif ou au moyen, le français et l'italien diffèrent. En français il n'est pas possible de mettre un prédicat complexe causatif au passif (17b), alors que cela est possible pour l'italien (17b'). Pour le moyen, il est difficile en français, possible en italien (18b) (Skytte et al., 1991:502-503).

(17) a. Jean fit tomber le livre.
a'. Jean fece cadere il libro.
b. *Le livre a été fait tombé (par Jean).
b'. Il libro fu fatto cadere (da Jean).

(18) a. Il proprietario vuole vendere queste case a caro prezzo.
Le propriétaire veut vendre ces maisons au prix fort.
b. Queste case si vogliono vendere a caro prezzo.
Ces maisons SI veulent vendre au prix fort.

La formation d'un prédicat complexe causatif est donc légèrement plus claire en italien qu'en français. Mais dans les deux langues, la construction causative est bi-clausale du point de vue de certaines propriétés comme le liage.

Le traitement en TAG nécessite des schèmes spécifiques pour l'infinitif entrant dans une construction causative, étant donné que le sujet de l'infinitif (le causé) ne porte pas la fonction sujet dans la construction causative (et donc n'a pas les mêmes positions syntagmatiques possibles).
Une solution en TAG est de définir des arbres élémentaires spécifiques pour la construction causative, ancrés par l'infinitif, mais comprenant le verbe causatif comme co-tête et les positions adéquates pour le causateur et le causé.

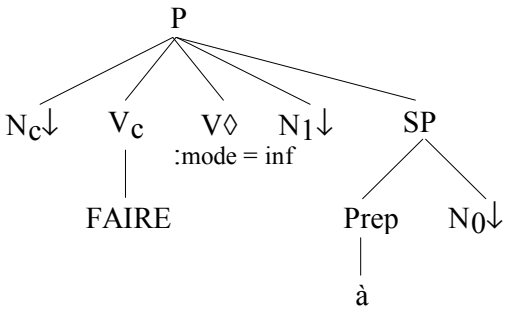


Figure 45. Schème pour une construction causative

La variation sur le verbe causatif (FAIRE et LAISSER sont possibles, ainsi que certains verbes de perception) devrait être en théorie gérée par une disjonction sur la co-tête. En pratique dans le système XTAG, on doit prévoir un noeud à substituer pour le verbe causatif. Cela n'est pas conforme au PCPA, puisque celui-ci requiert que les positions à étendre (les noeuds pied et les noeuds à substituer représentent des arguments sémantiques de l'ancre).
Les schèmes pour la construction causative sont représentés par une redistribution dans la MGf et la MGi (plus exactement plusieurs redistributions, selon la transitivité de l'infinitif).

1.3 La copule

Etant donné les principes sémantiques associés aux arbres élémentaires en TAG, pour représenter le passif en TAG, on utilise des schèmes où l'ancre principale est le participe passé, et "l'auxiliaire du passif" apparaît en théorie comme co-tête. C'est-à-dire que l'on considère cet "auxiliaire" et le participe comme un prédicat complexe. Cependant, il y a variation possible sur cet "auxiliaire", dans le cas d'un passif personnel. Avec l'interprétation stative en effet, tout verbe d'état peut entrer dans un prédicat complexe passif. On choisit pour cette raison de traiter "l'auxiliaire du passif" par substitution, de manière similaire au verbe causateur (cf. la section précédente).
La copule, et tous les verbes d'état projettent un arbre élémentaire pour un verbe "nu" : sans complément, et de racine V. Des traits :cop et :caus permettent de contrôler la substitution de ces arbres dans les arbres appropriés (causatifs ou passifs).

1.4 Les verbes restructurants en italien

1.4.1 Les données

En italien, certains verbes enchâssants prenant une infinitive (certains verbes à contrôle et certains verbes à montée) peuvent porter de manière optionnelle les clitiques compléments de l'infinitif.
Rizzi (1978) en a donné une analyse dans le cadre de la théorie standard étendue, et les a nommés "verbes restructurants" (ci-après VR). Il a montré que le cas avec clitique(s) sur le verbe enchâssé et celui avec clitique(s) sur le verbe enchâssant diffèrent par de nombreuses propriétés. On doit ainsi distinguer, au moins en syntaxe, l'emploi restructurant (19a) de l'emploi non restructurant (19b) (ici il s'agit d'un verbe à montée, mais cela pourrait être un verbe à contrôle).

(19) a. Mario lo può fare.
Mario ACC.3ms peut faire

b. Mario può farlo.
Mario peut faire-ACC.3ms

Les VR se classent en trois catégories. Il s'agit de (Skytte et al., 1991)[6] :

- verbes modaux :
/DOVERE/ (devoir), /POTERE/ (pouvoir), /SAPERE/ (savoir), /VOLERE/ (vouloir), /AVERE-A/ (avoir à), /SOLERE/ (avoir l'habitude de)
- verbes aspectuels
/COMINCIARE-A/ (commencer à), /INCOMINCIARE-A/ (commencer à, inchoatif), /RICOMINCIARE-A/ (recommencer à), /CONTINUARE-A/ (continuer à), /FINIRE-DI/ (finir de), /SEGUITARE-A/ (continuer à), /STARE-PER/ (être sur le point de)
- verbes d'état ou de mouvement

[6] Certains verbes sont incontestablement restructurants, pour d'autres les données peuvent varier d'une région à une autre, ou même au sein d'un groupe de locuteurs homogène.

/ANDARE-A/ (aller), /TORNARE-A (retourner), /VENIRE- A/ (venir), /STARE- A/ (rester)

La propriété des VR de porter les clitiques est corrélée à d'autres indices de formation d'un prédicat complexe. Nous détaillons ci-dessous les arguments de Rizzi. On confronte systématiquement le cas où le VR porte un clitique complément de l'infinitif, indice de l'emploi restructurant, et celui où l'infinitif porte un clitique.

- l'infinitif ne peut pas être nié :

(20) a. *__Gli__ vorrei non parlare.
DAT.3ms je-voudrais NEG parler

b. Vorrei non parlar__gli__.
je-voudrais NEG parler-DAT.3ms
(Je voudrais ne pas lui parler)

- le VR et l'infinitif peuvent apparaître dans une construction moyenne, comme s'il s'agissait d'un seul verbe :

(21) a. Queste case **gli** si vogliono vendere a caro prezzo.
Ces maisons DAT.3ms RF(moy) veulent vendre au prix fort.
(On veut que ces maisons se vendent au prix fort.)

b. *Queste case si vogliono vender__gli__ a caro prezzo.
Ces maisons RF(moy) veulent vendre-DAT.3ms au prix fort.

- l'infinitif et ses compléments ne forment pas un constituant :
Si le verbe enchâssant porte un clitique complément de l'infinitif (i.e. si le verbe enchâssant est en emploi restructurant), alors l'infinitive ne peut pas être clivée (contraste (22a)/ (22b)), ne peut pas entrer dans la construction pied-piping (contraste (23a)/(23b)), ne peut pas être postposée à un circonstanciel (contraste (24a)/(24b)).

(22) a. * E` proprio riportare i soldi che **gli** vorrei.
C'est justement rapporter les sous que DAT.3ms je-voudrais

b. E` proprio riportar__gli__ i soldi che vorrei.
C'est justement rapporter-DAT.3ms les sous que je-voudrais
(C'est justement lui rapporter les sous que je voudrais.)

(23) a. *Questi argomenti, a parlare dei quali **ti** verrò al più presto.
Ces questions, à parler desquels DAT.2 je-viendrai le plus rapidement

b. Questi argomenti, a parlar__ti__ dei quali verrò al più presto.
Ces questions, à parler-DAT.2 desquels je-viendrai le plus rapidement
(Ces questions, à te parler desquelles je viendrai le plus rapidement possible)

(24) a. *__Ne__ ho cominciato da Gianni a discutere.
GEN. j'ai commencé chez Gianni à discuter

b. Ho cominciato da Gianni a discutere__ne__.
J'ai commencé chez Gianni à discuter-GEN
(J'ai commencé à en discuter chez Gianni.)

- L'auxiliaire de temps, dans le cas restructurant, où le VR sélectionne ESSERE est toujours ESSERE (Rizzi, 1978). Ainsi, en (25), le complexe VR + Vinf sélectionne ESSERE, imposé par le RV /STARE-PER/.

(25) Sono/*ho stato per fare une schiocchezza.
Je-suis/*j'ai été pour faire une bêtise.
(J'ai été sur le point de faire une bêtise.)

Mais, l'auxiliaire de temps, dans le cas restructurant, ou le VR sélectionne AVERE, est celui sélectionné par l'infinitif, plus précisément, par le dernier infinitif entrant dans le complexe restructuré, celui-ci pouvant en effet comprendre plusieurs VR en cascade.

En (26a) l'auxiliaire est celui de PRENDERE (c'est-à-dire AVERE). En (26b), PRENDERE n'est pas dans le complexe restructuré (il porte un clitique). ANDARE, est lui-même un RV potentiel, qui prend pour auxiliaire ESSERE, alors que VOLERE sélectionne AVERE. Si ANDARE est dans son emploi restructurant, l'auxiliaire du complexe VOLERE + ANDARE est celui de ANDARE; sinon, l'auxiliaire de VOLERE est son auxiliaire habituel, AVERE.

(26) a. Mario li avrebbe voluti andare a prendere lui stesso.
Mario ACC.3mp aurait voulus aller à prendre lui-même.
(Mario aurait voulu aller les chercher lui-même)

b. Mario avrebbe / sarebbe voluto andare a prenderli lui stesso.
Mario aurait/serait voulu aller à prendre ACC.3mp lui-même.
(Mario aurait voulu aller les chercher lui-même)

Mais, si l'auxiliaire dans le cas restructurant est obligatoirement celui du dernier infinitif, on peut parfois trouver un changement d'auxiliaire dans des cas où l'infinitif porte un clitique, et où donc le verbe enchâssant n'est pas dans son emploi restructurant (27a) (Skytte at al., 1991:517).

(27) a. Gianni è voluto andarci.
Gianni est voulu aller-loc
(Gianni a voulu y aller)

b. Gianni ci è voluto andare.
Gianni loc est voulu aller
(Gianni a voulu y aller)

Rizzi donne également le test du Right Node Raising, mais le contraste entre emploi VR et emploi non VR est mis en doute par Monachesi (1995:174, note 12). Il signale enfin que les VR peuvent être utilisés sous une construction *adj da Vinf* (équivalent du français *adj à Vinf*), alors que normalement, cette construction ne permet pas d'enchâssement entre l'objet extrait et l'infinitif. Mais les données sont délicates, et Rizzi ne les croise pas avec la propriété de placement des clitiques.

Enfin il est intéressant de noter que le prédicat complexe verbe restructurant + infinitif prend la transitivité de l'infinitif (Skytte at al., 1991:519), ce qui est mis en évidence par le test du causatif (cf. Chapitre 4, la réalisation du sujet de l'infinitif dans une construction causative dépend de la transitivité de l'infinitif). Ainsi on peut comparer (28a) où /ANDARE-A/ n'est pas dans son emploi restructurant, et est intransitif pour le causatif (*Mario* est objet), et (28b) où /ANDARE-A/ est restructurant et est transitif (parce que /PRENDERE/ est transitif) pour le causatif (*Mario* est a1-objet). Le clitique gouverné

par l'infinitif, se place normalement sur le verbe restructurant, mais comme ce dernier est en plus dans un complexe causatif, le clitique se place en fait sur FAIRE.

(28) a. Piero farà andare Mario a prender**lo**.
Piero fera aller Mario à prendre-ACC.3ms
(Piero fera aller Mario le chercher)

b. Piero **lo** farà andare a prendere a Mario.
Piero ACC.3ms fera aller à prendre à Mario.
(*Piero le fera aller chercher à Mario)

1.4.2 La sémantique de la restructuration

La caractérisation sémantique des verbes à infinitive pouvant entrer dans l'emploi restructurant a fait l'objet de beaucoup de recherches (voir par exemple Napoli, 1981, et les références citées). A première vue, il est difficile de trouver une homogénéité sémantique. Les VR peuvent être aussi bien des verbes à montée (i.e. dont le N'est pas un argument sémantique) comme /STARE-PER/, que des verbes à contrôle (dont le sujet est un argument) comme /VOLERE/. De plus, des verbes à sémantique très proche, mais différant par leur fréquence d'emploi peuvent diverger quant à la sélection du comportement restructurant : /DESIDERARE/ (désirer) n'est pas un VR, au contraire de /VOLERE/ (vouloir).
Mais, le fait que les verbes restructurants peuvent alternativement entrer dans l'emploi restructurant ou pas, pousse à chercher une explication sémantique. Napoli (1981) va même jusqu'à considérer que la restructuration n'est pas lexicalement déterminée, mais est déclenchée par le contexte sémantique. De manière grossière, on peut dire que les verbes dans leur emploi restructurant ont une sémantique "affaiblie" : ils n'expriment pas un état ou une action par eux-mêmes, mais ajoutent de l'information sur un état ou une action dénotée par l'infinitif (Napoli, 1981:860). Ils se rapprochent en cela de la sémantique d'un auxiliaire : il s'agit de foncteurs, mais la sémantique de la combinaison foncteur + argument (VR + infinitif, auxiliaire + participe) est un type de la sémantique de l'argument (l'infinitif ou le participe). Nous introduirons au Chapitre 7 le terme de modifieur sémantique pour de tels foncteurs, et nous verrons que les modifieurs sémantiques sont systématiquement adjoints en TAG.

1.4.3 L'analyse en TAG

Rizzi montre clairement que les verbes restructurants ont deux structures syntaxiques possibles, un emploi non restructurant (comme verbe à montée ou comme verbe à contrôle, selon les verbes), et un emploi restructurant, où le RV et le ou les infinitifs forment un prédicat complexe.
Etant donné le statut de modifieur sémantique des VR dans leur emploi restructurant, nous proposons d'utiliser exactement le même mécanisme que celui pour les auxiliaires de temps, c'est-à-dire une adjonction du verbe restructurant sur l'arbre de l'infinitif. Les phrases avec VR +Vinf, ou bien avec V simple sont dérivées à partir des mêmes schèmes, avec ou sans adjonction d'un VR, de la même manière que l'on a les mêmes schèmes pour les formes verbales simples, et les formes verbales à un temps composés.

Les seules différences entre l'auxiliaire de temps et le verbe restructurant sont (i) la différence de mode du verbe "enchâssé" (participe versus infinitif) et (ii) que le VR se conjugue avec l'auxiliaire de temps de l'infinitif, alors que l'auxiliaire de temps ne peut pas être lui-même à une forme composée. On donne Figure 46 l'exemple de l'adjonction du verbe restructurant pour dériver une phrase comme (29).

(29) Mario ci vuole andare.
Mario LOC veut aller
(Mario veut y aller.)

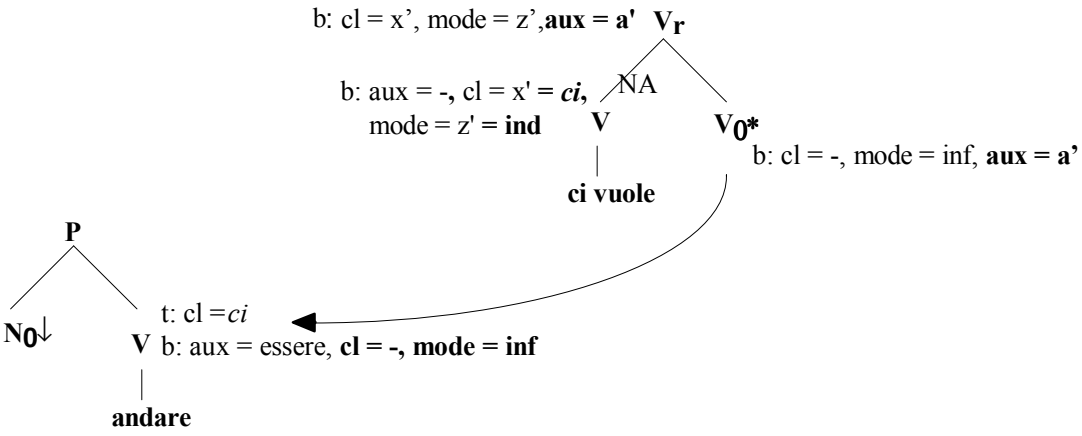


Figure 46. Adjonction d'un verbe restructurant sur un verbe dont un actant est réalisé comme clitique (locatif).

On anticipe sur le traitement des clitiques comme affixes flexionnels, qui sera détaillé au chapitre 5. Un trait :cl code les caractéristiques morphologiques des clitiques devant être réalisés. La valeur de ce trait a son origine dans l'arbre pour l'élément qui déclenche la présence des clitiques (pour (29), la présence de *ci* est déclenchée par *andare*), mais, par le jeu d'adjonction, ce trait peut "passer" sur un autre verbe, en l'occurrence le VR, d'où la possibilité d'adjoindre la forme *ci vuole*.

Pour un VR à un temps composé, les auxiliaires de temps s'adjoignent à la racine de l'arbre du VR. Dans le cas d'un VR qui dans son emploi non restructurant sélectionne l'auxiliaire AVERE, le complexe *VR + Vinf* sélectionne l'auxiliaire de temps du *Vinf* (le *Vinf* le plus à droite dans le cas de plusieurs *Vinf*). Cette sélection est faite par coindexation des traits aval :aux entre le noeud pied et la racine de l'arbre pour un VR. Ainsi on peut vérifier que dans l'arbre dérivé résultant de l'adjonction donnée Figure 46, *ci vuole* récupère le trait aval :aux de l'infinitif, en l'occurrence ESSERE. Pour les VR qui sélectionnent ESSERE dans leur emploi non restructurant, au lieu de cette coïndexation, un trait :aux = essere est simplement spécifié sur la racine de l'arbre du VR.

Les VR dans leur emploi non restructurant sélectionnent les arbres pour les verbes à montée ou les verbes à contrôle selon les cas. Cela revient à considérer que ces verbes n'ont pas la même sous-catégorisation dans leur emploi restructurant et leur emploi non restructurant. Rizzi (1978) avait opté, lui, pour une analyse par règle, dérivant l'emploi

restructurant de l'emploi non restructurant. Dans la MG, on aurait pu opter pour une règle de type redistribution sur la sous-catégorisation dans le cas non-restructurant.
Rizzi donne comme argument en faveur d'une règle, le fait qu'un verbe dans son emploi restructurant et non-restructurant a les mêmes caractéristiques sémantiques et ne montre pas de différence dans la sélection d'un complémenteur.
Nous avons vu au contraire (suivant Napoli) que la restructuration met en jeu une différence de perspective sémantique. La solution pour nous est de considérer deux entrées lexicales.

1.5 Les verbes à montée

On rappelle que les constructions avec verbes à montée plus infinitives peuvent être caractérisées en sémantique, par le fait que le verbe à montée est un sémantème à un argument, dénotant un événement au sens large. Cela se traduit en particulier par le fait que le verbe à montée n'impose pas de restriction de sélection sur son sujet, et peut donc prendre en particulier un sujet idiomatique faisant partie d'une expression figée avec l'infinitif, ou un sujet impersonnel, requis par l'infinitif. On classe donc par la sémantique les verbes à infinitive, en verbe à montée du sujet ou verbe à contrôleur sujet, selon que le sujet apparent du verbe enchâssant est un argument sémantique de ce verbe enchâssant.

(30) a. Il semble pleuvoir.
b. *Il veut pleuvoir.

On parlera simplement de verbes à montée pour les verbes à montée du sujet. On verra à la fin de cette section le cas des verbes à montée de l'objet.
L'analyse en TAG des verbes à montée (cf. (Kroch et Joshi, 1985) pour l'anglais, (Abeillé, 1991) pour le français) code les verbes à montée comme des arbres auxiliaires s'adjoignant sur un noeud VP (pour l'anglais) ou V (pour le français, étant donnée la structure plate).

1.5.1 L'analyse pour le français

On donne Figure 47 un exemple d'adjonction d'un verbe à montée, avec les traits de mode. On détaillera le placement des clitiques au Chapitre 5.

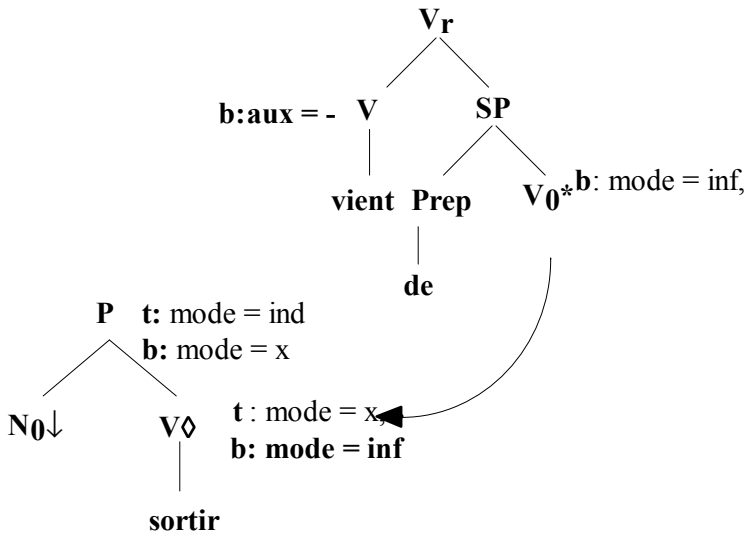


Figure 47. Adjonction d'un verbe à montée sur un intransitif (*N0 [vient de] sortir*)

Cette analyse est en accord avec le PCPA, puisque le verbe à montée, un sémantème unaire, ne contient alors qu'une seule position à étendre (le noeud pied verbal). On peut noter cependant une légère disparité, puisque l'argument sémantique du verbe à montée est représenté en syntaxe par un noeud verbal, alors qu'il s'agit en sémantique d'un événement, comprenant l'argument réalisé comme sujet. On verra Chapitre 7 que cette disparité peut poser des problèmes dans l'interprétation sémantique de l'arbre de dérivation, dans le cas d'interaction avec d'autres adjonctions enchâssantes.
On peut faire un parallèle avec l'analyse de la grammaire générative, qui considère que le verbe à montée n'a pas de sujet en structure profonde. Mais ici c'est légèrement différent. Le sujet du verbe à montée, qui n'en est pas un argument sémantique, est localisé dans l'arbre de l'infinitif. Cependant, du point de vue syntaxique, on représente par une contrainte d'adjonction obligatoire le fait que ce sujet ne peut pas être directement le sujet de l'infinitif.

Par ailleurs, du point de vue syntaxique, l'analyse proposée ne donne pas le statut de syntagme à l'infinitive et ses compléments. Pour le français, Abeillé (1998) montre que les verbes à montée ont d'un point de vue syntaxique un statut hybride avec des propriétés indices de la formation d'un prédicat complexe et d'autres indices d'une structure bi-clausale.
Les propriétés de type prédicat complexe sont que l'infinitif et ses compléments échouent aux tests typiques de constituance comme la clivée ou la pronominalisation; qu'un verbe à montée, pour certains locuteurs, peut se trouver sur un infinitif dans une construction en *adj à Vinf* (qui n'admet pas d'enchâssement en français); que certains groupes verbe à montée (aspectuel) plus infinitif peuvent être passivés.
Les propriétés des verbes à montée typiques d'une structure bi-clausale sont que le verbe à montée ne porte pas les clitiques compléments de l'infinitif, et que la négation peut se placer sur l'infinitif, même si cela est moins naturel que sur le verbe à montée. La négation de l'infinitif (31a) a, pour beaucoup de verbes à montée, un sens différent de la négation sur le verbe à montée (31b).

(31) a. Ils recommencent à ne pas répondre au téléphone.
b. Ils ne recommencent pas à répondre au téléphone.

L'analyse par adjonction permet de rendre compte des deux types de propriétés, en jouant sur les traits imposés sur le noeud pied (clitiques possibles/impossibles; négation possible/impossible).

1.5.2 L'analyse pour l'italien

Pour l'italien, l'analyse est parallèle à celle du français. L'interaction avec les clitiques est légèrement différente, puisque le seul clitique nominatif en italien (le SI-nominatif) n'a pas la distribution des clitiques nominatifs en français. On renvoie au Chapitre 5 pour le traitement du placement des clitiques.

Il faut noter que pour l'italien comme pour le français, les emplois alternatifs de certains verbes à montée, avec une complétive (*il semble que P/sembra che P*) sont considérés comme des sous-catégorisation initiales différentes de l'emploi comme verbe à montée.

1.5.3 Les verbes à montée de l'objet

On a mentionné au Chapitre 1, le cas des verbes à montée de l'objet, comme difficulté pour une théorie qui pose des contraintes strictes sur la localité des arguments sémantiques d'un lexème.
La distinction contrôle/montée est faite essentiellement selon le statut d'argument sémantique ou pas d'un actant du verbe à montée. Ce statut peut être testé lorsque cet actant éventuellement "monté", est non référentiel, en particulier s'il fait partie d'une expression figée. Pour un verbe comme /VOIR/, on peut trouver des cas d'enchâssements d'expressions à sujet figé inacceptables (32a), douteux (32b), mais également acceptables (32c).

(32) a. *J'ai vu ciel et terre être remués par certains candidats
b. ?J'ai vu la moutarde lui monter au nez.
c. Après des années de poisse, Paul a vu la chance lui sourire.

On a également des exemples comme (33), où l'objet d'entendre n'est pas littéralement l'entité qui fait le bruit.

(33) J'entends le prof se faire chahuter par les élèves. (Muller, 1995)

Pour l'italien, Skytte et al. (1991:509-510) arguent en faveur d'une analyse où l'objet du verbe de perception n'en est pas un argument sémantique, du fait de phrases comme (34a) où l'objet de la vue n'est pas visible, et (34b) où l'objet du verbe de perception forme une expression figée avec l'infinitive.

(34) a. Ho visto il vento muovere le foglie.
J'ai vu le vent remuer les feuilles
b. Ho visto mari e monti essere mossi per certi candidati.
J'ai vu mers et monts être remués par certains candidats
(J'ai vu certains candidats qui remuaient ciel et terre.)

L'analyse des verbes à montée du sujet, se fait par adjonction : le sujet "de surface" des verbes à montée, est localisé dans l'arbre de son gouverneur sémantique, et pas dans l'arbre du verbe à montée.
Malheureusement, l'analyse parallèle des verbes à montée de l'objet pose problème, étant donné que l'on ne peut pas alors représenter les régularités liées au statut syntaxique d'objet de l'argument "monté", comme la réalisation possible comme un clitique accusatif (cf. (35a) pour l'italien, (35a') pour le français), et dans certains cas rares, la possibilité de trouver le verbe à montée au passif, avec l'actant "monté" en position sujet, comme cela est régulier pour un objet (initial) (cf. (35b) pour l'italien, mais (35b') est douteux en français).

(35) a. Maria l'ha sentito suonare il pianoforte.
a'. Maria l'a entendu jouer du piano.
b. Piero è stato sentito suonare il pianoforte da Maria. (Skytte et al., 1991:509).
b'. ?Piero a été entendu jouer du piano par Marie.

Cela dit, malgré la possibilité d'enchâssement de certaines expressions figées, on peut tout de même remarquer qu'il y a un continuum entre la situation claire où l'argument "monté" n'est pas un argument sémantique, et celle où un argument est partagé par deux prédicats (c'est le cas verbe à contrôle).

Ainsi, si en (33), on peut évidemment mettre en cause le fait que 'prof' soit un argument sémantique d' 'entendre' (puisque ça n'est pas physiquement le prof que l'on entend), il nous semble tout de même qu'il existe une nuance entre (33) et (36a), alors que l'analyse stricte à montée prévoit une équivalence, et une nuance encore plus nette entre (33) et (36b).

(36) a. J'entends les élèves chahuter le prof.
b. J'entends que le prof se fait chahuter par les élèves.

Tout se passe comme si, la position comme actant grammatical d'un verbe poussait l'interprétation de cet actant comme un argument de ce verbe.
Il faudrait donc considérer plusieurs lexèmes pour le lemme ENTENDRE, dont l'un d'entre eux aurait pour objet, non pas l'entité physiquement source du bruit, mais plutôt l'entité qui est la cause du bruit. La distinction entre (33) et (36a), serait qu'en (33), on insiste sur le responsabilité du professeur pour la situation de chahut, ce qui n'est pas le cas en (36a).

Prenons également l'exemple de /EMPECHER/, que Bashung (1992) considère comme un verbe à montée. Si l'on considère la série (37), il nous semble également que, dans le cas d'un objet humain en tous cas, il y a une nette nuance sémantique entre la construction avec infinitive et celle avec complétive. En (37d), l'action "d'empêchement" peut être indirecte, ce qui n'est pas possible en (37c).

(37) a. J'empêche la tasse de tomber.
b. J'empêche que la tasse tombe.
c. J'ai empêché Jean de sauter.
d. J'ai empêché que Jean saute.

Dans la MGf et la MGi, nous ne distinguons pas de verbes à montée de l'objet, c'est-à-dire que nous considérons que ces supposés verbes sous-catégorisent un objet qui en est

un argument sémantique. Il faut alors pour l'interprétation sémantique distinguer l'emploi avec infinitive (38a) et l'éventuel emploi avec complétive (38c), ou sans infinitive (38b).

(38) a. Paul entend le prof se faire chahuter.
b. Paul entend le prof.
c. Paul entend que le prof se fait chahuter.

2 Les verbes pleins

Par contraste, on distingue des verbes déficients, les verbes pleins, qui dans le cas non marqué ont des actants grammaticaux référentiels (non figés) qui correspondent à leurs arguments sémantiques. Nous détaillons la projection que peut avoir un verbe dans les TAG du français et de l'italien, puis nous étudions Section 2.2 ce qu'on appelle l'hypothèse inaccusative, afin de déterminer la sous-catégorisation initiale considérée pour les verbes intransitifs en français et en italien. Nous pourrons alors fournir Section 2.3 les différentes fonctions grammaticales distinguées et les différents cadres de sous-catégorisation initiale prévus, pour la MGf et la MGi.

2.1 La projection d'un verbe : "l'artère" dans un arbre à ancre verbale

Nous exposons dans cette section les possibilités de projection d'un verbe, c'est-à-dire les types de structures que peut ancrer un verbe, dans la MGf et la MGi, et les LTAG sous-jacentes.
On rappelle que l'ancre est le noeud dominant l'élément lexical qui sélectionne le schème. Dans le cas d'ancres multiples (pour une expression figée, ou pour un verbe et une préposition régie par exemple), on distingue l'ancre principale, comme celle la plus saillante en syntaxe.
Nous reprenons les choix linguistiques de (Abeillé, 1991), qui a montré en particulier l'inadéquation d'un syntagme verbal pour le français. Les arbres élémentaires à ancre verbale ont donc une structure plate, sans noeud SV.
Pour pouvoir donner une description de l'artère d'un schème à ancre verbale, il faut remarquer les points suivants :

- L'ancre verbale est souvent la tête syntaxique de la phrase minimale représentée par l'arbre élémentaire, mais ce n'est pas toujours le cas. Les arbres de la MGf et MGi représentant un passif comportent une position prévue pour l'auxiliaire du passif, l'ancre au participe passé étant placée à droite de l'auxiliaire. Ceux représentant un complexe causatif, *faire + Vinf*, sont ancrés par l'infinitif et comportent une position prévue pour le verbe causatif, qui porte les inflections (nous renvoyons au Chapitre 4 pour l'étude du passif et du causatif en français et en italien). On doit donc distinguer la notion d'ancre de la notion de tête syntaxique de l'arbre élémentaire.

- D'autre part, en TAG, la complémentation phrastique est en grande partie traitée par adjonction du verbe enchâssant dans la structure du verbe enchâssé. Il s'agit d'adjonction enchâssante, c'est-à-dire que dans l'arbre adjoint, la tête syntaxique n'est pas le noeud pied (représentant le complément phrastique) mais l'ancre, qui est le verbe enchâssant. Donc après adjonction enchâssante sur un arbre à ancre verbale, la tête syntaxique de l'arbre dérivé est la tête syntaxique de l'arbre adjoint, et pas celle de l'arbre ayant reçu l'adjonction. Comme des contraintes exprimées dans un arbre élémentaire peuvent — par le jeu des unifications de traits dans le cas d'adjonction — passer sur un noeud de l'arbre adjoint, il faut également distinguer au sein d'un arbre élémentaire les contraintes qui vont rester locales, et celles qui peuvent concerner un éventuel arbre adjoint.

On appelle **tête locale** le verbe qui, s'il n'y a pas d'adjonction enchâssante, sera la tête de l'arbre dérivé.
La constante ?**Vanc** désigne le noeud dominant l'ancre verbale (ce noeud est marqué graphiquement par un signe ◊).
La constante ?**Vmarq** désigne le noeud dominant le verbe tête locale. Hors des cas de passifs et de complexes causatifs, ?Vanc et ?Vmarq désignent le même noeud. Dans le cas de passifs (respectivement de complexes causatifs), ?Vmarq désigne le noeud dominant la copule (respectivement le verbe causatif), qui a une position prévue dans le schème. On donne Figure 48 des exemples de schèmes pour un transitif à arguments nominaux, à l'actif simple, au passif et au causatif, en indiquant le noeud nommé par ?Vmarq. Ils peuvent être utilisés respectivement pour dériver les phrases *Marco suit Chiara / Chiara est suivie par Marco / Paola fait suivre Chiara par Marco.*

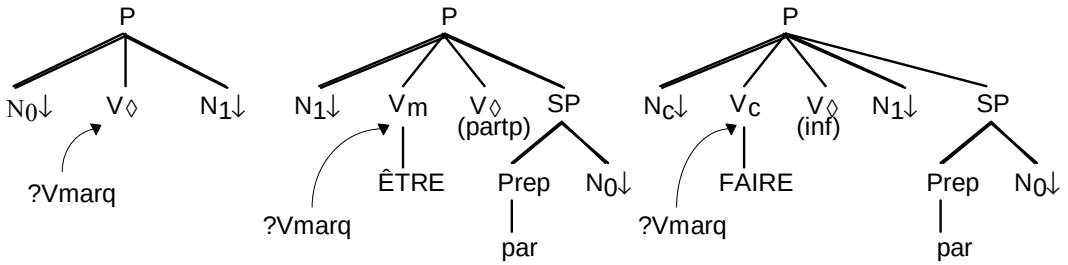


Figure 48 : Schèmes pour un transitif à l'actif, au passif, au causatif (faire-par)[7]

Dans la MGf et la MGi, le verbe projette dans la plupart des cas une phrase. On a vu en effet que l'on traitait les infinitives comme des phrases à sujet vide. Cela nous est en partie imposé par le traitement des dépendances non bornées[8], mais également pour

[7] La figure fait apparaître des co-têtes pour la copule "auxiliaire du passif", et pour le verbe causatif. Cependant, d'autres verbes peuvent jouer ces rôles (tout verbe d'état pour un passif avec interprétation stative) et LAISSER par exemple peut entrer dans la construction causative. En pratique, pour gérer cette variation, on utilise une position à substituer (cf. Sections 1.2 et 1.3), qui ne joue pas le rôle des positions à substituer argumentales dans l'interprétation sémantique.
[8] Du moins dans le cas des TAG standards, qui imposent l'égalité catégorielle entre racine et noeud pied d'un arbre auxiliaire.

garder une certaine homogénéité dans l'interface sémantique lexicale / syntaxe, entre la réalisation d'un verbe comme infinitif, ou comme forme finie. Dans les deux cas, si le verbe a un sujet qui représente un argument, cet argument doit pouvoir être représenté localement, ne serait-ce que par une trace.
Cependant, dans le cas d'un participiale adnominale, on considère que le réalisation du sujet du participe est le nom modifié, qui peut être localisé dans l'arbre représentant la participiale. Ainsi pour dériver la participiale présente de (39), on utilise le schème de la Figure 49, où le verbe ne projette pas de phrase.

(39) (les machines) produisant ces pièces (coûtent 500 000F chacune).

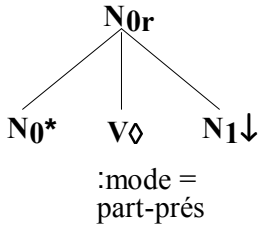


Figure 49. Schème pour une participiale présente adnominale

On appelle **artère** la séquence de noeuds en relation de dominance immédiate comprenant ?Vmarq et ses projections jusqu'au premier noeud non verbal, que l'on nomme par la constante **?P**. Comme on n'utilise pas de constituant SV, le verbe est toujours immédiatement dominé par un noeud non verbal. C'est-à-dire que l'artère est réduite à ?Vmarq dominé par ?P[9]. ?P nomme dans la plupart des cas un noeud phrastique, sauf comme on vient de le voir, dans le cas d'une participiale adnominale, où il s'agit d'un noeud nominal.

Les adjonctions de verbes enchâssants ont lieu sur l'artère. En cas d'adjonction, la tête syntaxique de l'arbre dérivé est l'ancre de l'arbre adjoint. On appelle **tête finale** le verbe tête syntaxique après toutes les adjonctions. S'il y a plusieurs adjonctions enchâssantes le long de l'artère (par exemple un auxiliaire de temps et un verbe à complétive) il peut y avoir une ou plusieurs têtes intermédiaires entre la tête locale et la tête finale.

Les relations de dominance vraies de tous les arbres élémentaires à ancre verbale sont représentées par la description partielle de la Figure 50, compte tenu d'un traitement des clitiques comme affixes.

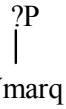


Figure 50. Description partielle subsumant tout schème à ancre verbale,

[9] Le fait que le verbe n'ait pas de projections verbales provient du fait que nous proposons de traiter les clitiques comme des affixes flectionnels, en morphologie. Dans le cadre d'un traitement des clitiques en syntaxe (ce que comprend l'implémentation actuelle), le verbe peut projeter un ou deux noeuds verbaux, selon qu'il y a ou pas un ou plusieurs clitiques compléments, et selon qu'il y a ou pas un clitique sujet. La représentation dans la MGf est donc plus complexe, elle utilise plusieurs constantes verbales en relations de dominance large, qui nomment le même noeud dans le cas où le schème ne comporte pas de clitiques.

dans la MGf et la MGi.

2.1.1 La représentation de l'artère verbale dans la MGf et la MGi

On vient de voir la description partielle représentant l'artère verbale dans la MGf et la MGi. Une classe a-une-ancre-verbale en dimension 1 introduit les constantes ?Vanc, ?Vmarq, et ?P (Figure 33). Elle est héritée par toute sous-catégorisation initiale pour une ancre verbale[10], pour la MGf.
La constante ?P a pour catégorie soit P, soit N, ce pour représenter le cas particulier des participiales (cf. supra), où l'on ne considère pas que le verbe projette une phrase.

A-UNE-ANCRE-VERBALE superclasses: nil			
traits et metatraits associés aux constantes :	description:		
?Vanc:pos = V ?Vmarq:pos = V ?P:pos = P/N			

Figure 51. La classe a-une-ancre-verbale

On a vu que dans les arbres au passif et ceux comprenant un complexe causatif, l'ancre n'est pas la tête locale de l'arbre élémentaire, c'est-à-dire dans la MGf et la MGi, que ?Vanc et ?Vmarq ne nomment pas le même noeud dans les arbres passifs et causatifs.
On utilise un champ "tête" spécifique, pour spécifier quelle est la constante à unifier avec ?Vmarq. Lors de la construction d'une classe croisée, après que toutes les redistributions ont eu lieu, ?Vmarq est unifiée (= est déclarée égale) avec la valeur du champ "tête", c'est-à-dire la constante représentant la tête locale.
Par défaut ce champ "tête" vaut ?Vanc, ce qui a pour effet d'unifier les constantes ?Vmarq et ?Vanc, et les traits et équations associés. Mais la redistribution pour le passif, ou le causatif, spécifie une autre valeur du champ "tête". Pour un passif, le champ "tête" prend comme valeur la constante ?Vpass, pour l'auxiliaire du passif. Pour une construction causative, le champ "tête" prend pour valeur la constante ?Vcaus, pour le verbe causatif. Ainsi, toutes les équations et traits faisant référence aux contraintes sur l'artère de l'arbre élémentaire, peuvent être spécifiées que l'ancre soit la tête locale ou pas.

On peut remarquer que l'on utilise un traitement procédural (avec l'utilisation d'une valeur par défaut) pour représenter la tête d'un arbre élémentaire, mais que cela est localisé en dimension 2, au niveau des redistributions (le passif et le causatif sont en effet traités par redistributions). Or la dimension 2 contient déjà des informations procédurales par définition, avec les redistributions qui révisent la valeur du métatrait :fonction.

[10] Nous avons choisi d'introduire toute l'artère jusqu'au noeud P en dimension 1. On pourrait imaginer que seules les projections verbales soient introduites, la présence de la projection phrastique dépendant de la réalisation du sujet. Mais en l'absence d'un noeud SV, on a besoin d'une projection englobant ?Vmarq et ?Vanc.

Ce traitement procédural de la tête locale permet de ne pas avoir à spécifier une morphologie particulière dans le cas de non passif ou non causatif. Sans l'utilisation d'un défaut, on aurait en effet à spécifier explicitement dans le cas non passif non causatif, que ?Vmarq et ?Vanc nomment le même noeud. Plus important, cela permet également de représenter la composition de redistributions (causatif o passif) ou (causatif o moyen), possible en italien (cf. Chapitre 4), en utilisant les mêmes classes que pour le passif simple ou le causatif simple. Le passif spécifie que la valeur du champ "tête" est ?Vpass, ce qui peut être éventuellement révisé par le causatif.

Disposant maintenant de la notion d'artère, nous pouvons détailler le mécanisme de gestion du mode dans les TAG sous-jacentes à la MGf et la MGi.

2.1.2 La gestion du mode

La construction syntaxique montrée par un arbre élémentaire impose le mode du verbe principal, c'est-à-dire de la tête finale. Le trait :mode est un trait de tête. La contrainte de mode sur la tête finale est imposée comme trait amont sur le noeud ?P de l'artère et le mode se propage par des équations de tête jusqu'à ?Vmarq. Toute adjonction d'un verbe enchâssant sur l'artère a pour effet de dévier cette propagation du mode vers l'ancre de l'arbre adjoint. Nous prenons comme exemple de référence, le cas d'un schème pour un ditransitif, au réfléchi datif, avec objet relativisé et sujet clitique[11]. Il peut être utilisé pour dériver par exemple la relative de (40).

(40) (la voiture) qu'il s'offrira

On donne ce schème n0Vn1an2-cl0-r1-réfl2, Figure 52, en ne donnant que le trait :mode.

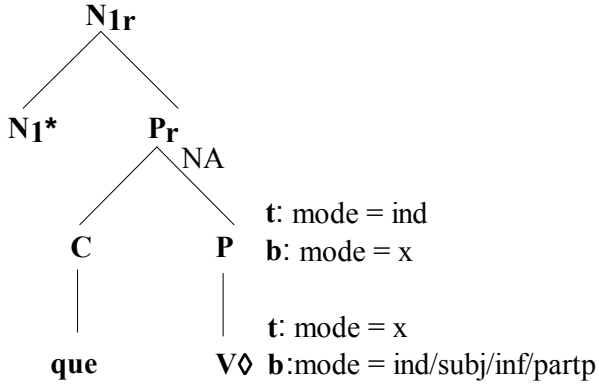


Figure 52. Le schème n0Vn1an2-cl0-r1-réfl2, avec le trait :mode.

La contrainte de mode imposée sur ?P dépend de la construction syntaxique décrite par le schème. Pour le français et l'italien, le mode est en partie imposé par le type de réalisation grammaticale du sujet. Dans la MGf et la MGi, les différentes contraintes de mode (fini pour une phrase simple, participe pour une participiale, infinitif pour une infinitive, impératif pour une impérative...) sont introduites dans les classes, en dimension 3, qui représentent les différentes réalisations grammaticales du sujet.

[11] Les clitiques sont représentés en morphologie, la contrainte sur leur existence est représentée par des traits omis ici (voir Chapitre 5, pour le traitement des clitiques pronominaux).

Dans le cas du schème de la Figure 52, comme il s'agit d'un cas avec sujet réalisé (bien que comme clitique), le mode indicatif est imposé au noeud ?P. La tête finale peut être :

1. l'**ancre** elle-même, s'il n'y a aucune adjonction enchâssante
   (la voiture qu'il s'offrira)
2. un **verbe à montée** (adjonction d'un verbe à montée)[12]
   (la voiture qu'[il-vient de] s'offrir)
3. un **verbe à complétive** (adjonction d'un verbe à complétive)
   (la voiture que [Marie parie qu'] il s'offrira)
4. un **auxiliaire de temps** (adjonction d'un auxiliaire de temps)
   (la voiture qu'[il-s'est] offerte)

En italien on a à considérer deux cas supplémentaires : l'adjonction d'un verbe restructurant, qui est similaire au cas de l'auxiliaire de temps pour ce qui est de la montée des clitiques (cf. infra) mais qui impose le mode infinitif au verbe sur lequel il s'adjoint; et également l'adjonction de STARE à valeur progressive (cf. supra), qui impose le mode gérondif sur lequel il s'adjoint. Pour ce qui est de la montée des clitiques, nous ne considérons que le cas où les clitiques montent, et donc où STARE se comporte comme un auxiliaire.

Un arbre auxiliaire enchâssant s'adjoignant sur l'artère peut lui-même avoir reçu une adjonction enchâssante, la tête finale est alors l'ancre de l'arbre auxiliaire le plus extérieur.

Le trait mode aval de ?Vmarq (concernant la tête locale) est instancié à la lexicalisation du schème, mais le schème prévoit une disjonction pour les quatre valeurs possibles (trait aval sur V b:mode = ind/inf/subj/partp, Figure 52). Ni le participe présent, ni l'impératif ne sont possibles dans ce contexte. Mais, le mode peut être l'indicatif (cas n°1 ci-dessus, s'il n'y a aucune adjonction enchâssante, où bien cas n°3 avec adjonction d'un verbe à complétive compatible avec l'indicatif), infinitif (cas n°2, avec adjonction d'un verbe à montée), subjonctif (cas n°3, avec un verbe à complétive compatible avec le subjonctif) ou participe (cas n°4). Pour l'italien, il faut ajouter la possibilité du mode gérondif (en dehors des schèmes pour l'impératif et le passif, cf. supra). Si la tête finale et la tête locale sont incompatibles (ici par exemple pour le mode), cela impose lors de la dérivation une adjonction rétablissant la compatibilité.
L'adjonction d'une complétive se fait au noeud ?P. L'adjonction d'un verbe à montée ou d'un auxiliaire (ou d'un verbe restructurant en italien) se fait au noeud nommé par ?Vmarq.
Nous donnons ci-dessous un exemple d'adjonction d'un verbe à complétive imposant le mode subjonctif. La Figure 53 donne l'arbre élémentaire pour un verbe à complétive (ancré par la forme *veulent*), avec le trait mode uniquement (comme d'habitude, les traits venant des lexiques sont en gras).

[12] Un verbe à contrôle est traité différemment d'un verbe à montée (cf. supra), et ne s'adjoint que sur des arbres où le sujet est non réalisé. Dans les exemples, on note entre crochets la séquence insérée par adjonction prédicative. La présence des clitiques est imposée par le schème de /OFFRIR/, mais peuvent être in fine portée par l'ancre d'un arbre adjoint (l'auxiliaire ou le verbe à montée dans les exemples).

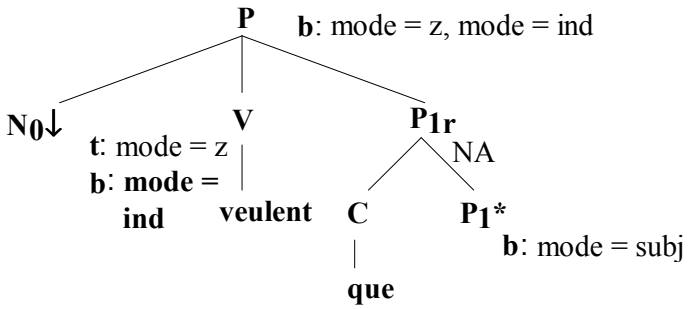


Figure 53. Arbre élémentaire n0Vs1subj[*veulent*] (trait mode uniquement)

Pour dériver *(la voiture) que les enfants veulent qu'il s'offre*, il faut utiliser l'adjonction illustrée Figure 54. n0Vn1an2-cl0-r1-réfl2 est ancré par *il s'offre*, forme ambiguë entre indicatif et subjonctif.

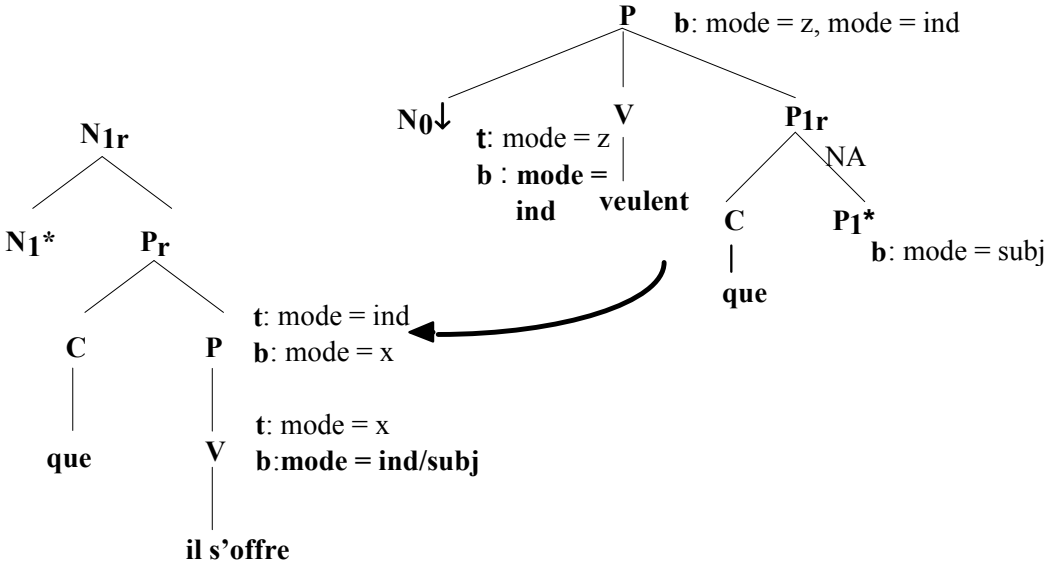


Figure 54. Adjonction de n0Vs1subj[*veulent*] sur n0Vn1an2-r1-réfl2[*il s'offre*] (*(N1) que N0 veulent qu'il s'offre.*)

Nous venons de préciser la représentation de l'artère verbale pour les schèmes à ancre verbale. Avant de détailler les différentes sous-catégorisations initiales retenues pour les verbes, nous commençons par étudier ce qui est connu sous le nom d'hypothèse inaccusative, afin de déterminer la sous-catégorisation initiale retenue pour les verbes intransitifs en français et en italien.

2.2 L'hypothèse inaccusative

L'hypothèse inaccusative a été initialement proposée dans le cadre de la Grammaire Relationnelle par Perlmutter (1983), et a été reprise par Burzio (1986) dans le cadre de GB. Cette hypothèse pose qu'il existe deux classes d'intransitifs, les inaccusatifs et les

inergatifs, *syntaxiquement* distinctes c'est-à-dire associées à des représentations syntaxiques différentes, au moins à un niveau profond. Levin et Rappaport (1995) (dorénavant LR95), s'appuyant sur leurs travaux antérieurs, distinguent en effet des propriétés d'inaccusativité de surface et d'inaccusativité profonde. Les premières ne sont possibles que lorsque l'argument d'un inaccusatif est objet en surface (pour nous objet final). Les secondes sont vérifiées par les inaccusatifs que leur argument se réalise en surface comme un objet ou un sujet. Pour LR95, une langue peut ne montrer que des propriétés d'inaccusativité profonde, c'est à dire que l'argument d'un inaccusatif supposé peut se réaliser toujours comme un sujet en surface.

Cette distinction syntaxique entre inaccusatifs et inergatifs est représentée de diverses manières selon les théories linguistiques. Pour la grammaire relationnelle, la distinction se situe au niveau de la relation grammaticale initiale associée à l'argument de l'intransitif. Les inaccusatifs ont un argument initial 2 (i.e. en gros un objet initial). Les inergatifs ont au contraire un argument initial 1 (i.e. un sujet initial, comme pour les verbes transitifs).
Pour GB, la distinction est exprimée en termes configurationnels en D-structure. Les inergatifs ont comme les transitifs un argument externe, les inaccusatifs n'en ont pas. Les inaccusatifs sont ainsi nommés car ils sont incapables d'assigner le cas accusatif à un argument interne qui dans le cas d'un verbe transitif reçoit le cas accusatif (Burzio, 1986).
Dans le cadre de LFG (étendu, comprenant une théorie de la correspondance sémantique lexicale/syntaxe, telle que décrite dans (Bresnan et Kanerva, 1989)), Bresnan et Zaenen (1990) utilisent des traits syntaxiques (tels que "fonction objective") pour marquer les arguments, ce qui leur permet de définir une notion équivalente à celle de l'argument externe en GB.
Pour HPSG, Dini (1995) représente l'argument des inergatifs comme valeur du trait SUJET, alors que l'argument des inaccusatifs apparaît comme premier élément de la liste COMPS.
Dans la MG, la distinction peut être représentée en termes fonctionnels. Nous disposons comme pour la grammaire relationnelle de plusieurs niveaux de description fonctionnelle (à la différence de LFG). On peut représenter les inaccusatifs comme sélectionnant un objet initial, et les inergatifs comme sélectionnant un sujet initial.

L'italien est bien connu pour bien se conformer à l'hypothèse inaccusative, alors que pour le français la question est nettement plus controversée. Dans le cadre de la construction de grammaires parallèles pour le français et l'italien, la question est donc d'importance.
Nous discutons de la validité de l'hypothèse inaccusative d'une manière générale avant de donner les critères habituellement retenus pour caractériser (syntaxiquement) les formes inaccusatives, pour l'italien et pour le français, et conclure que l'on retient l'hypothèse inaccusative pour l'italien mais pas pour le français.

2.2.1 L'inaccusativité, syntaxe ou sémantique?

Pour Perlmutter l'inaccusativité est déterminée sémantiquement et a un impact en syntaxe, c'est-à-dire qu'elle est représentée syntaxiquement (cf. LR95). Initialement pour

Perlmutter, l'hypothèse inaccusative a été introduite dans le contexte de la "Universal Alignment Hypothesis", qui faisait des prédictions très fortes sur la correspondance entre arguments et actants grammaticaux. Cette hypothèse suggérait en effet qu'il est toujours possible de déterminer l'expression syntaxique des arguments sur la base du sens du verbe. L'Agent devait être un initial 1, et le Patient/Thème devait être un initial 2. D'où l'hypothèse inaccusative pour les intransitifs, selon que leur argument est Agent ou Patient/Thème.
Depuis la question de l'interface sémantique lexicale/syntaxe est très étudiée (cf. par exemple les travaux de Grimshaw (1990), Jackendoff (1990), Levin et Rappaport (1995)), et si l'expression syntaxique des arguments d'un prédicat peut être largement prédite sur des bases sémantiques, ce n'est certainement pas avec un appariement aussi simple. Cette complexité a fait se dessiner deux approches pour l'inaccusativité, une approche syntaxique, contre la détermination par la sémantique du caractère inaccusatif ou inergatif d'un intransitif; et une approche sémantique, qui stipule que le caractère inaccusatif ou inergatif peut être complètement représenté en sémantique, et rend inutile sa représentation en syntaxe. LR95 font un rappel des arguments qui ont été avancés pour les deux approches. Elles-mêmes concluent en faveur de l'hypothèse de départ de Perlmutter, c'est-à-dire que l'inaccusativité est déterminée par la sémantique mais doit être représentée en syntaxe.
La difficulté pour statuer sur le caractère syntaxique ou sémantique de l'inaccusativité provient du fait que l'on peut définir pour *certaines* langues des critères syntaxiques d'inaccusativité, et on peut leur donner cette appellation commune seulement si on se réfère à des critères sémantiques d'inaccusativité, sensés être universels.
En faveur d'une définition purement syntaxique de l'inaccusativité, on peut trouver dans la littérature les arguments suivants :

1. Disparité entre critères sémantiques et syntaxiques au sein d'une langue

   Déjà en 1984, Rosen (1984) remarque que les critères syntaxiques retenus pour l'inaccusativité pour une langue donnée peuvent classer différemment des verbes appartenant à une même classe sémantique. Par exemple, It. RUSSARE (ronfler) est inergatif et It. ARROSSIRE (rougir) est inaccusatif. Mais Levin et Rapapport (1995) objectent qu'on ne sait pas a priori quels critères sémantiques sont pertinents pour le comportement syntaxique inaccusatif. Aussi, le fait que les deux verbes cités peuvent être classés ensemble, selon le critère sémantique "activité du corps" peut être totalement non pertinent pour l'inaccusativité. En revanche, la propriété aspectuelle de changement d'état associée à ARROSSIRE semble pertinente pour l'inaccusativité.

2. Disparité entre critères sémantiques universels et critères syntaxiques pour différentes langues

   Inversement, les critères syntaxiques retenus pour l'inaccusativité dans différentes langues ne donnent pas la même classification pour des verbes à sémantique comparable. Par exemple 'mourir' est inaccusatif en italien, et inergatif en Choctaw. Levin et Rapapport citent des travaux sur le Choctaw qui mettent en doute le critère syntaxique d'inaccusativité (l'accord) utilisé pour classer 'mourir' en Choctaw comme inergatif. Plus généralement, LR95 reconnaissent que l'on peut trouver des idiosyncrasies dans la classification syntaxique des intransitifs d'une langue, et que

plus la langue marque morpho-syntaxiquement la classification, plus on peut avoir d'idiosyncrasies (c'est-à-dire de verbes pour lesquels les critères syntaxiques et sémantiques divergent). Les auteurs s'appuient sur des considérations sur l'acquisition du langage. Si l'inaccusativité est clairement marquée morpho-syntaxiquement, alors le caractère inaccusatif d'un verbe peut être appris facilement. Au contraire si le marquage est moins clair, alors la classification inaccusatif/inergatif doit être régulièrement prévisible par la sémantique.

3. Une caractérisation sémantique des inaccusatifs manque toujours.

   Rosen a objecté qu'il n'existe pas une propriété sémantique commune aux supposés inaccusatifs. LR95 rétorquent qu'une caractérisation sémantique peut ne pas être unique. Autrement dit, il se peut que plusieurs propriétés sémantiques distinctes et pas forcément cumulatives déclenchent l'inaccusativité. Cependant Burzio (1986) montre qu'en italien les propriétés syntaxiques d'inaccusativité ne sont pas seulement valables pour des verbes intransitifs, mais également pour toutes les formes verbales passives et moyennes. Or il est peu probable que l'on parvienne à une caractérisation sémantique de tout cet ensemble. Donc les propriétés syntaxiques communes à toutes les formes dites inaccusatives en italien sont un argument fort en faveur d'une approche syntaxique de l'inaccusativité.

En faveur d'une détermination par la sémantique et d'une représentation en syntaxe, on peut arguer que d'une manière générale, une différence de comportement syntaxique généralisée a de grandes chances d'être déterminée par la sémantique. Pour l'italien par exemple, on a fini par trouver une caractérisation sémantique d'une bonne partie des verbes inaccusatifs.
Dini (1995), dans son traitement en HPSG de l'inaccusativité pour l'italien, considère que le comportement syntaxique inaccusatif en italien est déclenché par une configuration particulière de la sous-catégorisation (un trait SUJET vide, le premier élément de la liste COMPS prenant alors le cas nominatif). Cette configuration peut soit résulter d'un processus grammatical, comme le passif ou le moyen, soit être associée une sémantique particulière. Dini donne une généralisation sémantique permettant de prédire l'inaccusativité d'une partie des verbes inaccusatifs (en dehors du cas d'ergatif, c'est-à-dire d'intransitif alternant avec une variante transitive dont l'objet est le sujet de la variante intransitive). Les verbes dénotant un procès sont inergatifs, et les verbes à aspect télique (qui comprend une borne finale), s'ils satisfont une contrainte supplémentaire sont inaccusatifs. Cette contrainte supplémentaire est que, si on décompose la sémantique de ces formes verbales comme un acte causateur et un état résultant, il faut qu'il y ait coindexation entre l'agent de l'acte causateur et l'individu dont l'état résultant est prédiqué[13].
Cette caractérisation permet de coller aux critères syntaxiques, dans le cas en particulier de verbes "duals" pour l'inaccusativité (initialement fournis par Rosen), c'est-à-dire qui peuvent avoir ou pas un comportement syntaxique inaccusatif, suivant leurs emplois.

[13] C'est ce qui lui permet d'expliquer pourquoi *vendemmiare* (vendanger) n'est jamais inaccusatif. L'état résultant ne s'applique en aucune manière aux vendangeurs. Et également pourquoi les transitifs à objet omis, même télique n'ont jamais un comportement inaccusatif. C'est que l'état résultant est prédiqué de l'objet omis et pas de l'agent.

Certains verbes sont "duals" pour l'inaccusativité, c'est-à-dire qu'ils peuvent être, suivant leurs emplois, classés comme inergatifs ou inaccusatifs. Ainsi les verbes de mouvement ont dans le cas général un comportement inergatif (ils prennent l'auxiliaire AVERE, cf. infra 2.2.2), mais à y regarder de plus près, si leur aspect est modifié et devient télique, ce qui est possible par exemple avec un modifieur de but, alors ils prennent un comportement inaccusatif (ils prennent l'auxiliaire ESSERE).

(41) a. Giovanni ha/*è corso per tre ore.
Giovanni a/*est couru pendant trois heures.

b. Giovanni *ha/è corso a casa in due ore.
Giovanni *a/est couru à maison en deux heures
(Giovanni a couru à la maison en deux heures.)

Dini classifie également les verbes statiques parmi les inaccusatifs. Il étudie le cas de VIVERE (vivre), qui est également dual pour l'inaccusativité ((42) montre que les deux auxiliaires de temps sont possibles, et (43) montre que le *ne* partitif n'est possible qu'avec ESSERE).

(42) Leo ha/è vissuto male.
Leo a/est vécu mal.
(Leo a mal vécu.)

(43) a. A Parigi di poeti ne sono vissuti molti.
A Paris des poètes NE-part sont vécus beaucoup-mp.)
(A Paris des poètes il en a vécu beaucoup.)

b. *A Parigi di poeti ne hanno vissuto molti.
A Paris des poètes en ont vécus beaucoup-mp.

Dini montre que le verbe est en fait ambigu entre un procès et un état, ce qui est mis en évidence par la possibilité d'utiliser le progressif, compatible avec un procès uniquement (44).

(44) Stiamo vivendo proprio male.
VIVRE-progressif.1p vraiment mal
(Nous vivons vraiment mal en ce moment.)

La caractérisation sémantique prédit que le procès soit réalisé comme un inergatif et l'état comme un inaccusatif, et donc que le *ne* partitif ne soit pas possible avec le progressif. C'est effectivement le cas (45).

(45) *A Parigi di poeti ne stanno vivendo molti.
A Paris des poètes en VIVRE-progressif.3pl beaucoup-mp.

On voit qu'une analyse fine du comportement des verbes et de leur sémantique permet de bien corréler critères sémantiques d'inaccusativité et critères syntaxiques pour une langue donnée. Cependant la classification des intransitifs/inaccusatifs se fait sur des critères syntaxiques (pour l'italien en tous cas), et on essaie ensuite de trouver une explication sémantique.
Notre position est la suivante. Il existe pour certaines langues, dont l'italien, un comportement syntaxique inaccusatif, c'est-à-dire que certains intransitifs ont un argument qui peut se réaliser avec des propriétés d'objet. Ce comportement est associé à une configuration de la liste de sous-catégorisation de la forme verbale. Nous donnons Section 2.2.2 une description du comportement inaccusatif en italien et son traitement

dans la MGi. Nous suivons Dini en considérant que des facteurs sémantiques peuvent imposer pour un verbe une telle configuration de sous-catégorisation, mais qu'elle peut également être liée à un phénomène grammatical tel que le passif. C'est donc une approche mixte de l'inaccusativité que nous prenons. L'inaccusativité est représentée en syntaxe, mais elle ne peut pas être complètement dissociée d'une "sémantique inaccusative", faite de différents critères sémantiques déclenchants, tels que l'aspect télique par exemple. On n'a cependant as la connaissance complète aujourd'hui de cette "sémantique inaccusative".
Pour d'autres langues, l'existence d'un comportement syntaxique inaccusatif n'est pas établie, en tous cas pas d'une manière qui met en évidence une partition de l'ensemble des formes intransitives, en accord avec la "sémantique inaccusative". C'est ce que nous allons voir pour le français, Section 2.2.3.

2.2.2 Caractérisation des formes inaccusatives en italien

On désigne par formes inaccusatives des formes ayant un comportement syntaxique inaccusatif. On appelle inergatifs les verbes intransitifs non inaccusatifs.
Il existe en italien certains intransitifs dont l'argument peut donner lieu à un *ne* partitif, ce que l'on considère typique de l'objet. On appelle forme inaccusative une forme intransitive montrant cette propriété. Les formes inaccusatives sont en italien :

(i) Les verbes dits inaccusatifs, c'est-à-dire des verbes dont le comportement inaccusatif peut être détecté indépendamment de leurs inflections (ils peut s'agir de verbes ergatifs, comme CASSER en français c'est-à-dire qui ont un correspondant transitif de même forme, dont l'objet correspond au sujet de l'ergatif)

(ii) Les formes passives

(iii) Les formes à réfléchi intrinsèque et les formes moyennes[14]

On a en effet[15] :

(46) a. Ne sono partiti molti. — i
NE-part sont partis beaucoup.
(Il en est parti beaucoup.)

b. Ne sono stati mandati tanti al fronte. — ii
NE-part sont ESSERE.partp.mp envoyés tant au front.
(Il en a été envoyé tant au front.)

c. Se ne sono suicidati molti. — iii réfléchi intrins
RF.3 NE-part sont suicidés beaucoup.mp.
(ll s'en est suicidé beaucoup.)

d. Se ne sono affitate molte. — iii moyen
RF(moy) NE-part sont louées beaucoup.fp.
(Il s'en est loué beaucoup.)

[14] On distingue parmi les formes verbales portant un réfléchi, les réfléchis intrinsèques (du type *s'évanouir*), les formes moyennes (du type *ce livre se vend bien*) et les "vrais réfléchis" c'est-à-dire les formes interprétées avec un sens réfléchi ou réciproque.
[15] Pour faciliter la lecture, on note à droite le type (i,ii,iii) de forme inaccusative qu'illustre l'exemple.

Cette propriété est typique de l'objet (46e), elle n'est pas possible pour un complément prépositionnel (46f), ni pour le sujet (46g).

(46) e. Ne ho visti molti.
NE-part j'ai vus beaucoup.
(J'en ai vu beaucoup.)

f. *Ne ho parlato a molti.
NE-part j'ai parlé à beaucoup.

g. *Molti$_i$ ne$_i$ hanno mangiato una torta.
Beaucoup$_i$ NE-part$_i$ ont mangé une tarte.

Une forme inaccusative s'accorde en genre et nombre avec la réalisation de son argument, même si cet argument est objet, comme en (46). L'argument des formes inaccusatives peut également se réaliser comme un sujet (en position préverbale par exemple), et dans ce cas le *ne* partitif n'est pas possible, comme il n'est jamais possible avec un sujet. On donne simplement l'exemple d'un verbe inaccusatif et d'un inergatif.

(47) a. *Molti ne hanno telefonato.
Beaucoup.mp NE-part ont téléphoné

b. *Molti ne sono partiti.
Beaucoup.mp NE-part sont partis

On peut également remarquer que le *ne* partitif n'est pas seulement déclenché par l'ordre des mots, puisque le quantifieur peut être en position extraite (Perlmutter, 1983)).

(48) a. Quante ne sono rimaste?
Combien-fp NE-part sont restées
(Combien en est-il resté?)

Les formes inaccusatives sont également associés à d'autres propriétés (Salvi, 1988:49s; Dini, 1995), mais nous considérons qu'elles ne sont pas des propriétés définitoires de la classe syntaxique des inaccusatifs, en ce sens qu'elles ne montrent pas une réalisation d'objet de l'argument de la forme inaccusative. Ces propriétés sont :

- l'auxiliaire de temps est ESSERE

(49) a. Maria è partita. — i
Maria est partie.

b. Maria è stata mandata al fronte. — ii
Maria est ESSERE.partp.fs envoyée au front.

c. Maria si è suicidata. — iii réfléchi intrins.
Maria s'est suicidée.

d. Le case si sono affittate bene. — iii moyen
Les maisons se sont bien louées.

Cette propriété n'est pas caractéristique, les formes comprenant un vrai réfléchi se conjuguant également avec ESSERE. Or si le classement des formes avec réfléchi accusatif comme inergatif est discutable, les formes avec réfléchi datif, qui peuvent être transitives, ne peuvent absolument pas être classées comme inaccusatives.

(49) e. Maria si è accusata di frode. — réfléchi acc
Maria s'est accusée de fraude.

f. Maria si è comprata l'auto. réfléchi datif
Maria s'est achetée l'auto.

De plus le sujet *si* impersonnel (que nous appelons SI-nominatif), qui peut être le sujet de tout type de verbe, impose également l'auxiliaire ESSERE (voir Chapitre 4).

(49) g. Si è lavorato bene.
SI-nominatif est travaillé bien
(On a bien travaillé).

- Le participe aux temps composés s'accorde avec le sujet (cf. la série (49))
- La forme verbale peut être la tête d'une participiale passée[16] adnominale.

(50) a. I soldati partiti per la guerra... i
Les soldats partis pour la guerre...

b. I soldati mandati al fronte (E + dal colonello)... ii
Les soldats envoyés au front (E + par le colonnel)...

c. Uno studente pentitosi delle sue colpe... iii réfléchi intrins.
Un étudiant REPENTIR.partp.ms-RF.3p de ses fautes...

d. *Le persone licenziatasi ... (Salvi, 1988:103) iii moyen
Le personnes licensiées-RF(moy)

Cette construction n'est pas possible pour les inergatifs, ni pour les formes moyennes (50d).
Pour le réfléchi, à la différence du *ne* partitif, la construction participiale est acceptée pour une forme avec réfléchi accusatif (alternant avec objet) (Salvi, 1988:52).

(50) e. *I soldati camminati furono fucilati. inergatif
Les soldats marchés furent fusillés

f. Un individuo accusatosi di frode... (Salvi, 1988:52) vrai réfléchi acc.
Un individu accusé-RF.3p de fraude...
(un individu s'étant accusé de frause)

g. *un giovanotto appena compratosi l'auto... (Salvi, 1988:52) vrai réfléchi datif
un jeunot à peine acheté-RF.3p l'auto...
(un jeunot s'étant à peine acheté l'auto)

Une autre propriété est parfois décrite pour les inaccusatifs : le sujet de la forme verbale inaccusative peut être exprimé dans une participiale passée adverbiale (dite alors absolue).

(51) a. Partita Maria, la mia vita cambiò. i
Partie Maria, ma vie changea.

b. Bevuto un bicchiere di vino, Maria partì.
Bu un verre de vin, Maria partit.
(Après avoir bu un verre de vin, Maria partit.)

c. Pentitasi anche Maria, ... iii refléchi intrins.

[16] On garde la terminologie usuelle pour les participes présent ou passé, pour le français et l'italien. On parlera de participiale passée ou présente selon que le verbe est au participe passé ou présent, bien que la distinction n'ait que peu à voir avec le temps.

Repentie-SIréfl.intrins. aussi Maria, ...
(Marie s'étant aussi repentie, ...)

d. *Licenziatasi Maria, ... (Salvi, 1988:103) iii moyen
Licenciée-RF(moy) Maria, ...

e. Accusatosi di frode Giacomo, ... vrai réfléchi acc
Accusé-RF.3p Giacomo, ...

f. *Appena compratosi l'auto il giovanotto,... vrai réfléchi datif
A peine acheté-RF.3p l'auto le jeunot, ...

g. *Telefonato Maria, Giorgio se ne andò. inergatif
Téléphoné Maria, Giorgio s'en alla.

En fait cette construction n'est pas possible pour une forme passive, pourtant inaccusative. En effet (51b) (avec un transitif) ne peut pas être interprété comme un passif, ce que montrent différents tests (Bertucelli, 1991:594) : (i) la pronominalisation se fait par un clitique accusatif (52a); (ii) le complément d'agent est (en général) exclu (52b); (iii) certaines expressions figées non passivables peuvent apparaître comme participiales adverbiales (53c).

(52) a. Maltrattatala, Carla partì.
Maltraitée-ACC.fp, Carla partit.
(L'ayant maltraitée, Carla partit.)

b. *Conosciuta Sandra da tutti, Gianni l'ha lasciata.
Connue Sandra par tous, Gianni l'a quittée.

(53) a. Gianni ha fatto mente locale sulla questione.

b. *Mente locale sulla questione è stata fatta da Gianni.

c. Fatta mente locale sulla questione, Gianni decise di partire.

Donc on ne peut pas simplement dire que le sujet (de surface) d'une forme inaccusative peut être exprimé dans une participiale absolue. Cela n'est en effet pas possible pour une forme passive.

Les cas (51b), (52a) et (53c) sont à classer comme des cas de participiale adverbiale sans sujet. En effet, dans ces cas transitif, le sujet ne peut pas être exprimé (Dini, 1995) (54). Ces cas ont la particularité de déclencher l'accord de l'objet avec le participe.

(54) *Bevuto un bicchiere di vino Mario, Maria partì.
Bu un verre de vin Mario, Maria partit.

Il semble donc que l'on doive distinguer deux types de participiales adverbiales :

1. Participiale adverbiale sans sujet
Une participiale adverbiale, qui ne comprend pas le sujet, et qui est régulièrement formée pour toutes les classes, illustrée par (51b), (52a), (53c) et la série (55). A noter en effet, que cette construction est possible pour certains inergatifs, sous certaines conditions sémantiques (Dini, 1994), d'où le contraste (55e)/ (55f).

(55) a. Partita comprare il pane, Maria incontrò Giorgio. i
Partie acheter du pain, Maria rencontra Giorgio.

b. Maltrattata da tutti, Carla partì. ii
Maltraitée par tous, Carla partit.

c. Pentitasi più volte, Paola continuò però a rubare quadri. iii réfl. intrins.
Repentie-RF.3p plusieurs fois, Paola continua pourtant à voler des tableaux.

d. Analizzatosi attentamente, Gianni riconobbe aver sbagliato.
réfléchi acc
Analysé-RF.3p attentivement, Gianni reconnut s'être trompé.

e. * Lavorato come un pazzo, Gianni andò a dormire. inergatif
Travaillé comme un fou, Gianni alla dormir.

f. Vendemmiato in fretta e furia, i contadini lasciarono la campagna. inergatif
Vendangé à la hâte, les paysans quittèrent la campagne. (Dini, 1995)

Le participe s'accorde avec le contrôleur si la forme verbale ne montre pas d'objet, mais si un objet est présent, il y a accord avec l'objet.

2. Participiale adverbiale absolue (avec sujet)
Et une participiale ajout de phrase, illustrée par (51a,c,e), où le sujet d'un inaccusatif peut être exprimé (mais pas pour un passif), et également celui d'un réfléchi. On ne retient donc pas cette construction comme diagnostic d'inaccusativité.

2.2.2.1 Statut des formes à vrai réfléchi

Comme les formes comportant un vrai réfléchi ont certaines propriétés associées aux formes inaccusatives, leur statut demande à être précisé. Dini (1995) englobe toutes les formes réfléchies dans la classe des inaccusatifs, mais sans en donner la justification. Nous avons considéré décisif pour le caractère inaccusatif le *ne* partitif. Or les avis divergent sur la possibilité du *ne* dans le cas de vrais réfléchis. Salvi (1988:50) exclut cette possibilité. Perlmutter (1983) donne un argument précis pour ne pas considérer les réfléchis comme des inaccusatifs. (56a) est ambigu entre un sens réfléchi (cas 1) et un sens moyen (cas 2). Or d'après Perlmutter, (56b) ne peut pas être interprété comme un réfléchi, le *ne* partitif force l'interprétation moyenne.

(56) a. Si sono denunciate molte persone.
RF.3p/RF(moy) sont dénoncées beaucoup de personnes.
(Beaucoup de personnes se sont dénoncées/Il a été dénoncé beaucoup de personnes.)

b. Se ne sono denunciate molte.
RF.3p NE-part dénoncées beaucoup.
(Il en a été dénoncé beaucoup)

Plus précisément, Perlmutter (1983) indique que certains locuteurs consultés ont des doutes, mais que ceux qui ne doutent pas rejettent tous l'interprétation réfléchie pour (56b). Mais parmi les locuteurs que nous avons interrogés (10 personnes), certains (3 personnes) acceptent l'interprétation réfléchie également.
On peut faire un parallèle avec le français, avec l'ambiguité de (57), entre une interprétation réfléchie (syntaxiquement impersonnel) et moyenne (syntaxiquement impersonnel)[17].

(57) Il s'est dénoncé beaucoup de personnes.

[17] Personnellement je préfère l'interprétation réfléchie, mais je fais partie des locuteurs A de (Abeillé et al., 98).

Les réfléchis constituent un cas hybride, du fait qu'au niveau sémantique, il y a fusion de deux arguments, chacun ayant une FG initiale distincte (cf. également Alsina, 1996, pour un résumé du statut hybride des vrais réfléchis pour le catalan). Pour un verbe à vrai réfléchi alternant avec un objet, le sujet devrait être considéré comme à la fois sujet initial et objet initial (c'est également l'analyse initiale de la grammaire relationnelle, cf. Rosen, 1982). L'hésitation sur l'interprétation de (57) pourrait s'expliquer par une priorité différente donnée par les locuteurs au statut de sujet initial ou d'objet initial au sujet d'une construction à vrai réfléchi.
Nous n'avons pas poursuivi plus avant cette voie. En simplifiant, nous considérons que les verbes à vrai réfléchi ne constituent pas des formes inaccusatives (i.e. que leur sujet est un sujet initial).

2.2.2.2 Représentation dans la MG

Le comportement inaccusatif est représenté dans la MGi par un type de sous-catégorisation. Une sous-catégorisation inaccusative comprend un objet mais pas de sujet. C'est le parallèle de la définition en Grammaire Relationnelle :

> "A stratum is UNACCUSATIVE if and only if it contains a 2-arc and no 1-arc (Perlmutter, 1983:151)."

Pour les verbes inaccusatifs, cette sous-catégorisation est la sous-catégorisation initiale. Pour les passifs et moyens, cette sous-catégorisation résulte de la démotion du sujet. Mais à la différence de la Grammaire Relationnelle, *nous considérons qu'une sous-catégorisation inaccusative peut être finale, bien qu'elle soit systématiquement en concurrence avec une sous-catégorisation dérivée par la promotion de l'objet en position sujet*. Par analogie avec la terminologie française, on qualifie d'impersonnelle une construction dont la sous-catégorisation finale n'a pas de sujet (donc en particulier les sous-catégorisations finales inaccusatives, qui ont un objet mais pas de sujet) (voir Section 2.2.3 pour une comparaison avec le français).
On donne les sous-catégorisations initiale et finale que nous retenons pour les différents types de formes verbales[18] :

- **actif impersonnel** = inaccusatif avec objet non promu (montré par le *ne* partitif)

(58) Ne sono arrivati tre.
NE-part sont arrivés trois

	arg0
sous-cat initiale et finale	objet

- **actif personnel** = inaccusatif avec objet promu sujet

(59) Giovanni è arrivato.
Giovanni est arrivé

[18] La grammaire relationnelle considère que l'objet est systématiquement promu sujet, et éventuellement démis vers la relation "chômeur", pour ce qui est traditionnellement appelé "inversion du sujet". L'analyse comme démotion de l'inversion du sujet est motivée par des contrastes sur la capacité d'un sujet inversé en italien à contrôler le sujet d'une infinitive. Nous considérons que ces contrastes peuvent être expliqués par des contraintes discursives. Et nous donnons une analyse beaucoup plus simple des inaccusatifs : une sous-catégorisation inaccusative peut être finale. Ainsi *Ne sono partiti tre* est analysé sans aucune redistribution, l'objet initial de *partire* est l'objet final.

	arg0
sous-cat initiale	objet
sous-cat finale	sujet

- transitif

(60) Paolo ha rubato tre torte.
(Paolo a volé trois tartes)

	arg0	arg1
sous-cat initiale et finale	sujet	objet

- **passif impersonnel** = passif avec objet non promu (montré par le *ne* partitif)

(61) Ne sono state rubate tre dal gatto.
Ne-part ESSERE-passé.composé3fpl volées trois par le chat.
(Il en a été volé trois par le chat)

		arg0	arg1
	sous-cat initiale	sujet	objet
(passif-impers)⇒	sous-cat finale	agt-objet	objet

- **passif personnel** = passif avec objet promu sujet

(62) Tre torte sono state rubate dal gatto.
Trois tartes ESSERE-passé.composé3fpl volées par le chat.
(Trois tartes ont été volées par le chat)

		arg0	arg1
	sous-cat initiale	sujet	objet
(passif-pers)⇒	sous-cat finale	agt-objet	sujet

- **moyen impersonnel** = moyen avec avec objet non promu (montré par le *ne* partitif)

(63) Se ne sono affitate molte ad un prezzo eccessivo.
Réfl.3p Ne-part sont louées beaucoup.fpl à un prix excessif.
(Il s'en est loué beaucoup à un prix excessif.)

		arg0	arg1
	sous-cat initiale	sujet	objet
(moyen-impers)⇒	sous-cat finale	∅	objet

- **moyen personnel** = moyen avec objet promu sujet

(64) Molte case si sono affitate ad un prezzo eccessivo.
Beaucoup.fpl maisons réfl.3p sont louées à un prix excessif.
(Beaucoup de maisons se sont louées à un prix excessif.)

		arg0	arg1
	sous-cat initiale	sujet	objet
(moyen-pers)⇒	sous-cat finale	∅	sujet

Voyons maintenant comment sont déclenchées les propriétés des inaccusatifs :

- Accord du verbe

Il nous faut poser que le verbe prend les traits de personne et nombre soit avec le sujet s'il existe, sinon avec l'objet. On rappelle que dans le cas d'une forme non finie, le sujet est contrôlé par un actant d'un verbe enchâssant, mais la forme non finie comprend bien un sujet dans sa sous-catégorisation, et le verbe s'accorde avec le sujet et pas avec un éventuel objet (et dans la MG le Non réalisé des formes non finies est représenté en dimension 3, comme une réalisation particulière du sujet).

- Déclenchement du *ne* partitif

Le *ne* partitif est représenté comme une réalisation d'objet, en dimension 3. Cette réalisation est possible dès lors qu'un objet nominal apparaît dans la sous-catégorisation. Mais on a vu que l'objet d'une forme inaccusative ne peut pas se réaliser comme un clitique accusatif. On doit donc ajouter la contrainte sur la réalisation comme clitique accusatif : elle n'est possible que si la sous-catégorisation comporte un sujet.

clitique accusatif $\Rightarrow \exists$ sujet

Pour les autres réalisations d'objet dans le cas d'une forme inaccusative), il y a systématiquement ambiguité avec une réalisation de sujet (cf. nous avons dit que l'objet d'une forme inaccusative peut systématiquement être promu sujet). En effet, à la différence du français, le pronom relatif accusatif/nominatif prend la même forme *che* (donc on ne peut pas trancher en (65a) entre une analyse où l'actant extrait est sujet ou objet; dans le cas où il est objet, l'équivalent français comporte un sujet syntaxique explétif). En (65b) également, il y a ambiguité syntaxique entre sujet interrogé ou objet interrogé. Et en italien, à la différence du français, il peut y avoir dépendance non bornée entre le sujet et son verbe, donc (65c) ne permet pas de désambiguiser[19].

(65) a. Le persone che sono rimaste...
Les personnes pro.rel sont restées
(Les personnes qui sont restées / Les personnes qu'il est resté)

b. Quante persone sono rimaste?
Combien-fpl personnes sont restées?
(Combien de personnes sont restées / Combien de personnes est-il est resté?)

c. Quante persone hai detto che sono rimaste?
Combien-fpl as-tu dit complémenteur sont restées?
(*Combien de personnes as-tu dit qui sont restées / Combien de personnes as-tu dit qu'il est resté?)

- Sélection de l'auxiliaire

La sélection de l'auxiliaire est déterminée par une propriété de la sous-catégorisation, ou bien par la présence d'un clitique *si*. Si le verbe porte un *si* (SI-nominatif ou réfléchi), ou si la sous-catégorisation ne comporte pas de sujet ou bien comporte un Non initial alors l'auxiliaire est ESSERE. L'auxiliaire de temps est traité comme une propriété révisable dans la MGi (et la MGf).

[19] Pour l'implémentation, on a bloqué l'analyse où l'actant est objet, pour éviter une multiplication des analyses.

Les verbes duals (tels que les verbes de mouvement, cf. supra) doivent comprendre deux entrées avec deux sous-catégorisations initiales différentes (mais on peut imaginer que le module de correspondance sémantique lexicale / syntaxe crée ces deux entrées à partir d'une spécification commune, comme cela est fait par Dini (1995)).

- Participiale passée adnominale

La construction est à traiter en dimension 3, comme un type particulier de réalisation du sujet, qui impose le mode participe, et fait de la structure ancré par le verbe un modifieur de nom. Elle est bloquée pour le moyen et les inergatifs).

- Participiale adverbiale sans sujet

On a vu que cette construction est régulièrement formée pour tout type de forme verbale. La construction est à traiter en dimension 3, comme un type particulier de "consommation" d'un élément de la sous-catégorisation (le sujet) : c'est un cas où l'actant n'est pas réalisé phonologiquement (comme pour une infinitive), qui impose le mode participe. Le verbe s'accorde avec l'actant qui contrôle le sujet, sauf si un objet est présent, auquel cas il y a accord avec l'objet.
On explique l'impossibilité pour une bonne partie des inergatifs (mais une partie seulement, cf. le contraste (55e)/ (55f)) par des facteurs sémantiques (Dini, 1994).

2.2.3 Le français n'a pas de propriétés d'inaccusativité de surface

Les critères généralement avancés pour l'inaccusativité en français sont (Tasmowski, 1984) :

- L'auxiliaire de temps est ETRE
- Le sujet peut donner lieu à un *en* génitif
- La forme verbale peut être la tête d'une participiale passée adnominale

Les deux premiers critères ne sont pas du tout caractéristiques. L'auxiliaire ETRE est également présent pour toutes les formes réfléchies. Le *en* génitif venant du sujet est possible avec certains verbes transitifs, et impossible avec des verbes ergatifs (c'est-à-dire variantes intransitives de transitifs, dont le sujet est l'objet de la variante transitive), qui sont sensés être inaccusatifs.

(66) a. La fin en est connue.
b. *La carcasse en cuira en trois heures dans le jus. (Tasmowski, 1984).

Pour la possibilité d'avoir une participiale passée adnominale, il y a effectivement une bonne corrélation avec l'auxiliaire ETRE, encore que l'on puisse trouver des cas marginaux de verbes normalement conjugués avec avoir.

(67) a. Cette eau a/*est jaillie des entrailles de la terre.
b. Cette eau jaillie des entrailles de la terre...

Mais nous ne considérons pas que le nom modifié est objet mais sujet (donc pour nous ces données ne montrent pas un comportement d'objet de l'argument de la forme supposée inaccusative). Nous tirons cette conclusion d'une comparaison avec les participiales présentes. En effet une participiale présente peut modifier le sujet de la forme verbale, de manière très régulière pour tout type de forme verbale (cf. la série

(68), a. inergatif supposé, b. inaccusatif supposé, c. réfléchi intrinsèque, d. vrai réfléchi, e. moyen, f. passif).

(68) a. Les personnes téléphonant ...
b. Les personnes arrivant ...
c. Les personnes se trouvant dans l'impossibilité de ...
d. Les personnes se lavant toujours à l'eau froide sont plus résistantes.
e. Les vêtements se lavant à l'eau froide se vendent moins bien.
f. Les biens étant acquis par les époux après la date du mariage seront considérés comme communs.

La participiale présente peut contenir un temps composé ((69e) est exclue pour une incompatibilité d'aspect avec le moyen).

(69) a. Les personnes ayant téléphoné ...
b. Les personnes étant arrivées ...
c. Les personnes s'étant trouvées dans l'impossibilité de ...
d. Les personnes s'étant toujours lavées à l'eau froide sont plus résistantes.
e. *Les vêtements s'étant lavés à l'eau froide sont moins abîmés.
f. Les biens ayant été acquis par les époux avant la date du mariage sont considérés comme communs.

Dans ce contexte, on considère la série (70) comme une variante de la série (69), où l'auxiliaire de temps ETRE, s'il ne porte pas de réfléchi, ou bien la copule auxiliaire du passif, est omis. Ce qui fait que (70f) peut s'analyser comme la variante de (69f) où la copule est elle-même composée, mais aussi comme la variante de (68f).

(70) a. *Les personnes téléphoné ...
b. Les personnes arrivées ...
c. *Les personnes trouvées lors de la manifestation
d. *Les personnes lavées à l'eau froide sont plus résistantes. (pas avec un sens réfléchi)
e. *Les vêtements lavés à l'eau froide sont moins abîmés. (pas avec un sens moyen)
f. Les biens acquis par les époux après/avant la date du mariage sont considérés comme communs.

On peut remarquer que pour un verbe à réfléchi intrinsèque, on peut trouver un emploi adnominal du participe, mais l'interprétation statique (où le participe se comporte comme un adjectif) est forcée. Ainsi (71b) est impossible (l'interprétation statique est en conflit avec le reste de la phrase) alors que (71c) est possible.

(71) a. Les personnes s'étant évanouies lors de la manifestation ont porté plainte.
b. *Les personnes évanouies lors de la manifestation ont porté plainte.
c. Les personnes évanouies sont capables d'entendre ce qu'on leur dit.

Donc, on n'a apparemment aucun critère d'inaccusativité de surface pour le français, c'est-à-dire qu'à la différence de l'italien, on n'a pas pour l'argument des inaccusatifs supposés une propriété typique d'objet. En italien on a le *ne* partitif, mais l'équivalent français, le *en* partitif est bloqué pour les actants de tous les intransitifs.

Plus précisément, on peut trouver ce *en* partitif dans le cas impersonnel, où justement c'est ce qui amène à considérer que le syntagme extraposé des constructions impersonnelles est objet.
On fait un parallèle entre les traductions italien/français suivantes :

(72) a. Tre persone sono arrivate.
a'. Trois personnes sont arrivées.

b. Sono arrivate tre persone.
b'. Il est arrivé trois personnes.

c. Ne sono arrivate tre.
c'. Il en est arrivé trois.

L'équivalent du cas italien de sous-catégorisation inaccusative (i.e. avec un objet mais sans sujet) serait en français une sous-catégorisation avec un sujet impersonnel et un objet. Il s'agirait d'un équivalent syntaxique et discursif, ayant pour effet de déplacer vers la fin de la phrase l'actant d'une forme intransitive. On pourrait donc considérer que le français montre également une propriété d'inaccusativité de surface, c'est-à-dire que l'argument unique de certains intransitifs peut être objet en surface mais comme cela implique une absence de sujet, le *il* impersonnel est inséré (cf. Chapitre 4). La différence entre le français et l'italien se ramènerait à la possibilité pour un verbe en italien de ne pas avoir de sujet syntaxique[20]. Si l'on suit l'hypothèse inaccusative pour le français, ARRIVER sélectionne un objet initial, et s'il est objet final, un sujet syntaxique explétif doit être inséré.
Malheureusement, cette hypothèse pour le français est nettement affaiblie par le fait que l'on peut en français avoir un impersonnel pour des verbes dont on ne veut pas dire qu'ils sont inaccusatifs, soit qu'il s'agisse d'un transitif employé intransitivement, soit qu'il s'agisse d'un intransitif dont la sémantique amènerait à un comportement inergatif en italien.

(73) a. Il mangeait une dizaine de personnes dans le réfectoire quand l'incendie a éclaté.
b. Il nageait quelques personnes dans le bassin quand le plafond s'est écroulé.

Avec de surcroît, la possibilité d'un *en* partitif pour l'actant référentiel (même si cela est plus difficile).

(74) a. ? Il en mangeait une dizaine dans le réfectoire quand l'incendie a éclaté.
b. Il en nageait quelques-unes dans le bassin quand le plafond s'est écroulé.

L'équivalent italien de (73), est analysé comme une inversion du sujet (et pas comme dans le cas inaccusatif, une réalisation comme objet de l'unique actant) car on n'a pas le *ne* partitif dans ce cas inergatif. En français on peut avoir le *en* partitif pour un supposé inergatif (74) comme pour un supposé inaccusatif (72c), les contraintes sur l'impersonnel étant à expliquer en sémantique.

(75) a. Mangiavano una decina di persone nella mensa quando scopiò l'incendio.

[20] La possibilité de ne pas avoir de sujet en italien est tout de même réduite. Elle est possible pour les verbes sans arguments sémantiques (les verbes atmosphériques). Pour les verbes avec un argument, cet argument a une nette tendance à se réaliser comme sujet. Il peut se réaliser comme objet (et prendre en particulier le *ne* partitif) mais pas comme objet clitique défini. D'où le nom des inaccusatifs, qui serait des verbes incapables d'assigner le cas accusatif (Burzio, 86).

b. Nuotavano una decina di persone nella piscina quando crollò il soffitto.

Enfin, l'impossibilité d'avoir un passif impersonnel pour un intransitif est parfois retenue comme critère d'inaccusativité. Mais, Rivière (1981) montre qu'il y a désaccord entre le critère de l'auxiliaire de temps et le passif impersonnel. Le passif impersonnel serait plus généralement impossible sans un argument agentif. De là, une nuance d'agentivité peut suffire à le rendre possible (Tasmowski, 1984).

(76) a. *Il a été tombé dans les mêmes erreurs.
b. Il a été retombé dans les mêmes erreurs.

On renonce donc à l'hypothèse inaccusative pour le français. On perd le parallèle avec l'italien, mais seulement pour ce qui est des intransitifs. En effet, pour les formes passives (respectivement moyennes) nous obtenons pour les deux langues la même sous-catégorisation finale, mise à part la présence du *il* explétif en français.
Une différence importante entre le français et l'italien reste donc qu'en italien un verbe peut sous conditions (cf. supra) apparaître sans sujet. On donne les séquences de sous-catégorisations obtenues pour le passif et le moyen (ce qui sera repris Chapitre 4 sur la dimension 2).

- transitif

Paul a volé trois tartes.

	arg0	arg1
sous-cat initiale et finale	sujet	objet

- passif impersonnel (i.e. passif avec objet non promu)

Il en a été volé trois par le chat.

		arg0	arg1	IL
	sous-cat initiale	sujet	objet	-
passif-impers ⇒	sous-cat finale	agt-obj	objet	sujet

- passif personnel (i.e. passif avec objet promu)

Trois tartes ont été volées par le chat.

		arg0	arg1
	sous-cat initiale	sujet	objet
passif-pers ⇒	sous-cat finale	agt-obj	sujet

- moyen impersonnel (i.e. moyen avec objet non promu)

Il s'en est loué beaucoup à un prix dérisoire.

		arg0	arg1	IL
	sous-cat initiale	sujet	objet	-
moyen-pers ⇒	intermédiaire	∅	sujet	-
impersonnel ⇒	sous-cat finale	∅	objet	sujet

- moyen personnel (i.e. moyen avec objet promu)

Beaucoup de maisons se sont loués à un prix dérisoire.

		arg0	arg1
	sous-cat initiale	sujet	objet
(moyen-pers) ⇒	sous-cat finale	∅	sujet

2.3 Les sous-catégorisations initiales prévues pour les verbes pleins

Nous pouvons maintenant décrire les différents cadres de sous-catégorisation initiale prévus dans la MGf et la MGi. Nous avons vu en effet que pour pouvoir disposer à l'exécution des schèmes d'une TAG, il faut avoir prévu ces différentes sous-catégorisations, données dans le lexique syntaxique.
Nous fournissons la liste des fonctions grammaticales retenues pour décrire les deux langues. Les différentes propriétés de ces fonctions seront détaillées au Chapitre 5, lors de la description des classes pour les réalisations grammaticales de ces fonctions.

2.3.1 Les fonctions grammaticales distinguées

Pour représenter la sous-catégorisation initiale dans la MGf et la MGi, nous avons retenu les fonctions syntaxiques suivantes :

- sujet (77);

(77) a. **Paul** dorme.
a'. **Paul** dort.

- objet (78)[21];

(78) a. Paul mangia **una mela**.
a'. Paul mange **une pomme**.

- a1-objet (79);

(79) a. Paul parla **a Marie**.
a'. Paul parle **à Marie**.

- a2-objet (80);

(80) a. Paul pensa **a Marie**.
a'. Paul pense **à Marie**.

- gén-objet (81);

(81) a. Paul sogna **di Marie**.
a'. Paul rêve **de Marie**.
b. Ciò risulta **da un'incomprensione immensa**.
b'. Cela résulte **d'une immense incompréhension**.

- obl-objet (82);

(82) a. Paul conta **su Marie**.
a'. Paul compte **sur Marie**.

- locatif (locatif de destination ou de situation) (83);

(83) a. Paul va **a Parigi**.

[21] Nous employons le terme "objet" dans son sens restreint, d'objet direct.

a'. Paul va **à Paris**.

- loc-prov (locatif de provenance) (84);

(84) a. Paul viene **da Parigi**.
a'. Paul vient **de Paris**.

- manière (85);

(85) a. Paul si sente **male**.
a'. Paul se sent **mal**.

- inf-but (infinitive de but) (86);

(86) a. Paolo é corso dal panificio **a comprare pane**.
a'. Paolo a couru à la boulangerie **acheter du pain**.

- inf-per (infinitive de perception) (87);

(87) a. Paolo vede Maria **entrare nella stanza**.
a'. Paolo voit Marie **entrer dans la pièce**.

- interrogative (88).

(88) a. Paolo chiede a Maria **se Chiara è arrivata/chi ha chiamato**.
a'. Paolo demande à Maria **si Chiara est arrivée/qui a appelé**.

Pour la sous-catégorisation finale, on doit introduire une fonction supplémentaire pour le complément d'agent du passif, que l'on désigne par agt-objet.

Nous considérons également une fonction attribut (du sujet et de l'objet), pouvant être remplie par différentes catégories (SN, SA, SP). Cette analyse n'est cependant pas en accord avec les exigences concernant l'interface syntaxe/sémantique (cf. le PCPA), puisqu'il faudrait considérer dans le cas du verbe copule, que l'attribut est le gouverneur sémantique, et que donc il doit être l'ancre (nous développons cette analyse alternative au Chapitre 7).

En ce qui concerne les fonctions a1-objet et a2-objet, étant donné leur grandes similitudes (cf. Chapitre 5), on considère une fonction sous-spécifiée a-objet qui subsume les deux cas. Dans le lexique, lorsqu'un lexème sélectionne une famille avec a-objet (comme n0Van1), le lexème doit spécifier le trait :type-cl (valant a1 ou a2) qui définit s'il s'agit d'un a1-objet ou d'un a2-objet. Dans le cas de verbes pouvant prendre les deux comme RESSEMBLER, on peut sous-spécifier ce trait.

Les fonctions retenues sont celles qui peuvent être, d'après la terminologie de Lazard (1994a), **requises**, i.e. nécessairement présentes dans la sous-catégorisation initiale d'un verbe donné ou à défaut **régies**, i.e. dont la forme est contrainte par le verbe mais pas forcément la présence. On doit en outre distinguer une fonction **obligatoire**, i.e. nécessairement présente dans la sous-catégorisation initiale de tout verbe, d'une fonction simplement requise.
Le sujet est obligatoire en français, mais dans la sous-catégorisation finale. Pour certains verbes impersonnels en effet, nous considérons une sousècatégorisation initiale sans sujet, avec insertion de l'impersonnel ou insertion d'un causateur obligatoire (cf.

CHapitre 4). Au contraire en italien, pour ces verbes sans argument sémantique, nous considérons que l'italien n'impose pas de réalisation de sujet grammatical que ce soit en surface ou dans la sous-catégorisation initiale.
L'objet peut être requis en français et en italien (fr. *rencontrer*, it. *incontrare*), ainsi que l'a1-objet (fr. *appartenir à*, it. *appartenere a*). Pour le gén-objet, on trouve difficilement des verbes où il est requis (fr. *résulter de*, it. *risultare da*), il est en tout cas régi (fr. *parler de*, it. *parlare di*). Le locatif de destination peut être requis, mais il n'est en général pas régi (on peu tout de même citer fr. *se diriger*, it. *avviarsi*). Le locatif de provenance est régi mais pas forcément requis (fr. *revenir de*, it. *tornare da*).
On utilise une fonction oblique (obl-objet) pour les compléments prépositionnels requis n'ayant pas de propriétés syntaxiques particulières (pas de pronominalisation, ni de pronom relatif spécifique).
L'infinitive de but et l'infinitive de perception sont des fonctions dont la réalisation est uniquement infinitive. Les fonctions sujet, objet, a2-objet, gén-objet e obl-objet peuvent avoir une réalisation nominale ou phrastique.

Les MGf et MGi ne prennent pas en compte actuellement les compléments symétriques. Ainsi pour le verbe en (89a) on considère que la sous-catégorisation est (sujet,obl-objet), en (89b), on considère que l'obl-objet est omis. L'intégration d'une fonction spécifique est prévue. Une représentation par redistribution est également envisageable.

(89) a. Paul discute avec Marie.
b. Marie et Paul discutent.

Une autre lacune pouvant être citée concerne les prépositions discontinues, comme en (90).

(90) Le cortège est allé du panthéon à la colonne trajane.

2.3.2 Familles de schèmes et sous-catégorisations initiales

La sous-catégorisation initiale a été définie comme une liste d'actants grammaticaux devant être réalisés localement dans le cas non marqué, plus leurs catégories possibles, ainsi que pour chaque actant grammatical non référentiel, la forme de sa réalisation.
Les MGf e MGi engendrent des TAG (française et italienne), dont les schèmes sont régroupés en familles, chaque famille correspondant à un cadre de sous-catégorisation initiale.
Nous avons affiné la classification en différents cadres de sous-catégorisation initiale, en prenant en compte différents paramètres supplémentaires :

- Pour le français, les sous-catégorisations pour les intransitifs (au sens large, c'est-à-dire les verbes ne comportant pas d'objet) sont systématiquement sous-divisées selon l'auxiliaire de temps sélectionné par le verbe. La sélection de l'auxiliaire ETRE est en outre corrélée à la possibilité de former une participiale "passée" (cf. supra, Section 2.2.3).
  Pour l'italien, la sélection de l'auxiliaire ESSERE pour les intransitifs est impliquée par la sous-catégorisation inaccusative (i.e. avec objet initial, et pas de sujet initial).
- Pour les fonctions sujet, objet, a2-objet et gén-objet, nous avons considéré des cadres de sous-catégorisation initiale (et donc des familles) différents, selon que la réalisation est seulement nominale (et clitique) ou bien peut être phrastique.

La propriété pour un verbe d'admettre une réalisation phrastique pour une de ces fonctions est relativement idiosyncratique. La sémantique du verbe peut bloquer la réalisation phrastique (c'est-à-dire que si un verbe a un argument sémantique de type entité, cet argument ne peut pas se réaliser comme une phrase, renvoyant à un événement). Mais même pour des verbes dont la sémantique le permettrait, il peut y avoir blocage (91a), et l'on a alors recours en français à un SN complexe, dont la tête est *fait* (91b). De plus, des verbes à sémantique proche peuvent différer sur ce point.

(91) a. ?Paul a critiqué que tous soient partis si vite.
b. Paul a critiqué le fait que tous soient partis si vite.

En revanche, un verbe acceptant un dépendant phrastique, accepte généralement un dépendant nominal (de même FG). Nous n'avons cependant pas regroupé ces cas dans une même famille, pour des raisons pratiques, car cela reviendrait à avoir une famille (celle avec réalisation nominale uniquement) incluse totalement dans une autre (celle avec réalisations phrastique et nominale).

Dans le cas de réalisation phrastique, pour les fonctions citées (sujet, objet, a2-objet et gén-objet), un verbe admet généralement à la fois la réalisation comme infinitive et celle comme complétive (tensée). Les familles regroupent alors les deux types de réalisation. Cependant, on doit distinguer différentes familles, selon le complémenteur requis pour la réalisation infinitive. Cette réalisation infinitive peut en effet se faire sans complémenteur, ou bien avec complémenteur *de* et *à* pour le français, ou *di* seulement pour l'italien (cf. Chapitre 5 pour le détail de la représentation des actants phrastiques). Les complémenteurs étant placés dans les arbres des verbes enchâssants, on a distingué plusieurs familles de verbes enchâssants selon ce critère.
Enfin pour le français, les familles ont été distinguées selon que le mode de la complétive est, dans une phrase *assertive*, le subjonctif ou l'indicatif. Le mode est en effet corrélé à différentes propriétés syntaxiques (possibilité pour le verbe d'apparaître comme incise, possibilité pour le sujet du verbe enchâssant et celui de la complétive d'être coréférents)[22].

2.3.3 Liste des familles pour le français

Les cadres de sous-catégorisation initiale prévus sont pour la plupart limités à trois actants. Les familles avec argument phrastique sujet objet, a2-objet et gén-objet groupent les réalisations de l'actant phrastique comme complétive et celles comme infinitive, les verbes ne sélectionnant que l'infinitive (et plus rarement que la complétive) impose le mode approprié dans le lexique syntaxique.

[22] Nous verrons Chapitre 5 que nous considérons les clitiques pronominaux comme des affixes flexionnels, donc traités en morphologie. Cependant, dans l'implémentation, ils sont traités en syntaxe. Aussi, différentes familles ont-elles été distinguées, selon la présence ou l'absence de clitiques figés, réfléchi ou pas, la topologie des arbres étant alors différente selon qu'il y a des clitiques figés ou pas. Nous ne faisons pas apparaître ces familles ici. Avec un traitement des clitiques en morphologie, la distinction est faite au niveau des lexiques : les lexèmes portant des clitiques figés le spécifient au niveau du lexique.

Le détail des réalisations et une justification plus poussée de cette classification sont donnés au Chapitre 5, à la description des différentes réalisations grammaticales pour les fonctions.
Nous donnons Table 4 la liste des familles à actants nominaux, avec les fonctions initiales sous-catégorisées[23]. La Table 5 fournit des phrases exemples. Les familles pour les expressions figées ne sont pas données ici. Leur intégration dans la MGf et la MGi n'est pas achevée. On a tout de même certaines familles avec actant non référentiel : le *il* impersonnel. On verra cependant au Chapitre 4, sur les redistributions, que l'on considère pour certains verbes (atmosphériques) que le sujet impersonnel *il* est inséré par redistribution, on a donc une famille V, sans aucun actant initial. Pour d'autres verbes ou locutions impersonnelles, le *il* est spécifié dans le lexique sous la forme de traits (cf. on représente les clitiques en morphologie).
La famille Vn1 constitue un cas particulier en français, pour les locutions impersonnelles transitives, ou bien les verbes particuliers avec "sujet incorporé" *voici*, *voilà*.

[23] Les conventions pour le nommage des différentes familles sont les suivantes. Les noms de famille sont faits de la concaténation de:

- la catégorie de l'ancre, en majuscule
- la liste des arguments, représentés par la catégorie (principale) de leur réalisation syntaxique, plus un indice, qui suit un ordre d'oblicité syntaxique décroissant des fonctions initiales (0 pour le sujet initial s'il existe, ou bien l'objet initial, ou bien l'argument qui est le gouverneur syntaxique des modifieurs). Le sujet initial, s'il existe, est noté à gauche du caractère valant pour l'ancre, pour rappeler l'ordre canonique SVO en français. Donc par exemple, pour un transitif à arguments nominaux, on note la famille "n0Vn1". Au contraire, pour une famille représentant des inaccusatifs en italien, l'argument 0 (objet initial) est noté à droite. Donc par exemple pour les intransitifs stricts en italien, on notera n0V les inergatifs, et Vn0 les inaccusatifs.

Les catégories de réalisation syntaxique des arguments sont notées n (pour nominal ou clitique), s (pour phrastique fini ou infinitive), a (pour les attributs de l'objet ou du sujet, dont la catégorie est variable mais en particulier adjectivale). Ainsi par exemple, s0Vn1 vaut pour la famille à sujet phrastique et à objet direct nominal. "loc" vaut pour le locatif, "deloc" vaut pour le locatif de provenance.
Les compléments non directs sont précédés de : "a" pour le a1-objet ou l'a2-objet; "d" pour le gén-objet; "p" pour l'obl-objet, suivis de la catégorie (n ou p) de réalisation.
Ainsi n0Vdn1 vaut-il pour la famille d'un transitif avec sujet et gén-objet nominal. La famille n0Vas1 vaut pour un verbe à sujet nominal et a2-objet phrastique (*chercher à faire qqch)*.
Les arguments optionnels (dont la réalisation syntaxique peut se faire ou pas, en dehors des cas de redistribution des fonctions) sont notés entre parenthèses.
Pour les familles intransitives, la distinction sur l'auxiliaire est notée par un suffixe -a pour l'auxiliaire AVOIR, -e pour l'auxiliaire ETRE.

	fonctions initiales										
famille	SUJET	OBJET	A-OBJET	GEN-OBJET	OBL-OBJET	LOCATIF	MANIERE	LOC-PROV	ATTR-SUJ	ATTR-OBJ	**exemple**
V											pleuvoir
Vn1	figé	N									voici, (*il*) y avoir
n0V-a	N										dormir
n0V-e	N										arriver, s'évanouir
n0Van1-a	N		N								parler (a1), penser (a2)
n0Van1-e	N		N								parvenir (a2)
n0Vden1-e	N			N							mourir, se moquer
n0Vden1-a	N			N							rêver
n0Vden1pn2-a	N			N	+						discuter avec
n0Vpn1-a	N				+						compter sur
n0Vpn1-e	N				+						revenir sur
n0Vloc1-a	N					+					déboucher
n0Vloc1-e	N					+					se trouver
n0Vad1-e	N						+				aller, se sentir
n0Vdeloc1-e	N							+			venir
n0Va1	N								+		paraître
n0Vn1	N	N									manger
n0Vn1(an2)	N	N	N								donner
n0Vn1(den2)	N	N		N							abreuver
n0Vn1(pn2)	N	N			+						admirer
n0Vn1(loc2)	N	N				+					trouver
n0Vn1ad2	N	N					+				traiter
n0Vn1(deloc2)	N	N						+			rapporter
n0Vn1a2	N	N								+	trouver

Table 4. Familles à arguments nominaux, pour le français

famille	exemple
V	*Il pleut.*
Vn1	*Voilà les enfants. Il y a la mer.*
n0V-a	*Paul dort.*
n0V-e	*Paul arrive. Paul s'évanouit.*
n0Van1-a	*Paul par à Chiara. Paul pense à Chiara.*
n0Van1-e	*Paul est parvenu à ses fins.*
n0Vden1-e	*Paul est mort d'amour. Paul s'est moqué de la mort.*
n0Vden1-a	*Paul a rêvé de l'étoile du berger.*
n0Vden1pn2-a	*Paul a discuté de sa venue avec la direction.*
n0Vpn1-a	*Paul a compté sur Chiara.*
n0Vpn1-e	*Paul est revenu sur sa décision.*
n0Vloc1-a	*Le sentier a débouché sur une clairière.*
n0Vloc1-e	*Paul se trouve à Singapour.*
n0Vad1-e	*Paul va mal. Paul se sent bien.*
n0Vdeloc1-e	*Paul est venu de Rome.*
n0Va1	*Chiara paraît heureuse.*
n0Vn1	*Chiara mange des glaces.*
n0Vn1(an2)	*Chiara donne une glace à Paul.*
n0Vn1(den2)	*Paul abreuve ses élèves de citations.*
n0Vn1(pn2)	*Paul admire Chiara pour sa patience.*
n0Vn1(loc2)	*Paul trouve une perle dans la coquille.*
n0Vn1(deloc2)	*Paul rapporte une perle de Tahiti.*
n0Vn1ad2	*Paul traite bien Chiara.*
n0Vn1a2	*Paul trouve ces arguments peu convaincants.*

Table 5. Familles à arguments nominaux : exemples

On donne Table 6, la liste des familles[24] à ancre verbale, avec un actant phrastique, pour le français, et Table 7, des phrases exemples. La MGf ne couvre que des sous-catégorisations initiales avec un seul actant phrastique.
Les verbes "déficients" décrits Section 1, n'entrent pas dans cette classification.

[24] Pour les familles à objet (direct) phrastique, on a note dans le nom de la famille un "c" si le complémenteuer de l'infinitive est *de*, un "ca" si le complémenteur est *à*, et rien si l'infinitive n'a pas de complémenteur. Ainsi n0Vcas1inf désigne la famille à objet phrastique, dont l'objet à l'infinitif comprend un complémenteur *à*.
On ajoute en outre le mode (ind/subj) dans le cas de réalisation par complétive, dans une phrase assertive (mais uniquement lorsqu'il existe les deux possibilités de mode). Ainsi on précise n0Vs1subj par exemple, pour *vouloir venir/vouloir que Paul vienne*. Mais pour les sujets phrastiques, toujours soit infinitif soit au subjonctif, on n'a pas à préciser.

	fonctions initiales										
famille	SUJET	OBJET	A-OBJET	GEN-OBJ	OBL-OBJ	LOCATIF	ATTR-SUJ	INF-PERC	INF-BUT	INTERR	**exemple**
s0V	P										importer
s0Vn1	P	N									angoisser
s0Van1	P		N								plaire
s0Vpn1	P				+						influer (sur)
s0Va1	P						+				sembler
*il*Vs1	*il*	P (∅ inf /subj /ind)									falloir, sembler
n0Vs1subj	N	P (∅ inf /subj)									aimer
n0Vs1ind	N	P (∅ inf /ind)									penser
n0Vs1ind(an2)	N	P (∅ inf /ind)	N								dire
n0Vs1ind(dn2)	N	P (∅ inf /ind)		N							retenir
n0Vcs1ind	N	P (*de* inf /ind)									décider
n0Vcs1ind(an2)	N	P (*de* inf /ind)	N								promettre
n0Vcs1subj	N	P (*de* inf /subj)									regretter
n0Vcs1subj(an2)	N	P (*de* inf /subj)	N								ordonner
n0Vcs1subj(dn2)	N	P (*de* inf /subj)		N							obtenir
n0Vcs1subj(pn2)	N	P (*de* inf /subj)			+						convenir
n0Vcs1inf	N	P (*de* inf)									se presser
n0Vcas1inf	N	P (*à* inf)									hésiter
n0Vcas1inf(an2)	N	P (*à* inf)	N								apprendre
n0Vds1-a	N			P							rêver
n0Vds1-e	N			P							se souvenir
n0Vn1ds2	N	N		P							avertir, féliciter
n0Vds1(an2)-e	N		N	P							se plaindre
n0Vas1-a	N		P								tenir
n0Vas1-e	N		P								parvenir, s'attendre
n0Vn1as2	N	N	P								inciter
n0Vn1sperc2	N	N						+			voir
n0V(loc1)(sbut2)-e	N					+			+		aller, se retirer
n0V(loc1)(sbut2)-a	N					+			+		courir
n0Vn1(loc2)(sbut3)	N	N				+			+		envoyer
n0Vint1	N									+	savoir
n0Vint1(an2)	N		N							+	demander

Table 6. Liste des familles avec un argument phrastique, pour le français

famille	**exemple**
s0V	*Que tout soit prêt à 5h importe au plus au point.*
s0Vn1	*Que Chiara ne soit pas rentrée angoisse Paul.*
s0Van1	*Que tout soit prêt plaît à Chiara.*
s0Vpn1	*Que Paul soit jeune influe sur la décision du jury.*
s0Va1	*Que Paul parte semble inévitable.* *Que Paul parte est une chance pour lui.*
*il*Vs1subj	*Il faut partir/que Paul parte.* *Il semble que Paul est/soit déjà parti.*
n0Vs1subj	*Paul aime faire du ski/que Chiara fasse du ski.*
n0Vs1ind	*Paul pense avoir bien agi/que Chiara a bien agi.*
n0Vs1ind(an2)	*Paul dit (à Pierre) être allé au cinéma /que Chiara est allée au cinéma.*
n0Vs1ind(den2)	*Paul retient (de la réunion) que le travail ne manque pas.*
n0Vcs1ind	*Paul décide de partir/que tous partiront.*
n0Vcs1ind(an2)	*Paul promet (à Chiara) de partir/que tous partiront.*
n0Vcs1subj	*Paul regrette de partir/que tous soient partis.*
n0Vcs1subj(an2)	*Paul ordonne (à Chiara) de partir/que tout soit prêt à 5h.*
n0Vcs1subj(den2)	*Paul obtient (de Chiara) de pouvoir partir/que tous partent.*
n0Vcs1subj(pn2)	*Paul convient (avec Chiara) de partir/que Pierre quitte le pays.*
n0Vcs1inf	*Paul se dépêche de finir sa valise.*
n0Vcas1inf	*Paul hésite à partir.*
n0Vcas1inf(an2)	*Chiara apprend à Paul à rêver.*
n0Vds1-a	*Chiara a rêvé de partir.*
n0Vds1-e	*Chiara s'est souvenue d'être déjà passée par là/que tout était calme ce jour-là.*
n0Vn1ds2	*Paul convainc Chiara de partir/que tous peuvent partir.* *Paul félicite Chiara d'être à l'heure/que tout soit déjà prêt.*
n0Vds1(an2)-e	*Paul s'est plaint de ne pas pouvoir venir/que tous soient bruyants.*
n0Vas1-a	*Paul a tenu à partir/à ce que tous partent.*
n0Vas1-e	*Paul est parvenu à partir/à ce que personne ne l'entende*
n0Vn1as2	*Paul incite Chiara à partir/à ce qu'elle accepte l'offre.*
n0Vn1sperc2	*Paul voit Chiara entrer dans le grenier.*
n0V(loc1)(sbut2)-e	*Paul est allé (à la boulangerie) (acheter du pain).* *Paul s'est retiré (dans sa chambre) (étudier ses manuscrits).*
n0V(loc1)(sbut2)-a	*Paul a couru (à la boulangerie) (acheter du pain).*
n0Vn1(loc2)(sbut3)	*Paul a envoyé Chiara (à la boulangerie) (acheter du pain).*
n0Vint1	*Paul sait quel auteur aura le Goncourt/si Chiara aura le Goncourt.*
n0Vint1(an2)	*Paul demande (à Pierre) quel auteur aura le Goncourt/si Chiara aura le Goncourt.*

Table 7. Liste des familles avec un argument phrastique : exemples

2.3.4 Liste des familles pour l'italien

On donne Table 8 la liste des familles pour les verbes à arguments nominaux et Table 9, une liste de phrases exemples. Les différences avec la classification faite pour le français proviennent essentiellement du fait que l'on considère certains intransitifs comme sélectionnant un objet initial et pas de sujet initial. [25]

Le cas de la famille Vn1 est particulier. Elle est utilisée pour *ecco*, équivalent de *voici/voilà.* Ecco ne sous-catégorise qu'un objet initial, mais n'est cependant pas un inaccusatif (Salvi, 1988:73). Il ne peut apparaître qu'en proposition principale. On doit considérer qu'il incorpore un sujet, qui bloque la promotion de l'objet initial en position sujet.

	Fonctions initiales								
Famille	SUJET	OBJET	A-OBJET	GEN-OBJET	OBL-OBJET	LOCATIF	MANIERE	LOC-PROV	**Exemple**
V									*piovere*
Vn1	-	N							*ecco*
n0V	N								*dormire*
n0Van1	N		N						*parlare (a1), pensare (a2)*
n0Vdn1	N			N					*sognare*
n0Vpn1	N				+				*pattegiare con*
n0Vloc1	N					+			*nuotare*
Vn0		N							*arrivare, riposarsi*
Vn0an1		N	N						*mancare (a1)*
Vn0dn1		N		N					*fidarsi (di) risultare (da)*
Vn0pn1		N			+				*coincidere con*
Vn0loc1		N				+			*cadere*
Vn0ad1		N					+		*stare, sentirsi*
Vn0dloc1		N						+	*venire*
n0Vn1	N	N							*mangiare*
n0Vn1(an2)	N	N	N						*dare*
n0Vn1(den2)	N	N		N					*coprire (di) svincolare (da)*
n0Vn1(pn2)	N	N			+				*trattare con*
n0Vn1(loc2)	N	N				+			*nascondere*
n0Vn1ad2	N	N					+		*trattare*
n0Vn1(dloc2)	N	N						+	*prelevare*

Table 8. Familles à arguments nominaux, pour l'italien

[25] Cela vient également du fait que l'on a considéré une seule fonction gén-objet en italien pour deux prépositions marqueurs possibles : *di* et *da*.

Famille	Exemple
V	***E/ha nevicato.*** *(Il a neigé)*
Vn1	***Ecco Maria.*** *(Voilà Maria)*
n0V	***Paolo ha dormito.*** *(Paul a dormi)*
n0Van1	***Paola parla a Piero*** *(a1)* ***Paola pensa a Piero*** *(a2)* *(Paola parle à Pierre* *Paola pense à Pierre)*
n0Vdn1	
n0Vpn1	***Non patteggeremo col nemico.*** *(Nous ne pactiserons pas avec l'ennemi)*
n0Vloc1	***Nuotammo in silenzio verso l'isola.*** *Nous nageâmes en silence vers l'île.*
Vn0	***E arrivato Giovanni.*** ***Si riposa.*** *(Giovanni est arrivé)* *(Il se repose)*
Vn0an1	***Gli manca la parola*** *(a1)* *DAT1.3ms manque la parole* *(Il lui manque la parole)*
Vn0dn1	***Chiara si è fidata di Paolo****.* *Chiara s'est fiée de Paolo* *(Chiara a eu confiance en Paolo)*
Vn0pn1	***Questo è coinciso colla visita del Papa.*** *(Cela a coincidé avec la visite du pape)*
Vn0loc1	***Paola è partita per Siena.*** *(Paola est partie pour Sienne)*
Vn0ad1	***Mi sento bene.*** *(Je me sens bien)*
Vn0dloc1	***Paolo viene da Napoli.*** *(Paolo vient de Naples)*
n0Vn1	***Chiara mangia un amaretto****.* *(Chiara mange un amaretto)*
n0Vn1(an2)	***Paolo da un bacio a Chiara.*** *(Paolo donne un baiser à Chiara)*
n0Vn1(dn2)	***Paolo ha coperto Chiara d'insulti.*** *(Paolo a couvert Chiara d'insultes)* ***E difficile svincolare uno stabile da un'ipoteca*** *(Il est difficile de libérer un immeuble d'une hypothèque)*
n0Vn1(pn2)	***Giovanni tratta la vendita con Maria.*** *Giovanni traite la vente avec Maria.*
n0Vn1(loc2)	***Giovanni nasconde in banditi in questa casa.*** *(Giovanni cache les bandits dans cette maison.)*
n0Vn1ad2	***Chiara ti tratta bene.*** *(Chiara te traite bien)*
n0Vn1(dloc2)	***I rapitori hanno prelevato il prefetto dal suo domicilio.*** *(Les ravisseurs ont enlevé le préfet à son domicile)*

Table 9. Familles à arguments nominaux, pour l'italien (exemples)

Pour les verbes avec un argument phrastique, on a également la distinction entre inergatifs et inaccusatifs. Pour les inaccusatifs à objet initial phrastique, on doit faire une distinction selon que cet objet initial a l'infinitif prend un complémenteur *di* ou pas (cf. Chapitre 5, la section sur les sujets phrastiques pour plus de détails).
Une infinitive objet en italien est soit introduite par le complémenteur *di* soit sans complémenteur, on n'a pas l'équivalent des infinitives objet introduites par le complémenteur *à* en français (et donc pas les familles n0Vcas1, n0Vcas1(an2)).
D'une manière générale, une infinitive objet sans complémenteur est sélectionnée par moins de verbes en italien. En effet, toute la classes des verbes déclaratifs en français le permettent, alors que l'équivalent italien impose le *di*.

(92) a. Chiara dit / raconte / affirme / assure / confirme / nie / ... être allée au cinéma.
a'. Chiara dice / racconta / afferma / assicura / conferma / nega ... **di** essere andata al cinema.

Aussi n'a-t-on dans la classification italienne qu'une seule classe avec infinitive objet sans complémenteur (n0Vs1). Il s'agit d'une classe pour des verbes comme AMARE (aimer), VOLERE (vouloir), DESIDERARE (désirer), ODIARE (détester)..., dont seul un sous-ensemble admet la complétive. Ils correspondent en français à n0Vs1subj, mais Elia (1984) note qu'ils peuvent admettre également l'indicatif, nous n'avons donc pas contraint le mode de la complétive dans les schèmes.

Une autre différence dans la classification vient du fait que nous n'avons pas distingué plusieurs familles à objet phrastique selon le mode de la complétive tensée. Elia (1984) note un certain flottement dans le choix du mode des complétives en italien. Il distingue trois classes : les verbes imposant l'indicatif, les verbes imposant le subjonctif, et les verbes "flottant", en donnant comme ordre de grandeur environ 200 pour les premiers, 150 pour les seconds et environ 100 pour les troisièmes.
Il remarque cependant que le mode varie suivant le niveau de langue, et que l'indicatif a une nette tendance à supplanter le subjonctif. Les tables de verbes compilées par Elia, pour les verbes à un actant phrastique ne sont pas clairement distinguées pour le mode de la complétive. Aussi, nous n'avons pas distingué différentes familles selon le mode de la complétive. Dans le cas où le mode est clairement contraint, il faut le noter dans le lexique.
Plus généralement, dans le cas d'un verbe avec plusieurs compléments (dont un phrastique) on ne peut pas prévoir quel est le contrôleur de l'infinitive (le sujet ou un complément), et il faut le noter dans le lexique, comme pour le français.

	fonctions initiales									
famille	SUJET	OBJET	A-OBJET	GEN-OBJ	OBL-OBJ	LOCATIF	INF-PERC	INF-BUT	INTERR	**exemple**
Vcs0		P (di inf / subj)								*avvenire (arriver)*
Vs0an1		P (∅ inf / subj)	N							*piacere (plaire)*
Vcs0an1		P (di inf / subj)	N							*parere (sembler)*
s0Vn1	P	N								*preoccupare (préoccuper)*
s0Vpn1	P				+					*influire (influer)*
n0Vs1	N	P (∅ inf / subj)								*volere (vouloir)* *amare (aimer)*
n0Vcs1	N	P (*di* inf /ind /subj)								*capire (comprendre)* *credere (croire)*
n0Vcs1(an2)	N	P (*di* inf /ind /subj)	N							*promettere(promettre)* *imporre (imposer)*
n0Vcs1(dn2)	N	P (*di* inf /subj)		N						*ottenere da (obtenir)* *dedurre da (déduire)*
n0Vcs1(pn2)	N	P (*di* inf /subj /ind)			+					*decidere con (décider avec)*
n0Vcs1inf	N	P (*di* inf)								*smettere (cesser)*
n0Vds1	N			P						*dubitare (douter)*
Vn0ds1		N		P						*pentirsi, lamentarsi*
n0Vn1ds2	N	N		P						*avvisare*
n0Vas1inf	N		P(inf)							*rinunciare (renoncer)*
Vn0as1inf		N	P(inf)							*arbituarsi (s'habiture)*
n0Vn1as2inf	N	N	P(inf)							*costringere (contraindre)*
n0Vn1sperc2	N	N					+			*vedere (voir)*
Vn0(loc1)(sbut2)		N				+		+		*andare (aller),* *correre (courir)*
n0Vn1(loc2)(sbut3)	N	N				+		+		*mandare* *envoyer*
n0Vint1	N								+	*sapere (savoir)*
n0Vint1(an2)	N		N						+	*chiedere (demander)*

Table 10. Liste des familles avec un argument phrastique, pour l'italien

famille	exemple
Vcs0	***Di dover decidere in fretta non avveniva da tempo.*** *(De devoir décider à la hâte n'était pas arrivé depuis longtemps)*
s0Vn1	***Mi preoccupa essere rimasto da solo.*** *ACC.1 préoccupe être resté seul* *(Être resté seul me préoccupe)* ***Mi preoccupa che Paolo sia rimasto da solo***. *ACC.1 préoccupe que Paul soit resté seul* *(Que Paul soit resté seul me préoccupe.)*
Vs0an1	***Mi dispiace che Gianni non sia potuto venire.*** *DAT1.1 déplaît que Gianni NEG soit pu venir* *(Je regrette que Gianni n'ait pas pu venir)* ***Mi è piaciuto andare al mare***. *DAT1.1 plaît aller à la mer* *(Aller à la mer me plaît.)*
Vcs0an1	***Mi è successo di dover decidere in fretta*** *DAT1.1 est arrivé de devoir décider à la hâte.* *(Il m'est arrivé de devoir décider à la hâte)* ***Spesso mi succede che la sorte mi sia favorevole.*** *Souvent DAT1.1 arrive que le sort me soit favorable* *(Souvent il m'arrive que le sort me soit favorable.)*
s0Vpn1	***Che Paolo sia giovane influisce sui giudici.*** *(Que Paolo soit jeune influe sur les juges)*
n0Vs1[26]	***Muriel vuole che Max canti/ che Max canta /cantare.*** *(Elia, 1984)* *Muriel veut que Max chante(subj) / que Max chante(ind) / chanter* ***Amo andare al mare/?Amo che ognuno faccia il proprio dovere***. *(Acquaviva, 1991:638)* *J'aime aller à la mer/?J'aime que chacun fasse son devoir.*
n0Vcs1	***Ena crede che Mario vincerà la partita / che Max sia bello / di essere bella.*** *Ena croit que Mario gagnera la partie / che Max soit beau / d'être belle.* ***Ena capisce che Mario vincerà la partita/ di dover vincere la partita.*** *Ena comprned que Mario gagnera la partir / de devoir gagner la partie*
n0Vcs1(an2)	***Marco impone (a Chiara) di gestire il negozio.*** *(Marco impose à Chiara de gérer le magasin.)* ***Il generale ordinò che i soldati facessero saltare il ponte.***[27] *Le général ordonna que les soldats fassent sauter le pont.* ***Marco promette (a Chiara) di venire / che tutto sarà pronto entro le 5.*** *(Marco promet à Chiara de venir / que tout sera prêt avant 5h.)*

[26] Cette famille regroupe le cas infinitive sans *di*, et complétive. La complétive n'est cependant admise que par un groupe restreint de verbes sélectionnant cette famille (cf. Acquaviva, 1991:638).

n0Vcs1(dn2)	***Mario ha dedotto dalla faccia allegra di Chiara che aveva vinto/ di aver vinto.*** *Mario a déduit du visage joyeux de Chiara qu'elle avait gagné/d'avoir gagné.* ***Chiara ottiene (da Marco) che il negozio le sia affidato/ di poter gestire il negozio .*** *(Chiara obtient de Max que le magasin lui soit confié/ de pouvoir gérer le magasin.)*
n0Vcs1(pn2)	***Ena decide (con Chiara) che Mario arbitrerà la partita / di vincere la partita.*** *Ena décide (avec Chiara) que Mario arbitrera la partie / de gagner la partie.*
n0Vcs1inf	***Paolo ha smesso di fumare.*** *Paolo a cessé de fumer.*
n0Vds1	***Max dubita che Chiara parta / di poter partire.*** *Max doute que Chiara parte / de pouvoir partir.*
Vn0ds1[28]	***Mi sono accorto che oggi i negozi sono chiusi.*** *RF.1 suis aperçu qu'aujourd'hui les magasins sont fermés.* *(Je me suis rendu compte qu'aujourd'hui les magasins sont fermés)* ***Gianni si è pentito di aver detto una sciocchezza.*** *(Gianni s'est repenti d'avoir dit une stupidité.)*
n0Vn1ds2	***Lo avvisai di stare in guardia/che qui non si fuma.*** *(Je l'avertis de rester sur ses guardes/qu'ici on ne fume pas)*
n0Vas1inf	***Gianni ha rinunciato a frequentare la mensa.*** *Gianni a renoncé à aller à la cantine.*
Vn0as1inf	***Mario si è abituato a frequentare la mensa.*** *Mario s'est habitué à aller à la cantine.*
n0Vn1as2inf	***Il presidente invitò / costrinse / indusse Gianni a essere più conciso.*** *(Acquaviva, 1991)* *Le président invita / contraint / poussa Gianni à être plus concis.*
n0Vn1sperc2	***Marco vede Chiara entrare nella stanza.*** *Marco voit Chiara entrer dans la pièce.*
Vn0(loc1)(sbut2)	***Mario è corso/andato al panificio a comprare pane.*** *Mario est couru/allé à la boulangerie acheter du pain.*
n0Vn1(loc2)(sbut 3)	***Abbiamo mandato Gianni al panificio a comprare pane.*** *Nous-avons envoyé Gianni à la boulangerie acheter du pain.*
n0Vint1	***Sai di che romanzo meraviglioso sta parlando?*** *Sais-tu de quel roman merveilleux PARLER.progressif.3s*
n0Vint1(an2)	***Spiegami che tipo di macchina vuoi.*** *Explique-moi quel type de voiture tu-veux.*

Table 11. Liste des familles avec un argument phrastique, pour l'italien (exemples)

[27] Pour les verbes d'ordre/permission, la construction avec complétive (alors au subjonctif) est plus difficile que pour les verbes déclaratifs (où la complétive est à l'indicatif), et si elle est possible, c'est sans le a1-objet (Acquaviva, 1991:642). En français, cela est possible (famille n0Vcs1subj(an2)) : *Paul demande à Chiara que tout soit prêt à son retour.*

[28] Pour cette classe également, une part non négligeable de verbes ne sélectionnent pas la complétive, mais seulement l'infinitive.

Chapitre 4 : Les redistributions des fonctions, pour les verbes

Nous décrivons dans ce chapitre la dimension 2 pour le français et l'italien, c'est-à-dire les redistributions des fonctions assignées aux actants grammaticaux initiaux.

On y décrit des phénomènes que l'on peut qualifier de marginaux, sachant que la représentation dans la MG permet d'adapter les LTAG à des domaines particuliers. On peut en opérant des "coupes" dans la hiérarchie, compiler facilement des grammaires à couverture plus restreinte.

Nous représentons par des redistributions le passif, l'impersonnel, le moyen, le causatif et le réfléchi. Ce dernier est représenté par redistribution, car il interfère avec la transitivité d'un verbe et donc avec les autres redistributions applicables à ce verbe.
Nous avons choisi de représenter des redistributions productives, au sens qui s'applique à une grande partie des verbes ayant une sous-catégorisation donnée et qui ne requièrent pas une sémantique "très particulière". Nous faisons une distinction entre **redistributions** et **alternances**, bien que nous n'ayons pas une définition claire de ce qui les distingue. Les deux phénomènes représentent la relation entre deux sous-catégorisations *finales* possibles pour un lexème.
Les propriétés des deux types de relation sont que :

- une redistribution est (i) plutôt productive, c'est-à-dire valent pour un grand nombre de verbes ou formes verbales dont la sous-catégorisation partage certaines propriétés; (ii) ne fait pas appel à une sémantique "très particulière". Ainsi par exemple en français, il est général qu'un verbe pouvant entrer dans une sous-catégorisation comportant un sujet, admet également une sous-catégorisation où ce verbe et un verbe causatif forment un prédicat complexe, dont la sous-catégorisation comprend un sujet causateur et un objet correspondant au sujet du verbe pris isolément. Pour le passif, bien que l'on puisse sans doute isoler des facteurs sémantiques permettant le passif ou le bloquant, ces facteurs sont généraux et font que le passif apparaît comme une possibilité régulière pour une sous-catégorisation avec objet.
- une alternance s'applique à une classe plus réduite de verbes, des facteurs sémantiques plus spécifiques entrant en jeu; une alternance est en général non marquée morpho-syntaxiquement. Par exemple, l'alternance locative (*Paul a chargé le camion d'oranges/Paul a chargé les oranges dans le camion*) s'applique-t-elle à un nombre très réduit de verbes. Nous la considérons comme une alternance. Un bon exemple de distinction alternance/redistribution est celui de la distinction entre moyen (par exemple en (c)) et neutre (b) en français (voir aussi infra Section 1.3). La possibilité d'entrer dans une construction moyenne est générale pour les verbes transitifs. La possibilité d'entrer dans une construction neutre apparaît plutôt comme

une propriété lexicalement déterminée, c'est-à-dire s'applique à une classe sémantique de verbes bien inférieure à l'ensemble des verbes transitifs.

(93) a. Le souffle de l'explosion a brisé la vitre.
b. La vitre s'est brisée.
c. Dans les trains, les vitres, ça se brisent avec un marteau.

Une autre distinction, entre les deux types de relations entre sous-catégorisations concernent la sémantique associée à la forme verbale. On considérera systématiquement comme une alternance une relation entre deux sous-catégorisations telles qu'au niveau sémantique, les deux prédicats n'ont pas les mêmes arguments. Là encore la distinction neutre/moyen fournit un bon exemple. Au niveau sémantique, au delà de différences sémantiques entre les deux constructions (en particulier de différences aspectuelles, cf. Ruwet, 1972) une différence très nette est que pour le moyen, l'agent (le sujet initial) est non exprimé, mais est interprétable (comme pour l'agent dans le cas d'un passif sans agent). Pour le neutre, l'agent n'est ni exprimé linguistiquement, ni interprétable. Dans ce cas il paraît justifié de considérer deux entrées distinctes pour la variante neutre et la variante transitive[1].

Pour une relation d'alternance, on considère que ces deux sous-catégorisations finales sont deux sous-catégorisations initiales de deux entrées lexicales distinctes (telles que nous les avons considéré pour la MG; on peut définir une entrée lexicale plus abstraite encore, représentant l'invariant commun à plusieurs entrées lexicales dans une relation d'alternance). Pour une relation de redistribution entre deux sous-catégorisations, on considère que ces deux sous-catégorisations finales dérivent d'une même sous-catégorisation initiale pour une seule entrée lexicale.

Nous faisons également une différence entre une **sous-catégorisation réduite** et l'**ellipse** d'un actant. Les actants autres que le sujet peuvent en effet en général être non exprimés. On distingue deux situations, selon que l'actant non exprimé est isolable ou pas.
Ainsi hors contexte, (94b) est ambiguë. Mais en (94c), un interlocuteur (ou un auditoire) est présent dans la situation dénotée par le phrase, on dira qu'il s'agit d'une ellipse. En (94d), le verbe a le sens "avoir la capacité de parler", un interlocuteur n'est pas présent dans la situation ici dénotée. On dira qu'il s'agit d'une sous-catégorisation réduite.

(94) a. Paul parle à sa soeur.
b. Paul parle.
c. Chut, Paul parle.
d. Mon petit frère Paul parle maintenant.

En (95), ce n'est pas que l'objet (le mangé) n'est pas présent dans la situation dénotée, mais plutôt qu'il n'est pas isolable. La situation porte sur l'activité de manger. On considère alors qu'il s'agit d'une sous-catégorisation réduite.

[1] Il faut remarquer qu'une forme avec vrai réfléchi (i.e. à sens réfléchi), met en jeu au niveau sémantique un argument de moins que la variante non réfléchie : le réfléchi est la marque de la fusion de deux arguments. Cependant, cela est différent du cas du neutre, en effet pour le réfléchi, c'est un argument qui joue le rôle sémantique de deux arguments, ces rôles restent présents dans la situation dénotée par un verbe réfléchi. Il n'y a donc pas contradiction à représenter le réfléchi comme une redistribution.

(95) Paul n'aime pas être dérangé quand il mange.

Dans le cas d'un verbe transitif, on peut tester la transitivité, par le causatif par exemple, du verbe obtenu par ellipse de l'objet, et celle du verbe obtenu par réduction de la sous-catégorisation. On s'attend à ce qu'un verbe avec objet omis reste transitif, alors le verbe à sous-catégorisation réduite soit intransitif. Ainsi, en (96), *lire* est intransitif (étant donnée la réalisation directe du causé (cf. infra Section 1.4)), l'interprétation préférentielle à notre avis est que *lire* est pris ici comme une activité, et n'a pas de thème spécifique, c'est-à-dire qu'il s'agit d'un cas de sous-catégorisation réduite. En (97) au contraire, on amène par le contexte un objet spécifique de lecture (*la thèse*), donc *lire* est employé avec un objet omis. Or il est intéressant de remarquer que la réalisation directe du causé est bloquée dans ce cas, alors que la réalisation en *par* est possible.

(96) J'ai fait lire les élèves.

(97) _ Ta thèse était parfaite.
_ Oui je sais. J'ai bénéficié de nombreux commentaires et j'ai fait lire (par + *E) plusieurs personnes.

Dans la MG, l'ellipse est traitée comme une réalisation particulière pour la fonction pouvant être omise (la non-réalisation grammaticale), la sous-catégorisation réduite est traitée en assignant des sous-catégorisations initiales différentes aux lexèmes concernés. Alors que l'ellipse est généralement toujours possible pour un complément, pourvu que le contexte permette l'interprétation du complément elliptique, la réduction de sous-catégorisation n'est pas acceptée par tous les verbes.

Nous pouvons maintenant commencer la description des redistributions considérées. Nous séparons la description des redistributions pour le français (Section 1), de celle de l'italien (Section 2).

1 Les redistributions pour les verbes en français

Nous nous référons à l'ordre d'oblicité des fonctions grammaticales (cf. Chapitre 2) pour la représentation des redistributions des fonctions du français et de l'italien. On parlera de démotion (en traduisant le terme *demotion* de Comrie (1977)) pour une redistribution partielle qui fait descendre un argument dans la hiérarchie (voire qui l'efface totalement) et de promotion pour le contraire. On parlera d'insertion dans le cas d'actants grammaticaux nouveaux (un causateur par exemple, ou un explétif).
Dans la MGf, nous traitons par des redistributions les phénomènes suivants : l'impersonnel, le passif, le moyen, le réfléchi et le causatif. Comme ils interagissent de manière assez complexe, les différentes sections que nous leur consacrons comportent de multiples renvois. Nous donnons un schéma récapitulatif à la fin de cette Section (p. 178).
Nous avons veillé à prendre en compte les contraintes sur l'interaction des différentes redistributions, et nous décrivons les techniques utilisées pour bloquer certaines compositions. Cependant, dans l'implémentation, les compositions de redistributions ne sont pas appliquées librement. Il faut définir une classe pour chaque composition de

redistributions (y compris le cas de non redistribution). Les compositions non valides sont donc bloquées simplement parce qu'elles n'ont pas de classe correspondante.

1.1 L'impersonnel

Le terme impersonnel en français est employé pour un ensemble de constructions qui font apparaître, avec éventuellement d'autres propriétés, un sujet syntaxique non référentiel dont la forme est *il*.
Parmi les différents emplois du *il* impersonnel, il faut distinguer le cas de verbes ou locutions dits impersonnels (98a,b,c,d,e), où on n'a pas de syntagme postverbal direct qui peut alterner comme sujet avec *il*, des cas où cette alternance est possible. L'actif impersonnel (99b) alterne avec un actif "simple"; le passif impersonnel (100b) alterne avec un passif personnel et le moyen impersonnel (101b) alterne avec un moyen personnel.

(98) a. Il pleut.
b. Il fait beau.
c. Il s'agit de la meilleure performance de l'année.
d. Il faut partir.
e. Il faut que Paul parte.

(99) a. Beaucoup d'eau a coulé sous les ponts depuis.
b. Il a coulé beaucoup d'eau sous les ponts depuis.

(100) a. 100000 actions ont été échangées aujourd'hui.
b. Il a été échangé 100000 actions aujourd'hui.

(101) a. En France, plus de 1000 tonnes de fromage se consomment chaque année.
b. En France, il se consomme plus de 1000 tonnes de fromage chaque année.

Il y a également alternance entre une construction avec sujet phrastique (infinitif ou tensé) et construction impersonnelle ou l'actant phrastique est postverbal. Par exemple un actif personnel avec sujet phrastique alterne avec une construction impersonnelle (pour un verbe intransitif (102), (103); pour un verbe d'état avec attribut (104), (105)).

(102) a. Que Paul n'en sache rien importe au plus haut point.
b. Il importe au plus haut point que Paul n'en sache rien.

(103) a. Renouer le dialogue importe avant tout.
b. Il importe avant tout de renouer le dialogue.

(104) a. Que Paul n'en sache rien est nécessaire pour la réussite du projet.
b. Il est nécessaire pur la réussite du projet que Paul n'en sache rien.

(105) a. Garder le secret est nécessaire pour la réussite du projet.
b. Il est nécessaire pour la réussite du projet de garder le secret.

Dans ces cas, la variante impersonnelle apparaît souvent meilleur que la variante personnelle, qu'on ne peut cependant pas complètement exclure. Cela est également vrai dans le cas du passif impersonnel (106b), (106d), ou de moyen impersonnel (107b) (cf. Zribi-Herz, 1982 par exemple).

(106) a. Que tous partiraient avait été décidé de longue date.

b. Il avait été décidé de longue date que tous partiraient.

c. Acheter ces actions nous a été conseillé par Chiara et Josette.
d. Il nous a été conseillé d'acheter ces actions par Chiara et Josette.

(107) a. Que le président va démissionner se chuchote dans tous les bureaux.
b. Il se chuchote dans tous les bureaux que le président va démissionner.

1.1.1 Les verbes et locutions impersonnels

On peut analyser les verbes ou locutions impersonnels dans la MG en considérant que leur sous-catégorisation initiale comporte un actant grammatical non référentiel, de forme fixée (la forme doit toujours être fixée dans le cas d'actants grammaticaux non référentiels, cf. Chapitre 2). Mais cette analyse va contre l'intuition que le *il* impersonnel a un caractère surfacique en français, et qu'il semble n'être permis que dans les cas où aucun autre actant ne porte la fonction sujet. Cette intuition est confirmée par un exemple comme (108), où un verbe atmosphérique fait partie d'un prédicat complexe causatif.

(108) Chanter faux fait pleuvoir.

Le causatif pour un verbe employé intransitivement (voir Section 1.4.1 pour une analyse du causatif) est analysé comme une redistribution comportant l'insertion d'un sujet (représentant l'argument causateur) et la démotion du sujet de l'infinitif vers la fonction objet. Dans le cas d'un verbe atmosphérique, il devrait y avoir démotion du sujet impersonnel vers la fonction objet, ce qui n'est pas le cas. Une possibilité serait de dire que dans le cas d'un sujet non référentiel, il y a non pas démotion mais effacement. Cependant, cela constituerait le seul cas d'effacement d'un impersonnel.
Du fait de cas comme (108), nous préférons analyser de manière générale le *il* impersonnel comme un sujet devant être introduit, en français, dès lors qu'une sous-catégorisation ne comprend pas de sujet. Il s'agit donc d'un cas particulier de redistribution, qui introduit un actant grammatical non référentiel de forme *il*, et qui ne peut s'appliquer qu'en dernier (c'est-à-dire qu'aucune autre redistribution partielle ne peut la suivre).

Un sujet impersonnel ne peut subir aucune démotion.

Techniquement, on impose pour cela que le métatrait :fonction associé au sujet (le *il* impersonnel, représenté par la constante ?const-sujet[2]) soit non révisable. D'où la classe suivante :

impers→sujet
superclasses: promotion-vers-sujet
action: introduction d'un sujet (**non révisable**)
constantes:
?const-sujet :cat = Cl :fonction = sujet (**non révisable**)

[2] Par convention, on réfère par ?const-*fonction* à la constante qui porte la fonction *fonction* dans la sous-catégorisation **finale**. Les classes de la dimension 3 utilisent cette convention. Donc celles représentant le sujet utilisent la constante ?const-sujet (avec le métatrait :fonction=sujet).

?Vmarq:t:clS = (:num=sing, :gen=masc, :pers=3) (marque morpho. pour la réalisation du clitique)

Ainsi, on associe à /PLEUVOIR/ une sous-catégorisation initiale vide. Cette sous-catégorisation peut subir l'insertion d'un causateur comme sujet (ce qui implique la morpho-syntaxe causative) ce qui correspond à un emploi comme (108); sinon, un sujet impersonnel est introduit. Ce sujet impersonnel ayant une catégorie clitique[3], il ne sera compatible qu'avec les réalisations de sujet comme clitique (cf. Chapitre 5 pour la description des réalisations de fonction), soit le clitique sujet préverbal (ce qui correspond à l'emploi (98a)), soit le clitique inversé, pour une question (*Pleut-il?*).

La classe impers→sujet est également utilisée dans plusieurs redistributions, à la suite d'une démotion du sujet. On distingue dans la MGf l'insertion spontanée d'un impersonnel par une classe impers-spontané, qui hérite de impers→sujet, mais n'ajoute aucune information.

Malheureusement, l'analyse est un peu compliquée par le fait que les autres verbes ou locutions impersonnels, qui comportent un autre actant grammatical que le *il*, n'admettent en général pas le causatif:

(109) a. *Dieu fit faire beau.
b. *La pauvreté faisait falloir partir.
b. *La pénurie faisait y avoir des rixes fréquentes devant les étalages.

Donc pour les verbes impersonnels autres que les atmosphériques, on considère des sous-catégorisations initiale où le sujet impersonnel est déjà présent (cf. Chapitre 3, par exemple la famille *il*Vs0subj, pour *il faut Vinf / que Psubj*).

1.1.2 La redistribution "impersonnel"

On parlera de "l'impersonnel" pour les constructions impersonnelles qui alternent avec des constructions personnelles (=à sujet référentiel) avec une relation de paraphrase. Dans une optique transformationnelle, le sujet de la phrase personnelle correspondante est extraposé et le *il* impersonnel inséré dans la position de sujet (cf. l'accord avec le verbe). Nous appelons ce syntagme "extraposé" *l'actant H* (suivant (Lazard (1994b)).

La relation de paraphrase entre la phrase personnelle et la phrase impersonnelle tend à faire considérer que le sujet de la phrase personnelle reste sujet à l'impersonnel malgré sa position postverbale et la présence d'un autre sujet syntaxique. Cependant d'un point de vue purement syntaxique, outre le fait que l'actant H ne s'accorde pas avec le verbe, on peut déduire des critères retenus pour l'objet direct (cf. Chapitre 5), qu'il s'agit d'un objet[4]. Nous retenons les critères de la pronominalisation, en LE, et pour un objet

[3] Nous proposons au Chapitre 5 un traitement des clitiques comme des affixes flexionnels (Miller, 1991), les clitiques n'y constituent pas des catégories sytnaxiques. La constante ?const-sujet ne sera pas intégrée à la description partielle du schème en construction, héritant de sujet→impers. On renvoie au Chapitre 5 pour plus de détails.

[4] Dans le cadre de LFG, (Grimshaw, 1982) considère également une règle lexicale qui change le sujet en objet. (Abeillé, 1997) retient les critères du *en* et du *pas de*, pour englober le syntagme extraposé de l'impersonnel.

nominal la pronominalisation en *en*, et la négation en *pas de*. Les deux derniers sont vérifiés par l'actant H nominal.

(110) a. Il arrive des gens nouveaux tous les jours.
b. Il en arrive tous les jours.
c. Il n'arrivait jamais de gens nouveaux pour égayer ses journées.

(111) a. Il a été échangé 100000 actions aujourd'hui.
b. Il en a été échangé 100000 aujourd'hui.
c. Il n'a pas été échangé d'actions à risques aujourd'hui.

(112) a. En France il se consomme énormément d'antibiotiques.
b. En France, il s'en consomme énormément.
c. Dans cette enceinte, il ne se consomme pas de substances illicites.

On n'a cependant pas la pronominalisation en LE. Cela a parfois été expliqué par une contrainte plus générale qui empêcherait d'avoir un actant H défini (et donc pronominalisation en LE). En réalité, cela n'est pas formellement exclu, notamment dans le cas cataphorique, où l'article défini ne renvoie pas à un syntagme précédemment défini dans le discours, mais défini par un ajout[5]. Or dans ce cas la pronominalisation en LE reste impossible (113c).

(113) a. *Il arrive l'homme.
b. Il est arrivé la catastrophe redoutée de tous.
c. *Il l'est arrivée.

On n'a en tous cas aucun critère syntaxique de sujet pour l'actant H, qu'il soit défini ou indéfini. Pollock (1986)[6] donne en exemple la possibilité de dépendance non bornée entre l'actant H questionné et son verbe (114b), ce qui n'est pas possible pour un sujet (115b).

(114) a. Il faudrait qu'il vienne plus de linguistes à nos réunions pour que nos théories soient mieux comprises.
b. Combien de linguistes faudrait-il qu'il vienne pour que nos théories soient mieux comprises?

(115) a. Il faudrait que viennent plus de linguistes à nos réunions pour que nos théories soient mieux comprises.
b. *Combien de linguistes faudrait-il que viennent pour que nos théories soient mieux comprises?

Les tests du *en* et de la négation en *pas de* ne sont pas applicables à un actant phrastique. Nous considérons tout de même l'actant H comme phrastique, étant donné que la pronominalisation en LE est souvent possible (elle est possible pour un passif impersonnel, un moyen impersonnel, plus difficile dans le cas d'un actif impersonnel).

(116) a. Comme il vous l'a été dit/déclaré, le colloque n'aura pas lieu.
((Pollock, 1984), cité par (Cinque, 1988a))

[5] (Rivière, 1981) fait une étude complète de la détermination du syntagme postverbal dans les constructions impersonnelles. Elle montre que les contraintes sont plus complexes que la simple dichotomie défini/indéfini. Par exemple, l'article indéfini est inacceptable s'il a une valeur générique. Il s'agirait plutôt de contraintes sur le discours, l'actant H devant être non connu.

[6] Pollock réfute la possibilité d'avoir une coindexation entre le *il* sujet et l'actant H. C'est ce qui lui permet d'expliquer certains contrastes avec les constructions à sujet inversé, où il voit une coindexation avec une catégorie vide préverbale.

b. Que le président allait démissionner, il se l'est répété dans tous les bureaux depuis un mois, et on ne voit rien venir.

Cela semble indiquer que l'impossibilité d'avoir LE dans le cas d'un actant H nominal est une contrainte discursive plus que syntaxique.
On peut remarquer que dans le cas d'un actant H infinitive, le complémenteur *de* apparaît à l'impersonnel (cf. Chapitre 5, une infinitive objet peut comporter un complémenteur *de*, *à* ou bien pas de complémenteur).

Contraintes
L'impersonnel est restreint aux verbes employés intransitivement ((117a) contraste avec (117b,c)), ce dont l'unicité fonctionnelle rend compte dans la MG.

(117) a. Il conduisait moins de femmes que maintenant.
b. *Il conduisait des voitures moins de femmes que maintenant.
c. *Il ne conduisait de voitures que peu de femmes.

Même pour les formes intransitives, de nombreuses contraintes existent. Les verbes transitifs employés intransitivement n'acceptent pas tous l'impersonnel actif (Rivière, 1981). Ainsi on accepte (117a), mais on rejette (118b) malgré (118c).

(118) a. * Il choque beaucoup de paroles.
b. Ses paroles choquent.

Traitement dans la MG
L'impersonnel est traité comme une redistribution faite de la suite de deux redistributions partielles, la démotion du sujet vers la fonction objet, puis l'introduction d'un sujet impersonnel. La démotion du sujet est notée "ce qui est sujet devient objet", qu'il faut interpréter comme une révision du métatrait :fonction : la constante qui porte le métatrait :fonction = sujet, prend le métatrait :fonction = objet). On ne contraint pas la catégorie associée à cette constante, la démotion s'applique donc à un sujet nominal ou phrastique. On impose cependant le complémenteur *de* (métatrait :compl-inf = *de*, qui n'est pertinent que dans le cas d'un objet de type infinitive, cf. Chapitre 5).
On étend cette analyse aux moyens impersonnels (voir Section 1.3) mais pas aux passifs impersonnels, qui sont analysés de manière directe (sans composition) ce que nous justifions infra (Section 1.2).

sujet→objet
superclasses: démotion-du-sujet
action: ce qui est sujet devient objet (:compl-inf = *de*)

impersonnel
superclasses: sujet→objet impers→sujet

Le fait que l'impersonnel ne s'applique pas à une forme transitive est rendu par la contrainte générale sur les redistributions, qui veut que pour un changement de fonction d'un argument, la fonction d'arrivée ne soit pas déjà portée par un autre argument.

1.2 Le passif

Le phénomène le plus "célèbre" mettant en jeu une redistribution des fonctions syntaxiques est certainement le passif, comme paraphrase d'une phrase active.

Pour le français, on analyse traditionnellement la construction passive comme étant en relation avec une phrase active, et comme étant caractérisé par la promotion de l'objet en position sujet.
On peut en effet décomposer le passif personnel comme la conjonction de trois phénomènes, exprimés par rapport au correspondant actif :

(I)- Démotion du sujet
L'argument sujet à l'actif n'est pas le sujet au passif
soit (Ia) l'argument sujet à l'actif n'est pas exprimé
soit (Ib) l'argument sujet à l'actif est exprimé comme un SP en *par*

(II)- Promotion de l'objet
L'objet de la phrase active est sujet au passif

(III)- Morphologie passive
Le verbe prend la forme du participe passé (éventuellement) raccordé au sujet par le verbe copule ETRE, à distinguer de l'auxiliaire de temps.

Par exemple:

(119) a. La solution de facilité est souvent adoptée dans ces cas-là. (Ia)+(II)+(III)
b. Des irrégularités dans le vote ont été dénoncées par les perdants. (Ib)+(II)+(III)

1.2.1 Interprétations du passif et morphologie passive

Le passif peut être ambigu, entre une interprétation processive et une interprétation stative. Une des deux interprétations peut être forcée par différents facteurs, outre la sémantique lexicale du verbe lui-même. Par exemple l'interprétation processive peut être forcée par un temps parfait, la présence d'adverbe agentifs. Au contraire la présence d'ajout de durée favorise l'interprétation stative.

(120) a. La porte est fermée.
b. La porte (est + demeure) fermée depuis 3 heures.
c. La porte fut fermée violemment.

Or si l'interprétation processive n'est possible qu'avec la copule comme "auxiliaire du passif", l'interprétation stative s'apparente à une construction à attribut du sujet, et permet comme auxiliaires les verbes d'état en général.

(121) La porte est + semble + demeure + reste fermée.

Donc pour la "morphologie passive", il faut élargir la définition.

(III)- Morphologie passive
Le verbe prend la forme du participe passé (éventuellement) raccordé au sujet par le verbe copule ETRE, ou un verbe d'état.

En outre "l'auxiliaire du passif" n'est pas toujours présent. En effet dans le contexte d'une participiale présente, il est optionnel, ce qu'illustrent les deux possibilités (122a,b) (cf. Chapitre 3, la section sur l'hypothèse inaccusative). A noter qu'il s'agit bien d'un emploi d'ACQUERIR comme forme verbale passive, ce que montre la présence de l'agt-objet interprété comme un agent.

(122) a. Les biens étant ultérieurement acquis par les époux sont considérés comme communs.
b. Les biens ultérieurement acquis par les époux sont considérés comme communs.

1.2.2 Prise en compte du passif impersonnel

L'analyse du passif comme la conjonction des propriétés I + II + III pose problème, étant donné qu'il existe des cas combinant (I) et (III), mais sans (II), les verbes en question n'ayant jamais d'emploi avec un objet à promouvoir:

(123) a. Il a été dansé toute la nuit.
b. Il a surtout été discuté du problème de l'emploi par les ministres concernés.
c. Il a été procédé à une enquête par les services compétents.

Pour tenir compte de ces cas, nous réduisons la définition intensionnelle du passif, comme la conjonction de (I) et (III) seulement (Gaatone (1993) donne ces deux critères comme définition opérationnelle d'une phrase passive).
Mais en français, la combinaison (I) + (III) donne une sous-catégorisation sans sujet, donc il ne peut s'agir d'une redistribution complète. Elle peut être complétée par l'insertion d'un *il* impersonnel en position sujet (cf. Section 1.1). Nous allons désigner cette propriété par (IIb), la promotion de l'objet devenant (IIa) :

(IIa)- Promotion de l'objet
(IIb)- Insertion de *il*.

(123a) et respectivement (123b,c) sont analysés comme la conjonction des phénomènes (Ia)+(IIb)+(III) et respectivement (Ib)+(IIb)+(III).
On peut remarquer qu'un argument en faveur de la séparation des deux phénomènes démotion du sujet et promotion de l'objet est le fait que certaines langues présentent des cas de "démotion spontanée" du sujet, c'est-à-dire où le sujet est démis sans insertion d'autre sujet et où l'objet reste en place. Lamiroy (1993) cite l'hébreu, et Comrie (1977) fait une étude approfondie du gallois[7]. En français, un sujet syntaxique doit être présent, et on a le *il* impersonnel.

Les cas comme (123), de verbes intransitifs admettant une construction impersonnelle passive sont extrêmement rares, avec une difficulté plus grande pour les intransitifs

[7] Dans cette langue, on peut avoir une construction avec un verbe transitif à la forme passive auquel est associé un syntagme, clairement objet direct quand il s'agit d'un pronom personnel, qui est alors à l'accusatif :

Fe'i lladdwyd (gan ddraig)
3p.acc a-été-tué (par dragon)
"Il a été tué par un dragon"

stricts. Pour les transitifs, le passif impersonnel est en revanche généralisé. Aussi, (124c) est traditionnellement analysé comme l'impersonnel de (124b), lui-même passif de (124a).

(124) a. Nos courtiers ont échangé 10000 actions aujourd'hui.
b. 100000 actions ont été échangées aujourd'hui.
c. Il a été échangé 100000 actions aujourd'hui.

Dans cette optique, (123) reste inexpliqué. Pour *Il a été procédé à une enquête*, on n'a pas de passif personnel correspondant et donc on n'a pas de phrase de départ où *une enquête* est sujet. Même si on pose une relation avec *on a procédé à une enquête*, on n'a pas extraposition d'un sujet, mais effacement de celui-ci avec en plus une morphologie passive inexpliquée. Nous préférons donc donner une analyse unifiée du passif impersonnel, indépendantes de la présence éventuelle d'un objet. On analyse ainsi (124c) comme (Ia)+(IIb)+(III), exactement comme pour (123a).

En ce qui concerne la propriété (I), on a distingué deux cas selon que l'agent était exprimé (comme un SP en *par*) ou pas. On considère donc une fonction supplémentaire, agt-objet, par rapport à celles utilisées pour décrire la sous-catégorisation initiale.
Il existe des arguments en faveur de la fusion des cas (Ia) et (Ib), c'est-à-dire pour considérer que c'est une propriété de la fonction agt-objet de pouvoir être omise. Premièrement, la démotion du sujet vers la fonction agt-objet a pour effet de rendre très oblique l'argument correspondant. Or il est en général possible d'omettre une fonction oblique. Deuxièmement, il existe un autre contexte faisant apparaître la fonction agt-objet, la construction caus-agt (où le sujet d'un verbe est démis agt-objet, la fonction sujet étant alors prise par un arguments causateur supplémentaire). On verra Section 1.5.1.3 que l'intéraction du réfléchi et des constructions causatives donne un argument décisif pour considérer que les cas sans agent peuvent être analysés comme comportant un agt-objet omis.

Le passif est alors analysé comme (I) + (II) + (III), avec :

(I)- Démotion du sujet vers agt-objet

(II) a. Promotion de l'objet
b. Insertion de *il.*

(III)- Morphologie passive
Le verbe prend la forme du participe passé (éventuellement) raccordé au sujet par le verbe copule ETRE, ou un verbe d'état[8].

La seule caractéristique stable du passif semble être la démotion vers la fonction agt-objet. Dans cette optique, on considérera que la construction caus-agt est un passif, où la morphologie du verbe est une morphologie causative (Section 1.5.1.3). Cette

[8] Nous remercions Laurence Danlos pour la remarque suivante : la possibilité d'avoir un "auxiliaire du passif" autre que la copule est limitée au passif personnel. Le passif impersonnel ne permet en effet que l'interprétation processive, qui impose la copule. Actuellement, des traits manquent dans la MGf pour gérer cette contrainte. Il faut, parmi tous les verbes qui sélectionnent un schème projetant un verbe "nu" (sans complément, et ne projetant pas une phrase), distinguer la copule des autres, et ne permettre dans les arbres pour le passif impersonnel la substituion que de la copule.

caractéristique n'en est cependant pas une, étant donné l'existence d'un passif en *de* possible pour certains verbes.

1.2.3 Passif en *de*

Il existe effectivement une réalisation alternative de l'agent démis, en *de*. Le passif en *de* est possible pour certaines classes sémantiques de verbes (notamment les verbes psychologiques, mais également certains statifs). Il force en général l'interprétation stative, mais pas toujours (cf. l'exemple (125b), clairement processive, et (125c), où le parfait favorise l'interprétation processive).

(125) a. Paul est admiré/envié/aimé de tout le monde ici.
b. Le turbot a été suivi/accompagné de morilles.
c. Paul a été surpris de la venue de Chiara.

Le passif en *de* n'admet pas le passif impersonnel.

(126) Il a été admiré peu d'écrivains (par + *de) ses contemporains.

Il est intéressant de remarquer que l'agent de ce passif en *de* porte la fonction gén-objet. Il peut en effet être pronominalisé par *en* et peut être relativisé ou pseudo-clivé en *dont*.

(127) a. Paul en est admiré/aimé/envié.
b. Les morilles dont le turbot a été accompagné étaient exquises.
b. Ce dont Paul a été surpris, c'est de la venue de Chiara, pas de l'absence de Marco.

Il faut considérer une démotion su sujet vers la fonction gén-objet. Nous ne traitons cependant pas ce passif dans la MGf actuelle. Nous parlerons de passif tout court pour le passif mettant en jeu une démotion vers la fonction agt-objet.

1.2.4 Passif et actants phrastiques

Le passif personnel peut mettre en jeu la promotion d'un objet phrastique (complétive ou infinitive), même si le passif impersonnel est plus courant dans ce cas. Nous redonnons les exemples déjà fournis pour l'impersonnel.

(128) a. Que tous partiraient avait été décidé de longue date.
b. Il avait été décidé de longue date que tous partiraient.
c. Acheter ces actions nous a été conseillé par Chiara et Josette.
d. Il nous a été conseillé d'acheter ces actions par Chiara et Josette.

Un grand nombre de verbes à objet phrastique n'admettent cependant pas le passif. Il est d'une manière générale plus fréquemment possible pour les verbes à objet infinitive avec complémenteur *de*. Mais il n'est pas formellement à exclure pour les verbes dont l'infinitive est sans complémenteur (129). En revanche dans le cas d'un complémenteur *à*, la promotion de l'objet semble systématiquement bloquée.

(129) Partir est ardemment désiré/*?voulu par cette jeunesse en mal de grands espaces.

En particulier les verbes prenant une infinitive objet avec complémenteur *à* ne l'admettent pas.

Inversement, un verbe transitif à sujet phrastique (complétive ou infinitive) peut apparaître au passif (130a,b), mais il s'agit alors d'un passif en *de*. L'actant phrastique démis a en effet des propriétés de gén-objet phrastique (cf. Chapitre 5), en particulier la réalisation en *de ce que* (130c). La pronominalisation en *en* nous semble également possible (130d), la pseudo-clivée en *dont* est plus douteuse (130e).

(130) a. Marie est angoissée de partir.
b. Marie est angoissée que Paul ne soit pas rentré.
c. Marie est angoissée de ce que Paul ne soit pas rentré.
d. Marie en a été angoissée pendant toute la soirée.
e. ? Ce dont Marie est angoissée c'est que Paul ne soit pas rentré.

1.2.5 Interactions avec les autres redistributions

Il faut éviter qu'à la suite d'un passif personnel s'applique l'impersonnel, puisqu'on représente le passif impersonnel comme une redistribution directe (directement). On peut l'éviter par une contrainte de cooccurrence négative.
Inversement le passif (impersonnel) (qui ne requiert pas une forme transitive) ne doit pas s'appliquer sur un moyen, ou même un passif. On peut représenter cela par une contrainte générale, qui est que la démotion du sujet vers agt-objet ne peut s'appliquer qu'à un sujet initial.

Seul un sujet initial peut subir sujet→agt-objet.

1.2.6 Description des classes utilisées

Nous donnons avant la description des classes utilisées pour représenter le passif dans la MGf, un exemple de schème que ces classes permettent de décrire (que l'on a déjà mentionné également au Chapitre 3). Pour gérer la variation possible sur l'auxiliaire, on utilise une position à substituer[9].

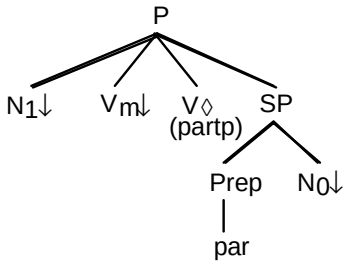


Figure 55. Schème pour un transitif strict au passif.

- La propriété (I) est représenté par la classe sujet→agt-objet, où le metatrait (:fonction sujet), pour la constante qui le porte, est changé en (:fonction agt-objet).

[9] Pour rester en accord avec le PCPA il faudrait adjoindre l'auxiliaire (il apporte une constribution de temps et aspect, voire dans le cas d'un verbe d'état une sémantique supplémentaire, mais toujours unaire). La solution par substitution est adoptée car plus efficace. Il faut considérer ces cas de substitution à part, par rapport à ceux représentant des positions argumentales. On aura un cas similaire pour le verbe causatif dans le cas d'un complexe causatif (cf. infra).

On impose que le sujet soit un sujet initial (cf. Section précédente) et que le trait :lie-réfl soit négatif. Cela permettra de bloquer l'application du passif, au sens général incluant la construction caus-agt, (cf. Section 1.4.1.1), sur un verbe portant un réfléchi (cf. Section 1.5.1.3 pour les contraintes sur le réfléchi). De plus l'agt-objet est nominal (cf. supra dans le cas d'un agent phrastique, il s'agit du passif en *de*).

sujet→agt-objet
superclasses: démotion-du-sujet
action: ce qui est sujet (et tel que :fonction-initiale=sujet, :lie-réfl=-, :cat = N) devient agt-objet

- (IIa) est représentée par objet→sujet, où le metatrait (:fonction objet), pour la constante qui le porte, est changé en (:fonction sujet). Pour (IIb), on a déjà vu la classe impers→sujet.

objet→sujet
superclasses: promotion-vers-sujet
action: ce qui est objet devient sujet

- Pour la propriété (III), on est obligé de distinguer deux cas, selon que la copule est présente ou pas.
  On considère la classe morpho-marquée, qui rassemble tous les cas de morphologies marquées introduites par redistribution, mais ne contient pas d'informations locales[10]. En particulier, une classe morpho-passive en hérite, qui impose le mode participe à ?Vanc, l'accord du participe (nommé par ?Vanc) avec le sujet final (désigné par ?const-sujet) en genre et nombre. ?Vanc ne peut pas porter de clitiques (donc en particulier la négation). On marque également ?Vanc d'un métatrait :morpho-passive=+ pour l'interaction avec le causatif (cf. Section 1.4.1.3).

morpho-passive	
superclasses: morpho-marquée	
constantes:	
?Vanc:mode = partp	
?Vanc:num = const-sujet.num	(accord sujet participe)
?Vanc:gen = const-sujet.gen	
?Vanc:neg = -	
?Vanc:morpho-passive = +	

Dans la classe morpho-passive-avec-cop, on spécifie que le verbe tête n'est pas l'ancre, mais la copule (cela est noté tête→?Vpass, cela signifie que la valeur du champ tête est révisée et prend la valeur "?Vpass"). ?Vpass est placé à gauche de ?Vanc[11].

[10] Les classes telles que morpho-marquée, qui n'existent que pour rassembler les sous-classes qui en héritent, sont notées dans les figures dans un cadre non arrondi.

[11] Normalement il faudrait stocker la valeur de ce qui était tête au moment de l'application du passif, et placer la copule à gauche de cette ex-tête, pour traiter par exemple en italien un passif qui s'applique à un causatif.

morpho-passive-avec-cop	
superclasses: morpho-passive	
tête→?Vpass	
constantes:	description:
?Vpass:cat = V	

Dans la classe morpho-passive-sans-cop au contraire, on ne modifie pas la valeur du champ tête (c'est-à-dire que l'on introduit pas de copule). Cette classe doit être exclusivement utilisée dans le contexte d'une participiale sans copule (on traite les participiales en dimension 3, comme des réalisations particulières du sujet, qui imposent des contraintes de mode, voir Chapitre 5), d'où une contrainte de cooccurrence positive avec la classe participiale-sans-cop.

morpho-passive-sans-cop
superclasses: morpho-passive
contraintes: ⇒participiale-sans-cop

Avec ces redistributions partielles, on peut construire trois redistributions, passif-personnel-avec-cop, passif-personnel-sans-cop et passif-impersonnel (qui impose la présence d'un auxiliaire de passif).
Nous donnons Figure 56 le schéma des classes intervenant pour la définition du passif, dans la MGf.

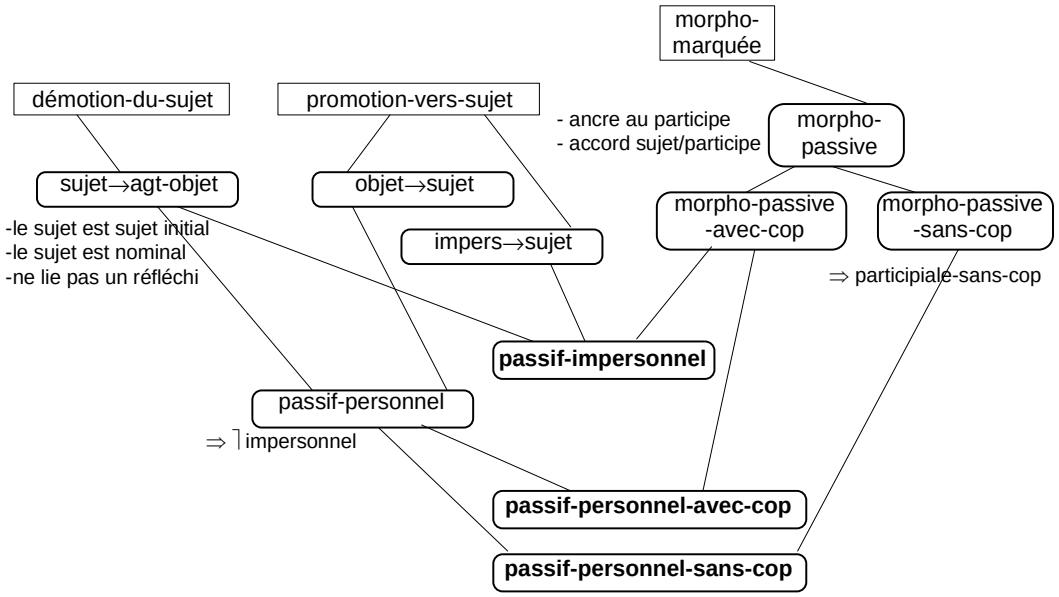


Figure 56. Les classes pour représenter le passif dans la MGf (redistributions terminales en gras)

1.3 Le moyen

Il existe en français d'autres constructions auxquelles on confère généralement "un sens passif" : les constructions moyennes. Elles sont en relation de paraphrase (partielle) avec une phrase active comprenant un objet. Elles ont les caractéristiques syntaxiques suivantes : le verbe porte un clitique traditionnellement appelé pronom réfléchi, dont la forme varie en nombre et personne, et que nous désignerons par SE; l'argument qui est sujet de la phrase active correspondante n'est pas exprimé; l'objet de la phrase active est sujet de la construction moyenne (131b), ou bien il est également objet de la phrase moyenne et le sujet est impersonnel (131c).

(131) a. Les Français consomment chaque année plus de 1000 tonnes de fromage.
b. Plus de 1000 tonnes de fromage se consomment chaque année.
c. Il se consomme plus de 1000 tonnes de fromage chaque année.

L'objet de la phrase active peut également être une complétive (132a), et l'on a alors pour la construction moyenne un sujet phrastique (132b), ou bien un sujet impersonnel (132c). Le moyen ne semble pas possible pour un objet infinitival[12].

(132) a. On chuchote dans tous les bureaux que le président va démissionner.
b. Que le président va démissionner se chuchote dans tous les bureaux.
c. Il se chuchote dans tous les bureaux que le président va démissionner.

Ces constructions partagent la propriété d'apparaître avec SE avec les vrais réfléchis (*Juliette se lave*) et les réfléchis intrinsèques (comme *s'immiscer*). Nous reportons la description de SE à la section 1.4 sur les réflexivation.
Les caractéristiques syntaxiques énoncées recouvrent des constructions dont le sens varie nettement. Ruwet (1972) distingue les constructions moyennes des constructions neutres. Dans la construction moyenne (comme en (131b,c)), le sujet initial n'est pas exprimé mais il est interprétable, comme c'est le cas pour le passif (Rivière (1981) parle de "l'impersonnel pronominal à sens passif", et Lamiroy (1993) parle directement de passif pour les constructions moyennes décausatives). Dans les constructions neutres (par exemple (133)), l'agent n'est ni réalisé syntaxiquement, ni interprétable sémantiquement (Lamiroy (1993) parle de tour autocausatif pour le neutre qu'elle appelle SE-moyen):

(133) Avec le temps, le bois de cette vieille poutre s'est fendu.

La construction moyenne est très productive, avec des contraintes syntaxiques et sémantiques générales. Zribi-Hertz (1982) distingue deux interprétations pour le moyen : il permet de prédiquer une propriété générique du sujet (de surface) (134a), ou bien il a une interprétation de procès (134b). Dans le premier cas il est incompatible avec un aspect ponctuel.

(134) Les chênes se plantent à basse altitude.
Il se mange sûrement des gâteaux au premier étage en ce moment.

Ruwet (1972) souligne également que les verbes à sujet initial non agentif ne peuvent pas entrer dans une construction moyenne:

[12] Zibri-Hertz (1982) par exemple, ne fournit pas d'exemples avec de verbe à infinitive permettant le moyen.

(135) a. Le gouvernement comporte 33 ministres.
b. *33 ministres se comportent.

Wehrli (1986), citant Burzio (1981) indique également que le neutre peut apparaître dans un ajout, alors que le moyen ne le peut pas[13].

(136) a. La fenêtre est tombée sans se casser.
b. *La partie s'est terminée sans se gagner.

La construction neutre apparaît plutôt comme une propriété lexicale du verbe, et a un sens inchoatif. De la même manière que certains verbes sont intrinsèquement pronominaux, d'autres admettent à la fois la construction transitive et la construction neutre.

Nous traitons de la construction moyenne étant donné son caractère productif. La construction neutre étant beaucoup plus idiosyncratique, nous considérons qu'elle doit être déclarée comme une sous-catégorisation initiale dans le lexique (cf. introduction de ce chapitre, la discussion sur la distinction redistribution / alternance). Ainsi BRISER serait le lemme de plusieurs lexèmes, un /BRISER/ avec sujet et objet initial, et un /SE-BRISER/ avec sujet initial et morphologie réfléchie (pour *la glace s'est brisée*). On peut concevoir un module supplémentaire pour le regroupement de différentes sous-catégorisations initiales. On peut remarquer que le transitif peut lui-même entrer dans une construction moyenne (clairement en (137), il ne s'agit pas du neutre).

(137) Ce pays est si riche que dans les trains, les vitres se brisent avec un marteau en or.

Du fait de sa ressemblance sémantique et syntaxique avec le passif (la démotion du sujet, la promotion de l'objet en sujet), certains auteurs regroupent sous le nom de passif les constructions moyennes (C'est le cas par exemple de Lamiroy (1993) qui parle de deux opérateurs passivant : *être V-é* et *se V*). Nous préférons garder comme caractéristique du passif la démotion vers la fonction agt-objet.
En effet le moyen est syntaxiquement bien distinct du passif, indépendamment de la différence de morphologie. Premièrement l'argument démis ne peut jamais être exprimé (comme un agt-objet ou autre).

(138) Les BMW, ça s'achète (E + *par les frimeurs).

Comme on a analysé, pour le passif, le cas sans agent comme un cas particulier de réalisation de la fonction agt-objet, on ne peut pas considérer que pour un moyen et pour un passif sans agent, l'agent a subi la même démotion.
Deuxièmement, les cas de moyens pour un intransitif sont extrêmement marginaux, et le sont plus que pour le passif (pour quelques intransitifs, le passif impersonnel est possible (cf. Section 1.2)).

[13] Wehrli (1986) argue également que le causatif sur un infinitif au moyen n'est pas possible (ii), alors qu'il l'est pour le neutre (i). Nous pensons que (ii) est mauvais du fait d'incompatibilité entre le moyen et l'aspect ponctuel. Nous citons infra dans la section sur les causatifs des exemples acceptables parallèles à (ii).
(i) Le chef a fait se réunir l'équipe.
(ii) *J'ai fait se nettoyer ma veste. (d'après Gross, 1968)

(139) a. Il s'est discuté/parlé longuement de la conception médiévale du monde.
b. Il se réfléchit à de drôles de choses (Gross, 1975, cité par Cinque, 1988a)

Donc on a deux solutions. Soit on considère que les cas de moyen impersonnel résultent de la composition de deux redistributions, le moyen (toujours personnel) et l'impersonnel (comme cela est fait traditionnellement, mais contrairement à ce que l'on a choisi pour le passif impersonnel). Soit on considère comme pour le passif impersonnel, que le moyen impersonnel n'est pas dérivé du moyen, mais est un moyen où la promotion de l'objet n'a pas lieu.
On choisit la première solution, comme plus adaptée aux données. Le moyen est caractérisé dans la MGf par la non expression du sujet initial (représenté par la démotion du sujet vers la fonction "∅") et la promotion de l'objet.

1.3.1 Interactions avec les autres redistributions

Il n'y a aucune interaction entre passif et moyen. Le moyen s'applique à un transitif or tous les passifs sont intransitifs, sauf le passif impersonnel. Mais comme on a déjà imposé que le *il* impersonnel soit le sujet final, on bloque bien ce cas.
On a déjà bloqué le passif (impersonnel) sur un moyen, en imposant que la démotion sujet→agt-objet ne s'applique qu'à un sujet initial.
Le moyen interagit également avec le causatif ce qui sera détaillé plus bas.

1.3.2 Description des classes utilisées

Pour la redistribution moyenne, on définit une classe terminale de la dimension 2, moyen, qui hérite d'une redistribution partielle sujet→∅, comme un cas particulier de démotion du sujet, puisque le sujet est effacé. Techniquement, on utilise la fonction spéciale "∅"[14], qui a pour particularité de ne pas être réalisée grammaticalement.
Nous donnons Figure 57 le schma des classes intervenant dans la définition de la redistribution "moyen", dans la MGf.

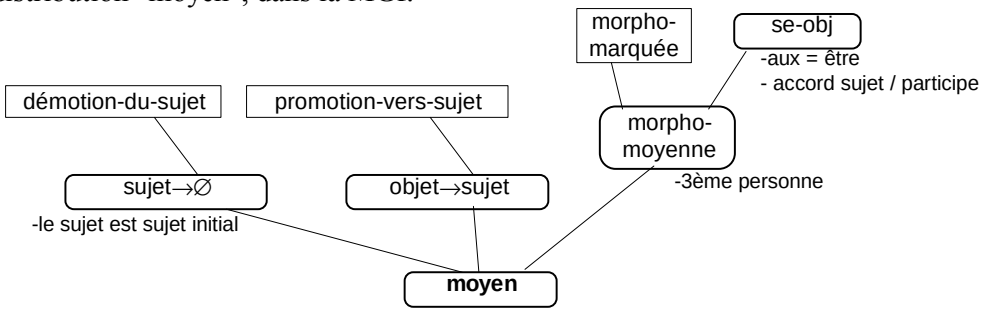


Figure 57. Les classes de la MGf pour le moyen (personnel)

Outre sujet→∅, la classe moyen hérite également de objet→sujet, déjà utilisée pour le passif personnel. La morphologie moyenne hérite de la classe SE-obj, que l'on décrit Section 1.4, qui introduit la morphologie réfléchie (l'auxiliaire de temps est ETRE, le verbe s'accorde en personne avec le réfléchi, le participe s'accorde en genre et nombre

[14] C'est l'équivalent du Φ en LFG. Cependant (Grimshaw, 82) mentione que la fonction Φ est soumise à la même contrainte d'unicité que les autres fonctions. Nous pensons que l'effacement de plusieurs actants n'est pas a priori à rejeter, donc pour nous plusieurs actants peuvent porter la fonction "∅".

avec le réfléchi) et qui est également utilisée pour les vrais réfléchis et certains réfléchis intrinsèques. On choisit de limiter la couverture de la grammaire en limitant le moyen à la 3ème personne (cf. on peut construire des exemples à la deuxième personne, en utilisant par exemple la métonymie comme dans *tu t'es bien vendu à la foire du livre* (Koenig et Jurafski, 1994)).

1.4 Le causatif

1.4.1 Analyse des données

Les constructions causatives en français sont bien connues, avec notamment l'étude de Kayne (1975), pour mettre en jeu un type particulier de complémentation infinitive. Nous suivons Abeillé et al. (1996b, 1997, voir également les références citées) qui distinguent, dans le cadre d'une analyse en HPSG, deux types de construction, dites à contrôle et à composition (analyse reprise également dans le cadre de LFG, par Zaenen et Dalrymple (1996)).

1.4.1.1 Construction à composition versus construction à contrôle

Dans la construction à composition, le verbe causateur et son infinitif se comportent comme un seul prédicat syntaxique, dit "prédicat complexe", dont la liste de sous-catégorisation comprend la réalisation des arguments de l'infinitif et un argument supplémentaire pour le *causateur* (Cette fusion de sous-catégorisations correspond à ce qui est appelé *causative clause union* en Grammaire Relationnelle). Ce causateur porte la fonction sujet, et le sujet initial de l'infinitif (dorénavant le *causé*) porte alors une fonction qui dépend de la transitivité de l'infinitif:

- si l'infinitif seul est intransitif, le causé est objet (on parlera de la construction caus-i)

(140) Paul fait rire Jean.

- si l'infinitif seul est transitif, le causé est soit un a1-objet, soit un agt-objet (ce qui inclut le cas où le causé est omis, cf. Section 1.2). On parlera des constructions caus-à et caus-agt (comme des sous-types de la construction à composition).

(141) a. Paul fait balayer la cour aux recrues.
b. Paul fait balayer la cour (E + par les) recrues.

Pour alléger la rédaction, on note [trans+] le cas d'un infinitif transitif, et [trans-] le cas d'un infinitif intransitif. Le *causateur* est l'argument supplémentaire agentif, le *causé* est le sujet de l'infinitif et on désigne par C(Vinf) les arguments de l'infinitif autres que le causé. L'analyse à composition est justifiée principalement par la position des clitiques devant le verbe causateur, non seulement pour la cliticisation du causé, mais aussi pour celle des C(Vinf).

(142) a. Paul la fait balayer (E + par les recrues).
b. Paul la fait balayer aux recrues.
c. Paul la leur fait balayer.
d. Paul le fait rire.

Les autres arguments en faveur de la formation d'un prédicat complexe sont :

- le fait que l'infinitif ne puisse pas être nié :

(143) a. *La vieillesse fait ne pas aimer le bruit aux recrues.
b. *Son sourire a fait ne pas enregistrer la sortie (E + par le surveillant).
c. *La rancoeur a fait ne pas céder à Paul.

- le fait que le verbe causateur et l'infinitif se comportent comme un seul verbe dans les constructions en *Adj à Vinf*. Dans cette construction, l'objet du *Vinf* n'est pas réalisé au niveau de son verbe, mais est contrôlé par le nom sur lequel prédique l'*Adj* :

(144) un problème facile à résoudre

En français, il ne peut pas y avoir d'enchâssement entre le *à* et l'infinitif dont l'objet n'est pas réalisé localement (145), or une séquence FAIRE *Vinf* est possible dans ce contexte (146a,b).

(145) *un problème facile à essayer de résoudre, mais plus difficile à résoudre.

(146) a. une loi impossible à faire respecter à ces adolescents en révolte
b. une loi impossible à leur faire respecter

Ces mêmes critères amènent à considérer une construction concurrente, pour laquelle le causé prend en général la forme d'un clitique accusatif sur FAIRE, possible même dans le cas [trans+] ((147a), (148a)). En effet, le placement des clitiques, la négation de l'infinitif et le tour en *Adj à Vinf* donnent des résultats différents dans ce cas[15]:

- Les C(Vinf) cliticisés se placent devant l'infinitif ((147b) et (148b,c)), et ne peuvent pas se placer sur FAIRE. C'est facilement compréhensible dans les cas qui donneraient lieu à des séquences de clitiques impossibles sur FAIRE (comme les deux accusatifs en (147c), mais c'est aussi impossible pour une séquence a priori permise de clitiques sur FAIRE, comme en (148d)).

(147) a. Paul les fait balayer la cour.
b. Paul les fait la balayer.
c. *Paul les la fait balayer.

(148) a. Le médiateur les fait rendre la clé au gardien.
b. Le médiateur les fait la rendre au gardien.
c. Le médiateur les fait lui rendre la clé.
d. *Le médiateur les lui fait rendre la clé.

- La négation de l'infinitif est alors possible (ou du moins meilleure):

(149) ?La jalousie le faisait ne pas voir la détresse qu'il engendrait.

- Le tour *Adj à Vinf* n'est pas possible ((150) contraste très nettement avec (146b)):

(150) *une loi impossible à les faire respecter

L'analyse à contrôle dans le cas [trans+] est restreinte aux cas où le causé est clitique (le causé ne peut jamais se réaliser comme un SN direct, ce qui suit la contrainte générale

[15] (Abeillé et al., 1996b) retiennent également un test de coordination pour justifier une différence entre les deux constructions.

du français de ne pas admettre deux SN objets). Ainsi, la distinction entre structures à composition et à contrôle est nette dans le cas [trans+].
Dans le cas [trans-], la distinction est moins évidente. En effet, dans le cas d'un intransitif strict comme dans (151a), les deux analyses sont a priori possibles (cf. Abeillé et al. (1997)). Le seul critère applicable pour trancher est la négation de l'infinitif. L'acceptabilité de (151b) montre que l'analyse à contrôle est possible, mais ne montre pas que l'analyse à composition est exclue. (151a) peut avoir deux structures, dont seule une admet la négation de l'infinitif.

(151) a. Justine le fait pleurer.
b. Seule la peur du ridicule l'a fait ne pas pleurer.

Pour un intransitif avec complément cliticisable, le placement du clitique doit fournir un test. Les deux positions (sur FAIRE et sur l'infinitif) semblent possibles, mais pas forcément pour les mêmes locuteurs, avec plus fréquemment un placement sur faire. Il faut retenir a priori les deux structures dans le cas où aucun C(Vinf) n'est pas cliticisé.

(152) a. Paul les a fait parler du suicide. (Contrôle/Composition)
b. Paul les a fait en parler. (Contrôle)
c. Paul les en a fait parler. (Composition)

Il existe des travaux qui mettent à part le cas des clitiques non marqués pour la personne (*en* et *y*). Pour Rouveret et Vergnaud (1980), Miller (1991) *en* et *y* peuvent se placer sur l'infinitif, même dans le cadre de ce que nous appelons ici la structure à composition. Ainsi pour Rouveret et Vergnaud, (153a) est acceptable (ainsi que (153b) où le clitique est sur FAIRE). Pour Miller, (153) et toute la série (154) sont acceptables. D'après les tests que nous avons pu faire, la cliticisation sur l'infinitif est nettement moins acceptée que sur *faire*. Mais on verra à la section suivante, qu'il faut indépendamment de ces données accepter la cliticisation sur l'infinitif de certains clitiques, même dans le cas à composition.

(153) a. Paul fera en acheter trois à Marie. (Rouveret et Vergnaud, 1980)
b. Paul en fera acheter trois à Marie.

(154) a. Paul fera y aller Marie. (Miller, 1991)
b. Paul y fera aller Marie.
c. Paul a fait en parler Marie.
d. Paul en a fait parler Marie.

Nous ne traitons pas ces cas de cliticisation sur l'infinitif de *en* et *y*. On peut analyser (154a,c) comme un cas de structure à contrôle (et on doit donc admettre que dans le cas [trans-] contrôle, le causé n'est pas forcément clitique). On peut d'autre part remarquer que cette hypothèse du placement "libre" (sur faire ou l'infinitif) de *en* et *y*, invaliderait l'analyse de (152b) comme un cas de structure à contrôle. On peut cependant avoir recours au test de la négation pour trancher. (155a) est inacceptable, ce qui est le résultat attendu (on a cliticisation du *en* sur FAIRE, et donc forcément composition). (155b) est douteux, mais en tous cas meilleur, ce qui est quand même un indice qu'on a deux structures différentes selon que le *en* est sur FAIRE ou sur l'infinitif.

(155) a. *La peur des représailles les en a fait ne pas parler pendant des années.
b. ?La peur des représailles les a fait ne pas en parler pendant des années.

1.4.1.2 Cas particulier des clitiques idiomatiques et réfléchis

Nous avons vu que le placement des clitiques est un critère essentiel de caractérisation des constructions causatives. Or tous les clitiques ne se comportent pas comme il a été illustré ci-dessus. Les clitiques idiomatiques et les réfléchis (au sens large, comprenant les réfléchis intrinsèques, les vrais réfléchis et le moyen) ont la particularité de devoir ou pouvoir apparaître sur l'infinitif, même dans le cadre de la construction à composition (Abeillé et al. 1997).
Pour les clitiques idiomatiques, la série (156) montre que le clitique ne peut pas se placer sur FAIRE (156b); il doit se placer sur l'infinitif (156c). (On remarque que caus-à est possible mais pas caus-agt (156d); or *en vouloir à* n'est pas passivable, cela concorde avec l'analyse de caus-agt comme un passif).

(156) a. Paul en veut à toute la société.
b. *Sa réclusion en fait vouloir à toute la société à Paul.
c. Sa réclusion fait en vouloir à toute la société à Paul. (adapté de (Abeillé et al., 1998))
d. *La réclusion fait en vouloir à toute la société par Paul.

Pour les réfléchis, on peut trouver sur le verbe infinitif un vrai réfléchi (ou réciproque) (157a), un moyen (157b) ou un réfléchi intrinsèque (157c) (voir Section 1.5.1.3 pour les contraintes sur l'interaction réfléchi/causatif).

(157) a. Paul et Virginie font se rencontrer leurs parents.
b. La crise fait se vendre les videos new age par milliers.
c. La chaleur a fait s'évanouir trois personnes.

De plus, dès lors qu'un clitique est sur l'infinitif, alors tous les C(Vinf) clitiques doivent être portés par l'infinitif (mais pas le causé, cf. le contraste (158d)/(158e)).

(158) a. La direction a fait s'acheter le dernier macintosh à tous les élèves.
b. La direction leur a fait s'acheter le dernier macintosh.
c. *La direction le leur a fait s'acheter.
d. La direction leur a fait se l'acheter.
e. *La direction a fait se le leur acheter.

Abeillé et al. (1998) fournissent une solution dans le cadre d'HPSG (et c'est à notre connaissance la seule). Cette solution repose sur une distinction entre deux listes de sous-catégorisation, une liste pour tous les actants grammaticaux (la liste ARG-ST), et une liste pour les compléments (non sujet) non cliticisés (la liste COMPS). Les éléments de la liste COMPS doivent être réalisés par des syntagmes. Le complexe causatif comprend dans son ARG-ST la liste COMPS de l'infinitif (et pas toute l'ARG-ST, comme c'est le cas pour un auxiliaire de temps par exemple), donc seuls les compléments non clitiques de l'infinitif peuvent se placer sur FAIRE. Cela revient à distinguer deux types de C(Vinf), les syntagmes, qui sont partagés avec le complexe causatif, et les clitiques qui ne le sont pas. En d'autres termes, les C(Vinf) ont pour gouverneur syntaxique FAIRE, mais dans le cas de la présence d'un clitique idiomatique ou réfléchi sur l'infinitif, ils ont pour hôte (au sens hôte d'un clitique) l'infinitif.
Dans notre architecture, on ne différencie pas au sein de la liste de sous-catégorisation les réalisations clitiques des réalisations syntagmatiques, mais on peut utiliser un

metatrait pour représenter l'hôte d'un actant grammatical dans le cas de sa réalisation clitique.
Dans le cas d'un clitique idiomatique (qui peut être réfléchi) ou d'un réfléchi, on spécifie que tous les actants grammaticaux, sauf le sujet, ont pour hôte le verbe qui est la tête *au moment où est spécifiée la présence du clitique idiomatique ou réfléchi* (cf. infra Section 1.5.2 pour plus de précision). Pour la réalisation clitique d'un actant, si un hôte est spécifié, la réalisation se fait sur l'hôte; sinon elle se fait sur le verbe tête après toutes les redistributions (qui est le verbe causatif dans le cas d'un complexe causatif).

1.4.1.3 Contraintes et interaction avec les autres redistributions

Il existe des constraintes complexes sur la combinatoire des différentes réalisations des actants grammaticaux du prédicat complexe. Mais, comme Tasmowski (1984), Miller (1991) et à la différence de Kayne (1975), Rouveret et Vergnaud (1980) nous pensons qu'il faut chercher une caractérisation discursive et non pas syntaxique de ces contraintes.

(159) a. *On lui fera téléphoner Pierre.
b. On lui fera transmettre le message au coursier.
c. On lui fera transmettre le message par le coursier.

En (159a) lui réfère à un C(Vinf) a1-objet, et on a inacceptabilité. En (159b), lui doit forcément référer au causé, il y aurait inacceptabilité si lui devait également référer à un C(Vinf). Au contraire, (159c) montre qu'avec un causé agt-objet, un C(Vinf) a1-objet peut se réaliser comme un clitique datif.

Tasmowski citant Barnes (1980) considère le sujet comme le thème de la phrase et le a1-objet, s'il existe, comme le thème du prédicat. Elle donne une liste de facteurs de "thématisation" :

- l'ordre des mots: le thème a tendance à venir très tôt dans la phrase,
- l'animation: le thème est plutôt animé qu'inanimé,
- les personnes: les personnes de discours sont très thématiques
- la détermination: le thème est généralement déterminé.

(159a) s'explique[16] (cf. Miller, 1991)) par un conflit entre plusieurs thèmes possibles pour le prédicat, le causé (*Pierre*) et le clitique, qui est très thématique d'après les facteurs cités. Une caractérisation syntaxique (qui empêcherait un datif complément de l'infinitif sur faire, dans le cas d'un causé exprimé) ne peut pas tenir si on tient compte de a,b). (160a) est parfaitement acceptable. On peut certes arguer une différence entre inergatifs et inaccusatifs, mais d'une part on a rejeté l'hypothèse inaccusative pour le français, d'autre part il faudrait encore expliquer pourquoi il peut y avoir inacceptabilité avec un autre inaccusatif supposé (comme en (160b)). De la même manière qu'on a situé la différence entre *téléphoner* et *parvenir* en sémantique (la différence primordiale concernant l'agentivité de l'argument unique), on explique le constraste (159a)/(160a) par la sémantique.

[16] Alors que (Kayne, 1975) proposait la *Specified Subject Condition* (critiquée par (Tasmowski, 1985), cité par (Miller, 1991)).

Un dernier argument en faveur d'une explication discursive est qu'on peut forcer l'acceptabilité (pour certains locuteurs en construisant le contexte discursif adéquat (160c,d).

(160) a. On lui fera parvenir votre lettre.
b. *Dieu lui fera arriver une catastrophe.
c. — Que faire pour amadouer le patron? Lui écrire? Lui faire téléphoner? Mais par qui?
— Moi je lui ferais téléphoner Marie. Elle est fine, elle le connaît bien, elle saura quoi lui dire. (Tasmowski, 1984)
d. Paul n'aime pas les jeux téléphoniques? Je lui ferai téléphoner tous les opérateurs possibles rien que pour l'embêter.

La caractéristique discursive de la construction caus-à serait de thématiser le causé (a1-objet). En effet, la différence sémantique entre caus-à et caus-agt est bien connue, entre caus-à où le causé est affecté par le procès et caus-agt, où il est seulement l'agent. Nous adaptons pour le français un exemple italien tiré de (Renzi, 1991:506).

(161) a. Je ferai voir ces malades au chef de clinique, ça peut l'intéresser.
b. ?Je ferai voir ces malades par le chef de clinique, ça peut l'intéresser.
c. Je ferai voir ces malades par le chef de clinique, je ne veux pas prendre la responsabilité d'un diagnostic.

Ainsi l'impossibilité d'avoir un C(Vinf) clitique datif concurrent à un causé à-objet (cf. l'interprétation forcée de (159b)), est-elle aussi expliquée par un conflit entre thèmes possibles : le causé a1-objet est thématique du fait des caractéristiques discursives de caus-à et le clitique datif est thématique du fait des facteurs de thématisation cités.

(162) a. Le patron m'a fait payer par l'agent comptable.
b. *Le patron m'a fait payer à l'agent comptable.

Dans le cadre d'une grammaire implémentée, nous ne pouvons que donner des heuristiques. Par exemple, on peut bloquer la construction caus-à dans le cas où le prédicat complexe porte un objet réalisé comme un clitique actanciel de 1ère ou 2ème personne (162b).

Une dernière remarque, sur la différence entre caus-à et caus-agt. L'analyse que nous suivons donne des structures très similaires pour les deux constructions. Or on sait bien qu'il existe une nette nuance de sens entre les deux constructions. La construction en caus-à fait du causé un thème, alors que la construction en caus-agt est focalisée sur le procès et le causé est un agent (exprimé ou pas). La nuance peut bien sûr impliquer dans certains contextes que seule une des deux constructions soit possible.

Voyons maintenant les différentes interactions avec les autres redistributions. Le causatif peut interagir avec le réfléchi, et nous renvoyons Section 1.5.1.3 pour une analyse.

Le causatif avec sujet impersonnel est impossible. Dans la plupart des cas, le prédicat complexe causatif comporte un objet, qu'il s'agisse du causé ou d'un argument de l'infinitif. L'impersonnel est alors bloqué par la contrainte générale qui interdit l'impersonnel lorsqu'est présent un objet.

Le seul cas où un complexe causatif apparaît sans objet, c'est lorsque l'infinitif est une forme intransitive (comme en (163a,b)) et que le causé est omis. L'impersonnel semble

cependant impossible dans ces cas (163c,d) (mais l'impersonnel est soumis à des contraintes discursives fortes).

(163) a. Les amours de Pauline font jaser.
b. Les bons instituteurs font lire à voix haute.
c. *Il faisait jaser des histoires salaces.
d. *?De nos jours, il fait lire à voix haute un instituteur sur dix tout au plus.
e. *?De nos jours, il ne fait plus lire à voix haute que quelques instituteurs.

Plus généralement on va représenter cela par la contrainte générale en français, qui bloque la démotion d'un causateur (cf. infra l'interaction avec le passif et le moyen).

Inversement, comme le *il* impersonnel doit être le sujet final, on bloque bien le causatif sur un impersonnel (passif ou moyen impersonnel, ou verbe impersonnel). On traite (164) comme l'introduction spontanée d'un causateur (cf. Section 1.1.1, classe impers-spontané).

(164) Chanter faux fait pleuvoir.

Le passif est incompatible avec le causatif en français. Plus précisément, d'une part, le complexe causatif ne peut pas être au passif (165b) (contrairement à l'italien).

(165) a. Pierre a fait tomber Paul dans l'escalier.
b. *Paul a été fait tomber dans l'escalier.

D'autre part, l'infinitif sous le causatif ne peut pas être un passif (le causatif ne peut pas s'appliquer à un passif) (166). Cela peut s'expliquer par une contrainte, imposée par la formation d'un prédicat complexe, qui bloquerait la morphologie passive sous le causatif, et plus généralement les formes composées.

(166) *L'incompréhension fait être rejeté Paul par/à tous ses amis.

On peut donc voir en la construction caus-agt, un cas de causatif passif, où cependant la morphologie passive habituelle n'est pas activée. On peut pour s'en convaincre comparer (167a) et (167b). Mais avec la construction caus-agt, l'interprétation stative du passif "classique" n'est pas possible.

(167) a. *Paul fait être balayée la cour par les recrues.
b. Paul fait balayer la cour par les recrues.

Donc on peut considérer que pour former une redistribution complète avec la redistribution partielle sujet→agt-objet, on a trois possibilités en français:

- passif personnel (sujet→agt-objet, objet→sujet) +morpho-passive
- passif impersonnel (sujet→agt-objet, impers→sujet) +morpho-passive
- caus-agt (sujet→agt-objet, causateur→sujet) +morpho-causative

Enfin il est possible de former un prédicat complexe avec faire et un verbe au moyen (168) (d'après Abeillé et al., 1998); inversement, le moyen peut difficilement s'appliquer à un causatif (169b) (contrairement à l'italien). Koenig (1994) fournit cependant des exemples acceptables (169c), qui utilise un sujet *ça*, favorisant l'interprétation typique du moyen comme propriété du sujet (promu).

(168) b. Le sexe fait se vendre ces magazines comme des petits pains.
(169) a. Mon crémier fait toujours goûter les fromages (E + par les clients + aux clients).
b. *Les fromages se font toujours goûter (E + par les clients + aux clients).[17]
c. Une toyota, ça se fait réviser tous les dix ans (Koenig, 1994).

On considère cependant ces cas comme marginaux, et on ne les traite pas dans la MGf. Ni l'impersonnel, ni le passif ni le moyen ne s'appliquant en français à un causatif, on peut dire qu'aucune démotion ne s'applique à l'argument causateur.

En français, l'argument causateur ne peut pas être démis.

1.4.2 Traitement en TAG

Pour traiter en LTAG les constructions causatives à contrôle, on utilise le mécanisme général des verbes à contrôle : le verbe *faire* ancre un arbre auxiliaire, portant une position pour un clitique accusatif (représentant le causé), cet arbre s'adjoignant sur une infinitive, dont un des arguments peut être extrait.

Pour traiter en LTAG les constructions à prédicat complexe, il suffit de localiser dans un même arbre élémentaire le prédicat complexe et des positions à étendre pour ses différents actants grammaticaux. Dans le cas du causatif, le prédicat complexe est fait du verbe causateur et de l'infinitif. Le verbe causatif est celui qui porte les inflections et les clitiques (sauf cas particuliers, décrits en détail Section 1.5.1.4). Pour représenter la variation de verbes causatifs, on prévoit un noeud à substituer pour celui-ci. Seuls les verbes causatifs et les verbes d'état ancrent des arbres initiaux de racine V, et parmi ceux-ci seuls des verbes causatifs (portant le trait :caus = +) peuvent être substitués dans les arbres élémentaires causatifs.
Comme pour les prédicats simples, on prévoit toutes les réalisations possibles des actants grammaticaux, y compris clitiques. La Figure 58 montre un schème pour un transitif strict, au causatif en *à* (caus-à) (pour dériver par exemple *Paul fait balayer la cour aux recrues*). On prévoit également les traits adéquats pour bloquer la négation sur l'infinitif. L'extraction hors du complexe prédicatif ne présente pas de caractère particulier puisque ce complexe est traité comme un seul verbe.

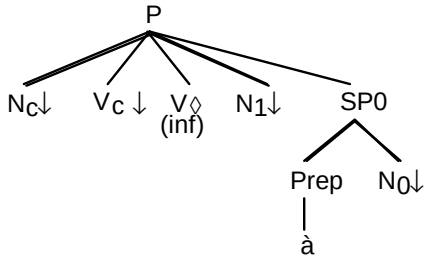
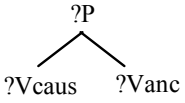


Figure 58. Schème pour la construction caus-à, pour un transitif strict.

[17] La phrase serait éventuellement acceptable avec une interprétation réfléchie (avec *les fromages* comme agent), mais l'interprétation moyenne est impossible.

1.4.3 Description des classes utilisées

On définit une classe morpho-causateur qui modifie la valeur du champ tête (cf. Chapitre 2), qui n'est plus l'ancre de l'arbre, mais le verbe causatif.
?Vcaus représente le verbe causatif, et est placé à gauche de ce qui était la tête auparavant (?Vanc). Le verbe causatif est à substituer, et porte le trait :caus=+ (Seuls LAISSER et FAIRE, et certains verbes de perception peuvent être substitués à ce noeud).
L'ancre (?Vanc) prend le trait :pers=3 (cf. Section 1.5.1.1, où on montre que l'infinitif dans le complexe causatif porte la personne non marquée, ce qui est mis en évidence quand l'infinitif porte un réfléchi, cf. (181)).

morpho-causateur	
superclasses: morpho-marquée	
tête→Vcaus	
constantes:	description[18]:
?Vcaus:cat = V :ind = c :qual = subst ?Vanc:pers=3 ?Vanc:mode = inf	

La classe causateur→sujet introduit le nouvel argument (constante argC), qui porte la fonction sujet, et cela est non révisable (ce qui représente que le causateur ne peut pas être démis en français). Le causateur est également un sujet initial (ce qui fait qu'une construction causative a toujours deux sujets initiaux, qui le sont par rapport à deux prédicats différents, l'infinitif et le complexe causatif.

causateur→sujet
superclasses: morpho-causateur
action: introduction d'un sujet **(non révisable)**
constantes:
argC:cat =N/Cl argC:fonction sujet (**non révisable**) argC:fonction-initiale sujet argC:ind "c"

Pour la démotion du sujet, on dispose déjà de sujet→objet et de sujet→agt-objet. On a besoin d'une classe sujet→à-objet. On peut alors construire les redistributions caus-à, caus-agt, caus-i et caus-spontané. On donne Figure 59 le schéma des classes de la MGf intervenant dans la définition de ces redistributions.

[18] Pour gérer les cas italiens de composition causatif ° passif, il faudrait non pas placer ?Vcaus à gauche de ?Vanc, mais à gauche de ce qui était la tête avant l'introduction de ?Vcaus.

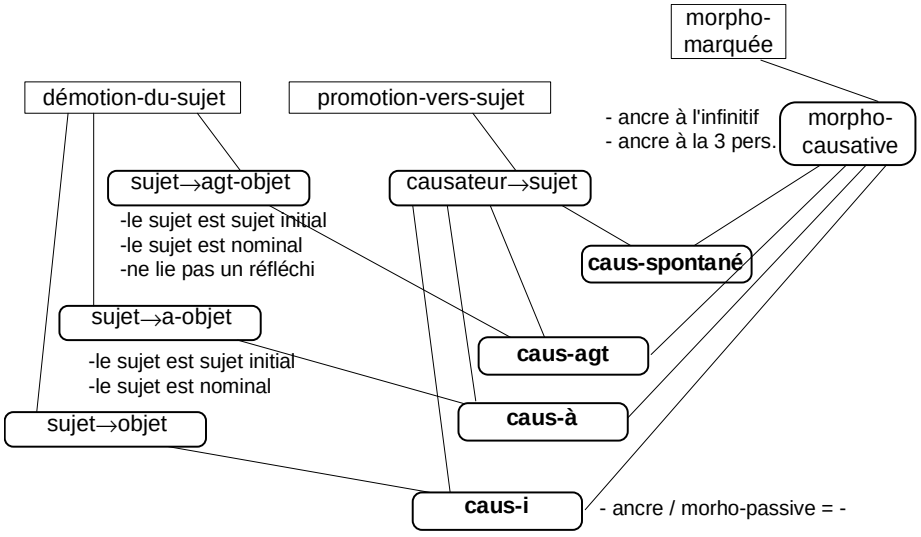


Figure 59. Les classes pour représenter le causatif dans la MGf

caus-i est incompatible avec la morphologie passive, ce qui bloque le causatif après le passif personnel (cf. supra). caus-spontané est utilisée pour l'introduction d'un causateur sur les verbes atmosphériques.

1.4.4 Remarques et points faibles

L'analyse que nous avons fournie comporte cependant plusieurs points faibles.

- Premièrement, l'analyse comme prédicat complexe viole le principe de minimalité sémantique, puisque les arbres élémentaires ainsi construits correspondent à un complexe sémantique et non pas à une seule unité sémantique. Face à une divergence forte entre syntaxe et sémantique, on est forcé de privilégier la syntaxe pour obtenir tout de même une analyse[19]. Cette analyse est diminuée, en ce sens qu'elle doit être sujette à un calcul sémantique supplémentaire par rapport aux autres analyses présentées.
- D'autre part, il semble que la formation d'un prédicat complexe causatif puisse s'appliquer récursivement (au moins deux fois) (cf. Kayne, 1975; Miller, 1991; Kulick, 1998). Ainsi (170b) peut être reliée à (170a) exactement de la même manière que (141a) peut être reliée à *Les recrues balaient la cour*.

(170) a. Son fils a fait sauter le pont.
b. Elle a fait faire sauter le pont (à + par + E) son fils.
c. Elle l'a fait faire sauter à son fils.
d. Elle le lui a fait faire sauter.
e. Elle l'a fait faire sauter le pont.

Or l'analyse que nous venons d'évoquer en TAG repose sur une redistribution opérée lexicalement (c'est-à-dire lors de la construction des structures lexicalisées, avant la

[19] On peut retourner le problème et conclure que ce sont les contraintes de bonne formation de départ, comme le principe de cohérence sémantique, qui ne sont pas tenables dans tous les cas.

dérivation). Comme, par définition, l'application des redistributions est finie, on ne peut traiter la récursion illustrée en (170) que par la compilation a priori de plusieurs formations de prédicats causatifs, c'est-à-dire nier toute récursion. La possibilité d'empiler des verbes causatifs remet en cause le placement des redistributions au niveau du lexique. La solution est sans doute de considérer que des structures lexicalisées peuvent être construites au moment de la dérivation.

- L'analyse fonctionnelle que nous donnons dans la MGf ne permet pas de considérer que plusieurs actants grammaticaux portent la même fonction. Or c'est ce qui est induit dans le cas d'une construction caus-à avec un infinitif comprenant lui-même un a1-objet:

(171) On lui fait remettre le message au secrétaire. (Tasmowski, 1984)

D'après ce qui précède, (171) est analysé avec deux a1-objets. Donc soit cette analyse fournit un contre-exemple au principe d'unité fonctionnelle[20], soit c'est l'analyse qui manque une différence entre le causé et le a1-objet nominal en (171).[21]
Tasmowski cite même des exemples de doubles datifs, ce qui confirmeraient la présence de deux a1-objets.

(172) ? Paul me la lui fait transmettre.

1.5 La "réflexivation"

Nous nous intéressons dans cette section aux constructions avec "vrais" réfléchis, c'est-à-dire les verbes portant le clitique SE, et dont le sens est réfléchi ou réciproque. Le SE est en distribution complémentaire avec un actant grammatical du verbe portant la fonction objet (173) ou a1-objet (174).

(173) a. Juliette lave Paul.
b. Juliette se lave.

(174) a. Paul donne une gifle à Juliette.
b. Paul se donne une gifle.

1.5.1 Analyse des données

1.5.1.1 Propriétés communes aux verbes réfléchis

Les "vrais" réfléchis partagent la propriété d'apparaître avec SE, avec les réfléchis intrinsèques (175) et les constructions moyennes (cf. supra Section 1.3). On parlera de "verbes réfléchis" pour désigner les verbes qui portent SE.

(175) Marie s'est évanouie.

(176) a. On fabrique maintenant des fourrures polaires avec du plastique recyclé.

[20] (Zaenen et Dalrymple, 1996) qui fournissent une analyse des causatifs en français dans le cadre de LFG, n'abordent malheureusement pas ce problème.
[21] Il existe des contraintes d'interprétation dans ces cas de doubles a1-objets. Pour (171), la seule interprétation possible est celle où le clitique datif est le causé, (Tasmowski, 1984)).

b. Les fourrures polaires se fabriquent maintenant avec du plastique recyclé.

Les verbes réfléchis ont les propriétés communes suivantes, que l'on impute à la présence du SE :

- L'auxiliaire de temps des verbes réfléchis est ETRE.
  Cette propriété est indépendante de l'auxiliaire de temps des verbes correspondants non réfléchis, quand ils existent.
- On considère généralement que le SE s'accorde en personne avec le sujet (177). Dans le cas de vrai réfléchi, on parle souvent de liage avec le sujet.

(177) a. *Tu s'évanouit.
b. *Tu s'est bien vendu au salon du livre.
c. *Tu se lave.
d. *Je se donne une gifle.

Dans le cas d'un impersonnel avec un réfléchi[22], cette règle d'accord avec le sujet semble adéquate (cf. (178b), (179b) et (180b)). Mais on peut remarquer que dans le cas de vrai réfléchi (180), le SE peut ne pas s'accorder avec l'argument auquel il est "lié sémantiquement", c'est-à-dire l'argument qui a fusionné avec un autre argument. Il faut donc distinguer l'actant grammatical qui représente *l'argument auquel le réfléchi est sémantiquement lié*, de l'actant grammatical avec lequel le réfléchi s'accorde éventuellement (en l'occurrence le sujet).

(178) a. Tu t'es évanouie à la foire du livre.
b. Il ne s'est évanoui que toi à la foire du livre.
c. *Il ne t'est/es évanoui que toi à la foire du livre.

(179) a. Tu t'es bien vendu à la foire du livre.
b. Il ne s'est bien vendu que toi à la foire du livre. (Koenig et Jurafski, 1994)
c. *Il ne t'est/es bien vendu que toi à la foire du livre.

(180) a. Tu t'es lavé dans cette eau.
b. Il ne s'est lavé que toi dans cette eau. (Abeillé et al., 1998)
c. *Il ne t'est/es lavé que toi dans cette eau.

L'interaction avec le causatif[23] pose un problème supplémentaire. En effet, dans le cas d'un complexe causatif, ni le causateur (le sujet du complexe causatif), ni le causé (qui représente l'argument auquel le réfléchi est sémantiquement lié) ne s'accordent avec un réfléchi placé sur l'infinitif, qu'il s'agisse d'un réfléchi intrinsèque (181a) ou d'un vrai réfléchi (181b,c).

(181) a. Tu n'a pas fait se/*me retourner que moi sur ton passage.
b. Pourquoi le metteur en scène n'a fait se/*me donner un coup qu'à moi?
c. Paul n'a fait se/*te regarder que toi dans la glace.

[22] Cf. Section 1.5.1.3 l'interaction du réfléchi et de l'impersonnel; on s'intéresse ici seulement à la question de l'accord.
[23] Cf. Section 1.5.1.3 l'interaction du réfléchi et du causatif et Section 1.5.1.4 pour le placement du réfléchi sur l'infinitif.

On ne peut pas imputer cela au seul mode infinitif, puisque dans le cas de simples constructions à contrôle ou à montée, un réfléchi sur un infinitif peut être marqué pour la personne (182).

(182) a. J'ai peur de m'/*s'évanouir.
b. Je suis sûre que tu va bien te/*se vendre.
c. Je veux me/*se laver.

Plus généralement Marandin (1997) remarque que le sujet inversé ne s'accorde pas en personne avec une forme verbale, y compris si elle est réfléchie (mais il y a bien accord en genre et nombre).

(183) a. A ce moment là sont arrivés Marie et toi, qui rentriez du cinéma.
b. A la fac se sont alors rencontrés Marie et toi, qui étiez tous deux étudiants en médecine. (Marandin, 1997)

Donc il faut plutôt considérer que le réfléchi s'accorde en personne avec le verbe (comme cela est suggéré par Abeillé et al (1998, note 26)), et que l'accord sujet-verbe tombe dans certains cas. Dans le cas d'un impersonnel on aura accord régulier avec le *il* et donc 3ème personne, dans le cas d'un infinitif entrant dans un complexe causatif (comme en (181a)) ou d'un sujet inversé (comme en (183b)), la forme verbale prend la 3ème personne (qui est la personne non marquée).

1.5.1.2 Les vrais réfléchis

Pour les vrais réfléchis, le SE alterne systématiquement soit avec un a1-objet, soit avec un objet (clitique non réfléchi ou SN). Le sujet du verbe avec vrai réfléchi porte les restrictions de sélection associées à la fois au sujet et à l'objet (respectivement l'a1-objet) de la forme verbale correspondante non réfléchie.

Les verbes réfléchis constitue un domaine très étudié de la grammaire française, et plus généralement de la grammaire des langues romanes (cf. notamment (Ruwet, 1972 ; Zribi-Hertz, 1982 ; Wehrli, 1986 ; Abeillé et al., 1998 pour le français; Alsina, 1996 pour le catalan, avec un panorama de la situation générale pour les langues romanes; Manzini, 1986; Monachesi, 1995 pour l'italien).
Du fait des propriétés communes à tous les verbes réfléchis, une analyse unitaire a été souvent recherchée dans la littérature. On peut trouver deux types d'analyse pour SE, comme marqueur ou comme simple réalisation alternative d'actant grammatical[24]. Comme Alsina (1996) le souligne, les différentes langues romanes ont chacune des arguments pour et contre les deux analyses.
L'analyse comme marqueur est naturelle pour les réfléchis intrinsèques comme pour les verbes au moyen. Pour les vrais réfléchis, l'alternance systématique avec un actant non réfléchi est plutôt un argument en faveur de l'analyse "simple" qui fait de SE une simple réalisation alternative de la fonction objet ou a1-objet. Cette réalisation alternative serait la marque de la sémantique particulière du réfléchi.
Mais Kayne (1975) a montré que les réfléchis où SE alterne avec un objet ont un comportement intransitif, argument net en faveur de l'analyse comme marqueur pour tous les cas de SE.

[24] Par exemple,

Dans le cadre de LFG, Grimshaw (1982) propose une analyse où SE est marqueur, aussi bien dans le cas de moyen que dans le cas de vrais réfléchis. Dans le cas du vrai réfléchi, le SE qui alterne avec un objet nominal est un facteur d'intransitivisation. SE ne porte donc jamais la fonction objet.
Wehrli (1986) considère que SE est toujours un marqueur de « l'absorption » d'un argument (en syntaxe dumoins, c'est-à-dire avec notre terminologie, absorption d'un actant grammatical) : il bloque toujours la réalisation d'un argument SN, cette absorption étant éventuellement lexicalisée (ce qui correspond aux cas de réfléchis intrinsèques). L'absorption du sujet donne lieu au moyen; celle de l'objet ou a-objet donne lieu au réfléchi/réciproque; celle de "l'extra datif" donne lieu au SE de possession inaliénable.
Nous allons suivre globalement cette approche. Nous prenons cependant en compte des données supplémentaires (mais nous ne traitons pas le SE de possession inaliénable). Il s'agit d'une part de rendre compte que les données montrant le caractère intransitif des vrais réfléchis ne sont en fait pas acceptées par tous les locuteurs (Abeillé et al, 1998). D'autre part, nous prenons en compte l'interaction du causatif avec le réfléchi, suivant Abeillé et al. (1998). Nous avons en effet vu Section 1.4.1.2, le placement sur l'infinitif d'un complexe *faire + Vinf*, de certains clitiques, dont le réfléchi.

Sémantique et vrai réfléchi

On ne peut en français interpréter (184) que comme (185a), les sens de (185b,c) étant inaccessibles (l'impossibilité d'obtenir le sens (185b) est plus nette). En anglais par exemple, les trois interprétations sont possibles (cf. Sells et al, 1987)).

(184) Jean se défend mieux que toi.

(185) a. Jean se défend mieux que tu ne te défends.
b. Jean se défend mieux que tu ne le défends.
c. Jean se défend mieux qu'il ne te défend.

Cette interprétation unique indique qu'au niveau sémantique, le prédicat réfléchi est un prédicat unaire. On peut remarquer que cela ne fournit d'arguments ni pour ni contre une analyse de SE comme marqueur. Sells et al (1987), faisant une étude sur la diversité des situations entre langues quand au statut morphologique, syntaxique et sémantique des constructions avec "vrais" réfléchis, citent des cas de réfléchi, comme le japonais *zibun*, où en sémantique, le réfléchi est unaire, et en syntaxe le réfléchi est transitif (en revanche ils excluent la possibilité d'avoir un prédicat binaire réalisé en syntaxe comme un intransitif).

L'accord du participe

Pour les vrais réfléchis, si le SE alterne avec un objet, le participe s'accorde avec le sujet (186a). Cet accord n'est cependant pas général à tous les emplois de SE. En effet, Godard et al. (mns) montrent que dans les cas où le SE alterne avec un a1-objet, la forme verbale suit le schéma général d'accord du participe passé avec un objet placé devant, d'où la différence d'accord entre (186b) et (186c)). Nous considérons que dans le cas où SE alterne avec un objet (dorénavant SE-obj), il s'accompagne d'un accord participe verbe, et dans le cas où il alterne avec un a1-objet (dorénavant SE-a-obj), il ne s'accorde pas avec le participe. Donc par la suite on distinguera deux SE, SE-obj qui déclenche l'accord

du participe avec le sujet, et SE-a-obj qui ne le déclenche pas. Le statut exact de la différence entre les deux reste à préciser.

(186) a. Juliette s'est lavée.
b. Les Dupont se sont construit une maison.
c. La maison que les Dupont se sont construite.

Pour les verbes à réfléchi intrinsèque, les grammaires font également une distinction entre ce que nous avons appelé SE-obj et SE-a-obj. Par exemple Grévisse (1969) indique une différence d'accord entre (187a) et (187b). On peut cependant considérer qu'il s'agit là de grammaire normative, et que l'accord est en fait généralisé, aussi bien pour *s'évanouir* que pour *se complaire* ou *se souvenir*.

(187) a. Marie s'est *évanoui/évanouie.
b. Elles se sont complu/*complues à ce jeu.

En ce qui concerne les verbes au moyen, (188) montre qu'ils font intervenir SE-obj.

(188) Trois villas se sont *construit/construites depuis janvier et deux sont en projet.

Le réfléchi est "intransitivant" pour certains

Les tests classiques pour la transitivité en français sont l'impersonnel (cf. supra l'impersonnel ne s'applique qu'à des transitifs) et le causatif (cf. supra, la distinction entre caus-i et caus-à ou caus-agt). De (189a,b) on peut déduire que les verbes moyens se comportent comme des intransitifs.

(189) a. Il s'est construit trois villas depuis janvier et deux sont en projet.
b. Le sexe fait se vendre ces magazines comme des petits pains.

Les vrais réfléchis avec SE-obj se comporteraient aussi comme des intransitifs (dans le cas de SE-a-obj, la transitivité dépend de l'existence d'un autre actant grammatical objet; *s'acheter une voiture* est transitif, *se parler* est intransitif).

(190) a. Il se lavait encore trois personnes dans la fontaine de Trevi quand la police est intervenue.
b. La direction a fait s'acheter un ordinateur portable à tous les élèves.

Mais Abeillé et al. (1998) montrent que la situation est plus complexe, étant donné que seuls certains locuteurs (dorénavant les locuteurs A) acceptent (190a,b), alors que d'autres (les locuteurs B) les rejettent[25]. Quoi qu'il en soit, les verbes au moyen sont intransitifs pour tous les locuteurs et les vrais réfléchis avec SE-obj le sont pour les locuteurs A.

Donc on a deux options :

(i) On définit la transitivité comme la présence d'un objet dans la sous-catégorisation, et alors on ne peut pas dire que SE-obj est objet; on doit considérer que SE-obj déclenche l'accord avec le participe comme un objet mais n'en est pas un.
(ii) On considère que SE-obj est un objet (et déclenche de manière régulière l'accord du participe avec l'objet réfléchi, qui s'accorde lui-même avec le sujet, cf. Godard et al., mns) et on doit restreindre la définition de la transitivité.

[25] (Abeillé et al, 1998) insistent sur le fait que (98a,b) sont soit tous deux acceptés ou soit tous deux rejetés.

Nous optons pour (i) étant donnée la possibilité d'avoir un clitique accusatif indéfini cooccurrent à SE-obj (191a), ou un *en* quantitatif (qui n'est possible que pour un objet) (191b,c).

(191) a. Il s'en pêche tous les jours ici, des poissons comme ça.
b. Il s'en lavait encore trois dans la fontaine de Trevi quand la police est intervenue. (pour locuteurs A seulement)
c. Faites les sortir avant qu'il ne s'en évanouisse une.

Donc, pour les locuteurs A, nous considérons SE-obj comme des marqueurs de construction, c'est-à-dire qu'ils ne portent pas de fonction grammaticale, avec cependant pour SE-obj la particularité de déclencher l'accord avec le participe, comme le ferait un objet antéposé au verbe. Pour les locuteurs B, SE-obj est un objet à part entière dans le cas des vrais réfléchis[26], un marqueur sinon.

Pour SE-a-obj, comme il n'influe pas sur la transitivité, on n'a pas de raison particulière pour considérer qu'il ne porte pas la fonction a1-objet, si ce n'est le souci d'une analyse unitaire avec SE-obj. Nous avons cependant un argument pour ne pas en faire un simple marqueur. En effet, s'il ne porte pas la fonction a1-objet, on ne peut pas expliquer simplement le contraste entre (192a) et (192b), puisque dans les deux phrases les formes verbales auraient la même sous-catégorisation finale.

(192) a. Pierre se fait des reproches plus souvent qu'à Paul.
b. *Pierre lit des poèmes plus souvent qu'à Paul.

On considère donc que SE-a-obj porte la fonction a1-objet, dans le cas de vrais réfléchis, et qu'il est un marqueur dans le cas de réfléchis intrinsèques.
Les réfléchis intrinsèques comportent directement ce marqueur. Pour les moyens, SE-obj est introduit lors de la redistribution moyenne; pour les vrais réfléchis, SE-obj est introduit en même temps que l'assignation de la fonction ∅ à l'objet. SE-a-obj est introduit comme réalisation alternative de la fonction a1-objet, mais on le représente tout de même en dimension 2, étant donné qu'il déclenche le placement de tous les clitiques sur le verbe causatif.

Donc, SE est toujours en alternance avec un actant syntaxique SN (cf. Wehrli, 1986), mais il ne constitue pas forcément une réalisation (comme clitique) d'une fonction sous-catégorisée. Il est marqueur pour les réfléchis intrinsèques, pour le moyen, pour le vrai réfléchi alternant avec un objet, mais seulement pour les locuteurs A.
L'on voit que la distinction nette entre marqueur/réalisation de fonction n'est sans doute pas très satisfaisante, et que SE a des propriétés hybrides, qui font qu'il peut ou pas constituer la réalisation d'un actant sous-catégorisé.

1.5.1.3 Contraintes et interactions avec les autres redistributions

[26] Pour les locuteurs B, les cas de réfléchis idiomatiques tels que *se payer du bon temps* ou *se serrer la ceinture*, se comportent comme des intransitifs (Abeillé et al., 1998). On doit donc considérer dans notre analyse qu'ils font dans tous les cas intervenir un SE qui ne porte pas la fonction objet.

SE dans le cas des vrais réfléchis indique la fusion de deux arguments sémantiques. Nous n'avons pas encore abordé les contraintes syntaxiques sur ces arguments. Un de ces arguments est associé à la fonction objet ou a1-objet (et reçoit la fonction "∅" ou reste a1-objet). L'autre argument (celui auquel le réfléchi est lié sémantiquement) doit être sujet initial. En effet, d'une part on n'a pas de réfléchi marquant la fusion d'un objet et un a1-objet (193)). D'autre part on n'a pas de réfléchi après le passif (194c), ce qui supposerait un réfléchi lié sémantiquement à un actant sujet, mais non initial.

(193) a *Tu s'_k achètera un $esclave_k$.[27]
b. *Tu_i s'_k achètera à un $esclave_k$.

(194) a. Gautier vend des esclaves à des esclaves.
b. Des esclaves sont vendus à des esclaves par Gautier.
c. *Des esclaves se sont vendus par Gautier.

Le réfléchi interagit avec l'impersonnel et le causatif, et les données montrent que SE doit être lié sémantiquement à un sujet initial, qui peut cependant être démis, vers la fonction objet, a1-objet mais pas vers la fonction agt-objet. En effet:

- On n'a jamais de passif après un réfléchi (195b), ni de caus-agt après un réfléchi (cela est vrai pour un réfléchi SE-a-obj idiomatique (196b,c) ou un vrai réfléchi (197b)).

(195) a. Paul se donne une gifle.
b. *Une gifle s'est donnée par Paul. *SE-a-obj(vr) o passif-pers

(196) a. La crise fait se serrer la ceinture à tout le monde. SE-a-obj(id) o caus-à
b. *?La crise fait se serrer la ceinture. *SE-a-obj(id) o caus-agt
c. *La crise fait se serrer la ceinture par tout le monde. *SE-a-obj(id) o caus-agt

(197) a. Le metteur en scène fait se_i donner une gifle à $Paul_i$. SE-a-obj(vr) o caus-à
b. *Le metteur en scène fait se_i donner une gifle (par $Paul_i$+ E)
*SE-a-obj(vr) o caus-agt

Ce point commun est un argument pour considérer que le passif est caractérisé par la démotion sujet→agt-objet, et que caus-agt est un passif.
On a la contrainte suivante:

Un argument auquel est lié sémantiquement un réfléchi ne peut pas subir sujet→agt-objet.

On représente cette contrainte par un trait :lie-refl sur l'argument auquel le réfléchi est lié, ce trait est introduit par la redistribution pour les vrais réfléchis. La démotion sujet→agt-objet est impossible si ce trait est positif.

On peut remarquer que le cas où le causé n'est pas exprimé montre la même inacceptabilité que lorsque le causé est agt-obj (197b). C'est un argument pour considérer que l'agt-objet peut être systématiquement omis.

- En revanche l'argument auquel le réfléchi est sémantiquement lié peut subir sujet→a1-objet ((196a) (197a)); et peut subir sujet→objet (c'est vrai dans le cadre d'un caus-i pour tous les SE-a-obj (intrinsèque, vrai réfléchi (198a,b)); pour tous les

[27] On note par un indice le partage du trait de personne.

SE-obj (intrinsèque, moyen, vrai réfléchi (199a,b,c)); ou dans le cadre d'un impersonnel pour tous les SE-a-obj (200a,b) et SE-obj (201a,b,c)).

(198) a. La lâcheté a fait se complaire Juliette dans la facilité. SE-a-obj(intrins) o caus-i
b. L'accident a fait se parler les vieux ennemis. SE-a-obj(vr) o caus-i

(199) a. La chaleur a fait s'évanouir trois personnes. SE-obj(intrins) o caus-i
b. La crise fait se vendre les videos new age par milliers. SE-obj(moy) o caus-i
c. Paul et Virginie font se rencontrer leurs parents. SE-obj(vr) o caus-i

(200) a. Il s'est complu de nombreux auteurs à plagier Proust. SE-a-obj(intrins)o impers
b. ? Il s'est parlé de nombreuses personnes

(201) a. Il s'est évanoui trois personnes lors de cette manifestation. SE-obj(intrins) o impers
b. Il s'est consommé trente tonnes de blé l'an dernier. SE-obj(moy) o impers
c. Il s'est lavé quelqu'un dans cette fontaine. SE-obj(vr) o impers

Enfin, il existe des contraintes sur la composition d'un causatif suivi d'un réfléchi. On peut très bien avoir un réfléchi après caus-agt, comme en témoignent (204a), (205a). Mais il est impossible d'avoir un réfléchi après caus-i ou caus-à, si le réfléchi ne marque pas la fusion du causateur avec le causé (on marque ce cas par "causé→∅" à côté des exemples), donc à côté de (202a) (206) tout à fait acceptables, (203b), (204b), (205b) sont bloquées.

(202) a. Paul se fait vomir. caus-i o SE-obj (causé→∅)

(203) a. Marie a réussi à faire téléphoner Pierre à Paul.
b. *Marie a réussi à se faire téléphoner Pierre. *caus-i o SE-a-obj

(204) a. Paul se fait coiffer par Jean. caus-agt o SE-obj
b. *Paul se fait coiffer à Jean. *caus-à o SE-obj

(205) a. Paul se fait porter des fleurs (E + par un coursier). caus-agt o SE-a-obj
b. *Paul se fait porter des fleurs à un coursier. (avec le sens (205a)) *caus-à o SE-a-obj

(206) Paul se le fait avaler. (Godard et al., mns) caus-à o SE-a-obj (causé→∅)

Des explications syntaxiques de ce phénomène sont souvent recherchées, et en effet le contraste est très net entre par exemple (204a) et (204b).

Abeillé et al. (1998) expliquent l'impossibilité d'avoir (204b) par leur interprétation du principe A dans le cadre de HPSG.
Le principe A est interprété comme une contrainte sur le liage de SE au sein de la liste ARG-ST. Le SE apparaît toujours dans la ARG-ST, et doit trouver à sa gauche un lieur, qui dans le cas de SE est obligatoirement sujet. Dans le contexte d'un complexe causatif, cette contrainte doit être respectée à la fois pour la liste ARG-ST de l'infinitif et pour celle de FAIRE. La liste ARG-ST de faire comprend le causateur, le causé, et les C(Vinf) et le

principe A est respecté pour tout réfléchi sur FAIRE coindéxé avec le causateur. Mais pour (204b), c'est la ARG-ST de l'infinitif COIFFER qui viole le principe A.

(204b) *Paul$_i$ se$_i$ fait coiffer à Jean$_k$ coiffer:ARG-ST=(NP$_k$, aff[ana]$_i$)

Dans une version antérieure non publiée (Godard et al., mns), la construction caus-agt est également traitée, comme un cas particulier de passif. Les ARG-ST de l'infinitif dans le cas d'un caus-agt comportent le agt-objet, en position oblique.

(204a) Paul$_i$ se$_i$ fait coiffer par Jean$_k$ coiffer:ARG-ST=(aff[ana]$_i$, par NP$_k$)
(205a) Paul$_i$ se$_i$ fait porter des fleurs$_m$ par un coursier$_k$. porter:ARG-ST=(NP$_m$, aff[ana]$_i$, par NP$_k$)

Pour expliquer pourquoi (204a) et (205a) sont possibles, il faut alors changer la formulation du principe A en:

> "Si un lieur possible (un sujet) apparaît à gauche de SE sur la ARG-ST, alors il doit lier SE." (en mes termes)

Mais cela n'est pas une solution aboutie, car cela n'expliquerait plus l'impossibilité d'avoir un réfléchi dans le cas d'un passif personnel "classique" (hors causatif).
Un esclave$_k$ s*$_{i/*k}$'est acheté par Paul$_i$. acheter:ARG-ST=(esclave$_k$, aff[ana]$_i$, par NP$_i$)

Pour expliquer l'inacceptabilité de (203b), (204b), (205b), nous suivons Miller (1991) en considérant que l'inacceptabilité de (203b), (204b), (205b) est due à des contraintes discursives, comme nous l'avons fait pour certaines contraintes sur le causatif indépendamment du réfléchi. De la même manière qu'on a le contraste (207a)/(207b) (cf. Section 1.4.1.3, la discussion sur la série (159)), on a le contraste (203b)/(208). (204b) est impossible de la même manière que (210c).
On a un continuum entre (209b) acceptable car le clitique accusatif n'est pas thématique au sens de Tasmowski (c'est un inanimé), (210b), moins bon car l'objet est animé, (210c) mauvais car l'objet est clitique personnel et donc très thématique. Pour (204b), on peut considérer le réfléchi comme hautement thématique.

(207) a. * Paul lui fait téléphoner Pierre
b. Paul lui fait parvenir votre lettre. (Tasmowski, 1984)

(208) Paul a réussi à se faire parvenir la lettre. caus-i o SE-a-obj

(209) a. Paul fait faire le gâteau à Marie.
b. Paul le fait faire à Marie.

(210) a. Le patron fait coiffer le nouveau client à Sylviane.
b. ? Le patron le fait coiffer à Sylviane.
c. * Le patron te fait coiffer à Sylviane. (adapté de (Tasmowski, 1984)).

1.5.1.4 Placement du clitique réfléchi

Pour compléter l'étude du réfléchi, il faut considérer la question du placement du SE dans le contexte d'un complexe causatif (ce qui a déjà été évoqué Section 1.4.1.2). Les séries (198) et (199) montrent que si le réfléchi lie sémantiquement le causé (=marque la fusion d'un argument de l'infinitif avec le causé), alors le SE se place sur l'infinitif. Les séries (202), (204), (205) montrent que si le réfléchi lie sémantiquement le causateur (=marque la fusion d'un argument du complexe causatif avec le causateur), alors le SE se place sur FAIRE.

On peut représenter de manière immédiate cette distinction étant donné la procéduralité utilisée pour représenter les redistributions. Le SE doit se placer sur le verbe tête au moment de l'introduction du SE (c'est-à-dire pour le moyen, lors de l'application du moyen; pour un vrai réfléchi, lors de l'application du réfléchi; pour un réfléchi intrinsèque, lors de la spécification de la sous-catégorisation initiale, en dimension 1). Pour (198) et (199), le réfléchi intervient avant le causatif, et le verbe tête est l'ancre. Pour (202), (204), (205), le causatif intervient avant le réfléchi, et le verbe tête est le verbe causatif.
Comme on l'a vu Section 1.4.1.2, si un des clitiques est sur l'infinitif, alors tous les autres clitiques éventuels correspondant à un C(Vinf) sont également sur l'infinitif (cf. supra, les exemples (156) et (158)). Ceci est traité par le trait :hôte, qui pour tous les C(Vinf) prend comme valeur le verbe tête au moment de l'introduction du SE. La réalisation d'un actant comme clitique (cf. Chapitre 5), en dimension 3, se fait sur l'hôte, où à défaut d'hôte spécifié, sur le verbe tête (au final).

1.5.2 Description des classes utilisées

Pour représenter le comportement de SE-obj et SE-a-obj, on utilise une classe commune se, qui hérite de clitiques (classe introduite dans la description de la dimension 3, Chapitre 5). Elle introduit l'auxiliaire de temps ETRE (on note → l'assignation d'une valeur qui masque éventuellement une autre valeur, ici l'auxiliaire ETRE masque l'auxiliaire AVOIR). L'accord en personne du SE avec le verbe est directement géré par le module morphologique (on considère en effet que les clitiques sont des affixes, ce qui sera détaillé Chapitre 5). Il faut pour cela considérer que l'infinitif est à la 3ème personne.

se
superclasses: clitiques
action: ∀ actant / ⌉(:fonction=sujet), :hôte=•tête
constantes:
•tête:aux→être arg0:fonction-initiale = sujet arg0:lie-réfl=+ •tête:clC:réfl ≠ - [28]

Cette classe spécifie également que pour tous les actants (c'est-à-dire les constantes qui ont une valeur pour le trait :fonction), l'hôte est la tête verbale courante (au moment de l'application de la réflexivation), représentée par la valeur du champ tête.
On note •tête la valeur du champ tête lors de l'héritage de la classe (on note simplement tête la valeur du champ tête interprétée après tous les héritages, et donc en particulier après toutes les redistributions).
Enfin, on doit représenter la contrainte que seul un sujet initial peut lier un réfléchi, et que ce sujet initial ne peut pas subir sujet→agt-objet. On sait que la constante sujet

[28] Cette équation signifie que le verbe tête au moment de l'héritage de cette classe, reçoit un trait imposant la réalisation d'un réfléchi. L'accord en personne avec le verbe est géré par la morphologie.

initial est toujours arg0 (l'inverse n'est pas vrai, cf. les intransitifs inaccusatifs en italien). Donc on peut ajouter le trait adéquat sur arg0 (arg0:lie-réfl = +).

On donne Figure 60 les classes héritant de se. La classe vrai-réfléchi-se-a-obj représente la réfléxivation d'un a1-objet. Elle hérite simplement de se, et stipule que la fonction a1-objet est remplie (cf. supra, on considère que SE-a-obj porte la fonction a1-objet). La classe se-obj ajoute l'accord du sujet en nombre et genre avec le participe. Elle est héritée par la morphologie moyenne, et par vrai-réfléchi-se-obj, qui représente la réflexivation de l'objet, et qui dément l'objet vers la fonction ∅.

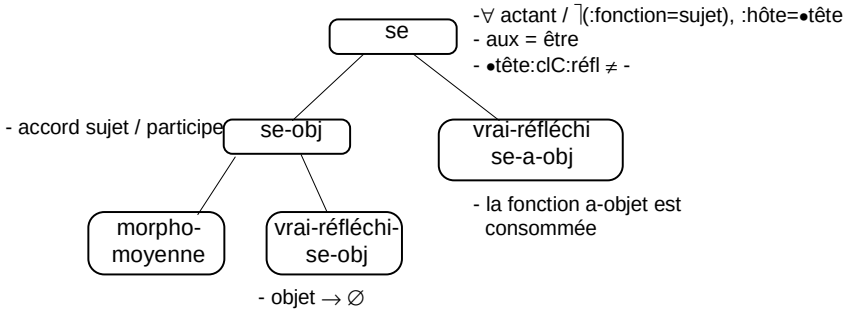


Figure 60. Classes de la MGf mettant en jeu un réfléchi

1.6 Récapitulatif

On donne Figure 61 le schéma récapitulatif des classes utilisées pour les redistributions en français (légèrement simplifié : on ne distingue pas morphologie passive avec ou sans copule). Les redistributions (non partielles) sont notées en gras.
Les compositions possibles n'y ont pas été représentées. Nous en donnons la liste ci-dessous, avec un exemple.

- moyen o impersonnel Il s'est vendu 3 millions de disques de Björk cette année-là.
- moyen o caus-i La crise fait se vendre les videos new age par milliers
- caus-agt o vrai-réfléchi-SE -obj Paul se fait coiffer par Jean.
- caus-agt o vrai-réfléchi-SE-a-obj Paul se fait porter des fleurs (E + par un coursier)
- caus-i o vrai-réfléchi-SE -obj Paul se fait vomir.

(si SE-obj marque la fusion du causateur avec le causé)

- caus-a o vrai-réfléchi-SE-a-obj Paul se le fait avaler.

(si SE-a-obj marque la fusion du causateur avec le causé)

- vrai-réfléchi-SE-a-obj o caus-a Le metteur en scène fait se donner une gifle à Paul.
- vrai-réfléchi-SE-a-obj o caus-i L'accident a fait se parler les vieux ennemis.
- vrai-réfléchi-SE -obj o caus-i Paul et Virginie font se rencontrer leurs parents.
- vrai-réfléchi-SE-obj o impers Il se lavait trois personnes dans la fontaine. (locuteurs A)

On bloque passif o impersonnel (cf. il existe une redistribution directe passif-impersonnel) par une contrainte de non-cooccurrence.
*passif-personnel o caus-i (metatrait :morpho-passive)
*caus-a o vrai-réfléchi-SE -obj Paul se fait coiffer à Jean (cf. le se ne peut pas marquer la fusion du causé)
*vrai-réfléchi-SE-a-obj o passif-personnel (metatrait :lie-réfl)
*vrai-réfléchi-SE -obj o passif-impersonnel (metatrait :lie-réfl)
*vrai-réfléchi-SE-a-obj o caus-agt (metatrait :lie-réfl) Le metteur en scène fait se donner une gifle (E + par Paul).

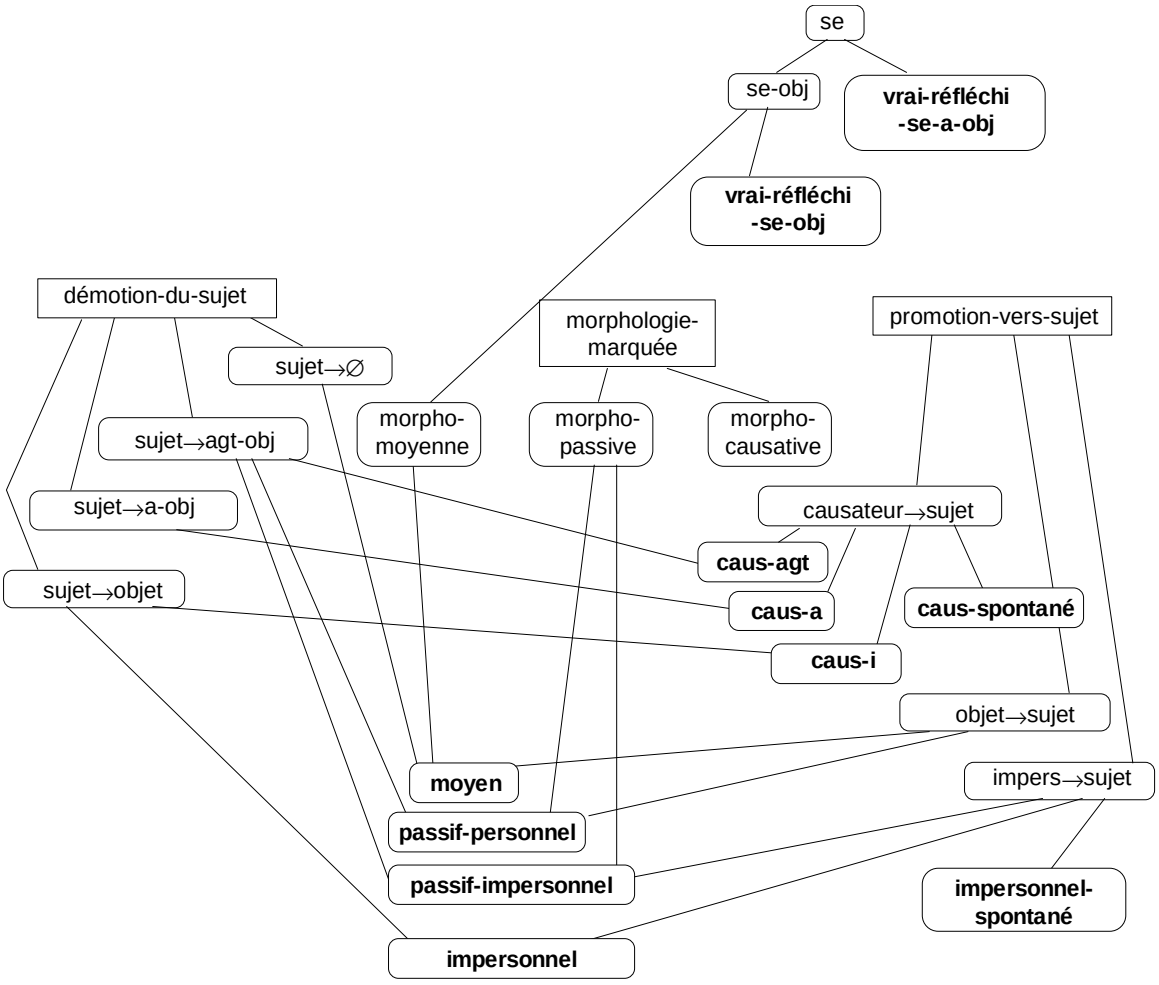


Figure 61. Les redistributions pour les verbes en français (simplifié pour la morphologie passive)

2 Les redistributions pour les verbes en italien

L'italien et le français sont des langues très proches, et on peut adopter comme méthodologie de partir des analyses données pour le français pour étudier leur pertinence pour l'italien. Ainsi, nous passons en revue les différentes redistributions qui ont été définies pour le français, et voyons comment elles s'adaptent à l'italien.
Nous donnons un schéma récapitulatif des classes pour les redistributions en italien à la fin de cette section (p. 200), avec également une version simplifiée, implémentée (p. 202).
On a déjà donné une différence pour l'analyse des verbes intransitifs en italien et en français, puisque nous retenons l'hypothèse inaccusative pour l'italien mais pas pour le français (cf. Chapitre 3).

2.1 Promotion spontanée de l'objet d'un inaccusatif

Comme on a retenu l'hypothèse inaccusative pour l'italien, il faut introduire une redistribution pour la promotion spontanée de l'objet d'un verbe intransitif inaccusatif, objet→sujet(spontané), qui hérite de objet→sujet (cette dernière classe étant par ailleurs utilisée comme en français pour le passif personnel et le moyen personnel).

2.2 L'impersonnel

2.2.1 Ce qui s'appelle impersonnel en italien

Ce qui s'appelle impersonnel en italien est une construction où pour tout type de forme verbale comportant un sujet — qu'il s'agisse d'une forme inergative (transitive ou pas), ou d'une forme inaccusative (verbe inaccusatif, passif) (Cinque, 1988a), sauf les formes moyennes, le sujet est un clitique de 3ème personne, dont la forme est *si*, homonyme du réfléchi (Salvi, 1988:110; Monachesi, 1995).

(211) a. Ora si parte.
Maintenant SI-nom part.
(Maintenant on part.)

b. Si lavora sempre troppo.
SI-nom travaille toujours trop.

c. Non si è mai contenti.
NEG SI-nom est jamais contents.
(On n'est jamais contents)

d. Si è spesso trattati male.
SI-nom est souvent traités mal.
(On est souvent mal traités)

e. Si mangia spesso i funghi porcini qui.
SI-nom mange souvent les cèpes ici
(On mange souvent des cèpes ici.)

A la différence du *il* impersonnel français, ce *si* est interprété avec référence arbitraire (Manzini, 1986). Un équivalent peut être trouvé en français avec *on*.
Nous suivons Monachesi (1995) en considérant que le *si* impersonnel est un clitique nominatif[29]. Ce *si* sujet est spécial puisqu'il est le seul clitique nominatif en italien. Pour éviter la confusion avec la notion d'impersonnel connue pour le français, et pour lequel on a un autre équivalent en italien (voir Section 2.2.2) on désigne désormais le *si* impersonnel par SI-nominatif.
Le SI-nominatif est une forme homonyme de la forme réfléchie[30] à la troisième personne, mais néanmoins distincte, ce qui peut être montré d'une part par le fait que le SI-nominatif est compatible avec une forme verbale passive (211d) alors que le réfléchi ne l'est pas, d'autre part par l'interaction avec un clitique accusatif[31]. En (212a) le SI-nominatif suit le clitique accusatif, en (212b) le *si* réfléchi précède le clitique accusatif.

(212) a. Lo si deve fare così.
ACC.3ms SI-nom faire comme ça.
(On doit le faire comme ça.)

b. Se lo feccero dare da Mario
RF.3 ACC.3ms ils-firent donner par Mario.
(Ils se le firent donner par Mario.)

Nous avons retenu le statut de sujet pour le *si* impersonnel. Manzini (1986) remarque cependant qu'en "Italien du Nord", dialecte qui comprend des clitiques sujets, le clitique sujet se place devant le clitique objet (comme en français)[32]. Monachesi (1995) fait la même observation avec le fiorentino. Elle explique cela en rappelant qu'en diachronie, les différentes interprétations de *si* proviennent toutes de la construction avec vrai réfléchi, pour laquelle jusqu'au XIIIème siècle, l'ordre était accusatif<réfléchi. Cet ordre a évolué vers l'ordre inverse (212b). Le SI-nominatif a pu dériver de la construction réfléchie alors que l'ordre était encore accusatif<réfléchi, et n'aurait pas subi ce changement d'ordre.

En tant que réalisation particulière de la fonction sujet, ce SI-nominatif s'emploie de manière régulière dans une phrase finie, avec tout type de forme verbale comportant un

[29] Le *si* en italien a été l'objet de très nombreuses études, dans des cadres théoriques variés. Dans le cadre de GB, on peut citer Hyams (1986) et Manzini (1986), cette dernière fournissant une analyse complète de tous les emplois de *si* (impersonnel, moyen, réfléchi). Pour le *si* impersonnel, les deux auteurs considèrent que *si* reçoit le cas nominatif. Manzini développe cependant l'idée que *si* impersonnel n'occupe pas la position sujet, mais y est lié en formant une chaîne avec lui. La position sujet est alors remplie par une catégorie vide.
[30] Comme pour le français, on emploie le terme réfléchi pour les cas de réfléchi intrinsèque, vrai réfléchi, moyen.
[31] Cette observation, que l'on trouve chez Monachesi par exemple, est absente dans l'article de Manzini (1986), qui propose une analyse unifiée de tous les emplois de *si*. Son analyse pêche donc en n'expliquant pas la différence de comportement du SI-nominatif et du *si* réfléchi (ni le fait qu'ils peuvent être cooccurrents, cf. infra).
[32] Cette observation plus celle concernant le Modenese, où "l'élement impersonnel peut et doit être cooccurrent à un sujet clitique" explétif *a*, amène Manzini à la conclusion que le *si* impersonnel en italien n'est pas en position sujet. Il s'agit d'une conception structurale de la notion de sujet. Nous considérons le *si* impersonnel comme un sujet, n'ayant pas la position structurale des clitiques sujets des langues proches (dialectes et français).

sujet dont le référent peut être humain, mais pas avec un moyen. Plus généralement, étudions l'interaction du SI-nominatif avec les autres *si*.

2.2.1.1 Interaction avec les autres constructions faisant intervenir un *si*

On a vu que le SI-nominatif est à distinguer du réfléchi. Les deux *si* peuvent d'ailleurs être cooccurrents, le réfléchi se plaçant avant. La séquence *si si* est impossible et on a à la place *ci si*.

(213) a. Ci si aspetterebbe di trovarli nelle pianure. — réfléchi intrinsèque
RF.3 SI-nom attendrait de trouver-ACC.3mp dans les plaines (Monachesi, 1995:106)
(On s'attendrait à les trouver dans les plaines.)

b. Non ci si vede mai. — réciproque
NEG RF.3 SI-nom voit jamais (Monachesi, 1995:106)
(On ne se voit jamais.)

En revanche, le SI-nominatif, dans le cas d'un verbe transitif, est incompatible avec une forme verbale moyenne. On peut donner comme explication qu'une forme verbale moyenne est difficile avec un sujet (initialement objet) humain.

(214) *Ci si dice.
RF(moy) SI-nom dit[33].

Le SI-nominatif est d'ailleurs dans le cas transitif en concurrence avec le moyen, qui comporte aussi un clitique *si*, avec inversion du sujet de surface du moyen. Ainsi (211e) est en concurrence avec (215), où le verbe s'accorde avec le syntagme postverbal. (211e) n'est d'ailleurs pas accepté par tous les italophones. En revanche, tous les italophones acceptent le verbe au singulier (et donc la construction SI-nominatif) avec un clitique accusatif pour l'objet (216).

(215) Si mangiano spesso i funghi porcini qui.
RF(moy) mangent souvent les cèpes ici
(Les cèpes se mangent souvent ici)

(216) Li si mangia spesso qui.
ACC.3mp SI-nom mange souvent ici.
On les mange souvent ici.

2.2.1.2 Traits morphologiques du SI-nominatif

Dans le cas de temps simple, la forme verbale avec SI-nominatif est à la 3ème personne du singulier.
Pour les formes composées, le SI-nominatif partage avec les réfléchis le fait d'imposer l'auxiliaire de temps ESSERE, même dans les cas où l'auxiliaire habituel est AVERE (voir Section 2.6.1 pour les propriétés des réfléchis). En outre, dans les cas où l'auxiliaire est ESSERE indépendamment de la présence du SI-nominatif, la forme verbale composée montre des traits d'accord particuliers : l'auxiliaire est comme aux temps simples à la

[33] On a utilisé RF(moy) dans la glose pour faciliter la lecture, mais en fait, le moyen fait intervenir un marqueur qu'il n'y a pas lieu de distinguer du réfléchi.

3ème personne, mais le participe est en général masculin pluriel, et éventuellement féminin pluriel si la référence est clairement à un groupe féminin seulement[34].

(217) a. Si è lavorato bene. inergatif
SI-nom est travaillé bien.
(On a bien travaillé)

b. Si è usciti. inaccusatif
SI-nom est sortis.

c. Quando si è donne, si è disposte a rinunciare a molte cose ... (Salvi, 1991:242)
Quand SI-nom est femme, SI-nom est disposées à renoncer à beaucoup de choses...

Les traits d'accord du SI-nominatif avec la copule (218a) ou une forme passive (avec l'auxiliaire copule (218b) ou l'auxiliaire VENIRE (218c), cf. Section 2.3) sont également masculin pluriel.

(218) a. Si è trascurati.
SI-nom est négligents.

b. Si era ammirati da tutti.
SI-nom était admirés par tous.

c. Si viene messi da parte.
SI-nom vient METTRE.partp.mpl à l'écart.
(On est progressivement mis à l'écart.)

Pour une forme verbale avec SI-nominatif et clitique accusatif, il y a accord du participe avec le clitique accusatif, ce qui ne diffère pas du cas général.

(219) Le si è comprate.
ACC.3fp SI-nom est achetées.
(On les a achetées)

Pour expliquer cet accord particulier, nous considérons que le SI-nominatif porte les traits masculin pluriel[35], 3ème personne, et que le sujet impersonnel ne s'accorde pas en nombre avec la forme verbale, celle-ci prenant le nombre non marqué singulier.
Il faut en outre remarquer que l'auxiliaire ESSERE imposé par le SI-nominatif n'influe pas sur l'accord du participe passé ou de l'attribut. Alors on peut voir que l'accord avec le participe passé ou l'attribut est totalement régulier :

- pas d'accord du participe dans le cas où la forme prendrait l'auxiliaire AVERE si le sujet n'était pas le SI-nominatif (217a);
- accord du participe avec le sujet dans le cas d'une forme prenant ESSERE indépendamment du SI-nominatif (217b);
- accord avec le sujet de l'attribut (218a) et du participe passif (218b,c);
- accord du participe avec l'objet dans le cas d'un clitique accusatif (219)).

On verra section 2.6.1 que cela est différent pour les réfléchis, qui imposent l'accord du participe avec le sujet.

[34] C'est une situation semblable à celle observée pour *on* en français (*On est partis*), ou inversement le *vous* de politesse en français, qui est pluriel syntaxiquement, mais peut être singulier sémantiquement, ce qui se traduit dans l'accord du participe, ou de l'attribut.
[35] Cela ne permet cependant pas de traiter le cas d'accord féminin pluriel.

Donc même si considérer que le SI-nominatif ne s'accorde pas en nombre avec la forme verbale peut être discutable, cette analyse permet de circonscrire la particularité du SI-nominatif à cette seule irrégularité. L'accord avec le participe passé est alors totalement régulier[36].

2.2.2 Ce qui s'appelle impersonnel en français

En étudiant l'hypothèse inaccusative pour l'italien (Chapitre 3), nous avons déjà établi un parallèle entre ce qui s'appelle impersonnel en français et certaines constructions italiennes.
Une forme verbale à l'impersonnel en français correspond soit à un verbe inergatif dont le sujet est inversé (220), soit à une forme verbale inaccusative, dont l'objet n'est pas promu sujet (221), (222), (223). On désigne désormais par impersonnel en italien, les cas où la sous-catégorisation finale comporte un objet mais pas de sujet. L'équivalent en français impose un sujet explétif (cf. Section 1.1).

(220) a. Il a téléphoné trois personnes pour Marco.
b. Hanno telefonato tre persone per Marco.

(221) a. Il est arrivé trois personnes.
b. Sono arrivate tre persone.

(222) a. Il a été examiné de nouvelles solutions.
b. Sono state esaminate nuove soluzioni.

(223) a. Il se construit de nouveaux HLM.
b. Si costruiscono nuove case popolari.

2.2.3 Représentation dans la MG

Le SI-nominatif italien est donc représenté dans la MGi comme une réalisation particulière de la fonction sujet, en dimension 3, avec une contrainte de cooccurrence négative avec le moyen.
Comme on l'a vu ci-dessus, il partage avec les autres réfléchis le fait d'imposer l'auxiliaire de temps ESSERE (on permet pour cela que la valeur du trait :aux soit révisable[37]). La MGi comprend donc une classe si qui introduit cette propriété. La classe si-nominatif en hérite (ainsi que la classe si-réfléchi, cf. Section 2.6.4), et elle hérite aussi de la classe sujet-réalisé (voir Chapitre 5, la hiérarchie pour les réalisations de sujet). Pour représenter le clitique, on ajoute un trait pour la réalisation morphologique du clitique (il s'agit d'un trait amont, ce qui permet de gérer l'éventuelle montée du

[36] Une alternative est d'utiliser une théorie de l'accord à deux niveaux, qui distingue un accord sémantique et un accord morpho-syntaxique (Pollard et Sag, 1994; Monachesi, 95). L'accord du participe passé relevant de l'accord sémantique en italien, et l'accord sujet-verbe de l'accord morpho-syntaxique.

[37] Il ne nous est pas possible de donner une spécification déclarative (sans révision) de la sélection del'auxiliaire. On peut remarquer que la solution de Dini (1995) par exemple, qui est déclarative, ne s'avère pas suffisante si on prend en compte toute la grammaire. Sa solution repose sur le fait que l'auxiliaire *avere* est possible pour une forme verbale dont le trait SUJET est non vide et est nominal. Cela bloque bien l'auxiliaire *essere* pour les formes inaccusatives, et les formes avec si impersonnel (le si clitique n'apparaît pas comme valeur de SUJET mais apparaît comme valeur du trait CL pour clitiques). Mais on voit mal comment cela va ne pas bloquer l'auxiliaire *avere* dans un cas de pro-drop.

clitique nominatif (ou autre) sur un auxiliaire ou un verbe restructurant, cf. Chapitre 5, la représentation des clitiques, et de leur montée).

si
superclasses: top
constantes:
⇒•tête.aux = essere

SI-nominatif
superclasses: si sujet-réalisé
contraintes: ⇒ ⌉moyen
constantes:
•tête.num = sing •tête:t:cl:nom = +

Pour les locuteurs n'acceptant pas (211e), on peut ajouter la contrainte de cooccurrence négative avec la classe objet-position-canonique.

2.3 Le passif

Le passif en italien est très similaire à celui du français. Une différence est cependant que pour l'interprétation processive, trois auxiliaires sont possibles, la copule, VENIRE et ANDARE. La construction avec la copule est comme en français ambiguë entre l'interprétation processive et l'interprétation stative. Les deux autres auxiliaires ne sont compatibles qu'avec l'interprétation processive (cf. (224b), (224c)). Ils sont donc souvent utilisés lorsque la copule favorise l'interprétation stative (Salvi, 1988:91). L'auxiliaire ANDARE est ambigu entre une interprétation progressive, et une interprétation déontique (224d). Dans les deux cas le complément d'agent est impossible. Mais pour la construction à sens normative, l'agent est tout de même interprétable, et peut contrôler le sujet d'une infinitive. Ce contrôle est impossible avec l'interprétation progressive. Cela est illustré par le contraste entre (225a) où l'interprétation normative est impossible, et (225b) où elle est possible.

(224)a. Francesco è amato (E + da Paola).
Francesco est aimé (E + par Paola)
(Francesco est aimé (E + de Paola))

b. La porta venne chiusa (violentemente) (E + da Piero).
La porte vint fermée (violemment) (E + par Piero)
(La porte fut fermée (violemment) (E + par Piero).)

c. Molti libri sono andati smarriti durante il trasloco.
Beaucoup de livres sont allés perdus durant le démenagement
(Beaucoup de livres ont été perdus durant le déménagement.)

c. La lettera andava consegnata a Maria, non a Giovanni (E + *dal postino).
La lettre allait consignée à Maria, pas à Giovanni (E + *par le facteur)
(La lettre devait être consignée à Maria, pas à Giovanni)

(225) a. *Il libro andò distrutto per compiacere l'Inquisizione.

Le livre alla détruit pour satisfaire l'Inquisition.

b. Il libro andrà distrutto per compiacere l'Inquisizione.
Le livre ira détruit pour satisfaire l'Inquisition.
(Le livre devra être détruit pour satisfaire l'Inquisition.)

Comme pour le français, le sujet d'une forme passive peut être phrastique. En français, dans ce cas le passif impersonnel est préféré au passif personnel. En italien, cela se traduit par une position post-verbale de l'actant phrastique. La position préverbale (et donc une vrai promotion en position sujet) n'est cependant pas à exclure tant dans le cas d'une complétive (Salvi, 1988:95) que dans le cas d'une infinitive (Acquaviva, 1991:672), même s'il existe des conditions limitantes le verbe doit être accompagné d'éléments pouvant servir de focus, en (227b), *tutti*).

(226) a. E stato annunciato che il treno avrà del ritardo.
ESSERE-passécomp.3ms annoncé que le train aura du retard.
(Il a été annoncé que le train aura du retard)

b. Che il treno avrà del ritardo è stato annunciato con l'altoparlante.
(Que le train aura du retard a été annoncé au haut-parleur).

(227) a. E stato deciso di partire.
ESSERE-passécomp.3ms annoncé que le train aura du retard
(Il a été annoncé que le train aura du retard)

b. Di partire presto è stato consigliato a tutti.
De partir tôt a été conseillé à tous

Comme pour l'équivalent passif impersonnel en français, le passif impersonnel en italien (avec l'actant phrastique post verbal) impose le complémenteur *di* dans le cas de l'infinitif (227a)[38]. Mais dans le cas du passif personnel (où l'actant phrastique est promu sujet, et apparaît en position préverbale, (227b)), le *di* apparaît également.

Le complément d'agent est un complément prépositionnel en *da*. On utilise le même nom de fonction que pour le français, c'est-à-dire agt-objet pour la fonction du complément d'agent. Il est optionnel comme en français.
Dans la MGi comme dans la MGf, le passif est caractérisé par la démotion du sujet vers la fonction agt-objet[39].
Comme on l'a vu lors de la discussion de l'hypothèse inaccusative (Chapitre 3), on considère que la promotion de l'objet est optionnelle, on a donc un passif impersonnel et un passif personnel, ce qui est parallèle au français, la différence étant que l'italien permet des sous-catégorisations finales sans sujet, et donc l'impersonnel italien ne

[38] Acquaviva (1991:674) note cependant certains cas où l'omission du *di* est possible. Elle impose qu'un ajout ne soit pas présent.

(i) Non mi fu concesso (E + di) partire.
NEG DAT1.1 fut concédé (E + de) partir.
(Il ne m'a pas été permis de partir)

(ii) Non mi fu concesso (*E + di) partire immediatamente.
NEG DAT1.1 fut concédé (E + de) partir immédiatement.

[39] Pour les cas avec l'auxiliaire ANDARE il faut plutôt les rapprocher de la construction moyenne, qui ne permet pas l'expression de l'agent, et qui est traitée par la démotion du sujet vers la fonction ∅.

nécessite pas l'insertion d'un sujet explétif. On redonne les exemples de passif impersonnel et personnel, avec leurs sous-catégorisations initiale et finale :

- **passif impersonnel** = (sujet→agt-obj)

(228) Ne sono state rubate tre dal gatto.
NE.part sont ESSERE-partp.3fp volées trois par le chat[40].
(Il en a été volé trois par le chat)

		arg0	arg1
	sous-cat initiale	sujet	objet
(passif-impers)⇒	sous-cat finale	agt-objet	objet

- **passif personnel** = (sujet→agt-obj, objet→sujet)

(229) Tre torte sono state rubate dal gatto.
Trois tartes sont ESSERE-partp.3fp volées par le chat.
(Trois tartes ont été volées par le chat)

		arg0	arg1
	sous-cat initiale	sujet	objet
(passif-pers)⇒	sous-cat finale	agt-objet	sujet

Le passif impersonnel est également possible pour quelques cas rares d'intransitifs (inergatifs) (Salvi, 1988:96).

(230) a. Fu proceduto al sequestro.
Fut procédé au séquestre.
(Il fut procédé au séquestre.)

b. Gli è stato detto del pericolo.
DAT.3ms est ESSERE-partp.3ms dit du danger.
Il lui a été parlé du danger.

Comme en français le passif impersonnel est en concurrence avec un sujet à référence arbitraire, c'est-à-dire *on* en français et SI-nominatif en italien.

(231) a. Picchiate e vi sarà aperto.
Frappez et cltq.dat2pl sera ouvert.
(Frappez et il vous sera ouvert.)

b. Picchiate e vi si aprirà.
Frappez et cltq.dat2pl SI-nominatif ouvrira.
(Frappez et on vous ouvrira.)

Le fait que le passif impersonnel soit impossible pour des intransitifs inergatifs (Salvi, 1988:96) peut être bien prédit par la généralisation suivante, à laquelle on a déjà abouti pour le français :

Seul un sujet initial peut subir sujet→agt-objet.

[40] La copule ESSERE "auxiliaire" du passif se conjugue avec ESSERE en italien.

2.3.1 L'équivalent du passif en *de* français

Belletti et Rizzi (1988) fournissent une étude des verbes psychologiques, et montrent l'existence d'un passif en *di* en italien.

(232) Piero è/fu sorpreso della tua venuta.
Piero est/fut surpris de ta venue.

L'agt-objet ne peut pas être phrastique en italien. En revanche le gén-objet peut l'être. Les passifs de verbes à sujet phrastique seraient donc des cas de passifs en *di*.

2.3.2 Interactions avec les autres redistributions

Comme en français, le passif est incompatible avec le vrai réfléchi. On verra Section 2.5.1 qu'il en est de même pour la construction causative italienne caus-agt (c'est-à-dire où comme en français, le causé est réalisé comme un agt-objet).

On a tout de même une différence avec le français, car le passif peut s'appliquer sur un complexe causatif, ce qui donne une preuve supplémentaire de la formation d'un prédicat complexe causatif en italien. On renvoie à la section 2.5.1 pour des exemples.

2.4 Le moyen

On a vu lors de l'étude des formes inaccusatives en italien (Chapitre 3), qu'il existe en italien une construction parallèle à la construction moyenne française, également est caractérisée par une morphologie réfléchie, et par la démotion du sujet. Il s'agit de formes inaccusatives, comportant une promotion de l'objet optionnelle, on parle donc de moyen impersonnel et de moyen personnel. Nous redonnons l'analyse en termes de sous-catégorisations initiales et finales pour les deux cas :

- **moyen impersonnel** = moyen avec avec objet non promu (montré par le *ne* partitif)

(233) Se ne sono affitate molte ad un prezzo eccessivo.
RF(moy) NE.part sont louées beaucoup.fp à un prix excessif.
(Il s'en est loué beaucoup à un prix excessif.)

		arg0	arg1
	sous-cat initiale	sujet	objet
(moyen-impers)⇒	sous-cat finale	∅	objet

- **moyen personnel** = moyen avec objet promu sujet

(234) Molte case si sono affitate ad un prezzo eccessivo.
Beaucoup.fp maisons RF(moy) sont louées à un prix excessif.
(Beaucoup de maisons se sont louées à un prix excessif.)

		arg0	arg1
	sous-cat initiale	sujet	objet
(moyen-pers)⇒	sous-cat finale	∅	sujet

Dans le cas impersonnel, le verbe s'accorde avec son objet (cf. Chapitre 3, section sur l'inaccusativité).
Le *si* montré par la construction moyenne diffère du SI-nominatif, car pour le moyen, on peut trouver des exemples à la 1ère ou 2ème personne. On considère que les formes moyennes comportent un réfléchi, au même titre que les formes avec réfléchi intrinsèque ou vrai réfléchi (voir Section 2.6.1 les propriétés communes aux formes verbales réfléchies).

(235) Mi trasporto facilmente, sono pieghevole, sono utilissima! (dice la valigia).
Refl.2 transporte facilement, je-suis pliable, je-suis utile-superlatif (dit la valise).
Je me transporte facilement, je suis pliable, je suis super utile! (adapté de Zribi-Hertz, 1982).

L'expression d'un complément d'agent étant impossible avec le moyen italien (comme français), le moyen est caractérisé par la démotion du sujet vers la fonction ∅. On peut s'attendre, comme pour la démotion vers agt-objet, à une contrainte limitant la démotion du sujet vers la fonction ∅ au sujet initial. Nous pouvons en effet prouver l'adéquation de cette contrainte, ce que nous faisons ci-dessous.

2.4.1 Contraintes et interactions avec les autres redistributions

Le moyen impersonnel en italien est ici analysé de manière directe, comme résultant de la démotion spontanée (=sans promotion de l'objet) du sujet d'un verbe vers la fonction "vide". Or pour le français, on a analysé le moyen impersonnel comme la composition de deux redistributions, le moyen (personnel) et l'impersonnel, étant donné qu'on n'a pas de moyen impersonnel pour des intransitifs. L'analyse "directe" du moyen impersonnel en italien est supportée par les données, puisqu'il y a en italien des cas de moyen pour des intransitifs. Mais cette affirmation doit être justifiée, étant donné que ces cas sont ambigus syntaxiquement avec une construction avec SI-nominatif.

(236) a. Si lavora sempre troppo.
RF(moy) travaille toujours trop / SI-nom travaille toujours trop.
(*Il se travaille toujours trop / On travaille toujours trop.)

b. Si arriva spesso in ritardo.
SI-nom arrive souvent en retard.
(On arrive souvent en retard.)

Nous affirmons que (236a) est ambigu entre moyen et SI-nominatif, pour certains locuteurs (dont Cinque), alors que (236b) ne l'est pas. Nous nous appuyons pour cela sur les dissymétries relevées par Cinque (1988a) pour les constructions en *si* des intransitifs, selon qu'ils sont inaccusatifs ou inergatifs, dans certains contextes comme sous un verbe à montée par exemple.

Reprenons l'examen des données de Cinque. Le SI-nominatif est possible pour toutes les classes de verbes, dont tous les intransitifs, avec un mode fini (cf. la série (211)), et est généralement exclu des propositions de mode non fini. Donc sous un verbe (ou locution) à contrôle (comme (237)), on n'a ni un verbe inaccusatif avec un SI, ni un inergatif, ni un transitif (qu'on analyse le SI comme un SI-nominatif ou comme un réfléchi marqueur de moyen) (238a,b,c).

(237) Sarebbe meglio scoprire il vero colpevole. structure à contrôle (sans SI)
Serait mieux découvrir le vrai coupable
(Il vaudrait mieux découvrir le vrai coupable.)

(238) a. *Sarebbe meglio scoprirsi il vero colpevole. transitif
Serait mieux DECOUVRIR-SI-nom/RF(moy) le vrai coupable

b. *Sarebbe meglio lavorarsi un po' di più. inergatif
Serait mieux TRAVAILLER-SI-nom/RF(moy) un peu plus

c. *Sarebbe meglio arrivarsi puntuali. inaccusatif
Serait mieux ARRIVER-SI-nom ponctuels

Mais Cinque montre que certains contextes avec mode non fini sont possibles avec un SI. Il s'agit des verbes à montée et de la construction Aux-to-Comp. Nous nous limitons aux verbes à montée ici. Cinque montre qu'un inaccusatif (239) ne peut pas porter un *si* sous un verbe à montée, mais qu'un verbe inergatif (240) ou un verbe transitif (241) le peuvent.

(239) *Sembra essersi arrivati troppo tardi. inaccusatif
Semble ESSERE-SI-nom arrivés trop tard.

(240) Sembra non essersi lavorato a sufficienza. inergatif (moy. impers.)
Semble neg. ESSERE-réfl.3 travaillé suffisamment.
(Il semble ne pas s'être travaillé suffisamment.)

(241) a. Sembr*ano* essersi vendut*e* poche automobile. transitif (moy. pers. ou impers)
Semblent neg. ESSERE-réfl.3 vendues peu.fpl autos.
(Peu d'autos semblent s'être vendues)

b. Sembrano essersene vendute poche. transitif (moy. impers.)
Semblent ESSERE- réfl3-partitif vendues peu.fpl.
(Il semble qu'il s'en est vendu peu.[41])

c. *Sembra essersi vendut*o* poche automobili. transitif (SI-nom)
Sembra neg. ESSERE-SI-nom vendu peu.fpl voitures.

d. *Sembra/*Sembrano esserlesi vendute a un prezzo eccessivo.
Semble/Semblent essere-acc.3fpl-réfl.3 vendues à un prix excessif.

Nous expliquons cette dissymétrie par le fait que pour les inaccusatifs, le moyen (impersonnel) n'est pas possible, car le moyen met en jeu la démotion d'un sujet initial uniquement.

L'impossibilité d'avoir un *si* (moyen ou nominatif) sous un verbe à contrôle est expliqué de la façon suivante:
Pour avoir une infinitive sous contrôle, il faut que la sous-catégorisation du verbe infinitif comprenne un sujet, qui n'est pas réalisé. La non-réalisation (phonologique) du sujet des infinitives est un cas particulier de "consommation" du sujet dans la liste de sous-catégorisation. Donc :

[41] On a traduit avec un autre emploi de SEMBLER en français, qui prend une complétive. En italien il s'agit toujours de l'emploi de SEMBRARE comme verbe à montée. L'équivalent *syntaxique* français serait *Peu semblent s'en être vendues.*

(i) L'analyse de l'infinitive en (238a) comme une construction moyenne personnelle (= avec objet promu sujet) est exclue, car *il vero colpevole* serait le sujet (réalisé) de l'infinitif, ce qui n'est pas possible sous un verbe à contrôle. (238b,c) ne peuvent pas non plus être des moyens personnels, puisqu'ils sont intransitifs au départ.

(ii) L'analyse de (238a,b,c) comme des moyens impersonnels (l'objet éventuel n'est pas promu sujet) est également exclue, car la sous-catégorisation de l'infinitif ne comporte alors pas de sujet, et donc le cas particulier de "non-réalisation" du sujet pour une infinitive sous contrôle ne peut pas s'appliquer.

(ii) (238a,b,c) ne peuvent pas non plus s'analyser comme des cas de SI-nominatif. En effet, le SI-nominatif étant une réalisation du sujet, il est en distribution complémentaire avec toutes les autres réalisations de sujet (cf. le principe d'unicité fonctionnelle), y compris le cas particulier de non-réalisation.

La dissymétrie sous un verbe à montée entre inaccusatifs (exclus) et inergatifs et transitifs (possibles), s'explique comme suit. Sous un verbe à montée, le SI-nominatif en tant que réalisation du sujet est impossible. Donc seul le moyen est éventuellement possible. Or le moyen, qui met en jeu la démotion d'un sujet obligatoirement initial, n'est pas possible pour un inaccusatif, d'où l'impossibilité de (239).
Au contraire (240) et (241a) mettent en jeu des verbes comportant un sujet initial, et peuvent donc s'analyser comme des cas de moyen. (240) est un moyen impersonnel, (241a) est ambigu entre un moyen personnel avec sujet inversé, et un moyen impersonnel. (241b) est seulement un moyen impersonnel, puisque l'argument postverbal donnant lieu au partitif ne peut pas être un sujet. On explique l'inacceptabilité de (241c) par le fait qu'il ne peut pas s'agir d'un moyen, car il devrait y avoir accord entre le verbe et le syntagme postverbal, donc il ne peut s'agir que d'un SI-nominatif, exclu sur une infinitive. Pour (241d), on a déjà vu que pour une sous-catégorisation sans sujet, l'objet non promu ne peut pas se réaliser comme un clitique accusatif.

Donc, en résumé, le SI-nominatif est exclu sur tous les contextes non finis où la réalisation d'un sujet est exclue (montée et contrôle). La différence entre montée et contrôle, c'est que le verbe à contrôle requiert un infinitif dont la sous-catégorisation comprend un sujet, "consommé" de manière particulière (la "non-réalisation"), alors que le verbe à montée, qui ne sous-catégorise pas de sujet, n'impose pas cette contrainte.
Reste à bien rendre compte de la différence entre montée (permettant une infinitive au moyen) et contrôle (ne le permettant pas). C'est immédiat si on rappelle le traitement des verbes à montée, comme verbes "déficients", qui à la différence des verbes à contrôle, ne sous-catégorisent pas de sujet (cf. Chapitre 3). Le verbe à montée permet donc une infinitive dont la sous-catégorisation ne comprend pas de sujet (et pour un verbe à montée, les clitiques éventuels se placent sur l'infinitif). Le verbe à montée s'accorde avec la forme verbale infinitive et, par transitivité, avec le sujet s'il existe, sinon avec l'objet.

Donc les données justifient la généralisation suivante, parallèle à celle du français :

Seul un sujet initial peut subir sujet→∅.

Il faut cependant noter que les cas que nous avons analysé comme des moyens impersonnels sous un verbe à montée ne sont pas unanimement acceptés par les locuteurs. Si certains font effectivement une différence entre inaccusatifs et inergatifs (contraste (239)/(240)), d'autres rejettent les deux cas.
Cela revient à interdire le moyen impersonnel pour les intransitifs. C'est le cas pour le français, pour lequel on n'a donné qu'une description du moyen personnel, le moyen impersonnel étant analysé comme la composition (moyen ○ impersonnel). Pour l'italien cela est plus délicat, nous préférons ajouter une contrainte sur la démotion sujet→∅, qui impose la présence d'un objet, cette contrainte n'etant pas nécessaire pour certains locuteurs.

Comme autre contrainte sur le moyen, on a déjà mentionné son incompatibilité avec le SI-nominatif. D'après les contraintes déjà mises en place, rien n'empêche que le sujet (objet promu sujet) d'une forme moyenne se réalise comme un SI-nominatif, comme cela est possible pour le sujet d'une forme passive. Or c'est impossible, et nous n'avons pas d'explication. On peut cependant ajouter une contrainte de cooccurrence négative sur les classes concernées pour bloquer cette interaction.

Le moyen interagit avec le causatif, ce que nous verrons Section 2.5.1.

2.5 Le causatif

La construction causative en italien est parallèle à celle du français. Dans ce qui suit on donne des exemples avec FARE, bien que LASCIARE, et certains verbes de perception entrent dans la même construction (Skytte et al., 1991:499). On a différents arguments pour opter pour une analyse avec prédicat complexe :

- Comme en français le placement des clitiques se fait sur le verbe causatif. Une construction concurrente (la structure causative à contrôle en français) ne semble pas être présente en italien. Le causé et les C(Vinf) se placent toujours sur FARE.

(242) a. Lo faccio uscire.
ACC.3ms je-fais sortir.
(Je le fais sortir)

b. Lo faccio mangiare a Piero.
ACC.3ms je-fais manger à Piero.
(Je le fais manger à Piero)

c. Gli faccio mangiare la minestra.
DAT.3ms je-fais manger la soupe.
(Je lui fais manger la soupe)

d. Gli farò scrivere una lettera da Giorgio.
DAT.3ms je-ferai écrire une lettre par Giorgio.
(Je lui ferai écrire une lettre par Giorgio)

e. Ci farò andare Giorgio.
LOC je-ferai aller Giorgio.
(J'y ferai aller Giorgio)

f. Ne farò uscire Giorgio.
NE.loc.prov je-ferai sortir Giorgio.

(J'en ferai sortir Giorgio)

g. Ne farò partire molti.
NE.part je-ferai partir beaucoup.
(J'en ferai partir beaucoup)

h. Ne farò mangiare molti a Giorgio.
NE.part je-ferai manger beaucoup à Giorgio.
(J'en ferai manger beaucoup à Giorgio)

- Le prédicat complexe peut être mis au passif et au moyen (Skytte et al., 1991:502-503), à la différence du français.
  La série (243) montre la succession causatif puis passif pour un inaccusatif; la série (244) le causatif puis passif pour un inergatif. Enfin la série (245) montre le causatif puis moyen.

(243) a. Il libro cadde.
Le livre tomba.

b. Paolo fece cadere il libro.
Paolo fit tomber le livre.

c. Il libro fu fatto cadere da Paolo.
Le livre fut fait tomber par Paolo

(244) a. Maria fa leggere il libro agli studenti.
Maria fait lire le livre aux étudiants

b. Il libro è stato fatto leggere agli studenti da Maria
Le livre est été fait lire aux étudiants par Marie

(245) a. I bambini escono prima.
Les enfants sortent en premier.

b. L'insegnante fa uscire i bambini prima.
L'instituteur fait sortir les enfants en premier.

c. I bambini si fanno uscire prima.
Les enfants RF(moy) font sortir en premier.[42]

Du fait de cette propriété et de l'absence de construction concurrente, on peut conclure que la formation du prédicat complexe est nettement plus claire en italien.
Mais comme en français, du point de vue de certaines propriétés, comme le liage, la construction causative est bi-clausale (Webelhuth, 1998). Le réfléchi non clitique (voir infra l'interaction avec le réfléchi clitique) s'il est C(Vinf) ne peut pas être lié au causateur.

(246) a. Con le minacce fecero accusare se stesso/*stessi a Giovanni.
Sous la menace ils-firent accuser lui-même/*eux-mêmes à Giovanni.

Parallèlement au français, on distingue parmi la construction causative les cas suivants, qui diffèrent par les redistributions en jeu (on donne entre parenthèses les séquences de redistributions partielles pour chaque redistribution) :

[42] En français *Les enfants se font sortir en premier* pourrait être acceptable mais avec une interprétation agentive des enfants. Il ne s'agit pas d'une construction moyenne comme en italien.

- **caus-i** : (sujet→objet, causateur→sujet)

N'est possible que si la fonction objet n'est pas occupée par un C(Vinf). Cela est bien représenté car la redistribution partielle sujet→objet ne serait pas applicable sinon.

(247) Paolo fa piangere Francesca.
Paolo fait pleurer Francesca

- **caus-a** : (sujet→a1-objet, causateur→sujet)

N'est possible que si la fonction sujet est déjà occupée par un C(Vinf). Ca n'est pas représenté tel quel, il faut ajouter la contrainte d'existence d'un objet.

(248) Paola fa mangiare la minestra a Piero.
Paola fait manger la soupe à Piero

- **caus-agt** : (sujet→agt-objet, causateur→sujet)

N'est possible que si la fonction sujet est déjà occupée par un C(Vinf). Ca n'est pas représenté tel quel, il faut ajouter la contrainte d'existence d'un objet.

(249) Farà correre i 100 m da Mario.
Il-fera courir les 100 m par Pierre

A la différence du français, certains intransitifs admettent cependant caus-agt (250b) et caus-a. Ainsi (250a) est ambigu entre une interprétation où *Piero* est le causé ou un C(Vinf) (Skytte et al., 1991:505).

(250) a. Farò telefonare a Piero.
Je-ferai téléphoner à Pierre
(Je ferai téléphoner à Pierre / Je ferai téléphoner Pierre)

b. Farò telefonare da Piero.
Je-ferai téléphoner par Pierre

Nous ne traitons pas ces cas dans la MGi.

Comme en français, si le passif n'est pas possible pour un verbe, caus-agt ne l'est pas non plus. Cela est bien représenté par l'utilisation de sujet→agt-objet comme redistribution partielle commune au passif (personnel et impersonnel) et à caus-agt (que l'on peut considérer comme un cas de passif). Aussi l'expression figée *sbarcare il lunario* (joindre les deux bouts) n'est pas passivable (251a), elle admet caus-a (251b) mais pas caus-agt (251c).

(251) a. *Il lunario è sbarcato da molta gente.
(*Les deux bouts seront joints par beaucoup de gens)

b. Ciò farà sbarcare il lunario a molta gente.
(Cela fera joindre les deux bouts à beaucoup de gens)

c. *Ciò farà sbarcare il lunario da molta gente.
(Cela fera joindre les deux bouts par beaucoup de gens)

- **caus-spontané** : (causateur→sujet)

On a vu pour le français une redistribution qui introduit un argument causateur "spontané", c'est-à-dire qui ne s'accompagne pas de la démotion d'un sujet. Pour le français, cette redistribution est limitée aux verbes atmosphériques qui ne sous-catégorisent pas de sujet-initial. Pour l'italien, elle est également utilisée pour les formes inaccusatives.

Ainsi, le causé dans le cas d'un verbe inaccusatif est objet initial, qui reste objet en surface, et un causateur est introduit.

(252) a. Il libro è caduto.
Le livre est tombé

b. Paolo ha fatto cadere il libro.
Paolo a fait tomber le livre

		arg0	argC
	sous-cat initiale	objet	-
(caus-spontané)⇒	sous-cat finale	objet	sujet

2.5.1 Contraintes

Il existe la même différence sémantique entre caus-agt et caus-a. Avec caus-a, le causé est "affecté" par l'évènement, ce qui n'est pas impliqué par caus-agt. On peut comparer (253a) et (253b).

(253) a. Farà correre i 100 m a Mario perchè se lo merita.
(Je ferai courir le 100 m à Mario parce qu'il le mérite.)

b. Farà correre i 100 m da Mario perché Piero è ammalato.
(Je ferai courir le 100 m par Mario parce que Piero est malade.

Le rôle discursif du causé a1-objet peut rendre impossible la réalisation clitique de certains C(Vinf). On retrouve les contraintes dont on a parlé pour le français (Section 1.4.1.3):

- un C(Vinf) a1-objet ne peut jamais être cliticisé avec caus-a

(254) Gli farò scrivere una lettera a Maria. (Gli = causé)
(Je lui ferai écrire une lettre à Marie.)

- un C(Vinf) a1-objet ne peut pas toujours être cliticisé avec caus-i (on a le contraste (255a)/(255b)). Mais en italien on dispose de la partition des intransitifs en inaccusatifs/inergatifs (ce que l'on n'a pas retenu pour le français). Donc dans la MGi, on pourrait représenter le contraste (255a)/(255b) en syntaxe, en imposant par exemple que pour un C(Vinf) clitique datif, l'objet du complexe causatif ne doit pas être un sujet initial. Cela bloquerait (255a), mais malheureusement pas une analyse de (254) où le clitique n'est pas le causé, alors que cette interprétation n'est pas possible. On voit que l'on ne peut pas renoncer à une explication discursive de certaines contraintes sur le causatif en italien, bien que l'on dispose de l'hypothèse inaccusative.

(255) a. *Gli farò telefonare Piero.
(*Je lui ferai téléphoner Pierre.)

b. Gli farò sembrare il compito più facile di quanto non sia.
(*Je lui ferai sembler le travail plus facile que ce qu'il n'est).

- On ne peut pas avoir un causé a1-objet avec en plus un clitique actanciel de 1ère ou 2ème personne ou un réfléchi (Skytte et al., 1991:508)

(256) a. *Paolo mi fece abbracciare a Maria.
(*Paolo me fis embrasser à Maria.)

b. *Piero si farà ascoltare a Giovanni.

(*Piero se fera écouter à Giovanni.)

2.5.2 Interaction avec les autres redistributions

On a donné comme arguments en faveur de l'analyse du causatif italien comme un prédicat complexe, le fait qu'il puisse être mis au passif et au moyen. On redonne les exemples, avec leur représentation en termes de sous-catégorisations successives.

(243) c. Il libro fu fatto cadere da Paolo.
Le livre fut fait tomber par Paolo

		arg0	argC
	sous-cat initiale	objet	-
(caus-spontané)⇒	intermédiaire	objet	sujet
(passif-pers)⇒	sous-cat finale	sujet	agt-objet

(244) b. Il libro è stato fatto leggere agli studenti da Maria
Le livre est été fait lire aux étudiants par Marie

		arg0	arg1	argC
	sous-cat initiale	sujet	objet	-
(caus-a)⇒	intermédiaire	a1-objet	objet	sujet
(passif-pers)⇒	sous-cat finale	a1-objet	sujet	agt-objet

(245) c. I bambini si fanno uscire prima.
Les enfants RF(moy) font sortir en premier.

		arg0	argC
	sous-cat initiale	objet	-
(caus-spontané)⇒	intermédiaire	objet	sujet
(moyen-pers)⇒	sous-cat finale	sujet	∅

Donc dans la MGi, on n'a pas comme en français la contrainte générale, qu'un argument causateur ne peut pas être démis. Dans la classe causateur→sujet de la MGi, la fonction du causateur introduit est révisable, au contraire de ce qui a été donné pour l'équivalent dans la MGf (Section 1.4.1.3).

L'interaction inverse n'est pas possible. Comme en français, on n'a pas de causatif sur un passif, ce qui peut être expliqué par l'impossibilité d'avoir la morphologie passive derrière un causatif. La morphologie du verbe suivant le verbe causatif est d'ailleurs plus contrainte qu'en français, puisque l'infinitif ne peut pas porter de réfléchi (cf. Section 2.6.2). Donc si le causatif est possible sur un moyen, comme le réfléchi est omis, on ne peut pas le distinguer d'un simple caus-i.

(257) a. *La crisi fa vendersi questi libri.
La crise fait VENDRE.RF(moy) ces livres
(La crise fait se vendre ces livres)

b. La crisi fa vendere questi libri.
La crise fait vendre ces livres

(La crise fait vendre ces livres / la crise fait se vendre ces livres)

2.6 La réfléxivation

2.6.1 Propriétés communes aux verbes réfléchis

- L'auxiliaire de temps des verbes réfléchis est ESSERE.
  Cette propriété est indépendante de l'auxiliaire de temps des verbes correspondants non réfléchis, quand ils existent. On a vu Section 2.2.1.2 que cette propriété est partagée également avec le SI-nominatif.
- Accord du réfléchi
  Pour ce qui concerne l'accord du réfléchi, on a du considérer pour le français (Section 1.5.1.1), que le réfléchi s'accorde en personne avec le verbe, du fait de son comportement sous l'impersonnel et le causatif en français.
  Pour l'italien, dans les constructions que l'on a considérées comme impersonnelles (sans sujet), on ne peut pas tester l'accord, car ces constructions empêchent justement la réalisation de leur argument comme un clitique accusatif. On ne peut donc pas donner des exemples équivalents à *Il ne s'est lavé que toi dans cette fontaine*. Mais comme pour une forme impersonnelle, il n'y a pas de sujet, on ne peut pas simplement dire que le réfléchi s'accorde avec le sujet. Il nous faut comme pour le français dire que le réfléchi s'accorde avec la forme verbale, et cette forme verbale, dans le cas impersonnel, s'accorde avec l'argument objet s'il existe.
- Accord du participe passé
  A la différence du français, tous les réfléchis, y compris les vrais réfléchis alternant avec un a1-objet (258b) s'accordent avec le sujet, sauf dans le cas de la présence d'un clitique accusatif (258d), qui "a priorité" pour l'accord avec le participe (Salvi, 1991:240-241) (l'objet extrait non clitique ne s'accorde pas avec le participe (258c). Donc on n'a pas besoin de distinguer deux SE.

(258) a. Chiara ha mangiato i biscotti.
Chiara a mangé les biscuits.

b. Chiara si è mangiata i biscotti.
Chiara s'est mangée les biscuits.

c. I biscotti che Paola si è mangiata.
Les biscuits que Paola s'est mangée.

d. Chiara se li è mangiati tutti.
Chiara se les est mangés tous.

Avec un accusatif indéfini (*ne*), il peut y avoir encore accord avec le sujet

(259) a. Di biscotti, Francesca non se ne è mangiati/mangiata per niente.
Des biscuits, Francesca neg. s'en est mangés/mangée du tout.

2.6.2 Le vrai réfléchi est-il intransitivant en italien?

Pour Monachesi (1995) dans les constructions à sens réfléchi ou réciproque, le réfléchi joue le rôle d'un actant grammatical, mais il est un marqueur dans le cas du moyen. Mais

comme pour le français, on a des arguments en faveur de cette analyse (le vrai réfléchi joue le rôle d'un actant) et également en faveur de l'analyse comme marqueur (où le réfléchi marque une réduction de la valence).
Pour le français, les arguments en faveur de l'analyse "marqueur", sont le comportement intransitif du réfléchi par rapport à l'impersonnel et au causatif.

L'équivalent d'une construction à l'impersonnel en italien est soit une sous-catégorisation sans sujet dans le cas d'une forme inaccusative, soit une inversion du sujet dans le cas inergatif. Cela ne constitue pas un test de transitivité pour l'italien.

Reste le test du causatif. D'une manière générale, on ne peut pas trouver un infinitif portant un réfléchi sous un causatif (Skytte et al., 1991:509). Cela est vrai pour un réfléchi intrinsèque (260b), pour un neutre (261b), et pour un vrai réfléchi (262b). En dehors du cas de réfléchi intrinsèque, l'interprétation réfléchie est en concurrence avec une interprétation non réfléchie.

(260) a. Giovanni si vergogna.
Giovanni réfl.intrinsèque a-honte
(Giovanni a honte.)

b. Le sue parole fecero vergognare Giovanni.
Ses mots firent avoir-honte Giovanni.

(261) a. La macchina si ferma.
(La voiture s'arrête.)

b. Paolo fece fermare la macchina.
Paolo fit arrêter la voiture.
(Paolo fit s'arrêter la voiture / Paolo fit arrêter la voiture.)

(262) a. Piero si lava.
(Piero se lave.)

b. La mamma ha fatto lavare Piero.
(La mamma a fait se laver Pierre / La mamma a fait laver Pierre)

L'important est qu'on peut avoir l'interprétation réfléchie avec le causé se réalisant comme un objet (et pas un a1-objet). On en conclut que, comme en français, une forme avec réfléchi accusatif est intransitif pour le causatif.
D'autre part l'interprétation réfléchie n'est pas possible avec caus-agt (comparer (263b) ambigu à (263c) non ambigu).

(263) a. Paolo si da uno schiaffo.
(Paolo se donne une gifle.)

b. Il regista ha fatto dare uno schiaffo a Paolo.
(Le metteur en scène a fait se donner une gifle à Paolo
/ Le metteur en scène a fait donner une gifle à Paolo)

c. Il regista ha fatto dare uno schiaffo da Paolo.
(Le metteur en scène a fait donner une gifle par Paolo)

Cela est parallèle au français, et vaut aussi pour le passif après un réfléchi. Donc on peut donner comme pour le français la généralisation :

Un argument auquel est lié sémantiquement un réfléchi

ne peut pas subir sujet→agt-objet.

2.6.3 Contraintes et interaction avec les autres redistributions

On vient de voir le cas du causatif avec un infinitif réfléchi (la généralisation qu'on vient de donner vaut aussi pour bloquer le passif pour un verbe portant un réfléchi).
Inversement, un prédicat complexe causatif peut porter un réfléchi (placé sur le verbe causatif).

- un C(Vinf) objet peut donner lieu à réfléchi (Skytte et al., 1991:501):

(264) Paolo si fa picchiare dai bambini più piccoli. caus-agt°réfl-acc
Paolo se fait frapper par les enfants plus petits

- comme en français, on ne peut pas avoir de réfléchi sur le verbe causatif dans le cadre de caus-a (Skytte et al., 1991:508), à moins que ce réfléchi soit le causé. On a déjà mentionné cela Section 2.5.1, et comme pour le français, on considère qu'il faut l'expliquer par des contraintes discursives.

(256) b. *Piero si farà ascoltare a Giovanni.
Piero se fera écouter à Giovanni.

- un C(Vinf) a1-objet peut subir réfléchi après caus-agt

(265) Piero si è fatto spedire un pacco dai genitori.
(Piero s'est fait expédier un paquet par ses parents.)

2.6.4 Représentation dans la MG

Etant donné les parallèles établis entre les données françaises et italiennes, on traite la réflexivation comme en français, par démotion soit de l'objet (ce qui rend intransitif le verbe) soit de l'a1-objet.
On a vu Section 2.2.3 la classe si, qui introduit l'auxiliaire ESSERE. On définit la classe si-réfléchi qui en hérite et qui ajoute (cf. Section 2.6.1) l'accord du participe avec le sujet (on n'a pas besoin comme en français de distinguer deux classes SE-obj et SE-a-obj). La constante représentant le sujet initial reçoit le trait :lie-réfl = +, ce qui permet de bloquer l'application de sujet→agt-objet sur un verbe portant un réfléchi.
Le traitement est plus simple qu'en français, parce qu'on n'a pas à gérer la réalisation de clitiques sur l'infinitif dans le cas d'un réfléchi. On donne Figure 62 le schéma des classes pour représenter le *si* (nominatif ou réfléchi). Le SI-nominatif est une classe terminale de la dimension 3 (i.e. elle représente une réalisation de sujet), qui hérite cependant de le classe si en dimension 2.

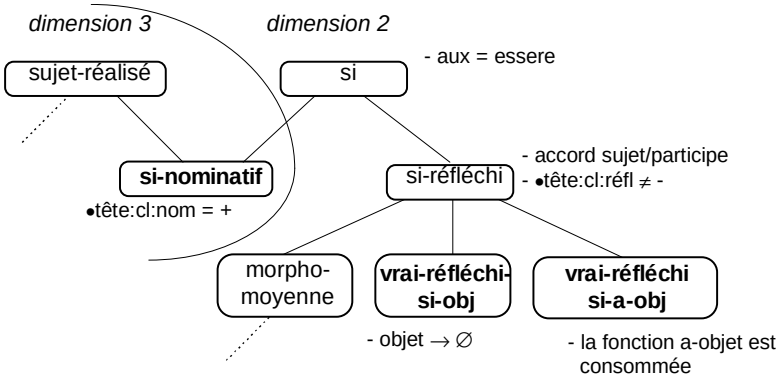


Figure 62. Les classes pour représenter le *si* (nominatif ou réfléchi), dans la MGi (avec en gras les classes terminales)

2.7 Récapitulatif

On donne Figure 61 le schéma pour les redistributions des verbes dans la MGi[43]. Les différences avec le français sont :

- On n'a pas besoin de distinguer entre deux réfléchis, dont l'un seulement déclenche l'accord du participe avec l'objet. En italien, on peut considérer que le réfléchi datif et accusatif déclenchent l'accord du participe avec le sujet, même si un éventuel accusatif clitique a priorité pour l'accord avec le participe.
- On n'a pas d'insertion d'un impersonnel comme sujet (classe de la MGf impers→sujet).
  Donc, d'une part on n'a pas de redistribution impersonnel. La démotion du sujet vers l'objet n'est utilisée que pour la construction caus-i.
  D'autre part, le passif impersonnel et le moyen impersonnel sont analysés simplement comme une démotion du sujet (vers agt-objet ou vers ∅) plus une morphologie marquée. Il y a cependant dans ce cas accord avec l'objet.
- Pour les locuteurs acceptant un *si* sous verbe à montée, pour les intransitifs inergatifs (exemple (240)), on conclut que le moyen impersonnel peut s'appliquer à des intransitifs (inergatifs seulement), contrairement au français, pour lequel on n'a que des moyens impersonnels pour des transitifs.
- On a une promotion spontanée d'un objet en position sujet, utilisée pour les verbes intransitifs inaccusatifs.
- Enfin, la classe caus-spontané n'est pas seulement utilisée pour des verbes atmosphériques dans un complexe causatif, mais également pour un verbe inaccusatif dans un complexe causatif.

En ce qui concerne les compositions de redistributions, elles sont similaires au français, sauf pour ce qui concerne les complexes causatifs passifs et moyens (encore que ces

[43] Le schéma est légèrement simplifié : on n'a pas noté la distinction entre morphologe passive avec ou sans copule.

derniers soient parfois possibles en français). Cela est représenté par le fait que le causateur introduit comme sujet, n'est pas obligatoirement le sujet final (son metatrait :fonction est révisable, contrairement à l'équivalent dans la MGf).
De plus, l'interaction du réfléchi avec le causatif est légèrement différente. Les contraintes sont exactement les mêmes, mais en italien, un infinitif dans un complexe causatif, ne porte pas de réfléchi, même s'il peut être interprété comme un réfléchi. Il s'agit d'une contrainte sur la morphologie causative.
Nous n'avons pas traité ces cas dans la MGi. Il faudrait pour le faire retarder le placement du réfléchi (intrinsèque, moyen ou vrai réfléchi) dans l'arbre décrit, après que toutes les redistributions aient eu lieu. Un trait :morpho-réfl=+/- après l'application des redistributions déclencherait ou pas le placement d'un réfléchi. La reflexivation stipulerait :morpho-réfl=+, mais ce trait pourrait être révisé par l'application d'une redistribution causative.

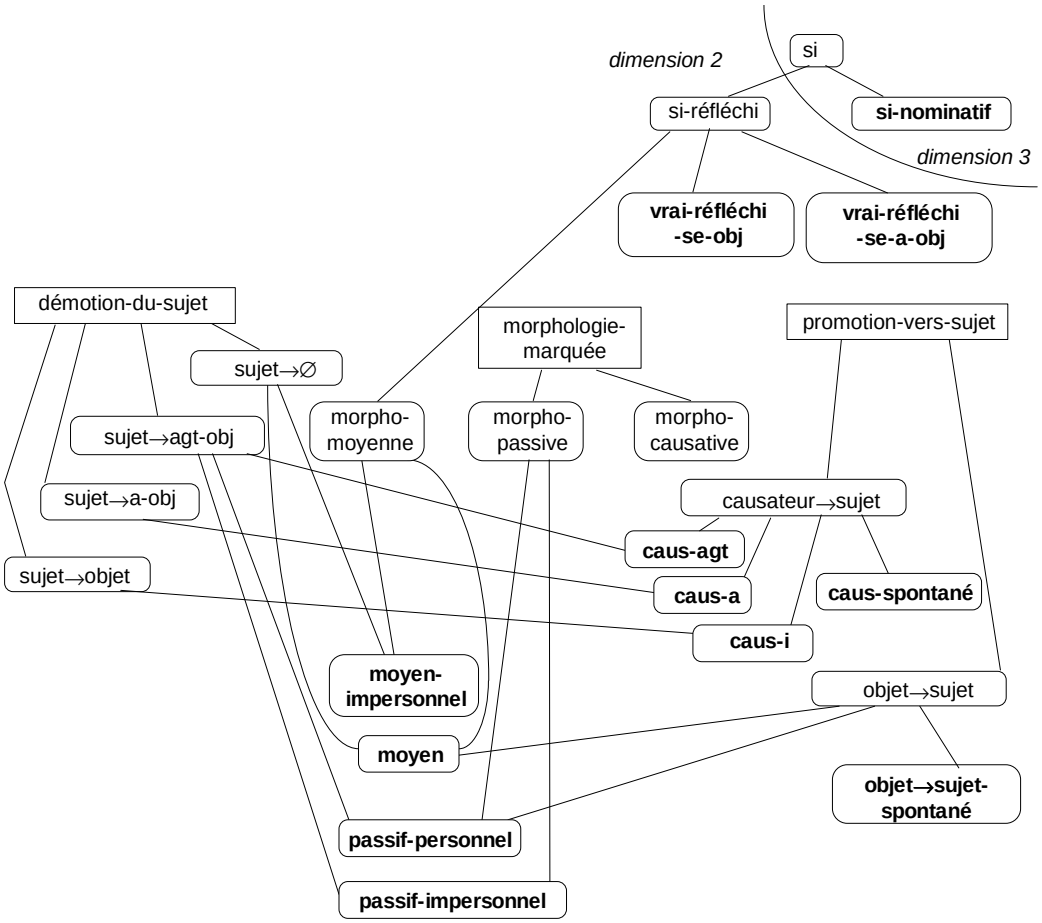


Figure 63. Les redistributions pour les verbes en italien

2.7.1 Implémentation simplifiée

Dans la section sur l'hypothèse inaccusative (Chapitre 3), on a considéré que les formes inaccusatives entraient systématiquement dans deux sous-catégorisations, une sous-

catégorisation sans sujet (que l'on a appelé impersonnelle), et une, dérivée de la première par promotion de l'objet en position sujet.
Cette analyse est justifiée par le fait que les formes inaccusatives ont un argument qui peut avoir des propriétés d'objet. Et également par le fait qu'il existe des cas de passifs impersonnels et de moyens impersonnels (en concurrence avec le SI-nominatif).

Mais cela rend le système assez complexe, car il faut (i) gérer que dans le cas d'une sous-catégorisation impersonnelle (= sans sujet) il y a accord avec l'objet s'il existe; (ii) gérer l'ambiguïté entre l'analyse personnelle (objet promu sujet) et impersonnelle, dans le cas où l'argument est extrait (cf. on ne peut pas trancher entre réalisation d'objet ou de sujet, cf. Chapitre 3, la section sur l'hypothèse inaccusative).

On a donc opté lors de l'implémentation pour une solution simplifiée. Où on considère que la promotion de l'objet est obligatoire dans le cas de sous-catégorisation sans sujet. Et pour rendre compte des propriétés d'objet de l'argument d'une forme inaccusative, on considère deux sous-types de la fonction sujet, un sujet(inacc-) et un sujet(inacc+). Le sujet(inacc+) permet toutes les réalisations de sujet(inacc-), plus le *ne* partitif (ce qu'il partage avec l'objet) et la réalisation particulière de sujet, où le verbe est tête d'une participiale (passée) adnominale (cf. Chapitre 5, la section sur les réalisations de sujet en italien).
La distinction inacc+/inacc- est représentée par la valeur de la fonction initiale. Le sujet(inacc-) est également le sujet initial. Le sujet(inacc+) est un objet initial. On donne Figure 64 le schéma des classes mettant en oeuvre cette solution simplifiée, implémentée dans la MGi.

La promotion de l'objet vers la fonction sujet précise que le sujet résultant est inacc+. La démotion du sujet au contraire, ne peut être qu'une démotion d'un sujet(inacc-). On a vu en effet que sujet→agt-obj et sujet→∅ implique que le sujet démis soit un sujet initial. Pour sujet→a1-objet, on ne voit pas de cas où le sujet démis serait non initial. Enfin, pour sujet→objet, le seul cas où le sujet démis ne serait pas initial serait celui d'un verbe infinitif au moyen entrant dans un complexe causatif. Mais on a vu que comme un infinitif sous un verbe causatif ne peut jamais porter de réfléchi (y compris moyen), on ne peut pas distinguer la composition (moyen-personnel o caus-i) de caus-i tout seul. On choisit dans la MGi de bloquer (moyen-personnel o caus-i), et donc on peut considérer que toutes les démotions du sujet s'appliquent à un sujet(inacc-).

Avec cette analyse simplifiée, on ne peut plus analyser comme on l'a fait Section 2.4.1, les cas de moyens impersonnels (possibles seulement pour les transitifs et les inergatifs). Mais on a vu que ceux-ci sont systématiquement ambigus syntaxiquement avec un SI-nominatif (qui est par ailleurs traité dans la MG), sauf dans le contexte d'un verbe à montée. La solution simplifiée ne permet donc pas de traiter (240) ni (241b). (241a) est bien représenté car il était considéré comme syntaxiquement ambigu avec un moyen personnel avec inversion du sujet. Etant donné que ces données sont marginales, et que de plus (240) n'est pas accepté par tous les locuteurs, cette solution simplifiée est adaptée à un cadre applicatif.

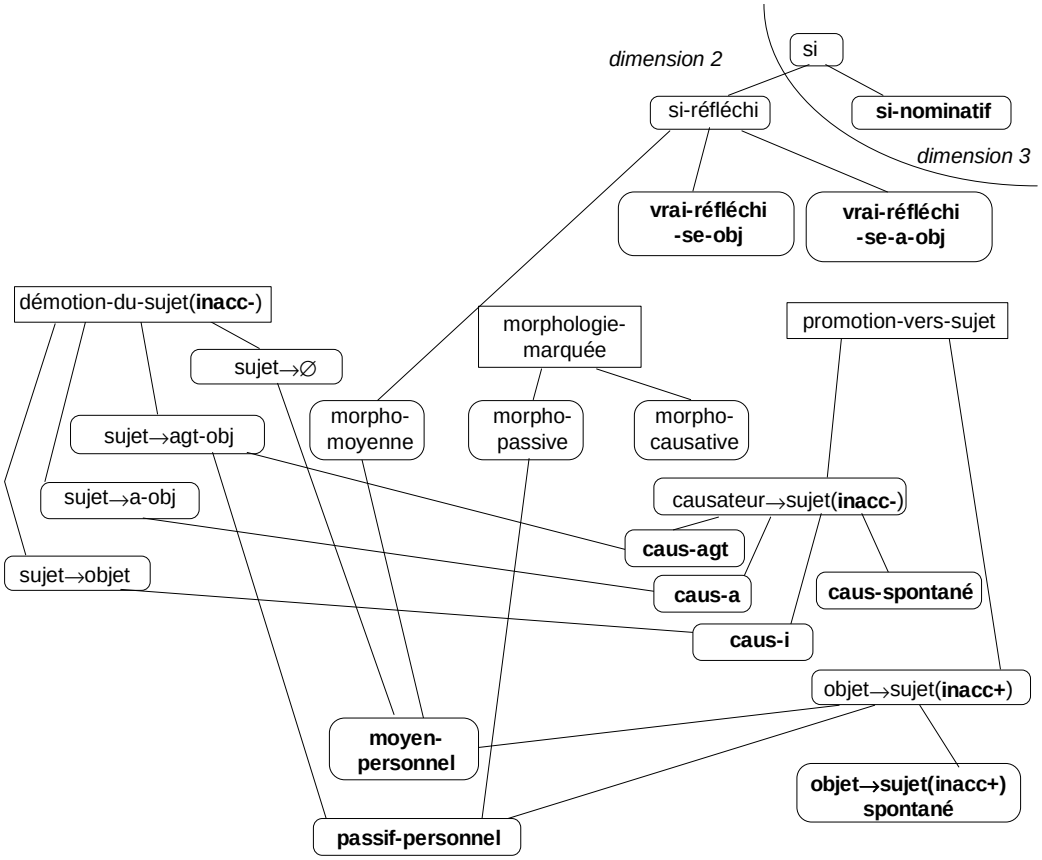


Figure 64. Implémentation simplifiée, pour représenter les redistributions des classes en italien

Chapitre 5 : La réalisation des fonctions dépendant d'un verbe

Nous décrivons dans ce chapitre les réalisations grammaticales des différentes fonctions que peuvent porter les dépendants du verbe.

Nous avons vu lors de la description de la dimension 1 (Chapitre 3), les différentes fonctions grammaticales distinguées pour représenter la sous-catégorisation initiale, que nous répétons ici : sujet, objet, a1-objet, a2-objet, gén-objet, obl-objet, loc (locatif de destination ou de situation), loc-prov (locatif de provenance), inf-but (infinitive de but), inf-per (infinitive de perception).
Pour la sous-catégorisation finale, on doit introduire une fonction supplémentaire pour le complément d'agent du passif, que l'on désigne par agt-objet.

La dimension 3 dans la MGf et la MGi est organisée selon deux paramètres : la fonction grammaticale et le type de construction (position canonique, position extraite, inversée, etc...). Ainsi, la relativisation du sujet est-elle représentée par une classe qui hérite des propriétés générales du sujet, et également des propriétés des relatives.

Nous passons en revue les caractéristiques des différentes fonctions, et nous donnons les classes qui sont utilisées pour leurs différentes réalisations grammaticales, sachant que les exceptions sont gérées à la lexicalisation : dans le lexique syntaxique, un lexème pointant vers une famille F donnée, peut par un système de traits ne pas sélectionner certains schèmes de F (cf. Chapitre 1, la sélection par un verbe d'un ensemble de schèmes, qui n'est pas forcément une famille entière).

Le détail des classes pour les constructions marquées (extraction, cliticisation ..) est donné dans des sections spécifiques pour ces constructions. De plus, nous séparons la présentation des réalisations nominales et des réalisations phrastiques (bien que certaines fonctions puissent avoir les deux types de réalisation).

Plus précisément le plan de ce chapitre est le sujet : description du sujet, de l'objet, des compléments prépositionnels, description des réalisations comme clitiques pronominaux, description des réalisations phrastiques, enfin description des positions nominales extraites (relatives, clivées, questions).

1 Les réalisations du sujet

La description du sujet est parmi les différentes fonctions, celles qui est la plus complexe. La réalisation du sujet dépend du mode du verbe. Dans la MGf et la MGi, on

a représenté la modalité à l'inverse, comme des types de réalisations du sujet. On a ainsi comme cas de réalisations du sujet, les participiales adnominales, les infinitives, les impératives (il s'agit en fait du cas particulier de "non réalisation" phonologique).
Nous décrivons d'abord les propriétés du sujet sous l'angle contrastif italien/français. Puis nous fournissons le schéma des classes utilisées pour représenter le sujet dans la MGf, puis dans la MGi. La plupart des cas seront détaillés ensuite dans les sections pour les différents types de construction (clitique, extraction ...). Avant de passer à la description de l'objet direct, nous détaillerons la gestion de l'accord sujet-verbe.

1.1 Divergences français / italien

L'actant grammatical sujet peut être nominal ou phrastique et a pour caractéristiques dans les deux langues l'accord avec le verbe, la position préverbale possible, la non réalisation pour les modes non finis et l'impératif.
La pronominalisation en français se fait par un clitique marqué pour le cas nominatif.
Pour l'italien, langue pro-drop, le sujet pronominalisé n'est pas réalisé phonologiquement, sauf dans le cas particulier du clitique SI-nominatif (cf. Chapitre 4).

(266) Il ragazzo biondo non ha salutato quando è uscito.
Le garçon blond NEG a salué quand est sorti.
(Le garçon blond n'a pas dit au revoir quand **il** est sorti)

En français, le sujet relativisé est caractérisé par le relatif *qui*, alors qu'en italien, le pronom est *che* (267a), qui est utilisé également pour l'objet (267b) et l'attribut.

(267) a. il ragazzo che vede la macchina
(le garçon qui voit la voiture)

b. la macchina che vede il ragazzo
(la voiture que voit le garçon)

1.1.1 L'inversion du sujet

En ce qui concerne l'inversion du sujet nominal (comme dans (267b)), elle est obligatoire en italien dans une interrogative directe à complément ou circonstant antéposé (268a). En français, elle est possible (268a') mais pas obligatoire, puisqu'il existe des constructions concurrentes où le sujet est préverbal : avec inversion complexe (un clitique sujet inversé double le sujet nominal) (268b) ou bien avec forme renforcée du pronom interrogatif (268c). Ces formes n'existent pas en italien.

(268) a. A chi parla Jean?
a'. A qui parle Jean?
b. A qui Jean parle-t-il?
c. A qui est-ce que Jean parle?

Hors interrogative directe, l'ordre sujet-verbe en italien est plus libre qu'en français. En français, l'inversion du sujet est productive dans le cas où il y a extraction d'un complément et où il n'y a pas d'objet en position canonique (postverbale) (cf. les

inacceptabilités (269a), (270a) versus (271a)). L'inversion locative, ou l'inversion "de lourdeur" sont plus difficiles à caractériser.
En italien, en première approximation, le sujet peut être inversé quand il fait l'objet d'un contraste avec une entité éventuellement sous-entendue (269a") ou porte une information nouvelle (269a') (Agard 1984:104). Ce cas recouvre l'inversion dans le contexte d'une extraction (où le sujet est pragmatiquement nouveau), indépendamment de la présence d'un objet postverbal (cf. (270a') et (271a')). Ce cas recouvre aussi les sujets post-verbaux indéfinis, et donc interprétés comme nouveaux (272a'), ou les sujets post-verbaux dans le contexte d'un complément clitique (273a') (Benincà et Salvi, 1988:123-125).

(269) a. *A téléphoné (une fille / Pierre) à ton frère.
a'. Ha telefonato una ragazza a tuo fratello.
(Il a téléphoné une fille à ton frère)
a". Ha telefonato Piero a tuo fratello.
(C'est Pierre qui a téléphoné à ton frère)

(270) a. *Ton frère, à qui a donné un baiser (une fille / Pia), ...
a'. Tuo fratello, al quale ha dato un bacio (una ragazza / Pia), ...

(271) a. Ton frère, à qui a téléphoné (?une fille / Pia), est euphorique.
a'. Tuo fratello, al quale ha telefonato (una ragazza/ Pia), è euforico.

(272) a. *A mangé les gâteaux un chat.
a'. Ha mangiato i dolci un gatto.

(273) a. *Les a mangés la grand-mère.
a'. Li ha mangiati la nonna.
(C'est la grand-mère qui les a mangés)

Dans le cas d'un sujet phrastique, on n'a jamais inversion en français, même dans le contexte d'une extraction (274). En italien, cela est possible (275), comme pour un sujet nominal.

(274) *Les nombreuses personnes qu'ennuit que Paul parte

(275) Mi ha sorpreso che Chiara sia andata alla festa di Marco.
ACC.1 a surpris que Chiara soit allée à la fête de Marco.
(J'ai été surpris que Chiara soit allée à la fête de Marco)

Pour le français, on doit considérer en outre les cas d'inversion du sujet clitique, ce qui n'existe pas en italien (même pour le seul clitique nominatif italien, le SI-nominatif).
On distingue l'inversion du sujet clitique, où le clitique apparaît seul sujet (276a), et l'inversion complexe (276b), ou le clitique double un sujet nominal. Bien que le statut du clitique diffère dans ces deux cas, les deux types d'inversion clitique sont trouvées dans les mêmes contextes : dans une interrogative directe, où le clitique inversé est alors porté par le verbe tête syntaxique.

(276) a. Vient-il?
b. Paul vient-il?

1.1.2 Différentes propriétés d'extraction en français et en italien

Enfin, le sujet montre des propriétés différentes dans les deux langues en ce qui concerne la possibilité de distance non bornée entre un sujet "extrait" et son verbe.
En français, pour un sujet interrogé par exemple, on a difficilement extraction à distance, et si elle est possible, le verbe pont apparaît avec un complémenteur *qui* (277a) et pas *que* (277b). En italien, on peut avoir extraction (277b'), exactement comme pour un complément.

(277) a. ? Qui crois-tu qui a téléphoné?
b. * Qui crois-tu qu'a téléphoné?
b'. Chi credi che abbia telefonato?

La situation est parallèle pour un sujet relativisé dans les deux langues. En français on a pour certains locuteurs dépendance non bornée (278a), mais elle ne suit pas le schéma général (278b), alors qu'en italien c'est possible (278b').

(278) a. ? l'homme que tu crois qui est venu
b. * l'homme qui tu crois que est venu
b' l'uomo che credi che è venuto

Pour le français, on représente la différence sujet/complément en donnant des structures différentes aux arbres pour le sujet relativisé et ceux pour les compléments relativisés. Pour l'italien, on a simplement pas cette différence de structure (cf. infra Section 7.1)
Pour le sujet interrogé, on utilise pour le français les mêmes schèmes pour un verbe avec sujet simple, et un verbe avec sujet interrogé (cf. infra Section 7.3). On n'a donc pas dans la MGf de classe spécifique pour le sujet interrogé. En revanche dans la MGi, cette classe existe.

1.2 Le sujet dans la MGf

On donne Figure 65 le schéma général pour les réalisations de sujet pour le français (certaines classes héritent d'autres superclasses, qui ne sont pas explicitées ici, par exemple le sujet clivé hérite d'informations générales sur les clivées, on renvoie à la section sur les constructions extraites). Les contraintes de cooccurrence positives (notées $\Rightarrow$) et négatives (notées $\Rightarrow\neg$) sont spécifiées sous les classes concernées.
La réalisation phrastique du sujet est ici simplifié, on renvoie à la Section 6.3.1 pour plus de détails (on doit en réalité distinguer le sujet infinitive et le sujet complétive au subjonctif).

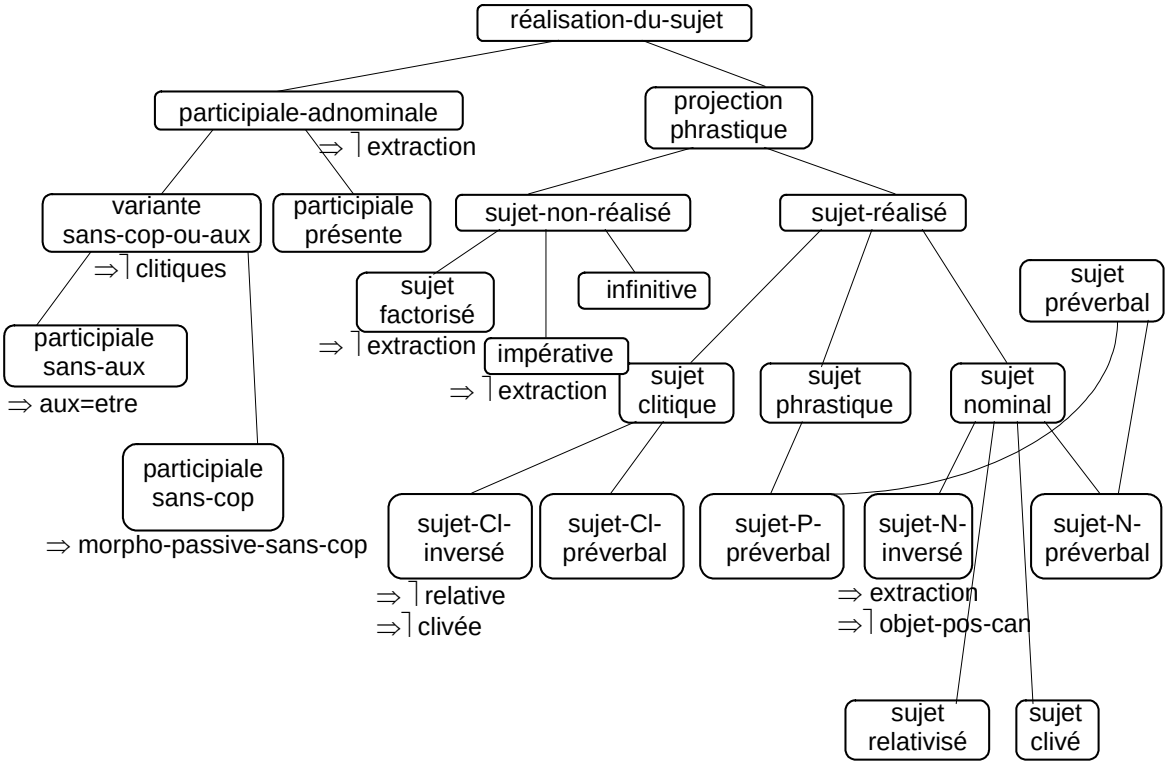


Figure 65. Les réalisations du sujet en français

La réalisation du sujet étant corrélée au mode du verbe, les différents modes (et les contraintes qui en dépendent) sont représentés par différentes classes pour la réalisation du sujet.
On distingue les cas où le verbe projette une phrase, des autres cas, c'est-à-dire les participiales adnominales (on a vu au Chapitre 3, la projection particulière utilisée pour une participiale adnominale). On distingue ensuite parmi les cas de projection phrastique les sujets réalisés de ceux non réalisés (phonologiquement). On rappelle en effet que les infinitives sont représentées par des phrases.
Nous détaillons ci-dessous les classes pour ces différents cas : les participiales, les cas de sujets non réalisés, les cas de sujets réalisés.

1.2.1 Les participiales

La MGf et la MGi couvrent actuellement les participiales adnominales.
Pour le français, on a vu lors de l'étude de l'hypothèse inaccusative (Chapitre 3), que l'on considère que la copule *étant* ou l'auxiliaire être *étant* comme optionnels, et que donc les participiales "passées" sont des variantes des participiales présentes. On redonne des exemples de participiale présente (série (69)), les variantes sans la copule ou l'auxiliaire (série (280)). Il n'y a pas optionnalité dans le cas où la copule ou l'auxiliaire de temps porte un clitique.

(279) a. Les personnes (téléphonant + ayant téléphoné) ...

b. Les personnes (arrivant + étant arrivées) ...
c. Les personnes (se trouvant + s'étant trouvées) dans l'impossibilité de ...
d. Les personnes (se lavant toujours + s'étant toujours lavées) à l'eau froide sont plus résistantes.
e. Les biens (étant + ayant été) acquis par les époux avant la date du mariage sont considérés comme communs.

(280) a. * Les personnes téléphoné ...
b. Les personnes arrivées ...
c. * Les personnes trouvées dans l'impossibilité de ...
d. * Les personnes toujours lavées à l'eau froide sont plus résistantes.
e. Les biens acquis par les époux avant la date du mariage sont considérés comme communs.

On donne Figure 49 et Figure 67, l'exemple des schèmes utilisés pour dériver respectivement (280e) et (69e). Le schème de la Figure 67 est utilisé à la fois pour la forme simple de la copule (auxiliaire du passif) *étant*, et la forme composée (*ayant été*), la forme composée étant obtenue par adjonction sur la copule[1].

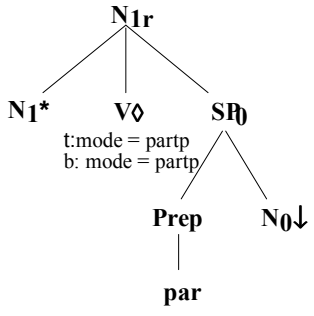


Figure 66. Schème pour dériver (280e)

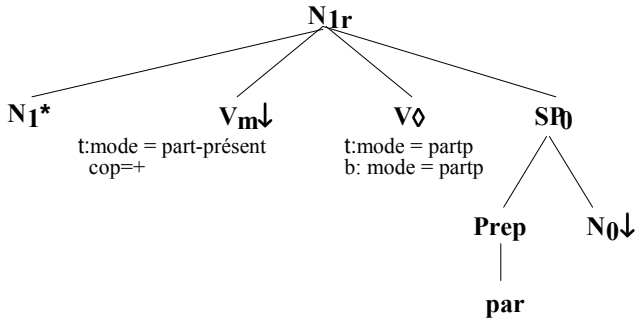


Figure 67. Schème pour dériver (69e)

On donne ci-dessous les classes pour les différents cas. On doit distinguer trois classes terminales : participiale-présente, participiale-sans-cop, participiale-sans-aux. Ces trois classes héritent d'un classe plus générale, participiale-adnominale, qui place le sujet (?const-sujet) comme noeud pied (nominal). La constante ?P, qui domine directement à la fois ?Vmarq et ?const-sujet, a la catégorie nom (dans tous les autres cas, il s'agit d'un noeud phrastique). Le croisement d'une participiale avec l'extraction d'un complément doit être bloqué, d'où la contrainte $\Rightarrow \neg$ extraction.

[1] "L'auxiliaire du passif" est représenté par un noeud à substituer, pour gérer la possibilité d'avoir comme auxiliaire tous les verbes d'état (on a alors l'interprétation stative).

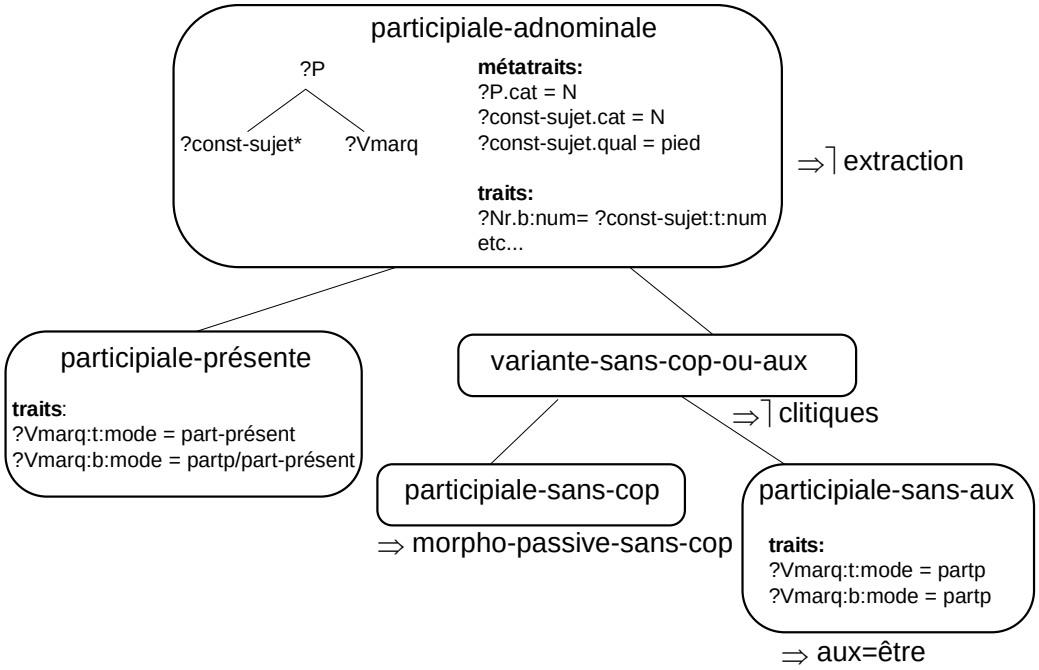


Figure 68. Les classes pour les participiales adnominales

La classe participiale-présente est utilisée pour tous les exemples de la série (69). L'alternance forme simple (par exemple *arrivant*) et forme composée (*étant arrivées*) est gérée par adjonction d'un auxiliaire de temps, comme pour les formes de mode fini.
Les variantes "participiales passées" héritent de variante-sans-cop-ou-aux. Elles ne sont pas possibles dans le cas de clitiques, (d'où la contrainte ⇒⌉ clitiques).
La classe participiale-sans-cop est utilisée pour les cas sans copule (donc au passif) comme (280e). Cette classe n'est utilisable que dans le contexte de la classe morpho-passive-sans-cop (cf. Chapitre 4).
Enfin, la classe participiale-sans-aux est utilisée pour les cas sans auxiliaire ETRE (comme (280b)). Elle n'est utilisable que si l'ancre sélectionne effectivement ETRE, d'où la contrainte ⇒aux=être, il s'agit d'une classe héritée par toutes les sous-catégorisations initiales pour lesquelles est spécifié l'auxiliaire de temps ETRE).

1.2.2 Les sujets non réalisés

On considère différents cas de "non réalisation" du sujet pour représenter les impératives (classe impérative), les infinitives (classe infinitive), la factorisation du sujet (classe sujet-factorisé) (ou coordination de syntagmes verbaux).
Ces derniers peuvent imposer le mode impératif ou infinitif[2]. La classe sujet-factorisé est utilisée pour représenter la coordination de syntagmes verbaux (c'est un abus de langage, puisque nous n'avons pas de syntagmes verbaux, il s'agit de coordination de non constituants) comme en (281). La conjonction et le deuxième membre de la

[2] On distingue les cas d'infinitive et d'impérative de manière à pouvoir représenter la différence de comportement quant à l'enchâssement : une impérative ne peut pas être enchâssée.

coordination sont adjoints, sont regroupés dans un arbre auxiliaire qui s'adjoint sur la phrase.

(281) Paul enfile sa veste [et sort].

On donne ci-dessous les schèmes pour un intransitif avec à gauche, la structure utilisée pour une infinitive, à droite celle pour un sujet factorisé (utilisable pour dériver (281)).

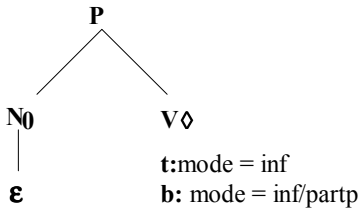


Figure 69. Schème pour une infinitive

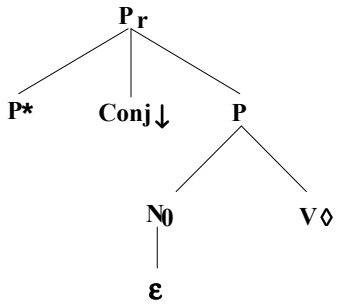


Figure 70. Schème pour un sujet factorisé

Le traitement de la coordination en général n'est que partiellement réalisé dans la MGf et la MGi. On peut remarquer que le schème pour le sujet factorisé enfreint le PCPA, d'une part parce qu'il comporte une position (le noeud pied) qui n'est pas un argument sémantique de l'ancre, d'autre part, parce que la conjonction est une unité sémantique en soi, qui n'est pas gouvernée sémantiquement par l'ancre. D'une manière générale, la coordination est difficilement représentable en grammaires de dépendances. Il n'est donc pas étonnant que dans l'arbre de dérivation, les dépendances obtenues soient erronées (ce qu'implique l'infraction au PCPA).

1.2.3 Les sujets réalisés

Quand il est réalisé, le sujet est soit phrastique, soit nominal, soit clitique.

Les sujets nominaux peuvent être en position préverbale ou inversée, relativisée ou clivée. Ces deux dernières réalisations héritent aussi de classes décrivant les relatives et les clivées (voir infra Section 7). On utilise des contraintes de compatibilité pour contraindre l'inversion du sujet nominal, qui est incompatible avec un objet en position canonique, et qui impose qu'un autre actant soit extrait (on ne traite que les cas d'inversion du sujet dans le contexte d'une extraction).

Les sujets phrastiques sont dans la MGf seulement préverbaux, car d'une part, il n'y a pas d'inversion en français (cf. supra), et d'autre part on ne traite pas actuellement les realtives phrastique (*Paul est parti, ce qui attriste Marie*), les clivées phrastiques (*C'est que Paul soit parti qui attriste Marie*), ni les pseudo-clivées (*Ce que Dalila souhaite c'est qu'Oscar revienne*).

Les clitiques sujet sont préverbaux ou postverbaux[3]. La classe pour les sujets clitiques postverbaux est utilisée pour le clitique sujet inversé (distingué de l'inversion

[3] On peut remarquer qu'en dimension 2, la classe IMPERS→SUJET spécifie que la sous-catégorisation finale comporte un sujet clitique *il* (non lié à un argument), ce qui ne sera compatible qu'avec les réalisations clitiques de sujet.

complexe). Elle est incompatible avec les clivées et les relatives, ce qui est représenté par des contraintes de cooccurrence négatives.
Pour traiter l'inversion complexe, on laisse libre les schèmes d'être ancrés par un verbe portant un clitique inversé (cf. infra Section 5, sur le traitement des clitiques comme des affixes pronominaux). L'inversion complexe doit être cependant bloquée dans le cas de relatives, de clivées et également dans le cas d'inversion du sujet nominal, ce qui est représenté par des traits.

1.3 Le sujet dans la MGi

Pour l'italien, on reprend quasi intégralement la hiérarchie pour le sujet de la MGf. Cependant les différences sont :

- Les sujets inversés n'ont pas les contraintes de cooccurrence du français[4]. Il suffit pour le représenter, de ne pas spécifier dans la MGi, les contraintes de cooccurrence ($\Rightarrow$ extraction et $\Rightarrow\neg$ objet-pos-can) qui sont associées à sujet-N-inversé dans la MGf.
- De plus, le sujet phrastique peut être inversé. On a donc une classe générale sujet-inversé, dont héritent deux classes sujet-N-inversé et sujet-P-inversé[5].
- On a une classe supplémentaire pour représenter le pro-drop, qui hérite des sujets non réalisés.
- La classe des sujets clitiques en français est maintenue en italien, mais précisée : il ne peut s'agir que du SI-nominatif.
- Il n'y a pas de sujet clitique inversé.
- Une classe pour le sujet interrogé est introduite, étant donné la possibilité d'extraire le sujet à distance (cf. supra).
- les contraintes sur les participiales adnominales sont différentes.

Avant de détailler les différences pour les participiales adnominales en italien, on donne la Figure 71, qui schématise tous ces points de divergences par rapport au français (les classes nouvelles sont en gras).

[4] On peut cependant pour une application ajouter la contrainte de non cooccurrence avec un objet post-verbal, étant donné que l'inversion est tout de même rare dans ce cas.
[5] Là encore, on a un peu simplifié, puisqu'il faut distinguer le cas sujet infinitive et sujet complétive, inversé ou pas. En outre pour l'italien, le sujet infinitive peut avoir le complémenteur *di* (cf. infra Section 6.3.1).

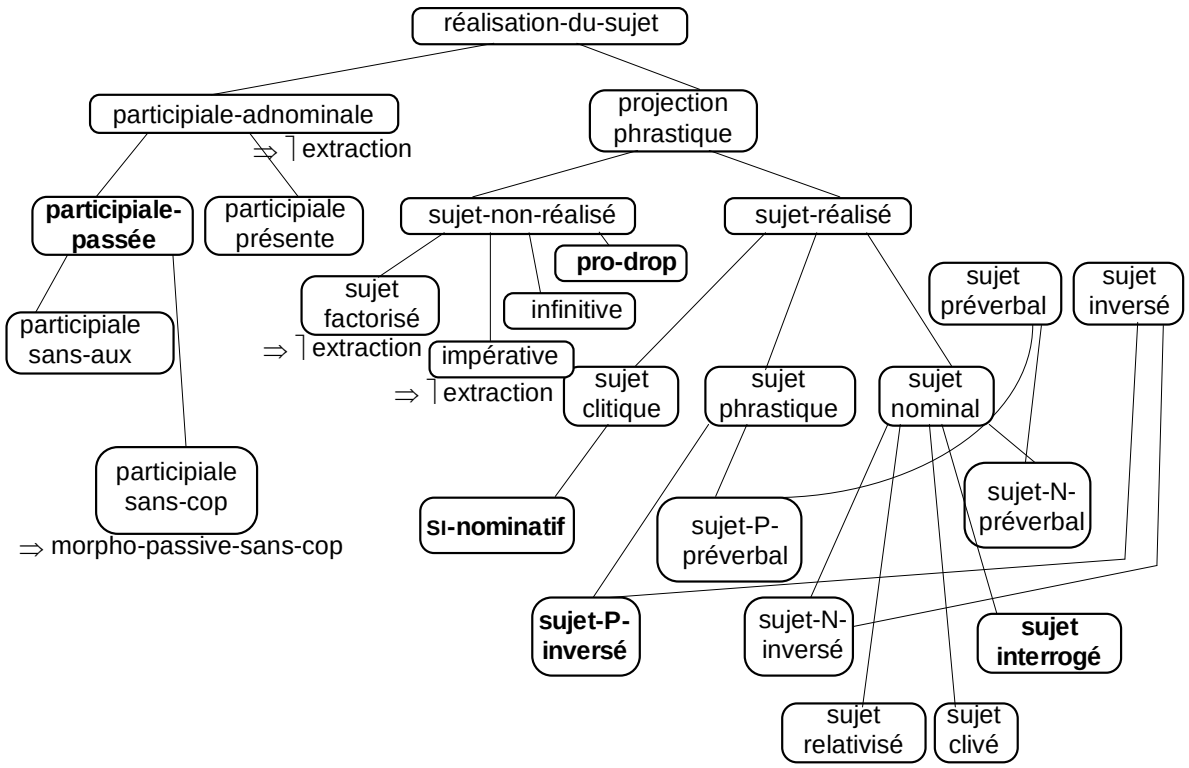


Figure 71. Les réalisations du sujet en italien,
(en gras les divergences avec le français)

1.3.1 Représentation des participiales adnominales

Pour les participiales adnominales en français, on a traité les "participiales passées" (c'est-à-dire avec un participe passé) comme des variantes de participiales présentes, en considérant la copule et l'auxiliaire de temps ETRE comme optionnels.
Pour l'italien, il existe une différence notable, de nature sémantique, avec des implications syntaxiques : le participe présent (dans son emploi verbal) désigne une propriété permanente (Benincà et Cinque, 1991). Donc, premièrement on ne peut donc le trouver à une forme composée (282a,b), et deuxièmement, il est de fait bloqué pour les verbes non compatibles avec cette restriction (283), c'est-à-dire pour la majeur partie des verbes.

(282) a. *Le persone aventi concorso al premio di poesia saranno convocate.
Les personnes ayant concouru au prix de poésie seront convoquées

b. *Il posto essente spettato alle auto non è più disponibile.
La place étant réservée aux autos n'est plus disponible.

(283) *I soldati uccidenti i nemici.
Les soldats tuant les ennemis

L'impossibilité de (283) est simplement représentée par la non existence dans le lexique morphologique de la forme participe présent pour ces verbes. Pour bloquer les formes

composées du participe présent, dans la classe participiale-présente, on ne permet comme trait aval sur ?Vmarq que part-présent (alors qu'en français on avait part-p/part-présent) ce qui bloque l'adjonction d'un auxiliaire de temps.

Mais la possibilité d'avoir des participiales "passées" (i.e. avec participe passé) pour ces mêmes verbes (284) rend difficile le lien que l'on a établi pour le français entre participiale présente et variante sans copule ou sans auxiliaire. Pour cette raison, au lieu de la distinction participiale-présente/variante-sans-cop-ou-aux, on garde la distinction plus traditionnelle participiale présente/participiale passée.

(284) I nemici uccisi dai soldati.
Les ennemis tués par les soldats

Ce que l'on a analysé comme des participiales sans auxiliaire de temps ETRE, sont en italien les participiales "passées" pour les verbes inaccusatifs. Il s'agit d'une propriété de "sujet inaccusatif" (cf. l'implémentation simplifiée proposée au Chapitre 4, pour traiter du comportement inaccusatif des formes inaccusatives). On substitue la contrainte ⇒ aux = être par la contrainte que le sujet du participe doit être un "sujet inacc+", c'est-à-dire doit être un objet initial (cela est représenté en interne par le métatrait ?const-sujet:fonction-i = objet).

(285) a. * le persone telefonato
les personnes téléphoné

b. le persone arrivate
les personnes arrivées

Une autre différence des participiales, est que le participe passé (dans son emploi verbal) peut porter des clitiques. La classe participiale-passée ne porte donc pas de contrainte ⇒ ⌉ clitiques.

(286) La merce consegnatami è avariata.
La marchandise consignée-DAT.1 est avariée
(La marchandise qui m'a été consignée est avariée.)

On donne ci-dessous le schéma des classes pour les participiales adnominales en italien, qui prend en compte toutes ces différences.

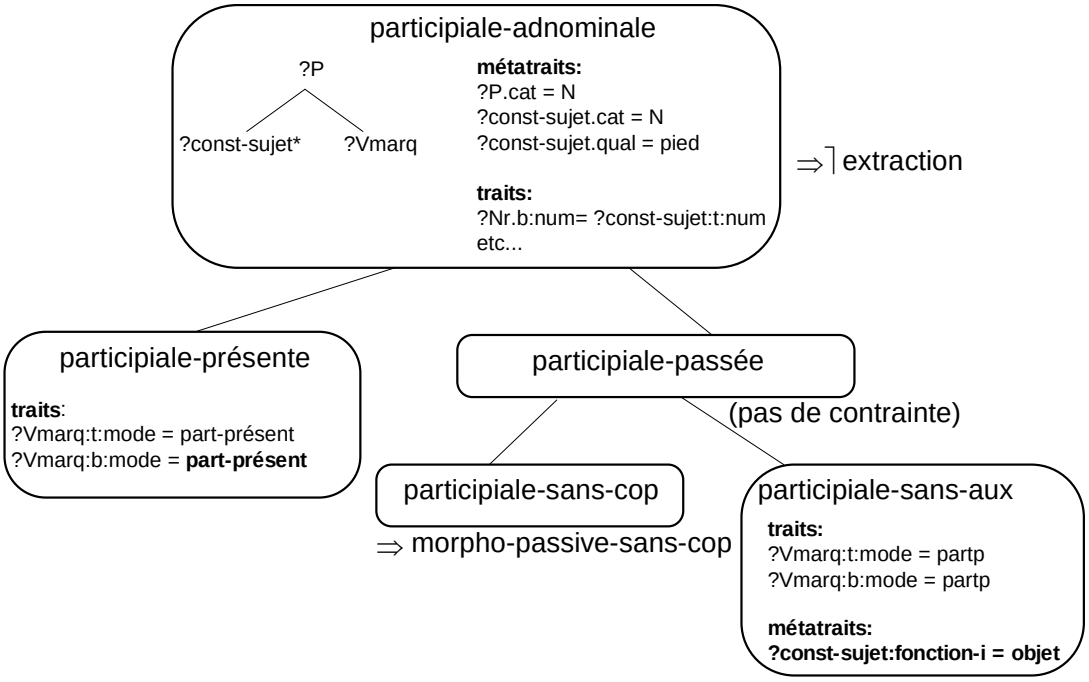


Figure 72. Les classes pour les participiales adnominales dans la MGi

Le travail pour représenter les participiales absolues (cf. la section sur l'hypothèse inaccusative) reste à faire pour le français e l'italien.

1.3.2 Les propriétés du "sujet inacc+"

Au Chapitre 3, nous avons mentionné pour l'italien une solution simplifiée pour le traitement des propriétés d'objet des formes inaccusatives. Une de ces propriétés est la possibilité d'être modifié par une participiale passée, et l'on vient de voir comment cela est représenté dans la MGi.

Une autre propriété concerne le *ne* partitif. On doit considérer une classe sujet-inversé-avec-partitif, qui n'est valable pour un sujet inaccusatif, c'est-à-dire qui est un objet initial (cela est représenté par le métatrait ?const-sujet:fonction-i = objet). On aura une classe similaire pour représenter le *ne* partitif dans le cas d'un "vrai objet". On renvoie à la section suivante, sur la description de l'objet direct, pour le détail de cette classe.

1.4 L'accord de la forme verbale

L'accord de la forme verbale avec le sujet est géré par la coindexation des traits amont :num et :pers de ?Vmarq[6] (qui est toujours le noeud verbal le plus haut de l'artère) et le sujet. Il s'agit d'équations de type traits d'accord, entre traits amonts. Il y a en outre

[6] Pour l'implémentation, les clitiques étant traités en syntaxe, on a plusieurs projections verbales dans le cas de clitiques. L'accord sujet-verbe se fait entre le noeud verbal le plus haut et le sujet. Dans la MGf et la MGi, l'artère est représentée par plusieurs constantes en relation de dominance, qui nomment des noeuds différents uniquement dans le cas où il y a des clitiques.

propagation (de type équation de trait de tête) vers les noeuds verbaux inférieurs s'ils existent.
Les traits d'accord du participe passé dans le cas d'un temps composé sont des traits spécifiques :partnum et :partgen, qui s'appliquent à la tête locale, et sont donc imposés comme traits aval sur ?Vmarq. On détaille le mécanisme de déclenchement de l'accord du participe Section 2.1.
Pour représenter l'accord sujet-verbe, on a deux cas de figure, selon que le sujet est réalisé ou pas, et s'il n'est pas réalisé, selon qu'il est contrôlé ou pas.

1.4.1 Si le sujet est réalisé

Si le sujet est réalisé, il y a équation d'accord (c'est-à-dire entre traits amont) entre les traits :num et :pers du noeud sujet et ceux de ?Vmarq. Avec l'analyse des clitiques sujet en morphologie, on n'a besoin de ces équations d'accord que si le sujet n'est pas clitique, l'accord clitique-nominatif/verbe pouvant être géré directement en morphologie.
Si aucune adjonction enchâssante n'a lieu, il y aura accord avec la tête locale. Dans le cas d'adjonciton enchâssante sur ?Vmarq, les équations d'accord s'appliqueront avec la nouvelle tête.
On donne , un arbre élémentaire pour un intransitif avec sujet nominal[7], avec les traits d'accord uniquement, recevant l'adjonction d'un verbe à montée. On peut voir que les traits d'accord seront propagés dans l'arbre dérivé, sur le verbe à montée.

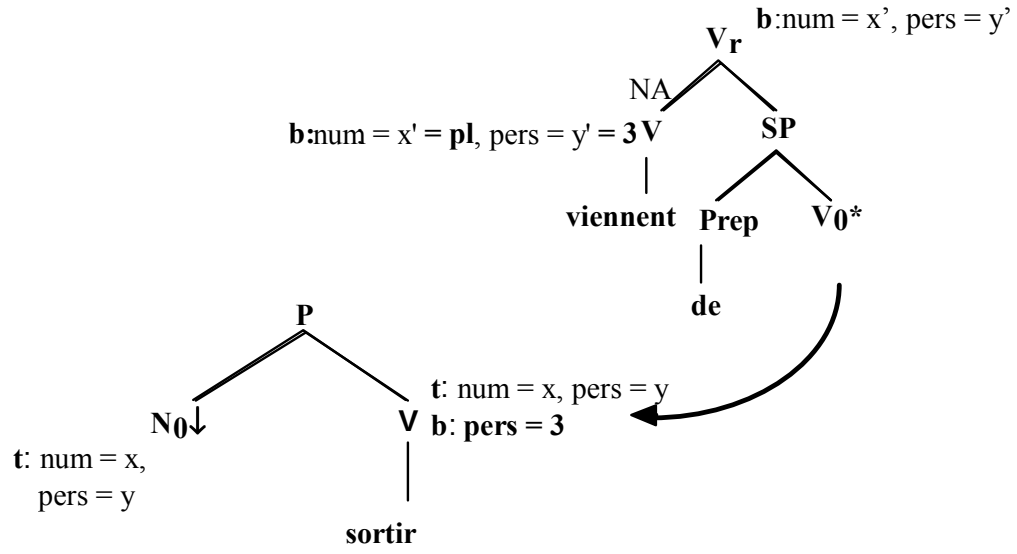


Figure 73 : Adjonction d'un verbe à montée sur un intransitif (traits d'accord).

1.4.2 Si le sujet n'est pas réalisé (localement)

Si le sujet n'est pas réalisé mais n'est pas contrôlé, par exemple pour une impérative, il n'y a pas d'accord à représenter. Les traits de nombre, genre et personne de la forme

[7] L'infinitif *sortir* porte le trait pers=3, ce qui a été justifié par l'interaction réfléchi/causatif, au Chapitre 4.

verbale sont simplement contraints par le type de construction. Pour une impérative, on impose la 1ère ou 2ème personne (la 3ème personne est syntaxiquement différente.

Si le sujet n'est pas réalisé, mais est contrôlé par un noeud non local (c'est-à-dire n'appartenant pas au même arbre élémentaire), on doit d'une manière ou d'une autre récupérer de ce noeud non-local (le contrôleur) les informations pertinentes pour les noeuds locaux, en l'occurrence les traits d'accord de la forme verbale.
On ne peut le faire en TAG que par le jeu de propagation de traits entre les noeuds locaux et les noeuds d'arbres avec lesquels il y a combinaison. Pour une infinitive dont le sujet est contrôlé par un actant d'un verbe à contrôle, l'arbre de l'infinitif reçoit l'adjonction du verbe à contrôle sur le noeud désigné par ?P. C'est à ce noeud que doit se faire la transmission des traits du contrôleur.
Un trait :cont pour contrôleur représente les traits pertinents du contrôleur (Abeillé, 1991)[8] est placé sur le noeud ?P de l'arbre de l'infinitive. Prenons un exemple d'infinitive au passif. Pour les arbres au passif, la copule est directement présente dans l'arbre. L'accord du participe passé (nommé par ?Vanc) se fait directement via les traits :num et :gen sur ?Vanc[9]. Dans le cas d'une infinitive, le sujet n'étant pas réalisé, les traits :num et :gen du participe doivent être coindicés avec les traits du contrôleur, via le noeud ?P (on n'a donné pour le contrôleur que le trait :num et :gen).

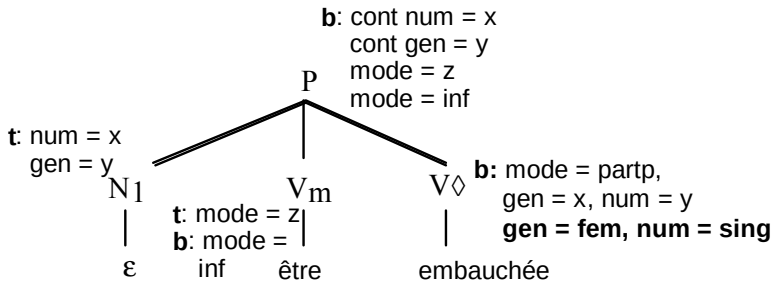


Figure 74. Un arbre élémentaire (n0Vn1-pass-sa-inf) pour une infinitive au passif

Dans l'arbre du verbe à contrôle, le noeud pied (représentant l'infinitive), contient un trait :cont avec pour valeur les traits du contrôleur.

[8] Pour réaliser une coindexation correcte entre le contrôleur et la position sujet non réalisée, Abeillé note qu'il faudrait pouvoir coindicer en bloc tous les traits d'un noeud, ce qui n'est pas faisable avec le système XTAG. On se contente ici de donner ici une coindexation trait par trait (traits :num, :gen).
[9] Cela est différent dans le cas de temps composés, où on utilise des traits spécifiques :partnum et :partgen, étant donné que l'on utilise les mêmes schèmes, que la forme soit simple ou composée (cf. la section suivante). Dans le cas de passifs, comme les schèmes sont de toutes façons distincts des arbres actifs, on peut représenter l'accord du participe directement.

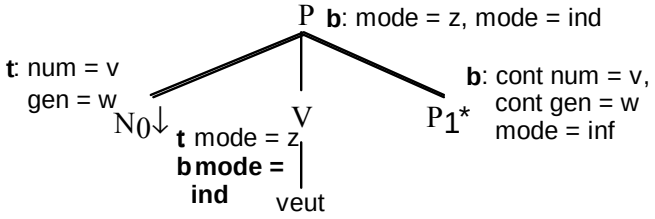


Figure 75. Un arbre élémentaire pour un verbe à contrôle

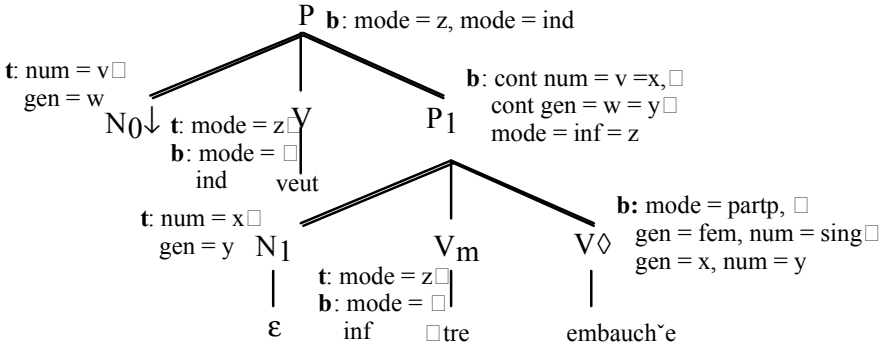


Figure 76. Arbre dérivé après adjonction d'un verbe à contrôle sur n0Vn1-pass-sa-inf[*embauchée*]

Si le contrôle est structural, c'est-à-dire si la sous-catégorisation du verbe à contrôle ne comprend qu'un contrôleur possible, le sujet, comme en (287a), le contrôleur peut être spécifié dans les schèmes. Si le contrôle est lexical (i.e. on ne peut pas prévoir si le contrôleur est le sujet ou un complément), comme en (287b,c), le contrôleur doit être spécifié via le lexique, pour chaque verbe à contrôle. Les cas de contrôle mixtes ne sont pas gérés.

(287) a. Jean veut partir.
b. Jean permet à Marie de sortir.
c. Jean promet à Marie de sortir.

2 Les réalisations de l'objet direct

Pour le français, nous suivons Abeillé (1997) qui retient comme caractérisation syntaxique[10], le cumul de deux critères : le *en* quantitatif (288) et la négation en *pas...de* (289).

(288) a. Paul a vu trois films d'affilée.
b. Paul en a vu trois d'affilée.

[10] (Gross 1969) a critiqué pour le français la notion traditionnelle d'objet direct, en montrant que les critères habituellement retenus (passivation, pronominalisation en *le,la,les*, question en *que...*) sont divergents pour beaucoup de verbes. Ce qui est clair, c'est que la caractérisation de l'objet s'appuie sur des critères moins nombreux et moins fiables que pour le sujet.

(289) a. Paul a vu des navets.
b. Paul n'a pas vu de navets.

Le *en* partitif permet de distinguer l'objet du complément indirect et du sujet, qu'il soit inversé (290) ou pas (291) :

(290) a. Trois personnes ont téléphoné.
b. *Trois en ont téléphoné.

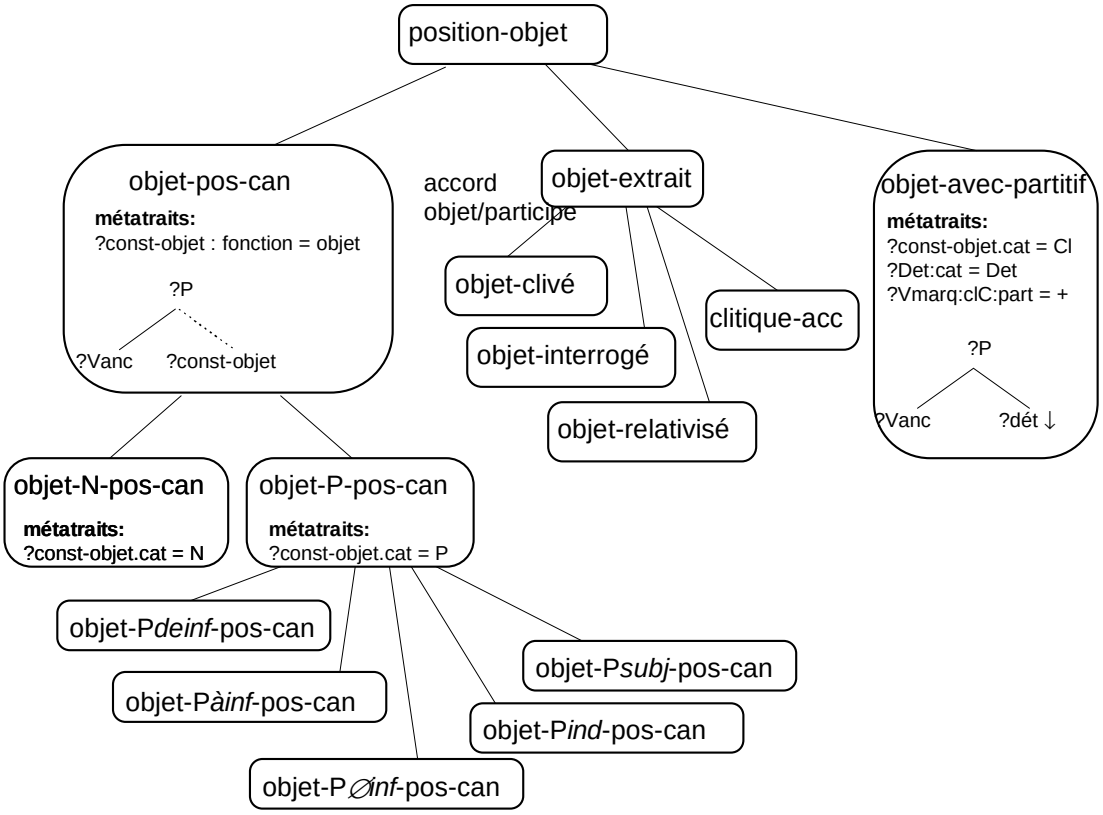
(291) a. Bien qu'aient téléphoné trois personnes, je me suis senti seul.
b. *Bien qu'en aient téléphoné trois, je me suis senti seul.

Et la négation en *pas...de* isole l'objet direct de l'attribut nominal. On peut remarquer que cette caractérisation englobe le syntagme extraposé des constructions impersonnelles (l'actant H), bien qu'il n'ait jamais la pronominalisation en *le, la les* :

(292) Des catastrophes, il m'en est arrivé trois cette année.
(293) Il ne m'est pas arrivé de catastrophes cette année.

La pronominalisation en LE n'est pas acceptée pour les actants H nominaux (mais elle l'est pour les phrastiques, cf. Chapitre 4).
L'objet direct en français a pour autres propriétés (pas forcément caractéristiques) le pronom interrogatif *que*, et le relatif *que*.
En français, dans le cas d'un objet extrait ou cliticisé, il y a accord avec l'éventuel participe passé.

On donne ci-dessous les classes de la MGf pour l'objet. Une classe objet-extrait est héritée de toutes les classes où l'objet est placé devant le verbe, et déclenche l'accord objet/participe. L'objet en position canonique peut être nominal ou phrastique, auquel cas on doit distinguer le cas complétive, infinitive sans complémenteur, infinitive avec *de*, infinitive avec *à* (cf. infra Section 6.3.2).



En italien, la caractérisation de l'objet direct est la série de clitiques accusatifs, et la possibilité du *ne* partitif (équivalent du *en* partitif en français) (294a).
L'accord du participe avec l'objet est facultatif dans le cas d'un clitique accusatif, et il n'y a pas d'accord dans le cas d'un sujet extrait.
On n'a pas de test équivalent à la négation en *pas...de*.
Comme on vient de le rappeler dans la section sur les sujet, les "sujets inaccusatifs" ont également cette propriété (c'est pourquoi en théorie, il faudrait les considérer comme des objets). On redonne en (294b) un exemple avec un verbe inaccusatif, mais cela est également vrai pour une forme inaccusative passive ou moyenne.

(294) a. Paul ne ha visto tre di seguito.
Paul NE-part a vu trois de suite.
(Paul en a vu trois d'affilée)

b. Ne sono arrivate tre.
NE-part sont arrivées trois.
(Il en est arrivé trois)

2.1 L'accord du participe passé

Nous détaillons dans cette section la gestion de l'accord du participe passé avec l'objet placé devant, et plus généralement de tous les cas d'accord du participe.

2.1.1 Pour le français

Dans le cas sans objet antéposé, ni réfléchi alternant avec un objet (dorénavant SE-obj)[11], l'accord du participe se fait selon l'auxiliaire habituel sélectionné par le verbe, indépendamment de l'auxiliaire effectif. Il y a accord en genre et nombre du participe avec le sujet, si l'auxiliaire habituel est être, sinon le participe est masculin singulier. En (295d), l'auxiliaire effectif est ETRE, du fait du réfléchi. Il s'agit du réfléchi SE-a-obj (qui alterne avec un a1-objet), qui ne déclenche pas d'accord particulier.

(295) a. Les enfants ont mangé la tarte.
b. Les enfants ont couru.
c. Les enfants sont accourus aussitôt.
d. Les enfants se sont parlé.

Si un objet est antéposé, il y a accord en genre et nombre du participe avec l'objet, ceci indépendamment de l'auxiliaire effectif (qui peut être ETRE, du fait de la présence d'un réfléchi).

(296) a. Quelle photo a-t-il prise?
b. Quelle voiture s'est-il offerte?
c. Il l'a prise hier (cette photo)
d. Il se l'est offerte (cette voiture)

Si un réfléchi SE-obj est présent, il y a également accord en genre et nombre du participe avec le réfléchi.

(297) Marie s'est regardée dans la glace.

On voit donc que l'accord avec le participe passé est indépendant de l'auxiliaire effectif de temps. Les schèmes doivent contenir l'information permettant de déclencher l'accord, dans le cas d'un auxiliaire habituel ETRE, dans le cas d'un réfléchi SE-obj, dans le cas d'un objet antéposé. Le genre et le nombre du participe sont stockés dans des traits spécifiques :partnum et :partgen, qui s'appliquent à la tête locale, et sont donc imposés comme traits aval sur ?Vmarq. Les schèmes partagés par les formes simples et les participes (attendant alors l'adjonction d'un auxiliaire de temps), stipulent soit :partnum = sing et :partgen = masc, dans les constructions où il n'y a pas d'accord du participe, soit coïndicent ces traits :partnum et :partgen avec ceux du noeud adéquat (le sujet ou l'objet).
Pour illustrer ce mécanisme, nous reprenons l'exemple d'un schème pour un ditransitif avec objet relativisé, réfléchi datif et sujet clitique (déjà fourni au Chapitre 3), c'est-à-dire pouvant être utilisé pour dériver *la voiture qu'il s'offre*, s'il est ancré par *offre*, ou bien encore *la voiture qu'il s'est offerte*, si le schème est ancré par *offerte*). Nous donnons Figure 77 ce schème lexicalisé par *offerte* (avec traits d'accord seulement, les traits venant de la lexicalisation sont en gras).
L'accord du participe avec l'objet (représenté par le noeud pied N1) et le fait que l'auxiliaire de temps soit ETRE sont des informations imposées par la construction syntaxique, et donc contenues dans le schème. Ce schème peut être ancré par une forme simple et alors ces informations ne sont pas pertinentes. S'il est ancré par un participe,

[11] On distingue SE-obj et SE-a-obj justement sur la base de l'accord qu'ils déclenchent, cf. Chapitre 4.

l'adjonction d'un auxiliaire de temps, de mode fini, permet de satisfaire la contrainte :mode=ind sur ?P (cf. supra la gestion du mode).

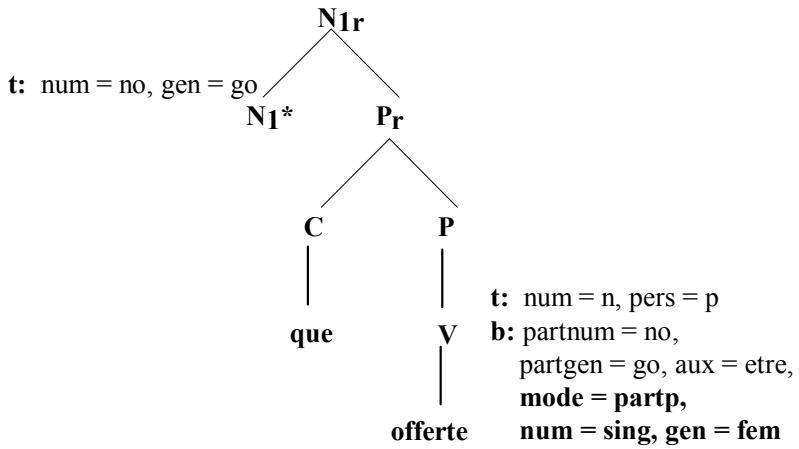


Figure 77 : Lexicalisation de n0Vn1an2-cl0-r1-réfl2 par la forme *offerte* (traits d'accord)

La sélection de l'auxiliaire effectif de temps est géré par les trait :aux (cf. Chapitre 3, la section sur les auxiliaires de temps). Le trait :aux vaut "avoir", sauf dans (i) les schèmes pour les verbes se conjuguant habituellement avec ETRE, (ii) les schèmes comprenant un réfléchi (SE-obj ou SE-a-obj), d'où le trait :aux = etre sur ?Vmarq, Figure 77.
Comme l'accord du participe ne dépend pas de l'auxiliaire effectif de temps, et est représenté dans les schèmes qui peuvent être ancrés par le participe, on a le même schème pour les formes de l'auxiliaire ETRE et celles d'AVOIR, qui représentent l'auxiliaire effectif de temps (Figure 78). Dans ce schème, on impose la coindexation des traits :num et :partnum (respectivement :gen et :partgen) sur le noeud pied, qui représente le participe.

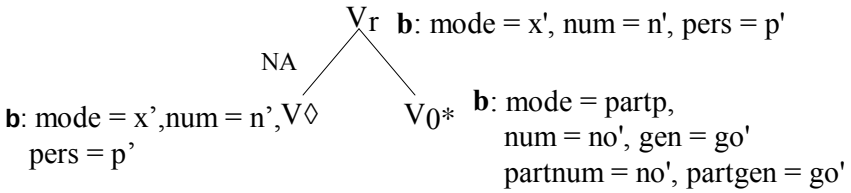


Figure 78 : Schème (Vv0-aux) pour les auxiliaires de temps
(traits d'accord et mode)

L'adjonction de l'auxiliaire au noeud ?Vmarq déclenche la coindexation pertinente entre les traits :num et :gen de ?Vmarq et ses traits :part-num et :part-gen.
Les arbres ancrés par ETRE ou AVOIR diffèrent seulement par le trait :aux qu'ils imposent au noeud pied. On donne Figure 79 l'arbre élémentaire pour la forme *il s'est*.

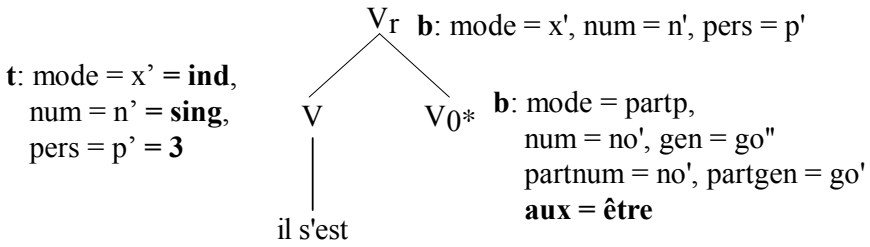


Figure 79 : Lexicalisation de Vv0-aux par la forme *il s'est* (traits d'accord et mode)

Dans le cas d'un prédicat complexe causatif en français, à un temps composé, le participe causatif est invariant.
Cette photo, Paul l'a fait/*faite refaire plusieurs fois avant d'en être satisfait.

Le verbe causatif est nommé par la constante ?Vmarq, dans les schèmes spécifiques pour la construction causative. Sans contre-ordre, il hériterait de l'accord régulier, dans le cas d'un objet clitique ou extrait. Il faut spécifier que les traits :partnum et :partgen du verbe causatif doit valoir sing et masc, même dans le cas d'un objet extrait.

2.1.2 Pour l'italien

L'accord du participe dans le cas de formes composées est légèrement différent en italien, étant donné que ESSERE déclenche de manière plus générale l'accord du participe avec le sujet, cf. il y a accord même dans le cas de réfléchi datif (cf. Chapitre 4). Mais d'une part dans le cas d'un clitique objet, l'accord avec l'objet a priorité. D'autre part, il y a des cas où l'accord est sensible à l'auxiliaire habituel de temps et non pas l'auxiliaire effectif, il s'agit des phrases à sujet *si* impersonnel (SI-nominatif, cf. Chapitre 4). Ce *si* porte les traits masculin pluriel, et force l'auxiliaire ESSERE, mais le participe ne s'accorde avec le *si* que pour les verbes dont l'auxiliaire habituel est ESSERE.
On garde donc pour l'italien le fait que l'accord soit codé dans l'arbre élémentaire ancré par le participe, et pas dans l'arbre de l'auxiliaire de temps ESSERE. Donc le mécanisme de l'accord est exactement similaire à celui qu'on vient de donner pour le français.

Une autre différence avec le français provient du fait que dans le cas de prédicat complexe causatif, le verbe causatif, à la forme composée, montre un accord régulier. On a le même accord régulier pour le verbe restructurant à une forme composée.

(298) a. Paolo **li** ha fatt**i** venire.
Paolo ACC.3mp a faits venir.

b. Maria **li** avrebbe volut**i** vedere lei stessa.
Maria ACC.3mp aurait voulus voir elle-même.
(Maria aurait voulu les voir elle-même).

L'accord du verbe causatif au participe suit le schéma régulier en italien, le verbe causatif est nommé par la constante ?Vmarq, et héritera des équations d'accord de manière régulière.

Le traitement de l'accord du participe verbe restructurant est différent, étant donné que les VR sont analysés par adjonction. Il faut donc transmettre les traits :partnum et :partgen de l'arbre recevant l'adjonction vers l'arbre s'adjoignant pour le VR. Ces traits sont portés par ?Vmarq, et le VR s'adjoint sur ?Vmarq, leur propagation est donc aisée.

2.2 Ajouts envisagés

Aux classes représentant l'objet dans la MGf et la MGi, nous envisageons d'ajouter des classes pour le traitement de :

- du quantifieur préverbal (*Paul a tout vu, Paul ne veut rien lire, Paul veut tout avoir lu, Paul veut avoir tout lu*)
- de la relative en fr. *à* (*un livre à lire, un exercice à leur donner)*, it. *da* (*un libro da leggere, un esercizio da dargli*)

Pour le premier cas, il faut considérer des réalisations particulières de l'objet, placé devant l'ancre au participe, et avec adjonction obligatoire d'un auxiliaire de temps. Il faut également une classe spécifique au placement devant l'infinitif.
Pour la relative en *à/da*, il faut définir une classe obligatoirement cooccurrente à la classe infinitive (cf. supra comme cas particulier de réalisation de sujet), qui place l'objet en position relativisée (cf. infra Section). Pour représenter le fait que la relative en *à* ne permet que peu d'enchâssement en français, il faut définir une structure plus plate que dans le cas d'une relative en qu-.

3 L'attribut

Comme nous l'avons dit au Chapitre 3, parmi les fonctions pouvant dépendre d'un verbe, nous considérons l'attribut du sujet et de l'objet, bien qu'une autre analyse, mieux en accord avec le PCPA, soit envisageable (cf. Chapitre 7).
L'attribut du sujet peut être adjectival, nominal ou même syntagme prépositionnel. Il peut être pronominalisé par fr. *le/l'* it. *lo/l'*, quel que soit le nombre et le genre de l'attribut.

(299) a. Paul semble en colère.
b. Pierre paraît épuisé.
c. Pierre est médecin.

L'attribut de l'objet ne peut pas être pronominalisé. Il est généralement placé après l'objet, mais on ne peut pas exclure qu'il apparaisse avant lui.

(300) Paolo crede Maria innamorata.
Paolo croit Maria amoureuse

L'interrogation sur l'attribut peut se faire par fr. comment, it. come, mais également par fr. que, it. che. La relativisation utilise le relatif fr. que, it. che (cosa).

(301) a. Comment semble Pierre?
b. Qu'est donc Pierre?

(302) a. Come pare Piero?
Comment paraît Piero?

b. Che cosa è questo?
(qu'est ce que c'est?)

4 Les compléments prépositionnels

On distingue dans les deux langues parmi les compléments prépositionnels les "prépositions" qui sont des marqueurs de cas des autres.
Pour le français et l'italien, on distingue deux marqueurs fr. *à*, it; *a* (qui marquent les fonctions a1-objet et a2-objet), un marqueur fr. *de*, it. *di*, qui marque la fonction gén-objet, un marqueur fr. *de*, it. *da* qui marque la fonction locatif-de-provenance (loc-prov). Ces fonctions donnent toutes lieu à une pronominalisation de tous le SP, ce qui est un indice du statut de marqueur de la préposition. Pour les autres prépositions régies (c'est-à-dire dont la forme est contrainte par le verbe), mais qui ne donne pas lieu à une pronominalisation clitique, on définit la fonction obl-objet.

4.1 a1-objet et a2-objet

L'a1-objet est réalisé par un syntagme prépositionnel — avec la préposition marqueur *à* en français, et *a* en italien — ou par un clitique marqué pour le cas datif1 dans les deux langues. On distingue l'a1-objet de l'a2-objet, qui peut être réalisé par un clitique datif2. On a donc un a1-objet en (303), où les syntagmes prépositionnels sont en concurrence avec un clitique datif1 (304).

(303) a. Paolo parla a Maria.
a'. Paolo parle à Marie.

(304) a. Paolo le parla.
a'. Paolo lui parle.

Mais en (305), il s'agit d'un a2-objet. Il y a concurrence avec un syntagme prépositionnel avec pronom fort dans le cas :animé + (306a,a'), ou avec un clitique datif2 dans le cas :animé - (306b,b').

(305) a. Paolo pensa a Maria.
a'. Paolo pense à Maria.

b. Paolo pensa alla sua tesi.
b'. Paolo pense à sa thèse.

(306) a. Paolo pensa a lei.
a'. Paolo pense à elle.

b. Paolo ci pensa.
b'. Paolo y pense.

En italien, l'a1-objet et l'a2-objet donnent lieu à une relative particulière avec le pronom relatif datif *cui*, qui apparaît sans préposition (Cinque, 1988b:457)[12] :

(307) a. l'autore cui è stato assegnato il premio
l'auteur REL-datif a été décerné le prix
(l'auteur à qui a été décerné le prix.)

b. la risoluzione cui hanno costretto la maggior parte di voi
la résolution REL-datif ils-ont contraint la majeure partie d'entre vous
(la résolution à laquelle ils ont contraint la majeure partie d'entre vous)

Etant donné la similitude des fonctions a1-objet et a2-objet, qui ne diffèrent que par la cliticisation, mais qui partage les mêmes structures pour tous les autres cas, on les représente par les mêmes classes dans la MGf et la MGi. En interne on utilise une seule fonction a-objet, mais qui comporte un trait :type-cl (valant a1 ou a2).
Ce trait est spécifié par les lexèmes, lors de la spécification de leur sous-catégorisation initiale. Par exemple, /PENSER/ sélectionne un sujet initial et un a-objet initial :type-cl=a2, alors que /PARLER/ sélectionne un sujet initial et un a-objet initial :type-cl=a1.
La valeur du trait :type-cl n'est pertinente que pour les réalisations clitiques. Le datif1 impose le trait :type-cl=a1, le datif2 impose le trait :type-cl=a2.
Outre la grande similitude entre le a1-objet et le a2-objet, on peut donner un autre argument en faveur de leur fusion, qui est que l'on n'a jamais cooccurrence des deux fonctions. Dorénavant on ne précisera a1-objet ou a2-objet que lorsque cela sera pertinent, autrement on utilise a-objet.

4.1.1 Représentation dans la MGf et la MGi

On distingue différents cas, selon que le a-objet projette un SP ou pas. On considère également l'ellipse, comme une réalisation possible de l'a-objet.
En effet une propriété de l'a-objet est de pouvoir être omis de manière très régulière[13] (on a défini au Chapitre 4, la distinction entre ellipse et sous-catégorisation réduite, et on a annoncé la représentation de l'ellipse comme des cas particuliers de consommation de la liste de sous-catégorisations).
On donne Figure 80 les classes pour les réalisations grammaticales de l'a-objet pour la MGf et la MGi. Les deux MG diffèrent uniquement par la valeur assignée à la co-tête pour la prépostion marqueur (on a *à* en français, et *a* en italien), et par le fait que l'italien dispose d'une relative particulière pour l'a-objet (en sus de la relative qui suit le schéma général pour les compléments prépositionnels).
La classe a-objet-omis "consomme" un actant dans la liste de sous-catégorisation, sans placer, par une description partielle, de position pour cet actant.
La classe a-objet-SP introduit la structure de SP. Le marqueur fr. *à*, it. *a* y est placé comme co-tête (constante ?a). Pour les positions canoniques (nominale ou phrastique, cf. infra la description des réalisations phrastiques), on place le SP en *à* à droite de l'ancre.

[12] Cinque parle de "complément de mouvement à lieu figuré" pour le a2-objet.
[13] Toutefois on utilise un trait pour bloquer éventuellement l'ellipse du a1-objet ou a2-objet pour certains lexèmes, ou pour certaines constructions, comme le caus-à par exemple (le causatif d'un transitif, sans causé, étant un cas de caus-par).

Les classes pour les positions non canoniques héritent également de la description générale de la construction grammaticale qu'elles représentent (par exemple a-objet-clivé hérite des clivées en général).
Pour l'italien, outre la relative typique des compléments prépositionnel (cf. infra Section 7) on a une relative en *cui* (cf. les exemples (307)), qui n'hérite pas de a-objet-SP, puisque l'a-objet ne projette pas un SP.

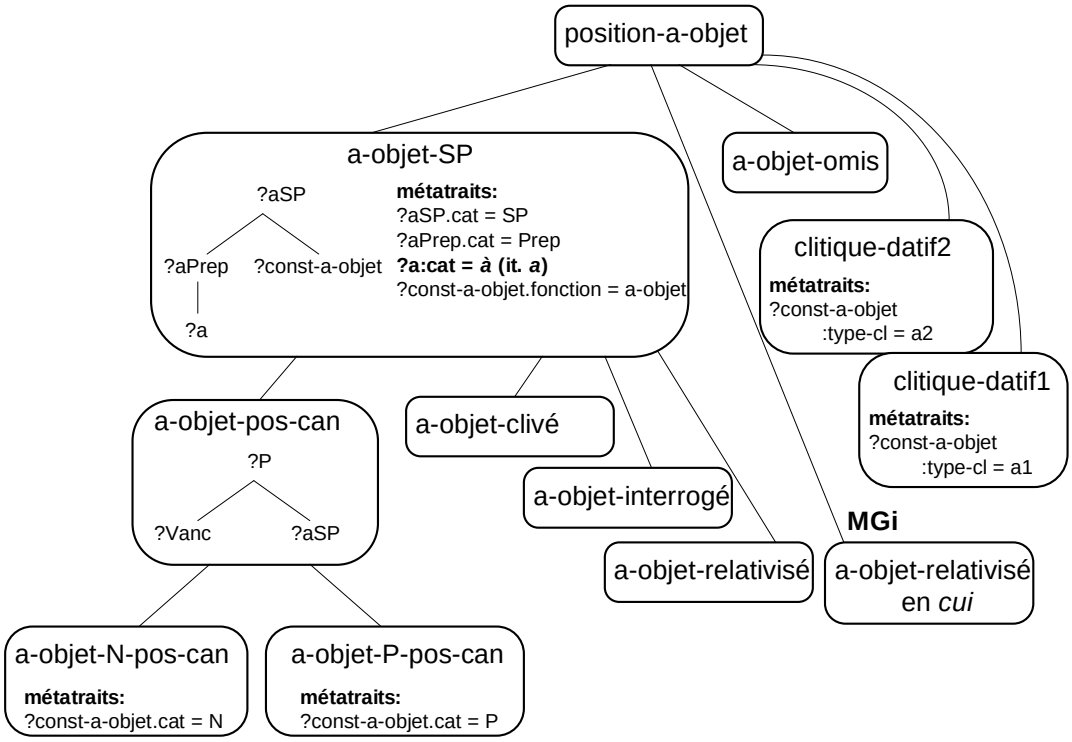


Figure 80. Les réalisations grammaticales de l'a-objet (MGf et MGi)

4.2 gén-objet

Le gén-objet est réalisé par un syntagme prépositionnel - avec la préposition marqueur fr. *de* et it. *di* **ou** *da* - ou par un clitique marqué pour le cas génitif (Fr. *en*, It. *ne*).dans les deux langues :

(308) a. Paolo approfitò subito **della**[14] patenta.
b. Paolo profita tout de suite **du** permis de conduire.

(309) a. Paolo ne approfittò subito.
b. Paolo en profita tout de suite.

(310) a. Maria dipende **da** Giovanni.
b. Maria dépend **de** Giovanni.

[14] Il s'agit de l'amalgame de *di* et *la.*

(311) a. Maria ne dipende.
b. Maria en dépend.

En français, le gén-objet donne en outre lieu à une relative spécifique, introduite par le pronom *dont*, en sus de la relative qui suit le schéma général des relatives pour les compléments prépositionnels (en *de* + LEQUEL) :

(312) Le permis de conduire, dont Pierre profita dès qu'il put, ...

D'où le schéma suivant pour les classes représentant le gén-objet dans la MGf et la MGi :

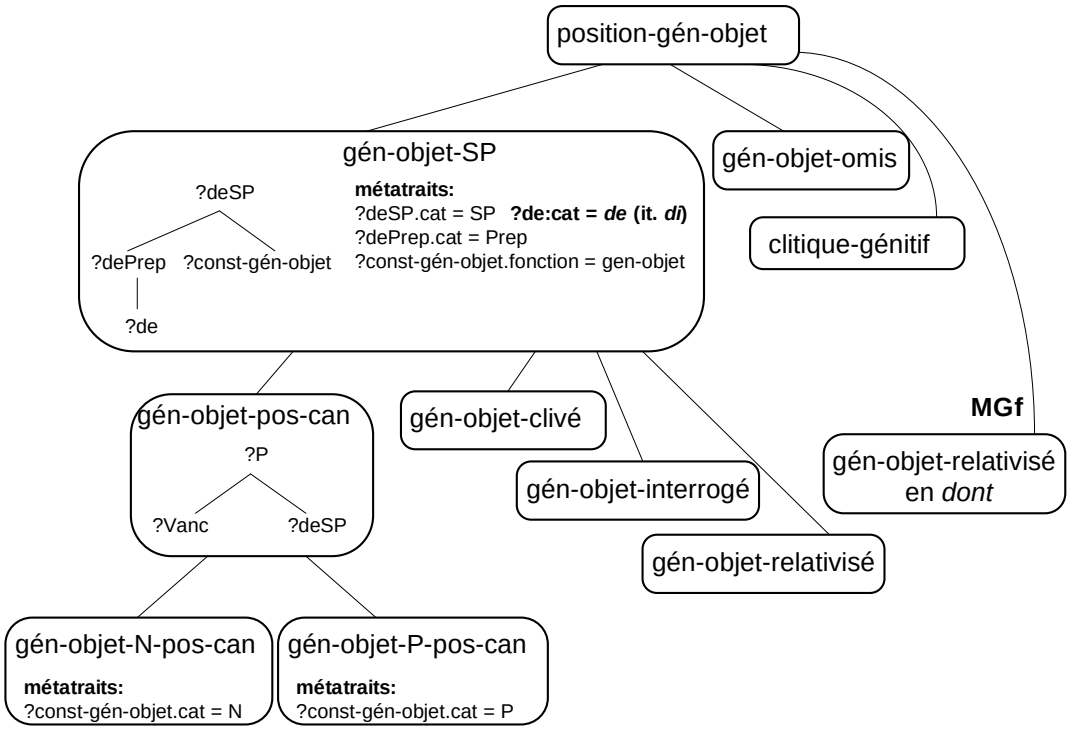


Figure 81. Les réalisations grammaticales du gén-objet (MGf et MGi)

Les prépositions fr. *de* et it. *da* marquent également la fonction "locatif de provenance", qui donne également lieu au clitique fr. *en*, it. *ne.* Mais cette fonction est caractérisée par des pronoms relatifs et interrogatif spécifiques (fr. *d'où*, it. *da dove*) (cf. infra Section 4.4). Pour certains verbes prenant un locatif de provenance, la lecture figurée bloque ce pronom interrogatif (Salvi, 1988:67). Nous considèrons alors qu'ils prennent deux sous-catégorisations initiales, une avec loc-prov (lecture littérale), et une avec gén-objet (lecture figurée).

(313) a. Piero si allontana da Roma / Piero se ne allontana / Da dove si allontana Piero?
a'. Piero s'éloigne de Rome / Piero s'en éloigne / D'où s'éloigne Piero?
b. Piero si allontana da Maria / Piero se ne allontant / Da chi si allontana Piero?
b'. Piero s'éloigne de Maria / Piero s'en éloigne / De qui s'éloigne Piero?

4.3 Les autres compléments prépositionnels : obl-objet et agt-objet

Les autres compléments prépositionnels à préposition régie sont l'agt-objet (pour le complément d'agent), et l'obl-objet, fonction utilisée pour tout autre complément à préposition régie (par exemple pour *compter sur N1*). Pour les cas avec obl-objet, la MGf et la MGi définissent des schèmes où la préposition n'est pas instanciée, elle soit être spécifiée dans le lexique syntaxique, pour chaque lexème sous-catégorisant un obl-objet. L'obl-objet n'est que nominal. On donne Figure 82 les classes pour le obl-objet.

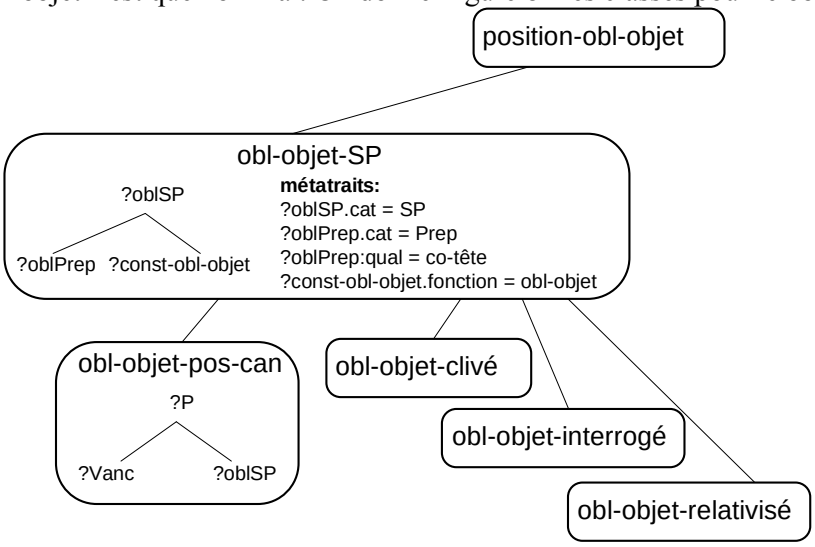


Figure 82. Les classes pour le obl-objet (MGf et MGi)

Pour le agt-objet, on a vu au Chapitre 4 que l'ellipse est une de ses propriétés. Les schèmes avec un agt-objet comprennent la préposition fr. *par*, it. *da* comme co-tête.
On a vu également que l'agt-objet n'est jamais phrastique (les cas de sujet phrastique démis au passif, sont des cas de passifs en fr. *de* et it. *di*, et ils mettent donc en jeu des gén-objets phrastiques).

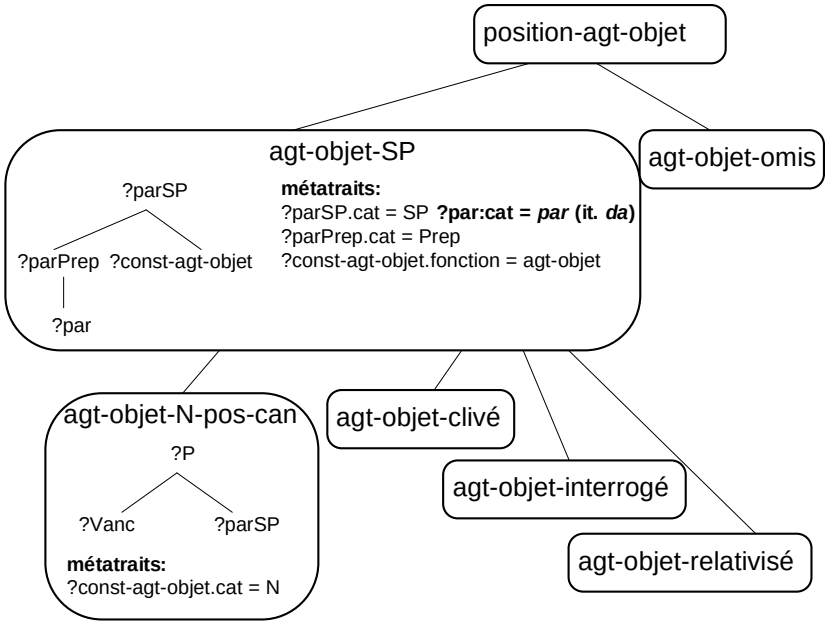


Figure 83. Les classes pour la réalisation du agt-objet (MGf et MGi)

4.4 Le locatif et le locatif de provenance

Enfin, on considère des compléments prépositionnels à sens locatif (de destination, de situation ou de provenance). Ils sont caractérisés par une pronominalisation, un relatif, un pronom interrogatif spécifique.
Outre leur fonction de modifieur possible (cf. Chapitre 6), ils peuvent en effet être complément requis de certains verbes (fr. *aller* it. *andare*, fr. *revenir* it. *tornare*).

Le locatif, de destination ou de situation, que nous désignons simplement par locatif, est caractérisé syntaxiquement par un clitique particulier (Fr. *y* , It. *ci*) (314b,b'), et par un pronom relatif et interrogatif spécifique (Fr. *où* , It. *dove*) (cf. série (315)) (outre la relative et l'interrogative régulière pour les compléments prépositionnels).

(314) a. Marco va a Roma.
a'. Marco va à Rome.
b. Marco ci va.
b'. Marco y va.

(315) a. Dove va Marco?
a'. Où va Marco?
b. il luogo dove va Marco...
b'. le lieu où va Marco...

Lorsqu'il n'est pas pronominalisé, il est réalisé par un syntagme prépositionnel dont la préposition peut varier (en fonction du verbe et du nom complément de la préposition) ou bien également par un adverbe locatif (par exemple fr. *là*, it. *qui*). On donne ci-dessous (Figure 84) un schème pour un verbe avec locatif, sans autre complément. Les prépositions locatives (Cf. Chapitre 6) ancrent à elles seules des arbres de racine SP,

avec une trace pour leur argument 0 (l'entité localisée). On a vu en effet au Chapitre 2, que l'on considère les compléments locatifs comme des cas de contrôle, l'argument 0 de la préposition est partagé avec le verbe sous-catégorisant le locatif (Figure 85). Les adverbes locatifs ancrent également des arbres de projection SP. Ces arbres de racine SP peuvent venir se substituer comme complément locatif.

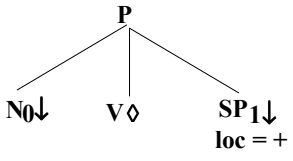


Figure 84. Schème pour un verbe avec locatif (par exemple ALLER)

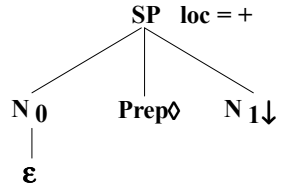


Figure 85. Schème pour une préposition locative (par exemple *dans*)

Le locatif de provenance est réalisé par un syntagme prépositionnel - avec la préposition *de* en français, et *da* en italien - ou par un clitique (Fr. *en*, It. *ne*). La relative et l'interrogative sont formées avec la préposition suivie du pronom Fr. *où*, It. *dove* (ou d'un SN relatif ou interrogatif) (série (317)).

(316) a. Pia si allontana da Roma.
a'. Pia s'éloigne de Rome.
b. Pia se ne allontana.
b'. Pia s'en éloigne.

(317) a. Da dove viene Pia?
a'. D'où vient Pia?
b. il paese da dove viene Marco...
b'. le village d'où vient Marco ...

Le syntagme prépositionnel peut être composé de la préposition suivie d'un SN (316a,a'), mais aussi d'un SP locatif, ou un adverbe locatif.

(318) a. L'animal vient de sous le lit.
a'. L'animale viene da sotto il letto.
b. Pierre vient de là.
b'. Pierre viene da qui.

5 Les réalisations comme clitiques pronominaux

Suivant Miller (1991), nous proposons un traitement des clitiques comme affixes flexionnels. La MGf et la MGi se chargent de déterminer (par des traits) quel est le verbe portant les clitiques compléments et celui portant le clitique sujet (les deux hôtes pouvant différer). Mais c'est la morphologie qui détermine l'ordre respectifs des différents clitiques, et le placement par rapport à l'hôte (proclitique ou enclitique).

La réalisation clitique constitue donc un cas particulier de réalisation pour un actant grammatical, qui ne s'opère pas en syntaxe mais en morphologie. Nous considérons tout de même qu'il s'agit d'un cas de réalisation, à distinguer de la "non réalisation", au sens phonologique (comme les cas d'ellipse, ou les cas de non réalisation du sujet, à l'infinitif ou à l'impératif).

Nous donnons dans cette section un rappel des faits linguistiques et des arguments en faveur de leur statut d'affixe, puis le système de traits pour le français et l'italien, pour gérer la détermination de l'hôte des clitiques, y compris en cas de montée, ainsi que le système pour l'interprétation sémantique.

L'implémentation de la MGi et la MGf n'intègre pas cette analyse, mais utilise un placement des clitiques sous la forme de noeuds à substituer dans les arbres élémentaires. On fournit Section 5.4.1 une correspondance systématique entre les arbres élémentaires avec clitiques tels que nous les aurons décrits dans toute cette section, et tels qu'ils sont implémentés.

5.1 Les faits linguistiques

Les clitiques pronominaux constituent une pierre d'achoppement aussi bien pour les théories linguistiques que pour l'implémentation de systèmes effectifs. Au niveau sémantique, ils peuvent soit représenter un argument sémantique de leur gouverneur, soit être non référentiel. Leur statut dans la grammaire se situe entre syntaxe et morphologie. Ils peuvent constituer la réalisation d'actants sous-catégorisés, et pour certains (*en*, *y* en français, *ci*, *ne* en italien) être la réalisation d'un ajout. Ils peuvent être portés par un hôte qui n'est pas leur gouverneur sémantique. Cette interaction avec la sous-catégorisation et le phénomène de montée des clitiques, comparable à l'extraction bien que les contraintes soient très différentes, semble relever du domaine de la syntaxe. Mais les études précises tant pour le français (Miller, 1991; Miller et Sag, 1997) que pour l'italien (Monachesi, 1995) montrent qu'ils ont les caractéristiques d'affixes flexionnels. D'un point de vue diachronique en français, les clitiques ont vraisemblablement joué le rôle de mots à part entière jusqu'à une certaine époque, mais ont été intégrés, ou sont en voie d'intégration à un système flectionnel (Miller, 1991).

5.1.1 Caractérisation des clitiques

Les clitiques pronominaux sont marqués pour le cas (et en français, il y a correspondance biunivoque entre le cas et la fonction grammaticale), et peuvent être marqués suivant les cas, pour la personne, le genre, le nombre, le trait "défini".
On sait bien que la conjonction de ces traits morphologiques ne permet pas de représenter de manière adéquate l'ordre respectif des clitiques dans le cas où il sont plusieurs. Pour représenter cet ordre, et également certaines incompatibilités entre formes, on peut classer les différentes formes clitiques pronominales en ensembles disjoints, mutuellement exclusifs, que nous nommerons *colonnes*, et que ce classement ne peut pas s'exprimer en utilisant les propriétés énoncées ci-dessus (de cas, personne ...). On redonne ci-dessous le tableau des colonnes, pour les formes proclitiques pour le

français. On ne désambiguise les formes que dans le cas où cela donne lieu à un placement différent (c'est le cas pour *nous, vous* selon qu'ils sont nominatifs ou pas).

colonne I	colonne II	colonne III	colonne IV	colonne V	colonne VI
je, tu *il, elle, on* *nous, vous* (nom.) *ils, elles*	*me te se* *nous* (acc/dat1) *vous* (acc/dat1)	*le, la, les*	*lui, leur*	*y*	*en*

Table 12 : Les clitiques pronominaux français classés par leur colonne

Les formes clitiques appartenant à la même colonne sont mutuellement exclusives, et les colonnes II et IV sont également mutuellement exclusives. L'ordre de placement, dans le cas proclitique en français se fait rigoureusement en fonction des colonnes auxquelles les clitiques pronominaux appartiennent, dans l'ordre I<II<III<IV<V<VI.
La colonne I est à distinguer des autres, car, d'une part elle est homogène pour le cas nominatif, et (i) le clitique *ne* peut s'intercaler entre un clitique nominatif et les autres clitiques pronominaux; (ii) les clitiques sujet et les clitiques compléments ne sont pas soumis aux mêmes règles de placement dans le cas de montée des clitiques (avec l'analyse des verbes à montée qui ne sous-catégorisent pas un sujet, on considère en TAG que le placement du clitique sujet sur le verbe à montée est un cas de montée des clitiques) et (iii) les clitiques sujet peuvent avoir portée large sur une coordination de groupes verbaux (cf. infra).

Pour l'italien, on a une situation parallèle, c'est-à-dire que l'ordre respectif ne peut pas s'exprimer en fonction de propriétés comme la fonction ou la personne (Monachesi, 1995). Il faut donc recourir à la notion de colonne comme pour le français. Monachesi fournit les colonnes suivantes pour l'italien (on désambiguise les formes dans le cas où cela donne lieu à un placement différent) :

colonne I	colonne II	colonne III	colonne IV	colonne V	colonne VI
mi, ti, gli, *le* (dat1)*, vi*	*ci*	*si* (réfl.)	*lo, la, li,* *le* (acc)	*si* (nominatif)	*ne*

Table 13 : Les clitiques pronominaux italiens classés par leur colonne

Les formes d'une même colonne sont mutuellement exclusives. Le groupe de clitiques se place devant le verbe hôte si celui-ci est à un mode fini non impératif, derrière le verbe hôte si celui-ci est à l'impératif (positif) ou à un mode non fini (infinitif, gérondif, participe) (Cordin et Calabrese, 1988). Et à la différence du français, il n'y a pas de différence dans l'ordre respectif ou dans la forme des clitiques, selon que le groupe clitique est proclitique ou enclitique.
L'italien ne dispose que d'un seul clitique nominatif (le SI-nominatif, cf. Chapitre 4), à sens impersonnel. Et à la différence du français, ce clitique ne se distingue pas des clitiques compléments pour ce qui est de la négation, ou de la coordination.

5.1.2 Les clitiques en syntaxe ou morphologie?

Il existe un ensemble d'arguments pour traiter les clitiques pronominaux dans les langues romanes comme des affixes[15] (cf. Miller, 1991 pour le français, Monachesi, 1995 pour l'italien). Les auteurs s'appuient sur les critères de Zwicky et Pullum (1983) permettant de conclure au statut d'affixe. On redonne brièvement ces critères :

- Le degré de sélection par rapport à l'hôte.
  Les clitiques pronominaux s'attachent toujours sur un verbe, même s'il dépendent sémantiquement d'un nom ou d'un adjectif.
- La rigidité de l'ordre respectif des clitiques.
  On a vu ci-dessus l'ordre fixe en italien et en français.
- L'impossibilité d'avoir une portée large dans le cas de coordination.
  Cet argument est cependant à nuancer pour le français dans le cas de clitiques sujets. Il est en effet tout à fait possible de trouver des exemples de coordination où le clitique nominatif a une portée large (319)[16].

(319) Nous apprécions et admirons son travail. (Miller, 1991)

- L'existence de "trous" dans les tableaux de combinaisons de clitiques.
  Comme on l'a vu pour le français et l'italien, les formes de la même colonne sont mutuellement exclusives.
- Les idiosyncrasies morphophonologiques.
  La forme des clitiques peut être altérée d'une manière explicable par des généralisations en phonologie. Ainsi par exemple les clitiques se terminant par -i ont une forme concurrente en -e utilisée obligatoirement et exclusivement devant un autre clitique commençant par n- ou l-.

(320) Martina te/*ti lo spedirà.
Martina DAT.2 ACC.3ms enverra.

Notre propos n'est pas d'apporter des arguments linguistiques pour ou contre cette analyse. En revanche, on peut vérifier l'impact de l'analyse en syntaxe ou en morphologie sur une grammaire de bonne couverture, en termes de couverture, d'efficacité et d'élégance de traitement. Ce qui est clair, c'est que la rigidité de l'ordre des clitiques, et leur classement en colonnes qui ne se ramène pas à un classement par fonctions syntaxiques, rend difficile leur gestion en syntaxe. D'un autre côté, leur prise en charge par la morphologie n'a pas été testée à notre connaissance pour une application réelle.
Il est à noter que même si un traitement en morphologie complexifie le module morphologique, un traitement en syntaxe implique de toutes façons un prétraitement morphologique (ou post-traitement en génération). En effet pour le français, les clitiques inversés sont accrochés au verbe par un trait d'union. Pour l'italien, les enclitiques sont

[15] Le terme clitique est alors incompatible avec celui d'affixe mais nous gardons le terme classique de clitique, par commodité.
[16] Miller (1991) estime qu'il ne s'agit pas du système français oral actuel, mais que ces cas relèvent du système écrit ou oral formel. Ils sont cependant bien attestés.

systématiquement amalgamés à la forme verbale de mode non fini (inifinitif, gérondif, participe présent, participe passé[17]).

(321) a. Vorrei finirlo.
Je-voudrais finir-ACC3ms.
Je voudrais le finir.

b. Avendolo finito, Paolo uscì.
AVERE-gérondif-ACC3ms fini, Paolo sortit.
L'ayant fini, Paolo sortit.

c. Gli argomenti riguardantici, ...
Les questions reguardant-acc1p, ...
(Les questions nous reguardant, ...)

d. Maltrattatala, Carlo uscì.
Maltraitée-ACC3fs, Carlo sortit.
(L'(ayant) maltraitée, Carlo sortit.)

5.2 L'analyse des clitiques pronominaux comme affixes

On utilise le terme clitiques compléments, pour les clitiques non nominatifs, qu'il s'agisse de clitiques pronominaux qui réalise un actant syntaxique complément, ou bien de clitiques figés, ou bien de réfléchis (intrinsèques, moyen ou vrai réfléchi).

Les clitiques (sujet et compléments) sont considérés comme des affixes flexionnels dans la MGf et la MGi (cf. infra Section 5).On reprend le terme classique d'**hôte** pour désigner l'élément sur lequel les clitiques sont réalisés en morphologie comme préfixe ou suffixe.
On désigne par le terme **gouverneur** d'un clitique, l'élément qui requiert la présence de ce clitique. Dans le cas où le clitique est la réalisation d'un actant sous-catégorisé, le gouverneur est le verbe qui sous-catégorise cet actant. Dans le cas où le clitique est un marqueur (comme clitique idiomatique, ou comme réfléchi), le gouverneur est le verbe qui déclenche la présence de ce clitique.

L'hôte d'un clitique complément peut être un verbe distinct du gouverneur du clitique, on parle alors généralement de "montée des clitiques", puisqu'en effet l'hôte est un verbe dominant syntaxiquement le gouverneur du clitique. La montée des clitiques est extrêmement contrainte, d'une manière qui diffère des contraintes régissant les dépendances non bornées entre un constituant extrait et son gouverneur.
En TAG, pour les cas de montée des clitiques, comme les cas de dépendances non bornées, il n'y a pas à proprement parler de mouvement. Les éléments extraits, les clitiques "montés" sont localisés dans l'arbre de leur gouverneur. A la dérivation, une adjonction peut venir ajouter une distance entre l'élément extrait ou le clitique "monté" et son gouverneur.

[17] L'italien permet en effet la cliticisation sur un participe passé. Celle sur un participe présent est en revanche plus limitée qu'en français, elle n'est permise que pour les formes qui sont datives1, ou homonymes d'un datif1 (Benincà et Cinque, 1991).

Dans le cadre du traitement des clitiques comme affixes flexionnels, ce qui est localisé dans l'arbre du verbe qui est le gouverneur d'un clitique, ce n'est pas une position syntagmatique pour le clitique, mais des traits représentant les caractéristiques morphologiques pour la réalisation des clitiques, ainsi que des traits pour l'interprétation sémantique des clitiques.
On donne ci-dessous le système de traits morphologiques pour la réalisation des clitiques pour le français et l'italien, et les contraintes de propagation pour gérer les cas où l'hôte n'est pas le gouverneur. Puis nous complétons la description de l'analyse en abordant le cas des clitiques ajouts (Section 5.2.3), puis le système pour interpréter la sémantique des clitiques lorsqu'ils sont la réalisation d'un argument sémantique (Section 5.2.4).

5.2.1 Le système utilisé pour le français

On a mentionné que l'hôte d'un clitique complément peut être distinct de son gouverneur. Pour le clitique sujet, certaines théories linguistiques (comme HPSG) considèrent que qu'il a toujours pour hôte son gouverneur. Dans le cas de TAG cela est vrai sauf dans le cas d'un verbe à montée. En effet l'analyse TAG des verbes à montée (cf. Chapitre 3) revient à considérer que le verbe à montée ne sous-catégorise pas de sujet, mais s'adjoint sur un noeud verbal, en en prenant le sujet (syntagmatique ou clitique).
Donc en TAG, pour le français, on est amené à distinguer pour un verbe gouverneur de clitiques, l'hôte du clitique sujet et l'hôte des clitiques compléments. Dans une phrase comme (322), le gouverneur est *fait*, l'hôte du clitique sujet est *semble*, et l'hôte du clitique complément est *avoir*.

(322) Il semble l'avoir fait.

Dans la MGf, on utilise l'attribut :clC pour spécifier les traits morphologiques des clitiques compléments. La valeur de l'attribut :clC est une structure de traits, organisée selon le cas des clitiques présents. Dans le cas d'aucun clitique on note en abrégé :clC = -. Les contraintes sur les groupes de clitiques possibles sont gérés par le module morphologique.

clC =	acc =	num =	pl/sing	*me te le la nous vous les en*
		pers =	1/2/3	
		gen =	fem/masc	
		déf =	+/-	
	dat1 =	num =	pl/sing	*me te lui nous vous leur*
		pers =	1/2/3	
	dat2 =	+/-		*y*
	réfl =	pers =	1/2/3	*me te se nous vous*
		num =	pl/sing	
	gén =	+/-		*en*
	loc =	+/-		*y*
	loc-prov =	+/-		*en*

Figure 86. La structure de traits valeur de l'attribut :clC, et les formes clitiques correspondantes

On considère le réfléchi à part, étant donné qu'un réfléchi généralement dit "accusatif" (provenant de la fusion de l'argument sujet et de l'argument objet, ou bien marqueur de moyen) peut être cooccurrent à un clitique accusatif indéfini *en*.

On utilise l'attribut :clS pour spécifier les traits morphologique du clitique sujet. La valeur de l'attribut :clS vaut soit - dans le cas où le verbe qui porte le trait ne porte pas de clitique sujet, sinon la valeur de l'attribut :clS est une structure de traits, contenant les attributs :num, :pers, :gen, :indéf (pour le clitique *on*). La valeur de clS comprend en outre un trait :dir (pour direction) qui vaut - dans le cas d'aucun clitique sujet, pre, dans le cas préverbal, et inv dans le cas sujet inversé.

Dans la MGf, les contraintes sur les valeurs pour les traits :clC et :clS sont imposées comme traits amont sur ?Vmarq. Prenons l'exemple des schèmes pour un verbe transitif strict.

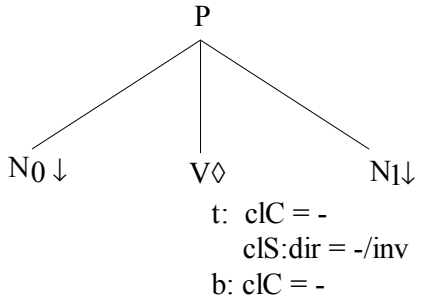


Figure 87. Schème pour un transitif, sans clitique argumentaux[18]

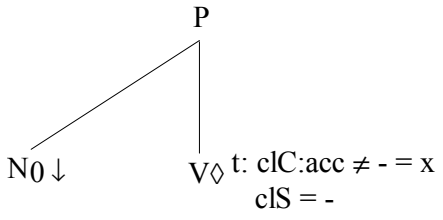


Figure 88. Schème pour un transitif avec accusatif

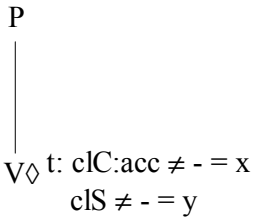


Figure 89. Schème pour un transitif avec accusatif et nominatif

L'adjonction d'un verbe à montée ou d'un auxiliaire peut venir rediriger ces contraintes vers l'ancre de l'arbre adjoint (i.e. le verbe à montée ou l'auxiliaire), le comportement du trait pour le clitique sujet étant différant dans les deux cas.

[18] On impose clS:dir = -/inv, pour bloquer le clitique sujet préverbal, mais pour permettre éventuellement la présence d'un clitique sujet inversé, comme inversion complexe, c'est-à-dire s'ajoutant à un sujet nominal.

Le fait d'utiliser un trait :clC, qui comprend les spécifications de tous les clitiques compléments, va bien représenter le fait que pour un gouverneur donné, les clitiques compléments ne peuvent pas être séparés, c'est-à-dire qu'ils ont tous le même hôte.
On donne ci-dessous le schème pour un auxiliaire de temps et celui pour un verbe à montée (traits clitiques uniquement).

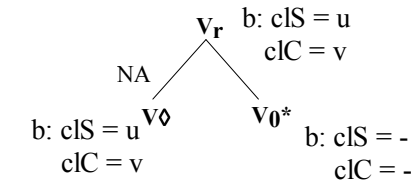


Figure 90. Schème pour un auxiliaire

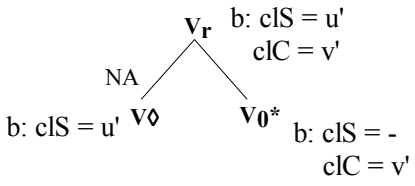


Figure 91. Schème pour un verbe à montée

Les deux schèmes ont la particularité de dévier la propagation du trait :clS. Pour le trait :clC, seul l'auxiliaire dévie sa propagation. Pour le verbe à montée, le propagation de :clC continue à se faire entre le noeud pied et la racine. On va voir par des exemples, que cela permet de déterminer l'hôte du clitique sujet et celui des clitiques compléments. L'analyse ne permet pas seulement de gérer ces cas, mais elle bloqie également les cas inacceptables, et ce en utilisant les mêmes schèmes pour les formes avec ou sans clitiques.
Les arbres lexicalisés construits à partir de ces schèmes s'adjoignent sur ?Vmarq. L'adjonction est bloquée si ?Vmarq domine une ancre portant des clitiques. Par exemple, l'arbre élémentaire de la Figure 92 n0Vn1-cl0-cl1[il le fait], ne peut recevoir ni de verbe à montée ni d'auxiliaire (outre l'incompatibilité de mode, qui n'est pas montrée ici, il y a incompatibilité sur les traits clitiques.

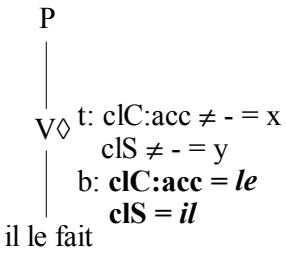


Figure 92. n0Vn1-cl0-cl1[il le fait]

On donne ci-dessous les figures pour la dérivation de *Il l'a fait* et *Il semble le faire*.

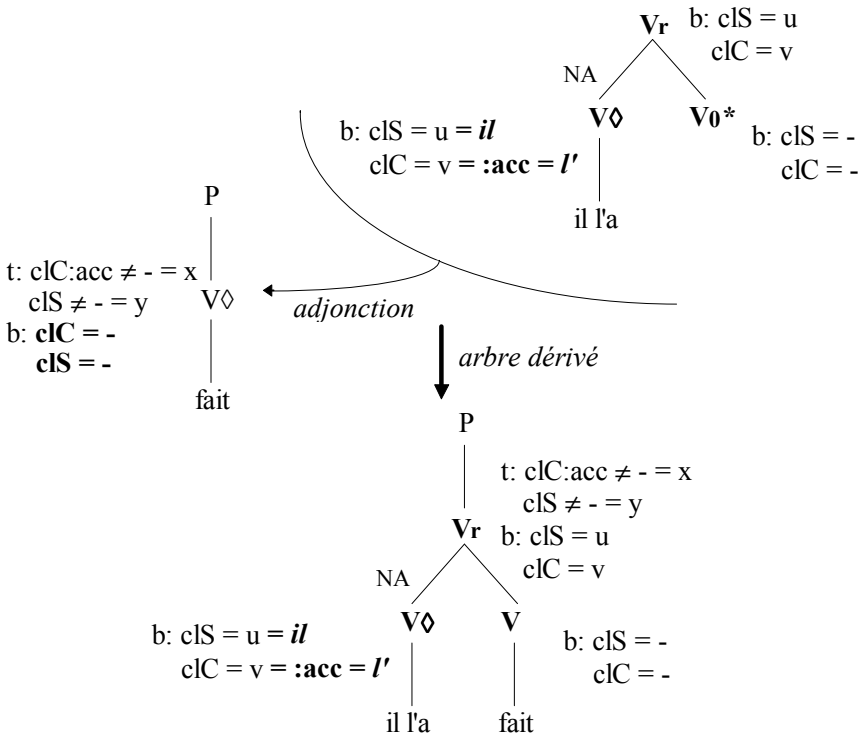


Figure 93. Derivation pour *Il l'a fait.*

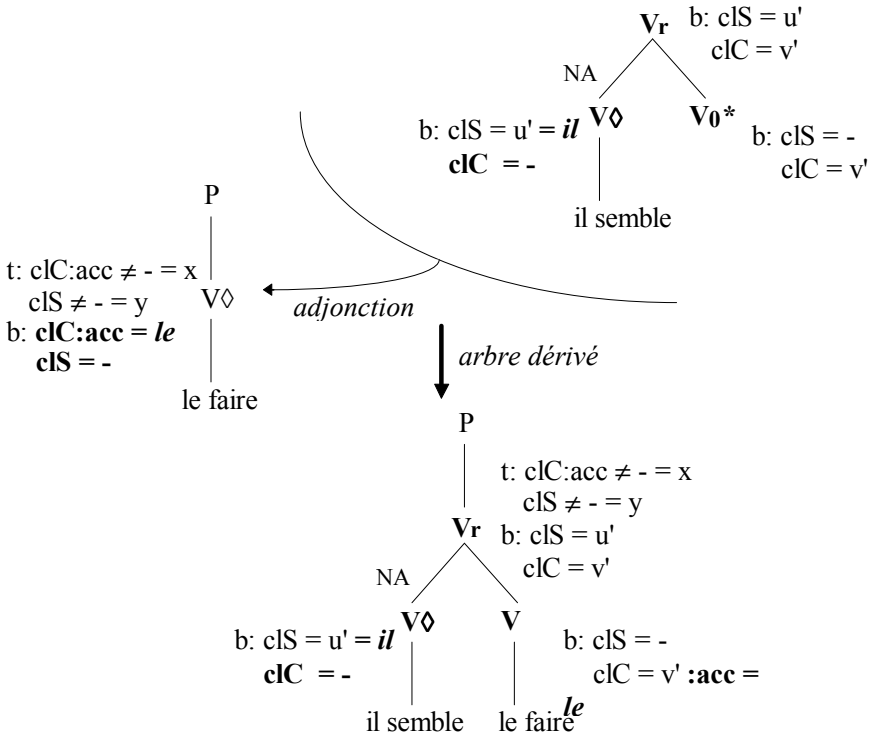


Figure 94. Derivation pour *Il semble le faire.*

Dans le cas d'interactions entre auxiliaires et verbes à montée, les adjonctions se font en cascade, sur la racine de l'arbre du verbe à montée ou du verbe auxiliaire. On donne ci-dessous l'exemple de la dérivation pour *Il semble l'avoir fait.*

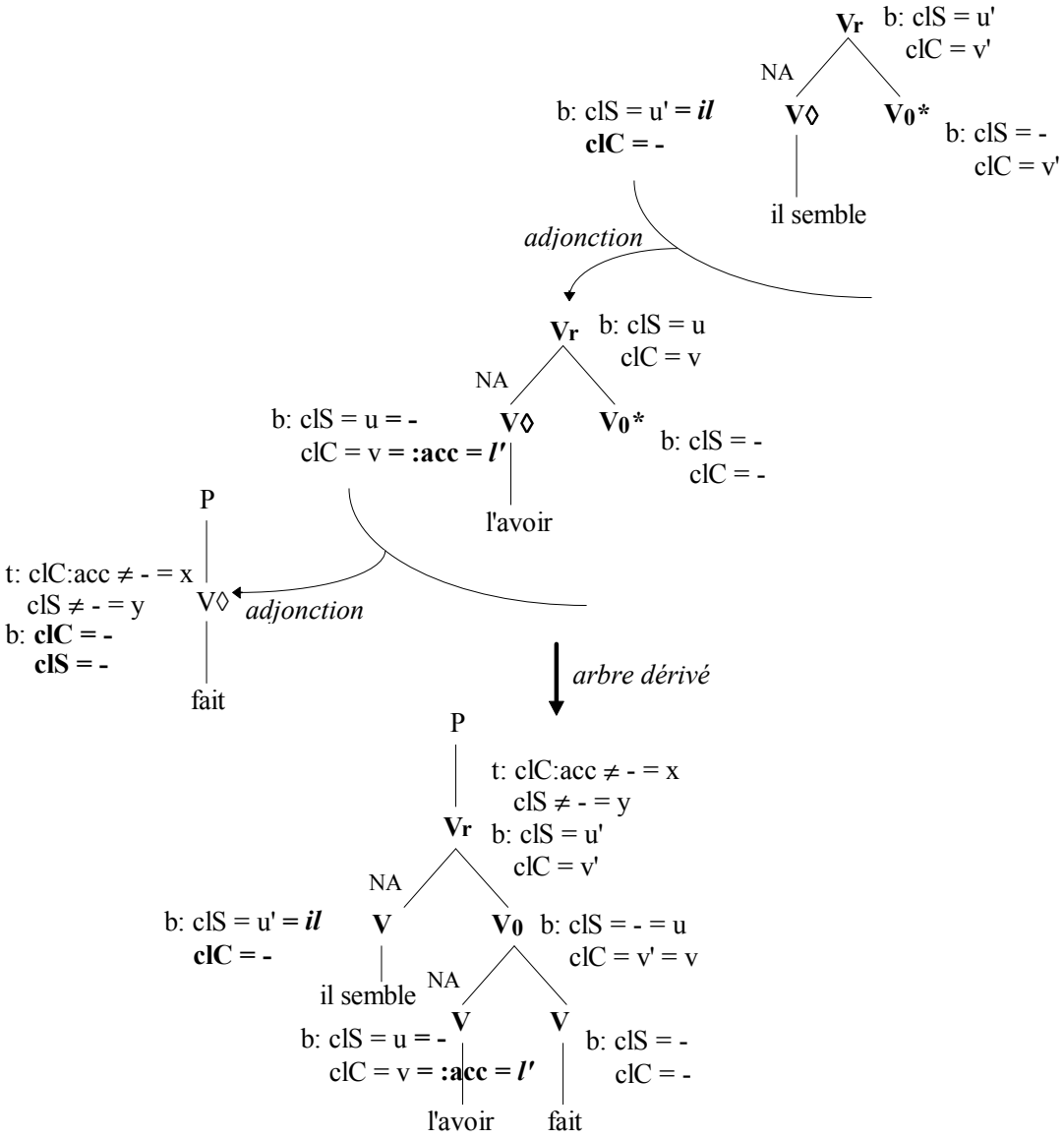


Figure 95. Dérivation pour *Il semble l'avoir fait.*

Il faut noter que les schèmes donnés pour les verbes à montée et auxiliaires imposent les contraintes appropriées sur les clitiques que peut porter une forme fléchie ancrant ces schèmes.

Dans le cas d'une adjonction d'un adverbe sur un noeud verbal, les traits pour les clitiques sont simplement propagés entre la racine et le noeud pied de l'arbre auxiliaire pour l'adverbe.

Le traitement comme affixes des clitiques pronominaux va de pair avec un traitement de la particule négative *ne* comme affixe. La gestion de la réalisation de *ne* en morphologie permet de représenter les cas d'haplologie, où *ne* est requis de par diverses sources indépendantes (comme dans *Personne ne donne jamais rien à personne*, Miller, 1991); ainsi que le fait que le *ne* ne peut pas avoir portée large sur une coordination.
Comme le montre Recourcé (1995), il faut cependant distinguer deux comportements grammaticaux de *ne* distincts, selon qu'il se combine avec un verbe fini, et il s'agit alors d'un affixe, ou avec un verbe infinitif, et il faut alors le traiter en syntaxe, ce que montre le fait que des éléments non clitiques peuvent apparaître entre *ne* et l'infinitif (323a), ou que ce *ne* a portée large sur une coordination (323b).

(323) a. Je crains de ne vraiment rien trouver d'appétissant dans ce plateau.
b. Jean préfère ne jamais discuter ou contester les ordres (Recourcé, 1995)

5.2.2 Le système utilisé pour l'italien

Le système est à la fois plus simple en italien, étant donné qu'on n'a pas de clitiques nominatifs, sauf le *si* impersonnel (que nous appelons SI-nominatif), mais l'italien comporte un pronom clitique, *loro* (datif1, 3ème personne du pluriel), qui se distingue de tous les autres clitiques par son comportement.

Le SI-nominatif n'est pas à distinguer des clitiques compléments

Le seul clitique sujet en italien est le SI-nominatif. Il n'a pas la distribution des clitiques sujet du français. On peut affirmer que l'hôte du SI-nominatif est aussi l'hôte des clitiques compléments, donc on n'a pas à distinguer une projection verbale supplémentaire pour le SI-nominatif.
En effet, le SI-nominatif et les clitiques compléments se placent sur l'auxiliaire de temps dans le cas de formes composées, et sur les verbes restructurants dans le cas de prédicats complexes verbe restructurant + infinitif.
Dans le cas de verbes à montée, comme en français, les clitiques compléments se placent sur l'infinitif et pas sur le verbe montée, mais à la différence du français, le clitique nominatif non plus. Plus précisément, le SI-nominatif est exclu sur un verbe à montée mais également sur l'infinitif sous un verbe à montée (alors que les clitiques compléments peuvent apparaître sur l'infinitif sous un verbe à montée) (Salvi, 1988:103-104). Ainsi, d'après Salvi, (324a) est un cas d'emploi restructurant de /POTERE/, où tous les clitiques montent, en l'occurrence le SI-nominatif. On peut mettre en évidence cela, par le fait que les verbes à montée qui n'admettent pas d'emploi restructurant, comme /SEMBRARE/, ne peuvent pas porter le SI-nominatif (324b).

(324) a. Si può mangiare bene.
SI-nom peut manger bien.
(Il est possible que l'on mange bien)

b. *Si sembra dormire troppo.
SI-nom semble dormir trop.
(Il semble que l'on dorme trop)

Donc, dans la MGi, on n'a besoin que d'un attribut :cl pour spécifier les traits morphologiques des clitiques. La valeur de l'attribut :cl est - dans le cas d'aucun clitique

(sur le verbe qui porte ce trait), sinon la valeur de l'attribut :cl est une structure de traits, organisée selon le cas des clitiques présents.

	nom = +/-		*si*
cl =	acc =	num = pl/sing	
		pers = 1/2/3	
		gen = fem/masc	*mi ti lo,la ci vi li, le*
		déf = +/-	
	dat1 =	num = pl/sing	*mi ti gli,le ci vi gli*
		pers = 1/2/3	
		gen = fem/masc	
	dat2 = +/-		*ci*
	réfl =	pers = 1/2/3	*mi ti si ci vi*
		num = pl/sing	
	gén = +/-		*ne*
	loc = +/-		*ci, vi*
	loc-prov = +/-		*ne*

Figure 96. La structure de traits valeur de l'attribut :cl, et les formes clitiques correspondantes

Dans la MGi, la valeur de l'attribut :cl est spécifiée comme trait amont sur ?Vmarq, et ne se propage pas plus haut. On a un seul cas d'artère, qu'un clitique soit présent ou pas, c'est-à-dire que :cl soit non nul ou nul. On donne un exemple de schème, pour un transitif, dans le cas d'un clitique accusatif et d'un clitique SI-nominatif. Il peut être ancré par *Li si mangia* (On les mange).

P

|

t: cl:nom = +, cl:acc = (pers = 3, num = pl, gen = masc)

V◊

Figure 97. Exemple de schème, pour l'italien, avec SI-nominatif, et clitique accusatif

Là encore, le fait d'utiliser un trait :cl, qui comprend les spécifications de tous les clitiques, va bien représenter le fait que pour un gouverneur donné, les clitiques ne peuvent pas être séparés.

Pour la montée des clitiques, on utilise le même mécanisme pour un auxiliaire de temps ou pour un verbe restructurant (dans son emploi restructurant) : une adjonction enchâssante, d'un auxiliaire de temps ou d'un verbe restructurant dévie la propagation du trait :cl, vers l'hôte des clitiques. On donne ci-dessous le schème pour un verbe restructurant avec marqueur *a* (il peut être ancré par exemple par *andare a*)

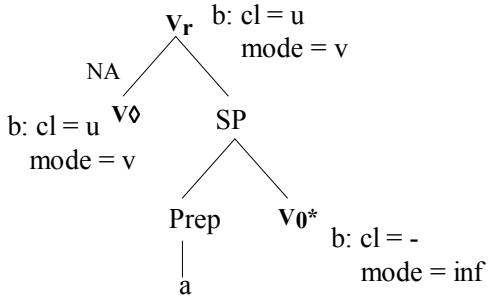


Figure 98. Schème pour un verbe restructurant, avec marqueur *a* (traits :cl et :mode)

Pour les verbes à montée, on avait en français déviation du trait :clS pour le clitique sujet. En italien, on a au contraire la contrainte qu'un verbe à montée ne peut pas s'adjoindre sur un verbe où un SI-nominatif est spécifié.
On donne ci-dessous le schème pour un verbe à montée en italien. On peut voir que le trait :cl spécifiant les clitiques ne passe pas sur l'arbre du verbe à montée, et reste transmis à l'infinitif. Pour représenter le fait que le SI-nominatif ne puisse pas être sur l'infinitif sous un verbe à montée, on impose le trait :cl:nom=- sur le noeud pied de l'infinitif.

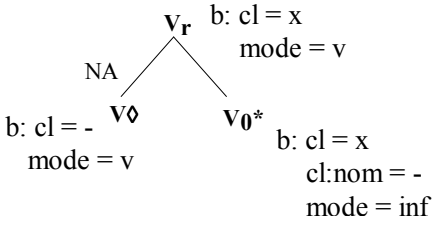


Figure 99. Schème pour un verbe à montée, en italien (trait :cl et :mode seulement)

Pour illustrer le mécanisme pour déterminer l'hôte des clitiques, on donne ci-dessous la dérivation pour (325), avec un verbe à montée (*sembra*) et un verbe dans son emploi restructurant, portant un clitique accusatif (*volerlo*).

(325) Paolo sembra volerlo fare.
Paolo semble vouloir-ACC3ms faire
(Paolo semble vouloir le faire.)

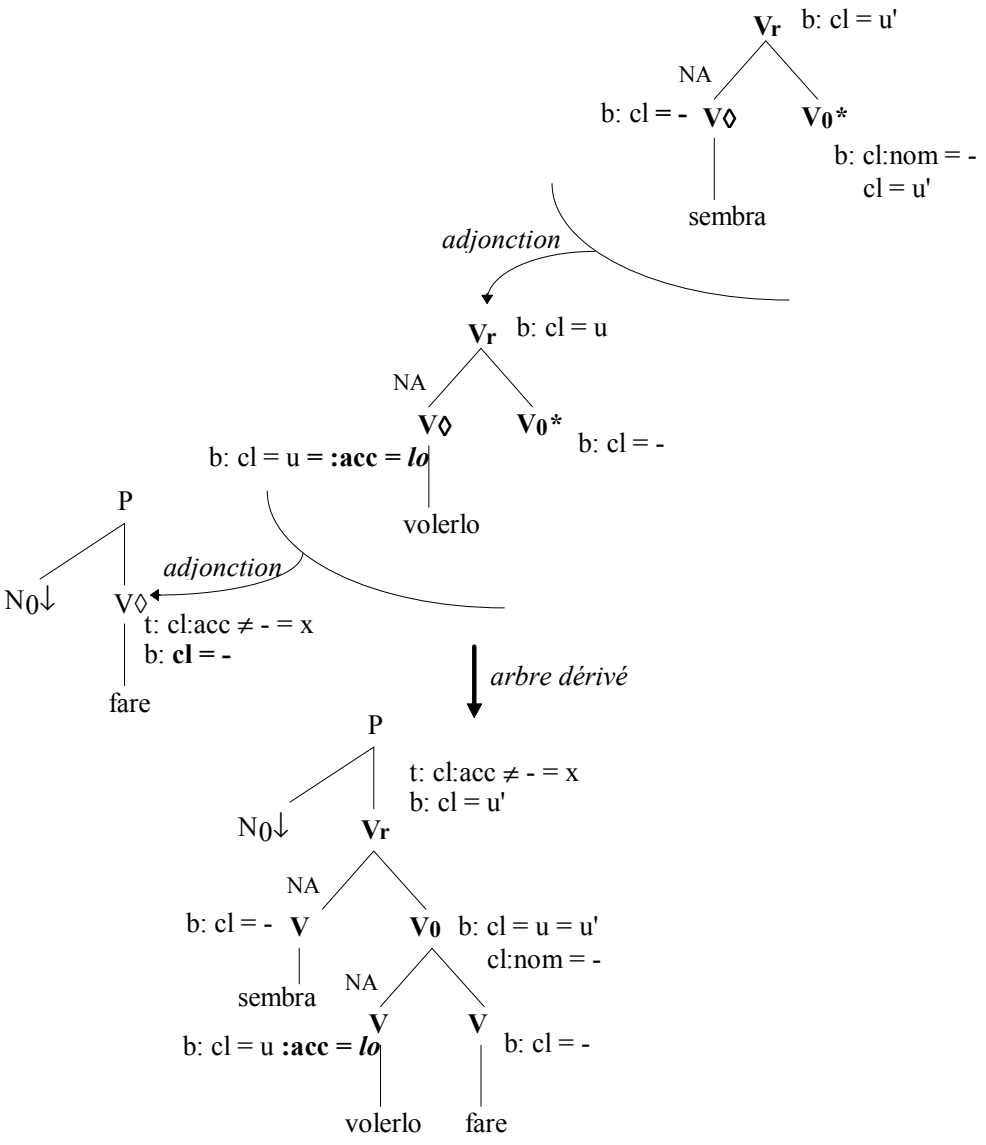


Figure 100. Dérivation pour *Paolo sembra volerlo fare*. (Paolo semble vouloir-ACC3ms faire)

Il faut noter que pour l'italien comme pour le français, les emplois alternatifs de certains verbes à montée, avec une complétive (*il semble que P/sembra che P*) sont considérés comme des sous-catégorisation initiales différentes de l'emploi comme verbe à montée.

Le clitique *loro*

Monachesi (1995) montre que le clitique datif *loro* est à distinguer des autres clitiques pronominaux, en cela qu'il n'a pas le comportement d'un affixe. Cette distinction est à relier au fait que *loro* est bisyllabique, ce qui a pu empêcher son intégration à un système flectionnel.

Monachesi insiste sur le fait que *loro* comme datif (il existe aussi la forme *loro* pour le pronom fort de troisièùe personne) est utilisé à l'écrit de manière productive, bien que son emploi à l'oral soit supplanté par la forme *gli* de datif masculin troisième personne.

Loro n'apparaît pas dans la même position que le complément plein correspondant, mais pas non plus dans la même position qu'un datif d'une autre personne. Il a en effet une série de caractéristiques qui le distingue nettement des autres clitiques, si bien qu'on peut parler d'un mot plein plutôt que d'un clitique.
Loro suit le verbe, quel que soit son mode (cf. (326a) indicatif et (326b) infinitif), mais peut de manière optionnelle le précéder dans certaines constructions, qui sont les relatives (326c), et les participiales[19] adnominales (326d) (les exemples sont de Monachesi, 1995:153).

(326) a. Martina spedisce **loro** la lettera.
Martina envoie DAT.3p la lettre.
(Martina leur envoie la lettre.)

b. Ammiratrici gli chiedono di dare **loro** un figlio.
Des admiratrices DAT.3ms demandent de donner DAT.3p un enfant.

c. La libertà di fantasia che **loro** appartiene.
La liberté de fantaisie qui DAT.3p appartient.

d. ... il potere d'acquisto **loro** fornito dalle banche.
... le pouvoir d'achat DAT.3p fourni par les banques.
(le pouvoir d'achat qui leur est fourni par les banques)

Loro peut être séparé du verbe par des adverbes, à la différence des autres clitiques. Monachesi suggère alors de qualifier plutôt *loro* d'adjacent au verbe plutôt que cliticisé.
Loro est toujours extérieur au groupe clitique. En effet, dans le cas d'enclitiques sur un infinitif par exemple, *loro* se place après (327).

(327) Martina deve spedirla loro.
Loro peut avoir portée large sur une coordination.

Enfin, en ce qui concerne la montée des clitiques, *loro* est également différent, puisque la montée sur l'auxiliaire est optionnelle, même dans le cas de présence d'un autre clitique qui lui doit monter[20].

(328) La scintilla, l'ha fornita loro una dichiarazione del ministero dell'Interno.
L'étincelle, l'a fournie DAT.3p une déclaration du ministère de l'Intérieur.
(L'étincelle, c'est le ministère de l'Intérieur qui la leur a fournie.)

Il apparaît clairement que *loro* n'est pas un affixe, et doit donc être traité en syntaxe. Dans la MG, une réalisation particulière de la fonction a-objet est prévue pour la

[19] A la différence du français, le participe passé en italien peut porter des clitiques. Cette impossibilité est représentée dans la MGf par une contrainte de cooccurrence négative entre la classe participiale et la classe clitique-complément, qui n'apparaît pas dans la MGi.
[20] Monachesi affirme aussi que la montée de *loro* est optionnelle dans le cas de verbes restructurants, mais les exemples qu'elle sonne ne comprennent pas d'autrec clitiques, ce qui fait qu'on ne sait pas si le verbe enchâssant est bien dans son emploi restructurant.

réalisation en *loro*. Une position à substituer est prévue, à droite de ?Vanc. On ne traite pas la montée de *loro*, mais on a vu qu'elle est optionnelle.

5.2.3 Traitement des clitiques ajouts

L'analyse que nous avons donnée implique qu'une ancre qui sous-catégorise un actant réalisé comme clitique sélectionne un arbre spécifique pour cela, où la réalisation de l'actant est représentée par des traits et pas par une position syntagmatique. Pour les clitiques ajouts, par cohérence on ne peut pas utiliser des arbres élémentaires ancrés par *en* ou *y*. Il serait difficile sinon impossible de calculer l'ordre respectif de différents clitiques si certains sont traités en morphologie et d'autres (les ajouts) en syntaxe. Une solution serait de considérer des arbres auxiliaires s'adjoignant sur un noeud verbal, ancrés par la chaîne vide, et qui ajouterait seulement les traits morphologiques pour la réalisation d'un clitique. Cette solution permet de conserver la compositionnalité sémantique, avec la contribution sémantique autonome d'un clitique ajout représentée par un arbre élémentaire pour ce clitique. Mais d'un point de vue computationnel, cela est très coûteux en analyse, voire ingérable, puisqu'une dérivation pourrait contenir un nombre infini d'adjonctions (c'est une solution cependant envisageable en génération). Nous proposons simplement de ne pas contraindre les attributs :clC:loc et :clC:part, laissant la possibilité pour un verbe de comprendre un *y* ou un *en*. Cette solution ne permet pas de récupérer la sémantique (puisque dans le cas de montée, le verbe qui porte les éventuels clitiques ajouts n'est pas le gouverneur), mais on peut remarquer qu'une analyse par adjonction des clitiques *y* et *en* sur l'hôte ne permet pas non plus de récupérer le gouverneur dans le cas de montée.

5.2.4 Analyse des clitiques et interprétation sémantique

Le PCPA énoncé au Chapitre 1 prévoit que les arguments réalisés en morphologie ne correspondent pas à des positions à étendre dans les arbres élémentaires. Nous considèrons en effet que les arguments sémantiques de l'ancre ont deux modes de réalisation, un mode syntagmatique, pris en charge par la syntaxe, un mode morphologique, pour les clitiques. Nous pensons que cette distinction est de toutes façons pertinente pour des langues à flexions plus riche que celle du français. On peut préciser le PCPA comme suit :

Principe de cooccurrence prédicat-arguments (précisé) : un arbre élémentaire contient pour chacun de ses arguments sémantiques exprimés, soit une position syntagmatique à étendre (sous forme de noeud à substituer ou de noeud pied), soit un trait pour la réalisation morphologique comme clitique, ainsi qu'un trait pour la sémantique du clitique.

Le principe de minimalité sémantique est respecté, dans la mesure où on considère que les clitiques sont des affixes. Le principe requiert qu'un arbre élémentaire soit ancré par un lexème non décomposable *en syntaxe* en plusieurs unités signifiantes. De la même manière que l'on considère comme ancre possible un verbe fléchi, portant des marques de temps et de mode, on peut considérer comme ancre possible un verbe "fléchi" pour des clitiques, c'est-à-dire portant des clitiques. La différence est tout de même que la

contribution sémantique des affixes de temps ou de mode est radicalement différente de celle des affixes pronominaux, qui peuvent être référentiels.

Même si l'on a formulé les principes linguistiques de manière à prendre en compte la réalisation des arguments comme des affixes, cette distinction pose un problème pour la procédure d'interprétation de l'arbre de dérivation. On verra Chapitre 7 que l'on peut interpréter les arcs de l'arbre de dérivation comme des dépendances sémantiques si les arguments sémantiques ont une position à étendre dans l'arbre de leur gouverneur sémantique. Avec un traitement des clitiques en morphologie, le lien entre le clitique et son gouverneur ne donne pas lieu à un arc dans l'arbre de dérivation et il faut donc prévoir un mécanisme supplémentaire pour l'interprétation sémantique des clitiques.

De manière à récupérer l'interprétation sémantique voulue, on s'inspire de (Miller, 1991) en distinguant les traits morphologiques de réalisation des clitiques des traits sémantiques. Nous proposons d'utiliser un trait :S-cl (pour sémantique des clitiques), localisé sur le gouverneur de tout clitique correspondant à un argument sémantique (il n'est donc pas pertinent pour les clitiques idiomatiques et les marqueurs de construction). On peut le faire puisqu'on a vu que les traits morphologiques pour la réalisation des clitiques peuvent se propager jusqu'à l'hôte approprié du clitique, mais ont leur origine dans l'arbre élémentaire ancré par le gouverneur du clitique.
Donc si un schème est défini avec pour un argument sémantique d'index *i*, une réalisation comme clitique, avec le cas *cas*, ?Vanc porte un trait **aval** :S-cl:arg*i*, dont la valeur est coindicée avec le trait :clC:*cas*, s'il s'agit d'un clitique complément, avec le trait :clS, s'il s'agit d'un clitique sujet.
Ainsi les traits sémantiquement pleins associés à un clitique sont stockés au niveau du gouverneur du clitique. Comme le trait :S-cl est un trait aval sur ?Vanc, il reste toujours local à l'arbre élémentaire où il a été spécifié, c'est-à-dire toujours local à l'arbre du gouverneur sémantique.
Le lexique syntaxique, qui associe des lexèmes fléchis à des schèmes élémentaire, ne fait pas référence au trait :S-cl, mais seulement aux traits :clC et :clS, étant donné que dans certains cas, on ne sait pas a priori si les clitiques portés par un verbe en sont des arguments sémantiques.
Par exemple, on donne Figure 101 l'exemple d'un schème avec un clitique datif1 et un clitique accusatif, correspondant tous deux à un argument sémantique.

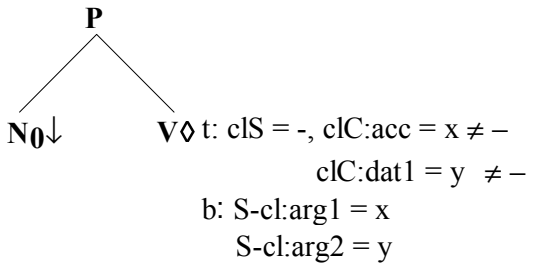


Figure 101. Schème avec clitique datif et accusatif

Ce schème peut être lexicalisé par exemple par les formes ci-dessous. La lexicalisation par *offre* donne un arbre avec une contrainte d'adjonction obligatoire sur V, étant donnés les traits aval et amont :clC. En tant que trait aval sur ?Vanc, le trait :S-cl ne peut que

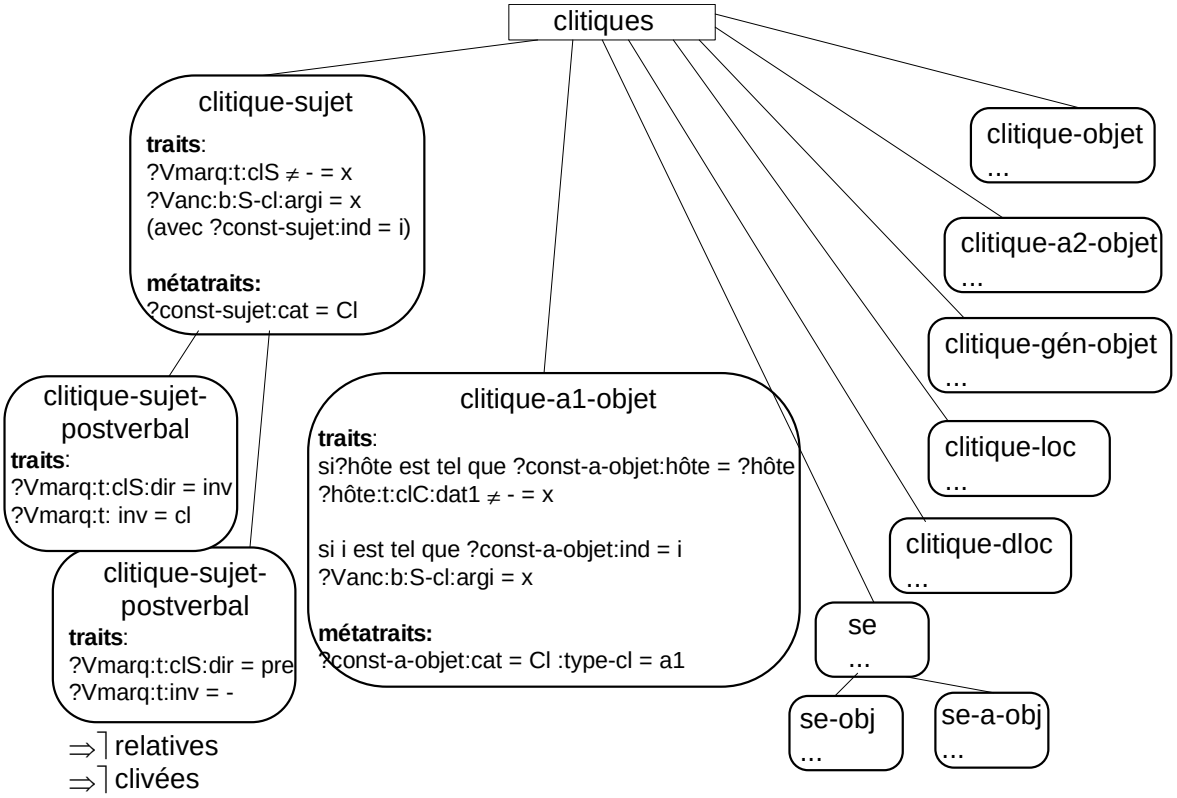
rester local à l'arbre élémentaire. On peut y avoir accès lors de l'interprétation sémantique de l'arbre de dérivation, ou plus simplement décider de le "remonter" au niveau de l'information générale sur un arbre élémentaire, pour ne pas avoir à aller chercher ce trait sur l'ancre.

le lui offre : clS = -, clC:acc(dˇf = +, num = sing, pers = 3)
clC:dat1(num = sing, pers = 3)
offre : clS = -, clC = -

On peut remarquer que la lexicalisation par une forme avec un seul clitique (*l'offre* par exemple), donne un arbre qui attend une adjonction sur V (d'un auxiliaire) amenant un clitique datif, mais d'après les analyses que nous avons données, une telle adjonction "prend" la valeur du trait :clC dans son ensemble, et impose sur le noeud site d'adjonction un trait aval :clC = -, incompatible avec la valeur non nulle de :clC:acc. Mais il n'est pas très satisfaisant qu'une telle lexicalisation, qui ne peut pas amener à un arbre dérivé valide, soit possible. Nous n'avons pas de solution à ce problème actuellement.

5.3 Les clitiques pronominaux dans la MGf et la MGi

Nous donnons ci-dessous les classes représentant la réalisation clitique d'arguments, suivant la fonction portée, pour le français.
Par exemple, la classe pour le clitique a1-objet spécifie que l'hôte de l'objet (stocké dans le métatrait :hôte, cf. Chapitre 4, sur le placement des clitiques dans le cas d'interaction réfléchi/causatif) porte un trait amont :clC:dat1 non vide.
De plus si la constante ?const-a-objet (qui porte la fonction a-objet) correspond à un argument n°i (i.e. si elle porte un trait :ind numérique), alors le trait sémantique :S-cl:argi est ajouté comme trait amont sur ?Vanc).
Pour le clitique sujet, la classe est légèrement plus simple, étant donné que l'hôte est toujours ?Vmarq. Les autres classes pour des clitiques compléments sont similaires à celle pour le clitique a-objet donné en exemple.
Pour le sujet, on doit distinguer deux cas, selon que le sujet est inversé (pour une question) ou pas.
Les classes pour la réfléxivation (cf. Chapitre 4) héritent également de la classe clitiques, de manière à pouvoir se référer à cette classe si l'on veut bloquer toutes les réalisations clitiques.



5.4 Les différents traitements des clitiques proposés pour la TAGf

Nous revenons sur les différents traitements qui ont été proposés pour les clitiques pronominaux en TAG.

Abeillé a fourni différentes propositions de traitement des clitiques pour le français, toujours dans le cadre d'un traitement en syntaxe. (Abeillé, 1991) proposait un traitement des clitiques en syntaxe, par adjonction des clitiques sur leur hôte. Chaque clitique pronominal ancrait un arbre élémentaire auxiliaire de racine et noeud pied verbal. Des traits sur le noeud pied permettaient de gérer l'ordre fixe des clitiques. Dans le cas de clitiques argumentaux (c'est-à-dire qui correspondent à un actant sous-catégorisé), Abeillé suggérait en cas d'adjonction d'un clitique, de consommer automatiquement par la chaîne vide le noeud à substituer correspondant.
Ce traitement n'est pas achevé, et en particulier la montée des clitiques n'est pas abordée. Il n'est pas adéquat pour l'interface avec la sémantique. Comme les clitiques s'adjoignent sur leur hôte et que cet hôte n'est pas leur gouverneur sémantique en cas de montée, on ne peut pas récupérer de manière directe le gouverneur sémantique du clitique. En particulier, la procédure d'interprétation de l'arbre de dérivation comme un graphe de dépendances sémantiques (cf. Chapitre 7) ne serait pas valide dans le cas de clitiques.

(Abeillé, 1992) décrit un traitement des clitiques dans le formalisme des TAG synchrones (Shieber et Schabes, 1990, cf. Chapitre 1). Abeillé a proposé de relier une TAG syntaxique avec une TAG sémantique, de "surgénérer" au niveau syntaxique en ce qui concerne la réalisation de clitiques, et de filtrer en parallèle par la dérivation sémantique synchrone pour ne retenir que les dérivations valides.
La surgénération au niveau syntaxique provient du fait que les arbres élémentaires de la TAG syntaxique comportent tous, à la fois les positions (sous forme de noeuds à substituer) pour la réalisation non pronominales des arguments, et celles pour la réalisation clitique (ordonnées par colonne). Ces positions, comme réalisation nominale ou clitique, sont notées comme optionnelles. Donc par exemple pour un transitif, la TAG syntaxique telle quelle (i.e. sans la TAG sémantique synchrone) permet de générer les séquences *N0 V N1*, *Cl0 Cl1 V*, mais aussi, *V*, *N0 Cl0 V*, *Cl1 V N1* etc... Le filtrage par la TAG sémantique repose sur le fait que plusieurs positions en syntaxe, en l'occurrence la position nominale et la ou les positions clitiques correspondantes, sont reliées à la même position en sémantique, qui ne peut être consommée qu'une fois (par substitution).
On prévoit dans tous les arbres élémentaires une position clitique optionnellement remplie (c'est-à-dire consommée 0 ou 1 fois) pour chaque colonne de clitique. Plus précisément, les colonnes I, V et VI sont toujours présentes, puisque les verbes pleins sous-catégorisent toujours un sujet, et les colonnes V et VI peuvent être remplies par un clitique adverbial. Les colonnes II, III et IV sont présentes si la sous-catégorisation du verbe comporte un objet ou un a1-objet.
Cette solution a plusieurs avantages. Une position pour une colonne en syntaxe peut être reliée à différentes positions en sémantique, de fait mutuellement exclusives étant donné que la position en syntaxe ne peut être consommée qu'une fois (par substitution). On représente ainsi les incompatibilités entre clitiques d'une même colonne. Ainsi par exemple la position clitique pour la colonne V (*y*) est reliée en sémantique à un argument (un datif2 en *y*) soit à un ajout (un locatif). La consommation unique de la position en syntaxe permet de bloquer une séquence *y y*. De plus, un relâchement des contraintes en sémantique, permet de générer des phrases à redoublement clitique, acceptables à l'oral (comme *je les lave, mes chemises*).

Cette solution permet en théorie de gérer à la fois le placement particulier des clitiques et leur interprétation sémantique correcte, du fait de l'utilisation de la TAG sémantique. Cependant, c'est au prix d'un traitement nettement complexifié. D'une part il faut considérer des arbres en syntaxe avec syntagmes optionnels, ce qui revient à multiplier énormément le nombre d'arbres élémentaires ancrés par un verbe. D'autre part, l'interprétation sémantique est prise en charge par les TAG synchrones, ce qui complexifie le système. Il ne s'agit d'ailleurs pas d'une proposition ayant fait l'objet d'une implémentation.

Enfin une autre solution (ébauchée dans (Abeillé, 1993)) est simplement de respecter rigoureusement le PCPA y compris pour la réalisation clitique des arguments sémantiques. Les clitiques sous-catégorisés ancrent alors un arbre élémentaire à eux seuls, et sont substitués dans l'arbre de leur gouverneur. Les clitiques ajouts ancrent des arbres auxiliaires et sont adjoints sur leur hôte. Les clitiques idiomatiques et les marqueurs (comme le réfléchi) sont directement placés comme co-têtes dans les arbres

élémentaires des verbes qui les requièrent. La montée des clitiques est soit traitée par adjonction de l'arbre de l'hôte entre le clitique et son gouverneur, soit traitée par formation d'un prédicat complexe hôte + gouverneur du clitique.
Cette solution n'est envisageable que si l'on dispose d'un outil de génération des arbres élémentaires, comme la MG par exemple. C'est d'ailleurs la solution que nous avons implémentée pour la MGf. Elle nécessite de distinguer de nombreux arbres élémentaires distincts. Par exemple pour un ditransitif dont les deux compléments sont cliticisés, il faut distinguer selon que les traits de personne des clitiques, les deux ordres possibles pour les clitiques accusatif et datif.
Figure: exemples de schèmes avec positions à substituer pour les clitiques (ditransitif, acc<dat; dat<acc).

Un gros avantage de cette solution, par rapport aux solutions citées plus haut, et même par rapport au traitement des clitiques en morphologie, est que l'interprétation sémantique des arguments cliticisés ne diffère pas de celle des arguments réalisés comme des syntagmes. On peut suivre le protocole général d'interprétation de l'arbre de dérivation, puisque les clitiques sous-catégorisés sont substitués, et les clitiques ajouts sont adjoints. Un problème peut cependant se poser si un clitique adjoint ne s'adjoint pas sur son gouverneur. On ne peut pas alors récupérer quel est ce gouverneur.
Mais cette interprétation est également possible dans le cadre du traitement des clitiques en morphologie (cf. infra).

Cette solution a cependant des inconvénients. Le placement correct des clitiques ajouts est mal géré. Le *en* doit s'adjoindre toujours à droite des éventuels autres clitiques, ce qui est facilement représentable. Mais cela est plus délicat pour le *y* ajout, qui doit être devant le *en* et après les autres clitiques. Même si la séquence *y en* (hors figement) est plutôt rare, cela nécessite en théorie de distinguer différents arbres à clitique accusatif, même dans le cas d'un clitique sous-catégorisé accusatif unique.
De plus, certains cas de montée des clitiques dépassent la capacité générative des TAG standards si on respecte le PCPA et si on adopte un traitement en syntaxe des clitiques. Nous passons en revue ces cas Section 5.5.
Enfin, nous montrons Section 5.6 que l'analyse des clitiques en syntaxe nécessitent de dupliquer certains arbres élémentaires, pour gérer le placement des clitiques avant ou après l'hôte.

5.4.1 Correspondance systématique entre analyse en morphologie et analyse en syntaxe

Nous avons donné les arbres élémentaires comme si l'on disposait d'un module morphologique fournissant l'analyse des formes fléchies avec clitiques. Cependant, pour l'implémentation, nous n'avons pas implémenté ce module. A l'heure actuelle, les LTAG pour le français et l'italien comportent un traitement des clitiques en syntaxe, tel que nous venons de le décrire. On utilise une catégorie syntaxique Cl spécifique, les clitiques qui peuvent être arguments ancrent des arbres de racine Cl, et sont substitués dans l'arbre de leur gouverneur. Les clitiques ajouts sont adjoints, et les clitiques idiomatiques ou marqueurs de construction sont directement placés dans l'arbre.

Pour l'italien, les enclitiques (pour les hôtes à l'infinitif, participe, gérondif, impératif) ne sont pas séparés à l'écrit par un espace. Donc, même pour un traitement en syntaxe des clitiques, un prétraitement est nécessaire pour séparer ces clitiques de leur hôte. De plus, que ce soit dans le cas d'enclitiques que de proclitiques, le clitique datif1 3ème personne *gli*, est amalgamé avec le clitique *lo, la, le, li,* ou *ne*, si ce clitique le suit (formes *glieli, glielo, gliela, gliele, glene*). Là encore dans le cadre d'un traitement des clitiques en syntaxe, un prétraitement doit séparer les deux clitiques.

Entre les arbres "théoriques" (qui correspondent à un traitement des clitiques en morphologie) et les arbres réellement implémentés, il existe l'équivalence suivante :
Pour un schème théorique à ancre verbale, dont un noeud verbal V_x porte un trait amont :clC (repectivement :clS), il existe un ou plusieurs schèmes réels, sans traits pour clitiques, et tels qu'un noeud verbal V_x' domine immédiatement le noeud V_x et a pour fils un ou plusieurs clitiques compléments (respectivement un clitique sujet). Ces fils sont soit des positions Cl à substituer, s'il s'agit de clitiques correspondant à des arguments, soit des noeuds Cl dominant une co-tête (pour un clitique idiomatique ou un marqueur).
Un schème théorique peut correspondre à plusieurs schèmes réels dans le cas où il faut prévoir en syntaxe différents ordres possibles pour les clitiques, selon leurs traits :pers ou :déf.

On donne Figure 102 un schème théorique pour le français, pour un ditransitif avec sujet, objet et a1-objet clitiques.

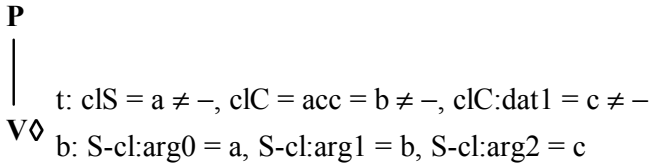


Figure 102. Schème "théorique" pour un ditransitif avec ses arguments cliticisés

Les arbres réels correspondants sont les suivants :

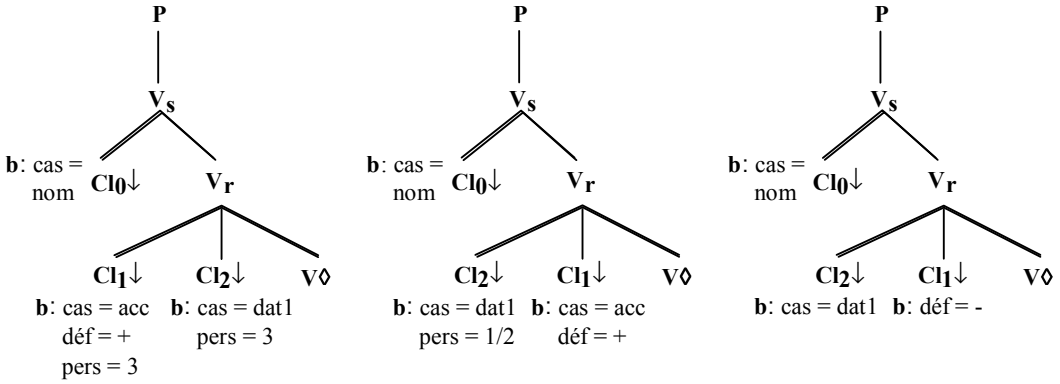


Figure 103. Les trois schèmes implémentés correspondants

5.5 La montée des clitiques et le pouvoir génératif de TAG

Si on se place dans le cadre d'une analyse des clitiques pronominaux en syntaxe, avec le respect du PCPA, c'est-à-dire que les positions syntagmatiques des clitiques doivent être prévues dans l'arbre élémentaire de leur gouverneur, alors on peut montrer que certains cas de montée des clitiques ne peuvent pas être traités en TAG.
En utilisant les données de l'espagnol, Bleam (1994) a étudié spécifiquement quels sont les cas de montée des clitiques qui peuvent être traités en TAG standard en respectant le PCPA et ceux requérant un formalisme étendu.
Pour la montée des clitiques sur un verbe restructurant (ou sur un auxiliaire), on peut adjoindre le verbe restructurant, même si le clitique est placé dans l'arbre de l'infinitif. Ainsi pour (329), on schématise en (329') cette adjonction entre le clitique et l'infinitif, en notant la séquence adjointe entre crochets.

(329) Mari la quiere ver.
Mari ACC.3fs veut voir.
(Mari veut la voir.)

(329') Mari la [quiere] ver.

Mais l'espagnol comporte également des verbes restructurants à contrôleur a-objet (ce qui n'est pas le cas de l'italien à notre connaissance). Ainsi *permito* en (330) est gouverneur du clitique datif, mais porte également le clitique accusatif, gouverné par l'infinitif.

(330) Te lo permito ver.
DAT.2 ACC.3ms je-permets voir.
(Je te permets de le voir.)

Avec un traitement en syntaxe des clitiques, pour analyser (330) il faut envisager des solutions qui sont tout à fait contraire aux principes généraux de syntaxe. Par exemple on peut adjoindre [te permito] sur le chemin entre la racine de [lo ver] et le clitique. Or on voudrait plutôt adjoindre sur le chemin entre la racine de [lo ver] et l'infinitif. Bleam définit alors une analyse dans le cadre des TAG ensemblistes (MC-TAG).
Avec le traitement en morphologie que nous avons décrit supra, on peut très bien analyser (330). Le verbe restructurant peut être adjoint avec les spécifications pour le clitique qu'il gouverne sémantiquement, plus la spécification du clitique gouverné par l'infinitif, dont le verbe enchâssant "hériterait" lors de l'adjonction. Il faut cependant également prévoir le cas où l'a-objet du verbe enchâssant est réalisé en syntaxe[21].

Un autre cas qui ne peut pas être traité en TAG standard, si l'on veut traiter les clitiques en syntaxe et respecter la localisation du clitique dans l'arbre de son gouverneur, est celui des clitiques complément de nom. Par exemple, pour le français, Abeillé (1998) montre qu'un traitement dépassant le cadre de TAG est requis. Pour l'exemple (331), on voit bien en effet qu'on ne peut pas trouver une dérivation TAG où *Paul* est localisé dans l'arbre de *aime* et *en* est localisé dans l'arbre de *fin*. Abeillé suggère un traitement dans le cadre des MC-TAG locaux, mais montre qu'un formalisme utilisant des descriptions partielles d'arbres (Vijay-Shanker, 1992) serait plus adapté. Ce formalisme

[21] Une analyse par adjonction du verbe enchâssant ne serait pas possible dans des cas équaivalents au français ?*Cela a permis de voir à tout le monde la cupidité de Frison.*

n'est cependant pas défini, plusieurs possibilités pouvant être envisagées sur la nature exacte des opérations de combinaison de descriptions.

(331) Paul en aime la fin.

Un traitement par la morphologie du placement des clitiques permet de rester dans le cadre de TAG. Il suffit pour cela que le nom transmette le trait clitique au verbe. Le verbe et le nom peuvent porter un trait non nul pour le clitique *en*, et la morphologie peut traiter la réalisation du clitique, le trait :clC n'étant pas pertinent pour un nom mais pour un verbe.

La contrainte que le *en* complément de nom ne peut provenir que d'un objet ou d'un sujet (pour certains verbes seulement), pourrait être simplement représentée par le fait que les arbres élémentaires de la grammaire ne font pas partager le trait clitique génitif entre un nom dans un complément prépositionnel et son verbe.

5.6 Cas problématiques pour un traitement en syntaxe

Nous montrons par quelques exemples que le traitement en syntaxe des clitiques nécessitent une multiplication des arbres élémentaires par rapport à un traitement en morphologie. Cela n'est pas étonnant en soi, puisque les traiter en morphologie signifie décharger le composant syntaxique de cette tâche.

L'analyse des auxiliaires par adjonction permet de représenter le placement des clitiques sur l'auxiliaire. Il suffit pour cela d'adjoindre l'auxiliaire entre les clitiques et le verbe. On voit que cette solution repose sur le fait que les clitiques soient préverbaux.
Cette analyse est donc problématique pour les cas de clitiques postverbaux avec auxiliaire. L'impératif est difficile avec un temps composé, mais on peut avoir aussi un clitique sujet inversé, sur un auxiliaire.

(332) a. Vois-tu Céline?
b. As-tu vu Céline?

Pour (332), le clitique constitue un argument du verbe et devrait être donc substitué dans l'arbre de son gouverneur. Or le clitique inversé est à droite de son gouverneur dans le cas d'un temps simple (332a), et à gauche de son gouverneur dans le cas d'un temps composé (332b). On ne peut prévoir sa position à substituer par rapport à l'ancre (le gouverneur du clitique) que si l'on sépare les arbres pour les formes simples de ceux des formes composées.

On a exactement le même problème dans le cas d'un clitique inversé sur un verbe à montée (332c). Avec l'analyse des verbes à montée par adjonction, on ne peut prévoir la position du clitique par rapport à son gouverneur que si l'on distingue deux cas, avec ou sans verbe à montée, ce qui n'est pas satisfaisant.

(332) c. Veux-tu lire Céline?

Une solution technique est de renoncer à prévoir une position pour le clitique inversé, et de l'adjoindre[22] (Abeillé, 1991). Mais bien sûr cela ne respecte pas le PCPA, ce qui se traduit par une interprétation sémantique plus difficile.

En italien, on a également des cas comme ceux que l'on vient de décrire, pour lesquels la position d'un clitique par rapport à l'ancre varie selon la présence de verbes enchâssants, qui sont analysés par adjonction.
En italien en effet, les clitiques sont postverbaux dans le cas d'un hôte à l'infinitif. Donc, pour une infinitive avec certains actants cliticisés, la position du clitique par rapport au gouverneur infinitif dépend de la présence d'auxiliaire ou de verbe restructurant. Ainsi en (333a) le clitique est à droite de son gouverneur *mangiare*, et en (333b), il est à gauche de son gouverneur, du fait que l'hôte est un verbe restructurant, en (333c) aussi, du fait que l'hôte est un auxiliaire. Donc pour l'italien, il faut distinguer les cas d'infinitives avec ou sans verbe restructurant ou auxiliaire de temps, dès que l'infinitif a un actant cliticisé (ou un clitique idiomatique).

(333) a. I bambini giurono di **mangiarlo**.
Les enfants jurent de manger-ACC3ms
(Les enfants jurent de le manger.)

b. I bambini giurono di voler**lo mangiare**.
Les enfants jurent de vouloir-ACC3ms manger
(Les enfants jurent de vouloir le manger.)

c. I bambini giurono di aver**lo mangiato**.
Les enfants jurent de avoir-ACC3ms mangé.
(Les enfants jurent l'avoir mangé.)

Cette duplication est également nécessaire pour les arbres élémentaires avec sujet réalisé, dès lors que l'ancre gouverne un clitique. En effet, d'après l'analyse des verbes restructurants et des verbes à montée que nous avons décrite supra, un même arbre élémentaire à sujet réalisé peut être utilisé tel quel pour une phrase de mode fini (334a), ou bien recevoir l'adjonction d'un verbe restructurant ou d'un verbe à montée, l'ancre de l'arbre recevant l'adjonction étant alors à l'infinitif (334b).

(334) a. Mario dorme.
Mario dort.

b. Mario può dormire.
Mario peut dormir (ambigu restructurant ou verbe à montée).

Mais on voit avec la série (335) qu'on ne peut pas placer le clitique par rapport à l'ancre sans prévoir qu'il y aura ou pas verbe à montée (335b).

(335) a. Mario lo fa.
Mario ACC.3ms fait

b. Mario può farlo.
Mario peut faire-ACC.3ms

[22] L'adjonction du clitique inversé est aussi utilisée dans le cas d'inversion complexe (*Pierre vient-il?*).

Ces duplications d'arbre élémentaires ne sont pas nécessaires avec un traitement en morphologie du placement des clitiques par rapport à l'hôte. C'est le mode de l'hôte qui détermine la position enclitique ou proclitique.
Mais dans l'implémentation actuelle, où les clitiques sont traités en syntaxe, les problèmes cités se posent. Pour le français et l'italien, les arbres élémentaires ne sont pas distingués selon que l'ancre est à un temps composé ou pas (on garde le traitement des auxiliaires de temps par adjonction). Pour le français, le clitique sujet inversé est adjoint, ce qui permet de gérer (332b,c). Mais pour l'italien, la MGi implémentée ne couvre pas les cas d'enclitiques sur un auxiliaire ou un verbe restructurant.
Pour les arbres avec sujet réalisé, et comprenant un clitique, on a placé le clitique à gauche de l'ancre, ce qui permet de traiter (335a), ou le cas d'un complexe verbe restructurant + infinitif (*Mario lo può fare*), mais on n'a pas dupliqué tous ces arbres pour traiter (335b), car cela est beaucoup trop coûteux en termes d'efficacité, puisque cela augmente considérablement le nombre de schèmes sélectionnés par un verbe. Il vaut mieux revoir le traitement des verbes à montée. On peut toujours en effet les traiter comme des verbes à contrôle, même si cela est moins satisfaisant pour l'interprétation sémantique.

6 Les réalisations phrastiques

Avant de détailler la représentation de l'extraction (relatives, clivées, questions), nous décrivons la représentation des réalisations phrastiques, à la fois pour les fonctions qui n'admettent pas de réalisation nominale, et celles qui admettent les deux possibilités (sujet, objet, a2-objet, gén-objet).

6.1 Généralités sur la représentation des actants phrastiques

6.1.1 Complémenteurs pleins et complémenteurs régis

Les propositions subordonnées peuvent dépendre syntaxiquement de lemmes prédicatifs de différentes catégories et porter différentes FG. La subordination peut être marquée par un complémenteur. Nous prenons ce terme dans une acception large de marqueur de subordination phrastique, qu'il soit sémantiquement vide ou plein. Ainsi, on considère les conjonctions de subordination traditionnelles (*quand, si, lorsque, puisque...*) comme des complémenteurs pleins.
De façon analogue à la distinction entre préposition régie et préposition pleine, le complémenteur régi (sémantiquement vide) n'est pas la tête de la structure, alors que le complémenteur plein l'est. D'après le principe de minimalité sémantique, les complémenteurs pleins, en tant qu'unité sémantique, doivent ancrer des arbres élémentaires à aux seuls (cf. Chapitre 6, pour la représentation des schèmes à ancre complémenteur). Au contraire, les complémenteurs régis doivent apparaître comme co-tête.

On emploie le terme complétive pour une proposition subordonnée tensée et infinitive pour une proposition avec verbe à l'infinitif. L'expression du sujet n'est possible que dans le cas tensé, et les complémenteurs diffèrent pour les deux cas.
Pour les infinitives, en français le complémenteur peut ne pas exister, sinon il est soit *à* soit *de*. Dans le cas d'une complétive, en français on doit avoir un complémenteur, plein ou régi (les deux complémenteurs régis étant *que* et *si* (dans son emploi avec interrogative indirecte). En italien les deux complémenteurs régis sont *che* et *se*. Certains verbes enchâssants permettent en outre une complétive sans complémenteur (cf. "l'omission du *che*", Section 6.3.2).

Dans la MGf et la MGi, les complémenteurs régis sont co-têtes dans l'arbre de l'élément enchâssant (Abeillé, 1991). Cela est motivé par le fait que c'est l'élement enchâssant qui détermine la nature du complémenteur, du moins en partie. Par exemple, pour un verbe à contrôle, l'absence de complémenteur, la présence du complémenteur *de* ou *à* en français est déterminée par le verbe enchâssant. Autre exemple, la possibilité d'avoir un complémenteur *si* (pour une interrogative indirecte totale) est déterminée par le verbe.
Pour l'italien, la possibilité "d'omission du *che*" est lexicalement déterminée par le verbe enchâssant. Cette omission est donc facilement contrôlable dans le cadre d'une analyse où le complémenteur est co-tête du verbe enchâssant[23].

6.1.2 La représentation des extractions à distance, hors des complétives et infinitives

Nous revenons sur le traitement de l'extraction par adjonction, qui a été déjà mentionné à plusieurs reprises. Notamment au Chapitre 1, nous avons décrit les adjonction enchâssantes, comme des adjonctions telles que la tête syntaxique après adjonction est l'ancre de l'arbre adjoint. Elles sont utilisées en particulier pour adjoindre un verbe à contrôle sur son infinitive et un verbe à complétive sur sa complétive.
TAG permet ainsi de localiser la dépendance entre l'élément extrait et son verbe, même dans le cas d'extraction à distance : les expressions apparaissant entre l'élément extrait et son verbe sont adjoints. Ainsi la phrase (336) peut-elle être dérivée par adjonction de *Paul pense-t-il que* dans l'arbre pour la question *quelle fille Pierre a vu* (

(336) Quelle fille [Paul pense-t-il que] Pierre a vu?

Le PCPA est respecté pour la structure ancrée par *vu* : elles comprend une position à étendre même pour son argument extrait.

[23] Cependant, un problème général lié à la représentation des mots fonctionnels comme co-têtes est celui de la coordination. Il faut gérer la réplication du complémenteur adéquat dans le cas de coordination de complétive ou infinitive.

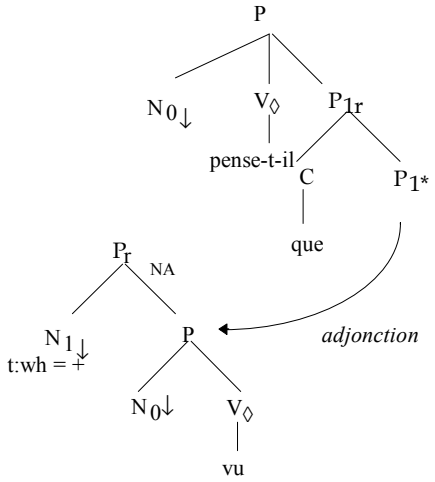


Figure 104. Adjonction d'un verbe à complétive sur l'arbre pour une question à objet antéposé

Les verbes à contrôle

On a déjà mentionné la différence sémantique faite entre verbes à contrôle et verbe à montée : les verbes à contrôle gouvernent sémantiquement le contrôleur de l'infinitive. Ce contrôleur apparaît comme la réalisation unique d'un argument sémantique partagé par deux prédicats, le verbe à contrôle et l'infinitif.
Cet argument sémantique partagé (le contrôleur) est représenté en TAG par un noeud argumental (à substituer) localisé dans l'arbre du verbe à contrôle, le verbe infinitif n'ayant alors pas de position pour cet argument, mais récupérant par des traits les informations pertinentes sur le contrôleur.

L'analyse de l'extraction hors d'une infinitive par adjonction de verbes pont impose que l'infinitive soit représentée comme une phrase, cela si l'on considère les TAG standards, avec la contrainte d'identité catégorielle entre noeud racine et noeud pied d'un arbre auxiliaire. En effet l'arbre pour le verbe à contrôle est de racine phrastique, il doit donc être de noeud pied phrastique. On donne Figure 105, l'adjonction du verbe à contrôle nécéssaire pour dériver (337).

(337) Quelle part Paul veut-il manger?

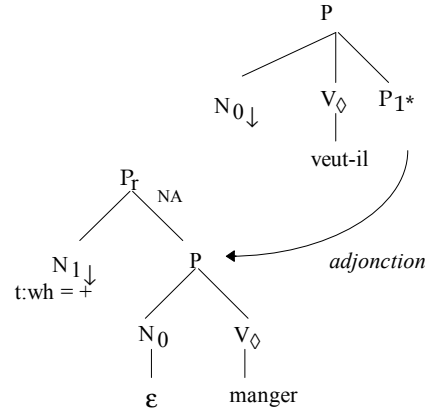


Figure 105. Adjonction d'un verbe à contrôle, pour dériver (337)

Pour envisager dans le cadre de TAG une analyse avec les infinitives comme SV (ou plus généralement comme constituant non phrastique), il faut pouvoir relâcher la contrainte sur l'identité catégorielle entre racine et noeud pied. Cela est possible dans le cadre de formalismes étendus, comme les D-Tree Grammars (Rambow et al., 95) qui utilisent des quasi-arbres, ou bien dans le cadre de TAG, en considérant que les catégories des noeuds des arbres syntagmatiques peuvent être complètement codées dans les traits aval et amont, laissant ainsi la possibilité d'une divergence de catégorie entre partie aval et partie amont d'un noeud (Kahane et al., en préparation).

Une solution compatible avec les choix de catégories syntaxiques effectués pour la MGf et la MGi, serait de considérer toutes les projections verbales (y compris la phrase) comme des noeuds V (distingués par des traits), comme on le fait pour les noms par exemple. Cette solution est intéressante, mais nécessite de repenser le système de traits pour contrôler la distinction actuelle entre les arbres auxiliaires pouvant s'adjoindre sur V et ceux pouvant s'adjoindre sur P.
Nous n'avons pas suivi cette voie, et avons donc intégré la contrainte que les infinitives doivent projeter des phrases. L'analyse des infinitives comme des phrases est une hypothèse qui est apparu aux tous débuts de la grammaire transformationnelle. Elle a été sévèrement critiquées depuis, qu'il s'agisse de critique externes à la grammaire générative (avec les analyses en LFG ou HPSG) ou internes (voir par exemple Wehrli, 1980, et les références citées). Nous n'avons pas la prétention d'entrer dans ce débat. Nous pouvons cependant noter, que du point de vue de l'interface syntaxe/sémantique, la représentation des infinitives comme des phrases peut être considéré comme un avantage, en cela qu'elle permet le positionnement d'un sujet vide dans une infinitive. On peut considérer que ce sujet vide est une représentation de l'argument sémantique requis par l'infinitif.

Les TAG sous-jacentes à la MGf et la MGi, comportent des arbres spécifiques pour les infinitives (qui peuvent recevoir l'adjonction d'un verbe à contrôle mais peuvent également fonctionner comme sujet infinitif) de racine phrastique, et comprenant non pas une position à substituer pour leur sujet, mais une trace. Dans la MGi et la MGi, cela permet également un partage des traits d'accord entre le cas d'un sujet réalisé et celui d'un sujet non réalisé pour une infinitive.

Les verbes à complétive

Nous avons donné , un exemple d'adjonction d'un verbe à complétive, permettant la représentation de l'extraction hors de la complétive.
Pour bloquer au contraire l'extraction hors d'une complétive, on représente cell-ci dans l'arbre du verbe enchâssant comme un noeud à substituer (cf. également Chapitre 1) (Abeillé, 1991). Cela peut être utilisé soit pour des verbes non pont (comme /S'ECRIER/) pour lesquels l'extraction est toujours mauvaise[24], ou bien dans le cas de verbes pont, dont un autre complément est extrait (cf. Chapitre 1).

[24] En pratique, dans l'implémentation, il faut distinguer plusieurs familles de verbes à complétive, selon qu'ils admettent ou pas l'extraction. Les TAG française et italienne actuelles ne comprennent que les familles pour les verbes permettant l'extraction.

La MGf comprend des familles de schèmes différentes (cf. Chapitre 3) selon le mode de la complétive (indicatif ou subjonctif) imposé par le verbe dans une phrase assertive. On donne Figure 106 un exemple de schème pour un verbe à complétive au subjonctif (avec les traits de mode uniquement).

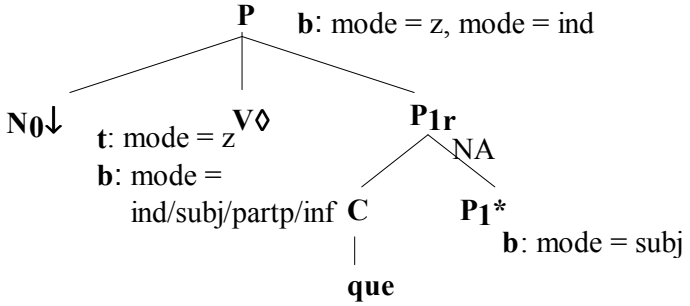


Figure 106. Schème pour un verbe à complétive (n0Vs1subj)

6.1.3 Structures des complétives et des infinitives

Nous venons de voir la représentation des infinitives comme des phrases à sujet vide. Elles peuvent comprendre éventuellement un complémenteur (fr. *de, à*, it. *di*).

On donne ci-dessous un exemple, pour l'italien, d'un arbre dérivé avec verbe enchâssant et infinitive avec complémenteur *di*. Le complémenteur provient de l'arbre du verbe enchâssant.

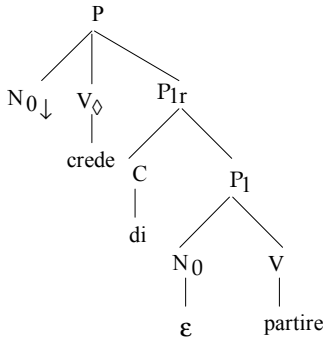


Figure 107. Exemple d'arbre dérivé avec infinitive et complémenteur *di*

Pour une complétive, le gouverneur syntaxique d'une proposition subordonnée peut être un verbe, un nom, un adjectif, une préposition ou un complémenteur plein. Dans tous les cas sauf ce dernier, la complétive est introduite en français par un complémenteur régi (*que* ou *si*), en italien par *che* ou *se*, ou pour quelques verbes, sans complémenteur.

Elia (1984:155s) note une différence importante entre le français et l'italien, en ce qui concerne les complétives prépositionnelles. En effet, en français elles peuvent être formées selon le schéma *prep ce que P*, alors que cela n'est jamais possible en italien.
Ainsi, pour une complétive gén-objet, on n'a qu'une réalisation directe en italien, alors que le français permet une réalisation en *de ce que* (soutenu) et une réalisation directe.

(338) a. Ena doute de ce que tu partes.
a'. *Ena dubita di ciò che tu parta.
b. Ena doute que tu partes.
b'. Ena dubita che tu parta.

6.2 Les fonctions à réalisation seulement phrastique

6.2.1 Infinitive de perception et infinitive de but

On a une infinitive locative par exemple en (339a) et une infinitive de perception en (339b).

(339) a. Paul court à la boulangerie acheter du pain.
b. Paul entend Pierre chanter.

Les deux fonctions sont distinguées car elles ont une réalisation différente en italien, où l'infinitive locative fait intervenir un complémenteur **a** ((339a'), alors que l'infinitive de perception n'en comprend pas (339b').

(339) a'. Paul corre al panificio **a** comprare pane.
b'. Paul sente Pierre cantare.

Infinitives de perception

Les verbes de perception comprennent une infinitive, dont le sujet est réalisé comme l'objet du verbe de perception. En effet, tant en français qu'en italien (Skytte et al., 1991:509), on a alternance avec un clitique accusatif, et alternance possible avec un passif (selon les verbes, cela est tout de même rare, en français ainsi qu'en italien).

(340) a. Maria ha sentito Piero suonare il pianoforte.
a'. Maria a entendu Piero jouer du piano.
b. Maria l'ha sentito suonare il pianoforte.
b'. Maria l'a entendu jouer du piano.
c. Piero è stato sentito suonare il pianoforte da Maria. (Skytte et al., 1991:509).
c'. ?Piero a été entendu jouer du piano par Marie.

On a discuté au Chapitre 3 du statut de verbe à montée de l'objet pour ces verbes, et on a conclu en faveur d'une analyse "à contrôle", leur objet pouvant être considéré comme un argument sémantique.
Pour les constructions avec infinitives des verbes de perception, nous utilisons la fonction inf-per, pour infinitive de perception, qui n'a que des propriétés "négatives" : pas de pronominalisation, pas de clivée, pas d'alternance avec un SN.
On a aucune propriété de prédicat complexe entre verbe de perception et infinitif : l'objet et l'infinitive peuvent permuter, les clitiques compléments (au sens strict, hors sujet) de l'infinitif se placent sur l'infinitif, l'infinitif peut être nié. L'analyse du verbe de perception comme verbe à contrôle ne pose donc pas de problème particulier.
On ne traite cependant pas dans la même famille les cas d'alternances avec une relative ou un attribut, possible en français et en italien.

(341) a. Maria ha sentito Piero che suonava il pianoforte.
a'. Maria a entendu Piero qui jouait du piano.
b. Maria ha visto Piero stanco.
b'. Maria a vu Piero fatigué.

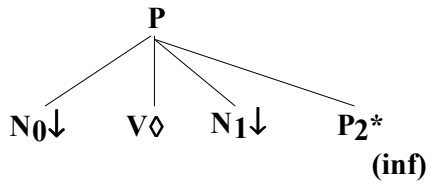


Figure 108. Schème pour un verbe de perception (dans son emploi avec inf. de perception)

L'infinitive de perception est représenté dans l'arbre du verbe de perception comme un noeud pied, ce qui permet de gérer l'éventuelle extraction à distance de l'objet (342).

(342) Qui Paul pense-t-il que Pierre a déjà entendu jouer du piano?

Infinitives de but

On donne le schème pour un verbe à locatif et infinitive de but en italien (l'infinitive est marquée par *a*).

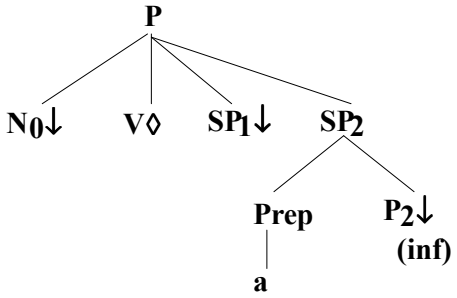


Figure 109. Schème pour un verbe italien avec infinitive de but et locatif

On considère le locatif comme optionnel, ainsi que l'infinitive de but. Seuls certains verbes cependant admettent l'ellipse des deux compléments (cf. le contraste (343d)). Nous avons choisi de bloquer dans les familles avec infinitive de but (n0V(loc1)(sbut2) et n0Vn1(loc2)(sbut3)) l'ellipse simultanée des deux compléments (ce qui est représenté par une contrainte de cooccurrence négative entre locatif-omis et inf-per-omis). Pour un verbe comme COURIR ou DESCENDRE, on doit fournir dans le lexique à la fois la sous-catégorisation n0V(loc1)(sbut2) et la sous-catégorisation intransitive (en l'occurrence la famille n0V-a pour COURIR, et n0V-e pour DESCENDRE, qui ne sélectionnent pas le même auxiliaire.

(343) a. Paul descend/va à la cave chercher des bougies.
b. Paul descend/va à la cave.
c. Paul descend/va chercher des bougies.
d. Paul descend/*va.

Pour l'italien, cette double sous-catégorisation est indépendamment rendue nécessaire, étant donné notre traitement de l'inaccusativité. On a vu en effet au Chapitre 3, que les verbes de mouvement sont inaccusatifs dans le cas de la présence d'un but (locatif ou infinitival, ou les deux), même si dans leur emploi intransitif strict certains sont inergatifs (comme CORRERE, courir). Pour représenter cela, CORRERE doit sélectionner deux familles : une famille inaccusative Vn0(loc1)(sbut2) pour (344a,b,c) et une famille inergative n0V pour (344d). Ce traitement implique que Vn0(loc1)(sbut2) ne comprenne pas les schèmes avec ellipse du locatif et de l'infinitive de but.

(344) a. Paul è/*ha corso al panificio a comprare pane.
b. Paul è/*ha corso al panificio.
c. Paul è/*ha corso a comprare pane.
d. Paul *è/ha corso. (cf. Dini, 1995)

L'ordre respectif du locatif et de l'infinitive de but est en général locatif < inf-but. Cependant l'ordre inverse n'est pas formellement à exclure. Le focus peut être placé sur le locatif (345). Cela pose un problème pour l'analyse syntaxique, étant donné que les cas comme (345) vont être systématiquement doublement analysés, selon que l'on attache le locatif au verbe enchâssant (ici *descend*) ou au verbe enchâssé (*chercher*). Pour cette raison, nous considérons pour (345) que le locatif est complément de l'infinitif.

(345) Paul descend en chercher à la cave.

Notons enfin que l'infinitive de but est représentée par un noeud pied, ce qui permet de représenter une éventuelle extraction à distance hors de l'infinitive, comme par exemple en (346b)[25]. On fournit l'adjonction nécessaire pour dériver (346a).

(346)a. Quel pain Jean est-il allé acheter?
b. Qu'est-ce-que Paul pense que Jean est allé acheter?

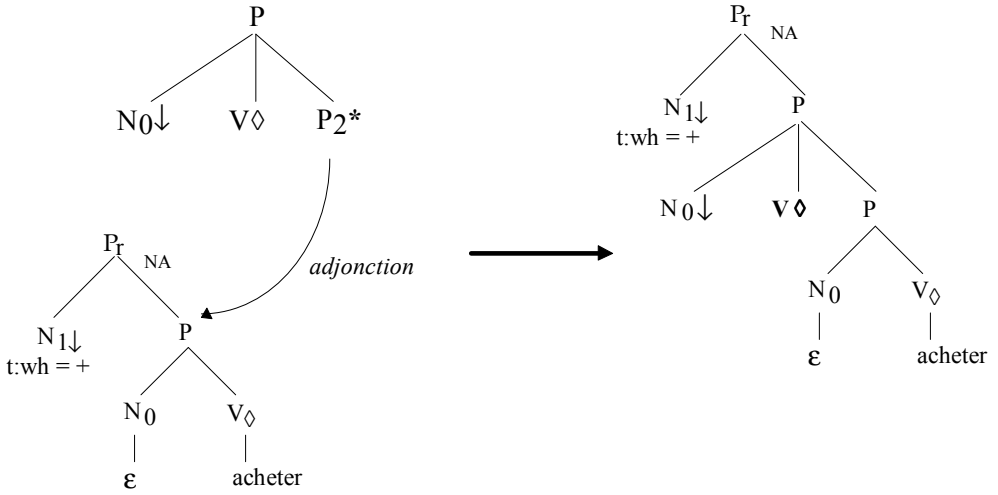


Figure 110. Adjonction du verbe enchâssant pour dériver (346a)

[25] L'extraction hors de l'infinitive est impossible si le locatif précède l'infinitive, c'est pourquoi l'on a un noeud pied pour l'infinitive, Figure 109.

6.3 Les fonctions alternant entre réalisation nominale et phrastique

6.3.1 Sujet phrastique

En français, certains verbes admettent un sujet phrastique (cf. les tables 4 et 5 de Gross, 1975, il s'agir surtout de verbes psychologiques). Dans le cas d'une phrase tensée, le mode est systématiquement le subjonctif. Dans le cas d'une infinitive, elle apparaît sans complémenteur en général. La présence d'un complémenteur *de* appartient à un registre très soutenu.

(347) a. Que Paul revienne réjouit Marie.
b. Partir réjouit Marie.
c. De partir réjouit Marie.

Le sujet phrastique subit de manière quasiment obligatoire une démotion, en particulier il peut apparaître comme objet, à l'impersonnel. Dans le cas d'une infinitive, un complémenteur *de* apparaît alors. Le passif est également possible, mais il s'agit d'un passif en *de* (cf. Chapitre 4), (qui met en jeu une démotion sujet→gén-objet).

(348) a. Il réjouit Marie de partir.
b. Il réjouissait tout le village que Paul soit revenu sain et sauf.

(349) a. Marie est angoissée de partir.
b. Marie est angoissée que Paul ne soit pas rentré.

Dans la MGf, on représente en imposant d'une part que la démotion sujet→agt-objet ne s'applique pas à des sujets phrastiques. D'autre part la démotion sujet→objet impose sur l'objet (l'actant H), le métatrait :compl-inf = de, qui n'est pertinent que pour une réalisation infinitive (cf. infra la Section sur les objets phrastiques, l'utilisation d'un métatrait pour contrôler le complémenteur de l'infinitive objet).

A l'inverse, un actant phrastique peut difficilement être promu sujet. Dans le cas de passif, on a en effet plutôt des passifs impersonnels (cf. (350a) versus (350b)). Dans la MGf, on contraint la catégorie de l'objet dans la promotion objet→sujet : l'objet doit être nominal. Le passif impersonnel est représenté de manière directe (sans utiliser la classe du passif personnel), et on ne contraint pas la catégorie de l'objet, on peut ainsi obtenir le schème pour (350b).

(350) a. ? Plier bagages a été ordonné à tous.
b. Il a été ordonné à tous de plier bagages.

Pour l'italien, il faut distinguer d'une part les sujets phrastiques des inaccusatifs de ceux des inergatifs. En (351) il s'agit de verbes transitifs. Le sujet phrastique est un sujet initial. A l'infinitif, il apparaît comme en français sans le complémenteur *di*.

(351) a. Che si propongano certi progetti mi disturba / secca / interessa.

(Qu'il se propose certains projets me perturbe / ennuie / intéresse.) (Acquaviva, 1991)

b. Ricevere certi progetti mi disturba /secca / interessa.
(Recevoir certains projets me perturbe / ennuie / intéresse.)

Parmi les verbes inaccusatifs à objet initial phrastique, il faut en outre distinguer plusieurs types. Acquaviva (1991:666-667) distingue les sujets phrastiques de verbes inaccusatifs de complément phrastique de verbe inaccusatifs.
Les premiers sont des sujets d'inaccusatifs, c'est-à-dire pour nous des objets initiaux promus sujet. A l'infinitif ils sont en général sans complémenteur (352).
Les seconds sont selon Acquaviva des verbes impersonnels. L'argument phrastique apparaît en position postverbale en général (353a), et ne peut apparaître devant le verbe que sous conditions (353b). En général le complémenteur de l'infinitive est *di*, et ce même en position préverbale.

(352) A Mario è bastato / dispiaciuto / servito far uso di determinati mezzi.
A mario est suffi / déplu / servi faire usage de moyens déterminés
(A mario il a suffi / déplu / servi de faire usage de moyens déterminés.) (Acquaviva, 1991)

(353) a. Mi pare / sa / importa (di + *E) aver capito bene.
(Il me semble / sait / importe d'avoir bien compris) (Acquaviva, 1991)

b. Di aver capito bene è parso a tutti.
D'avoir bien compris est semblé à tous.

Dans la MGi, nous traitons les deux cas comme des "sujets inaccusatifs" phrastiques, c'est-à-dire des objets initiaux promus sujet. Dans le lexique, on peut en spécifiant un trait :sujet-post ne sélectionner que les schèmes avec sujet post-verbal. Les familles sont distinguées selon la présence ou l'absence de complémenteur *di* pour l'infinitive (le complémenteur s'il existe devant être placé dans les schèmes).

Les réalisations phrastiques de sujet en italien peuvent dans le cas d'une infinitive comporter ou pas le complémenteur *di*, selon les verbes. Le cas avec complémenteur *di* n'est rencontré que pour un sujet infinitif qui est un objet initial (pour les verbes inaccusatifs donc).

Le sujet phrastique en italien peut être démis (mais on n'a pas la démotion vers la fonction objet, comme on l'a en français avec l'impersonnel). Il s'agit de la démotion mise en oeuvre dans le cas de passif.

(354) a. Che Max sia venuto sbalordisce Ena.
Que Max soit venu abasourdit Ena.

b. Ena è sbalordita che Max sia venuto.
Ena est abasourdie que Max soit venu.

c. Ena è sbalordita di dover partire.
Ena est abasourdie de devoir partir.

Comme pour le français, il s'agit alors du passif en *di*, même si on a moins d'argument pour le dire que pour le français. La cliticisation de "l'agent" (l'actant phrastique initialement sujet) est en *ne* (354d), comme un gén-objet. Mais l'indice clair en français

(la tournure en *de ce que*), n'est pas transposable à l'italien qui n'a jamais ce tour (354e) (Elia, 1984:156).

(354) d. Ena ne è sbalordita.
Ena en est abasourdie.

e. * Ena è sbalordita di ciò che Max sia venuto.
Ena est abasourdie de ce que Max soit venu.

Dans le cas d'un passif avec agt-objet pour ces verbes à complétives sujet, la seule solution est de recourir à un SN complexe Nom + complétive (355d,d') (Elia, 1984).

(355) a. Ena è sbalordita dal fatto che sia venuto. (Elia, 1984)
a'. Ena est abasourdie par le fait que Max soit venu.

Un objet phrastique peut être promu sujet par le passif en italien, mais le sujet reste alors souvent en position postverbale. Le complémenteur de l'infinitive est *di*. Cela est représenté dans la MGi, en ne contraignant pas la catégorie de l'objet dans la promotion objet→sujet, et en ajoutant le métatrait compl-inf = *di*.

6.3.2 Objet phrastique

En français, les infinitives objets[26] peuvent être introduites par un complémenteur *de*, *à* ou bien ne pas comporter de complémenteur. Lorsqu'il y a complémenteur *à* ou *de*, l'infinitive objet peut être confondue avec une infinitive portant la fonction a2-objet ou gén-objet.
Les complétives objets sont introduites par *que*. Comme les complétives gén-objet peuvent elles aussi être directement introduites par *que*, il peut également y avoir confusion.

On repère les objets phrastiques par les propriétés suivantes :
- pseudo-clivée en "ce que n0 V c'est s1"
- éventuelle pronominalisation en *le/l* (clitique) ou en *cela* (directe)
- éventuelle alternance avec un SN objet (pour un même sens du verbe)

On peut remarquer que ces tests peuvent être en contradiction avec le test du causatif, qui met en évidence un comportement transitif/intransitif selon la fonction prise par le causé. Par exemple pour un verbe comme AVOUER, à côté de (356a,b,c) qui le classe clairement comme transitif, on a (356d). On ne retient pas le test du causatif par la suite.

(356) a. Paul avoue que cette histoire le trouble.
b. Paul a avoué cela.
c. Ce que Paul a avoué c'est que cette histoire le trouble.
d. ? Paul a fait avouer devoir partir Marie.

(357) a. Paul cherche à partir.
b. Paul cherche à ce que tout soit en ordre.
c. Paul cherche (E + *à) cela.
d. * Paul y cherche.

[26] Comme pour les objets nominaux, on emploie le terme "objet" au sens strict, pour "objet direct".

e. Ce que Paul cherche, c'est que tout soit en ordre.
f. *Ce à quoi Paul cherche c'est que tout soit en ordre.

(358)a. Paul propose partir.
b. Paul propose que tout soit en ordre.
c. Paul propose (E + *de) cela.
d. *Paul en propose.
e. Ce que Paul propose, c'est que tout soit en ordre.
f. *Ce dont Paul propose c'est que tout soit en ordre.

Nous avons déjà mentionné au Chapitre 3, que l'on distingue les familles à objet phrastique selon que :

- Le mode de la complétive est, dans une phrase assertive, le subjonctif ou l'indicatif
- L'infinitive n'est pas introduite par un complémenteur, est introduite par *de*, est introduite par *à*.

Le premier critère n'est pas véritablement employé pour imposer directement dans les schèmes le mode de la complétive. En effet, le mode est soumis à des contraintes diverses, pas uniquement lexicales, comme par exemple la présence d'une négation (cf. (359a)/(359b).

(359) a. Je crois que Paul est parti.
b. Je ne crois pas que Paul soit parti.

Mais le mode imposé dans le cas d'une phrase assertive est relativement bien corrélé à deux propriétés (liées), associées aux verbes imposant l'indicatif :

- la possibilité pour le sujet de la complétive et celui du verbe enchâssant d'être coréférents (360).
- la possibilité pour le verbe enchâssant de constituer une incise (361)

(360) a. Paul_i dit qu'$\text{il}_{i/j}$ part.
b. Paul_i veut qu'$\text{il}_{*i/j}$ parte.

(361) a. Paul part, dit-il.
b. *Paul part, veut-il.

Ces différences syntaxiques ne sont pas représentées actuellement dans la MGf, mais la classification des familles prévoient cette distinction, pour une intégration future.

Pour l'italien, la différence est que l'infinitive objet est soit directe, soit introduite par le complémenteur *di*. On n'a pas l'équivalent du complémenteur *à* du français (voir par exemple Acquaviva, 1991).
Une autre différence est que pour l'italien, une sous-classe de verbes à objet phrastique permet la complétive sans complémenteur *che*, la complétive étant alors au subjonctif (Acquaviva, 1991:644). Il s'agit des verbes d'opinion (plus SPERARE (espérer) et TEMERE (craindre). Tous ces verbes, dans le cas d'une infinitive imposent le complémenteur *di*.

(362) a. Non credo si possa far molto per salvare la situazione.
NEG je-crois se puisse faire beaucoup pour sauver la situation.
(Je ne crois pas qu'il se puisse faire beaucoup pour sauver la situation)

b. Immagino Gianni sia già al corrente delle novità.
j-imagine Gianni soit déjà au courant des nouvelles
(J'imagine que Jean est déjà au courant des nouvelles.)

On prévoit donc dans la MGi une classe spécifique, avec complétive sans complémenteur. Les schèmes correspondants ne sont prévus que dans les familles où un complémenteur *di* est imposé pour l'infinitive (n0Vcs1, n0Vcs1(an2)). Un trait contrôle la sélection par le lexique de ces schèmes sans complémenteur.

6.3.3 Gén-objet phrastique

Certains verbes sous-catégorisent un gén-objet phrastique (complétive ou infinitif). Dans le cas de la complétive, la construction en *prep ce que* est parfois possible, mais est concurrencée par une construction directe (moins soutenue).

(363) a. Jean rêve de ce que tous les hommes soient égaux en droits.
b. Jean rêve de partir.
c. Jean rêve que tous les hommes soient égaux en droits.

Un gén-objet phrastique est repéré par les propriétés :
- éventuelle alternance avec un gén-objet nominal (pour un même sens du verbe)
- pronominalisation en *en*
- pseudo-clivée en "*ce dont N0 V, c'est que P1*"

(364) a. Jean rêve de l'égalité des chances pour tous.
b. Jean en rêve.
c. Ce dont Jean rêve, c'est que tous les hommes soient égaux en droits.

La MGf ne couvre pas les constructions en *de ce que*. Dans le cas de la complétive, on ne contraint pas le mode indicatif ou subjonctif. On peut avoir en effet variation, pour un même verbe ou pour des verbes différents.
Certains verbes n'admettent pas l'infinitive (par exemple AVERTIR).

(365) a. Paul avertit Pierre que tous arriveront à 5 heures.
b. Paul en avertit Pierre.

Pour l'italien, la différence vient de l'absence généralisée de tournure en *prep ce que* (cf. supra la série (338).

Il faut remarquer en outre qu'en ialien, les gén-objets en *da* ne peuvent pas donner lieu à une complétive autrement que via un SN complexe en *il fatto que (le fait que)* (Elia, 1984:148).

(366) a. Ciò deriva dal fatto che fumi.
Cela dérive du fait que tu fumes.
b. *Ciò dériva che tu fumi.
Cela dérive que tu fumes
c. *Ciò deriva da che tu fumi.
Cela dérive de que tu fumes

On a regroupé sous la fonction gén-objet les cas avec préposition *di* et ceux avec *da*, sur la base d'une même cliticisation, mais il est vrai que cette différence pour la réalisation phrastique peut être un argument pour séparer deux fonctions.
Les deux marqueurs *di* et *da* correspondent au seul *de* en français. Il est cependant intéressant de remarquer que les verbes prenant un gén-objet en français, dont la traduction est proche morphologiquement en italien, et qui est une traduction avec gén-objet en *da*, (par exemple DERIVER/DERIVARE, RESULTER/RISULARE) n'admettent pas non plus de complétive directe (cf. la traduction de b). Cela laisse supposer que c'est le regroupement du gén-objet en français qui est trop grossier.

6.3.4 A2-objet phrastique

Certains verbes sous-catégorisent un a2-objet phrastique, plus précisément une complétive au subjonctif ou une infinitive. Dans le cas de la complétive, seule la construction en *prep ce que* est possible, et pas la construction directe (à la différence du gén-objet).

(367) a. Jean tient à ce que Paul parte.
b. Jean tient à partir.
c. *Jean tient que Paul parte

En français, la complétive a2-objet est repérable par les propriétés :
- éventuelle alternance avec un a2-objet nominal (pour un même sens du verbe)
- clivée en "*ce à quoi N0 V, c'est que P1*"
- pronominalisation en *y*

(368) a. Jean tient au départ de Paul.
b. Ce à quoi Jean tient, c'est que Paul parte.
c. Jean y tient.

La MGf ne couvre actuellement que les a2-objets infinitives (d'une manière générale, les tournures en *prep ce que* ne sont pas couvertes).

Pour l'italien, la plupart des verbes avec a2-objet phrastique prennent une infinitive, qu'ils s'agissent d'inergatifs (369a) ou d'inaccusatifs (369b), mais pas de complétive (Acquaviva, 1991:651), et cela même si la sémantique du verbe le permettrait (370).

(369) a. Gianni ha cominciato / continuato / pensato / rinunciato / ripreso a frequentare la mensa.
Gianni a commencé / continué / pensé / renoncé / repris à fréquenter la cantine.
b. Mario si è abituato / deciso / impegnato / prestato a frequentare la mensa.
Mario s'est habitué / décidé / impliqué / prêté à fréquenter la cantine.

(370) *Mi sono habituato (a + E) che parli.
RF.1 suis habitué (à + E) que tu-parles.

Mais dans le cas où la complétive est possible, on a déjà annoncé l'impossibilité d'avoir l'équivalent de *prep ce que*, d'où (371a). On a en revanche une structure *a che*

impossible en français (371b). De plus, une complétive a2-objet introduite directement est possible (371c).

(371) a.* Gianni acconsente a ciò che la partita sia disputata.
Gianni consent à ce que la partie soit disputée.

b. Gianni acconsente/ ambisce / contribuisce / si è impegnato a che la partita sia disputata.
Gianni consent / ambitionne / contribue / s'est impliqué à que la partie soit disputée

c. Ho imparato / insegnato[27] che non bisogna pretendere più del dovuto.
(J'ai appris / enseigné qu'il n'est pas nécessaire de prétendre plus que son dû.)

On représente dans la MGi uniquement les cas avec infinitives.

6.3.5 Interrogatives indirectes

Enfin on considère une dernière fonction réservée aux interrogative indirectes.
On considère que cette fonction peut avoir une réalisation correspondant à une interrogation totale (avec le complémenteur fr. *si*, it. *se*) ou une réalisation correspondant à une interrogation partielle.

(372) a. Chiedeva (à Maria + E) se gli fosse permesso.
a'. Il demandait (à Maria + E) si cela lui était permis.

b. Sapresti dirmi quanti studenti lo hanno nominato rappresentante?
b'. Saurais-tu me dire combien d'étudiants l'ont nommé représentant?

Pour le premier, on représente des schèmes incluant le complémenteur fr.*si*, it. *se* (Figure 111). Pour le second cas, le verbe enchâssant ancre un schème avec une position phrastique à substituer, portant le trait wh = + (Figure 112).

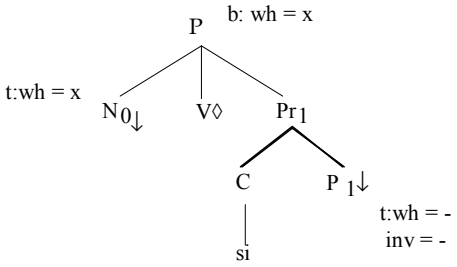


Figure 111. Schème pour une interrogation indirecte totale (MGf)

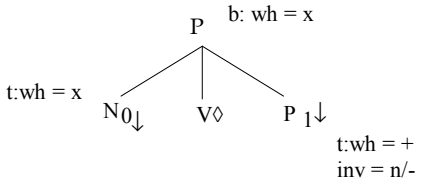


Figure 112. Schème pour une interrogation indirecte partielle (MGf)

Les phrases pouvant se substituer dans le schème de la Figure 112 correspondent soit à des interrogatives directes à actant antéposé (cf. supra le traitement des interrogatives directes, à actants antéposés), soit des interrogatives directes avec ajout interrogé (comme dans (373)). On donne les combinaisons nécessaires pour dériver (373a').

(373) a. Paolo sa quando vengono gli studenti.

[27] *insegnare* (enseigner) et *imparare* (apprendre) sont des cas limites, en cela que le nominal alternant avec l'infinitive en *a* est un objet (direct).

a'. Paul sait quand viennent les étudiants.

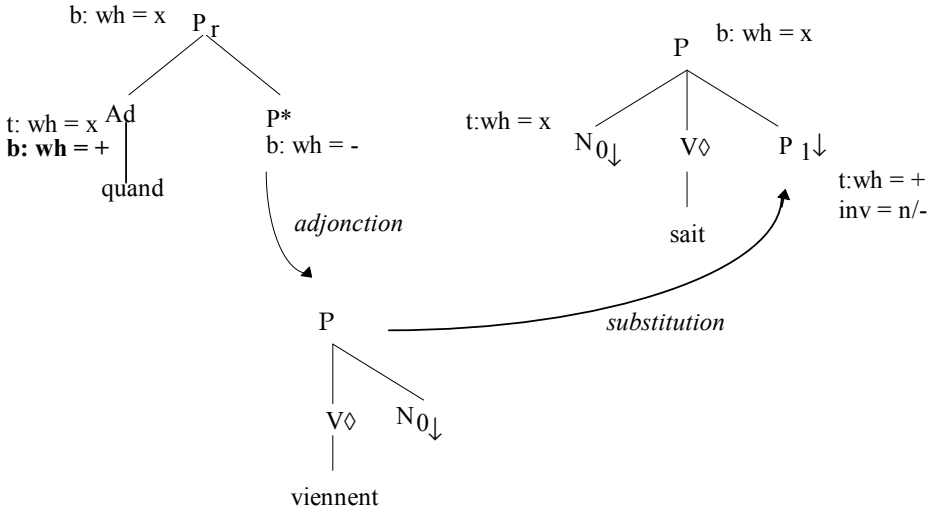


Figure 113. Dérivation d'une interrogative indirecte (avec ajout interrogé, (373a'))

Une différence entre les interrogatives indirectes et les directes est cependant en français que l'interrogative indirecte ne peut pas comporter de clitique sujet inversé (d'où la contrainte inv = n/- Figure 112, qui bloque le clitique inversé impliquant inv = cl).

(374) a. Quand Paul part-il?
b. * Dis-moi quand Paul part-il.

On considère également comme réalisation possible de l'interrogative indirecte la réalisation non phrastique, correspondant à une question cachée (plusieurs catégories sont possibles, Ad, N, SP).

(375) Je sais pourquoi/qui/à qui.

Ces réalisations (interrogatives totale, partielle, cachée) ne posent pas de problème particulier pour la représentation dans la MGf et la MGi. Nous ne détaillons pas les classes ici.

7 Les positions extraites : relatives, clivées, questions

Toutes les réalisations de fonctions extraites héritent d'une classe générale extraction.
En français et en italien, seul un dépendant d'un verbe peut être extrait, si l'on exclut les cas rares comme (376).

(376) ...les pommes que je ne sais pas à qui donner

On représente cela en déclarant la classe extraction comme une classe disjonctive (cf. Chapitre 2, les moyens expressifs pour la représentation des contraintes de compatibilité).

Les MGf et MGi couvrent les relatives, clivées et questions à actant extrait.

7.1 Les relatives

7.1.1 Les relatives en français

Nous ne traitons que des relativisations d'arguments nominaux et pas phrastiques comme dans *Marie est partie, ce qui attriste Jean*. On ne traite pas non plus les relatives substantives (comme dans *Qui dort dîne*).

La MGf traite les cas suivants :

(377) a.	(l'homme) qui chante	sujet relativisé
b.	(l'homme) que Paul vois	objet relativisé
c.	(l'homme) dont je parle	gén-objet relativisé (en *dont*)
d.	(l'homme) à qui/auquel Paul pense[28]	SP relativisé (pour a-objet, obl-objet)
e.	(la chose) à quoi/laquelle je pense	
f.	(la maison) où/dans laquelle va Paul	locatif relativisé, modifieur de nom
g.	(là) où Paul va	locatif relativisé, modifieur d'adverbe
h.	(l'endroit) d'où/duquel revient Paul	locatif de provenance relativisé

Sont également traités les cas de pied-piping (i.e. où le SP relatif est contenu dans un SP) en utilisant les mêmes schèmes que pour les relatives correspondantes où le SP relatif n'est pas contenu dans un SP).

(377) d'.(l'homme) à la fille de qui/laquelle Paul pense
f'. (la maison) dans la cave de laquelle va Paul
h'. (la maison) du grenier de laquelle ressort Paul

Suivant la distinction de Kayne (1976), reprise par Godard (1988), nous distinguons les pronoms relatifs (*qui*, *quoi*, LEQUEL) des complémenteurs relatifs, de catégorie C, marqués pour la fonction (*qui, que, dont*)[29]. Les pronoms ancrent des arbres de catégorie

[28] La grammaire traite les cas préposition + LEQUEL de manière analytique. Pour l'analyse, un prétraitement "défait" les amalgames à+LEQUEL et de+LEQUEL.

[29] Abeillé (1991) distinguait seulement le complémenteur relatif *que*. La distinction complémenteur / pronom relatif tient au fait que les complémenteurs pour les relatives introduisent aussi des complétives (*Je pense **que** Jean viendra, Qui crois-tu **qui** viendra?, Je cherche **à** venir*). Pour *dont*, l'argument est plus fragile : *dont* introduit les relatives "à pronom" : *un homme dont il est clair qu'il est intelligent*. Au contraire, les pronoms relatifs sont marqués pour le trait animé (*qui/quoi*), ou bien en genre et nombre (LEQUEL), et ils introduisent également les interrogatives.
Un avantage de distinguer complémenteur et pronom relatif, c'est que l'on explique pourquoi l'équivalent de l'anglais (i) est impossible en français (ii).
(i)*the girl whose mother came yesterday*
(ii) **la fille la mère de laquelle est venue hier.*
Pour dériver (ii), il faudrait une position nominale à la place du relatif. Comme on a un complémenteur, qui ne peut pas être "complexe", (ii) est bien bloquée.

N, portant le trait :wh = rel[30]. Les complémenteurs sont eux directement présents dans les schèmes verbaux, comme co-têtes.
Nous donnons ci-dessous les schèmes (sans traits) représentant différents cas de relatives. En tant que modifieurs de noms, il s'agit d'arbres auxiliaires de racine et noeud pied nominal. Seul dans le cas illustré en (377g), la relative est un modifieur d'adverbe, et on la représente par un arbre auxiliaire modifieur d'adverbe (locatif). Dans le cas de schèmes avec complémenteur, c'est le noeud pied (représentant l'antécédent) qui est considéré comme le noeud argumental (il porte donc un indice). Dans le cas de schèmes avec pronom relatif, c'est le pronom (où sa projection) qui est considéré comme position argumentale).

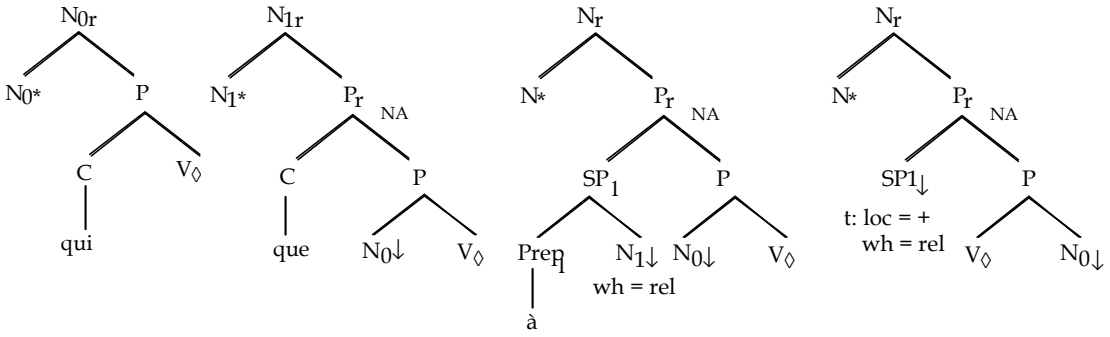
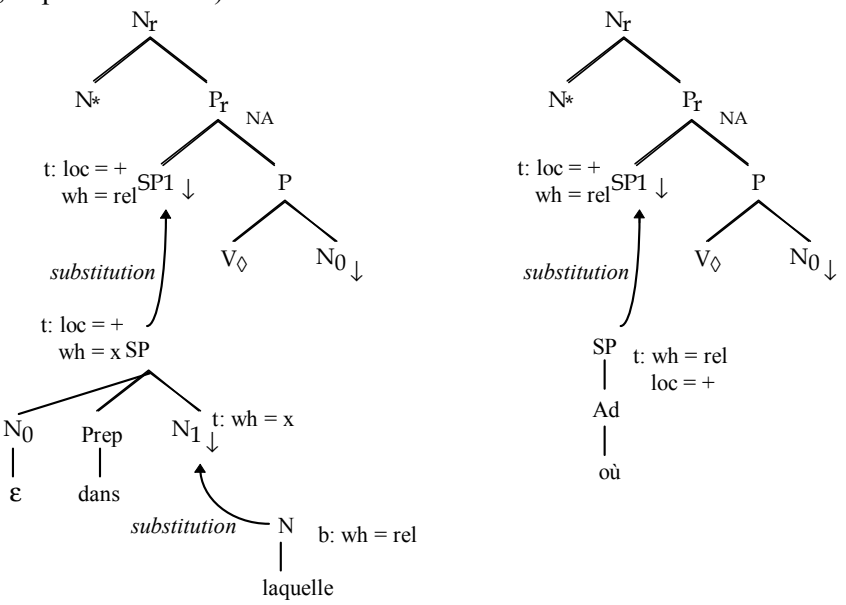


Figure 114. Schèmes pour (377a), (377b), (377d) et (377f)

L'adverbe relatif *où* est représenté par un arbre de projection SP, portant les traits :loc = + et :wh = rel, ce qui est parallèle au traitement des adverbes locatifs comme arbres à ancre adverbiale, mais de projection SP. On donne Figure 115 les dérivations pour analyser les deux cas de (377f) (soit pronom relatif *où*, soit SP étendu, avec préposition locative, et pronom relatif).



[30] L'attribut :wh peut avoir les valeurs rel (pour les pronoms relatifs), + (pour les pronoms interrogatifs), suj (pour représenter les relatives à sujet relativisé) et -.

Figure 115. Dérivations pour (377f)

Pour tous les cas illustré en (377), sauf le sujet relativisé (377a), la grammaire permet une distance non bornée entre l'antécédent et son verbe, par adjonction d'un verbe pont. En effet pour le français, il y a éventuellement une possibilité d'extraction à distance du sujet relativisé, mais elle ne suit pas le cas général, et n'est pas acceptée par tous les locuteurs (ce que l'on a déjà mentionné Section 1, exemple (278)).
Cela est représenté dans la grammaire par une structure plus plate pour l'arbre avec sujet relativisé, et un trait :wh = suj sur la seule projection phrastique pour bloquer l'adjonction d'un verbe pont, qui impose :wh = -. En revanche on permet l'adjonction de modifieurs de phrase à ce noeud.
Les arbres avec pronom relatif (i.e. sous un SP, cf. les deux arbres de droite, Figure 114) permettent d'analyser les cas de "pied-piping". Les arbres comprennent en effet un noeud à substituer nominal dans le SP antéposé. Ce noeud doit recevoir une substitution qui porte le trait :wh = rel. Le pronom relatif peut directement se substituer, ou bien, pour traiter le pied-piping (comme en (378) par exemple) le pronom peut être substitué dans un SN, remonter le trait :wh = rel, le tout se substituant dans l'arbre de la relative.

(378) La personne au père de laquelle Van Gogh avait donné un tableau

Nous reviendrons au Chapitre 7 sur cette analyse du pied-piping, qui ne permet pas de récupérer les dépendances sémantiques appropriées.

On ne traite pas les relatives où c'est un complément de nom qui est extrait (379). On peut montrer que pour le français, si l'on veut respecter le PCPA (c'est-à-dire si l'on veut que le complément extrait ait son origine dans l'arbre du nom dont il est le complément, alors il faut utiliser une extension de TAG (cf. Abeillé, à paraître, ou Candito et Kahane, 1998b pour une analyse en Graph Adjoining Grammar).

(379) Le gamin dont Paul connaît les parents.

Représentation dans la MGf

Les cas de relatives avec pronom relatif (sous une préposition) sont représentés en considérant que le noeud à substituer pour les pronoms relatifs est le noeud argumental qui porte la fonction par rapport au verbe. Dans le cas de relative avec complémenteur, on considère que c'est le noeud pied (représentant l'antécédent) qui est le noeud argumental qui porte la fonction. Dans une application où l'on voudrait récupérer des dépendances sémantiques lexicalisées (prédicat-argument), cela permet de récupérer de manière directe le lien antécédent-verbe.

Nous donnons Figure 116 le schéma des classes de la MGf pour les relatives avec dépendant du verbe relativisé (légèrement simplifié : la classe pour le locatif de provenance relativisé n'est pas représentée). Certaines de ces classes sont également héritées pour les relatives avec ajout relativisé (cf. infra Chapitre 6). Les classes terminales de ce schéma sont des classes terminales de la dimension 3, chacune pour une FG différente.

La classe objet-relativisé hérite de objet-extrait, et hérite donc des équations d'accord avec le participe. On peut éventuellement bloquer la relativisation de l'objet (par une contrainte de cooccurrence négative), pour l'objet d'une construction impersonnelle (cf. Chapitre 4, on considère l'actant extraposé des constructions impersonnelles comme un objet). En effet certains locuteurs n'admettent pas (380b) par exemple (mais nous l'acceptons).

(380) a. Il arrive de nouveaux blessés à toute heure.
b. ? les nouveaux blessés qu'il arrive à toute heure

Pour les classes avec pronom relatif dans un complément prépositionnel, elles sont distinguées selon la fonction de l'élément relativisé, et elles héritent chacune de la classe introduisant le SP étendu, pour chacune de ces fonctions, ce qui fournit la topologie complète du SP.

L'interaction des relatives avec l'inversion du sujet est traitée par des contraintes sur l'inversion du sujet. Pour l'inversion du sujet nominal, nous avons vu supra que la classe qui le représente doit être cooccurrente à une classe héritant de extraction, comme les relatives par exemple. Pour le clitique sujet inversé, au contraire il est impossible avec une relative ou une clivée (et porte la contrainte de cooccurrence négative en conséquence). L'inversion complexe est elle aussi impossible, ce qui est représenté par un trait :clS:dir=-/n, sur ?Vmarq.

Figure 116. Les classes de la MGf pour les relatives (dépendants du verbe relativisés)

7.1.2 Les relatives en italien

Les cas traités dans la MGi sont les suivants :

(381) a. (l'uomo) **che** canta. — sujet relativisé
l'homme qui chante

b. (l'uomo) **che** vedo. — objet relativisé
l'homme que je vois

c. (l'autore) **cui** è stato assegnato il premio — a-objet relativisé
...l'auteur PRO-datif a été décerné le prix
(...l'auteur à qui a été décerné le prix.)

d. (l'uomo) di (**cui** + **il quale**) parlo — gén-objet relativisé
l'homme duquel je parle — (idem a-objet, obl-objet)

e. (il luogo) **dove** va Paolo — locatif relativisé, modifieur de nom
le lieu où va Paolo

f. (là) **dove** va Paolo — locatif relativisé, modifieur d'adverbe
là où va Paolo

h. (il luogo) **donde** viene Paolo — locatif de provenance relativisé
le lieu d'où vient Paolo

Les pronoms relatifs sont *cui* (invariable) et IL QUALE (LEQUEL) variable en genre (pour IL) et nombre (pour IL et QUALE). IL QUALE est en deux mots[31], à la différence du français LEQUEL. Ils servent à la relativisation d'un complément prépositionnel.
Cependant l'emploi de *cui* est plus restreint que celui de IL QUALE (Cinque, 1988b), en cela que *cui* (derrière une préposition, cf. infra, il existe également d'autres emplois de *cui*) est exclu pour une relative appositive, et ne peut pas faire partie d'un SP complexe (c'est-à-dire que l'on ne peut pas avoir de pied-piping avec *cui*). De plus, pour les relatives appositives, le pronom IL QUALE peut être également utilisé pour relativiser un sujet ou un objet (Cinque, 1988b:447)

(382) La sua proposta, che/la quale è senza dubbio frutto di promesso ...
Votre proposition, CHE/laquelle est sans doute pleine de promesses...

On représente les pronoms par des arbres de racine N, portant le trait wh = rel (Figure 117 et Figure 118).

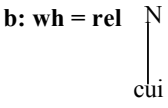


Figure 117

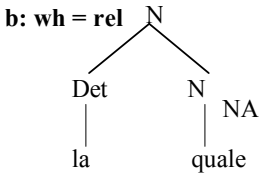


Figure 118[32]

[31] Un prétraitement sépare une préposition (*di, da, con, in*) et le déterminant IL, dans le cas où il y a amalgame.

Nous considérons que les complémenteurs relatifs en italien sont *che* et *cui*, en adaptant la distinction faite pour le français entre complémenteurs et pronomes relatifs. La motivation principale est que pour *che* et *cui* (dans son emploi comme relatif pour le a-objet), apparaissent directement, et non pas dans un SP, et surtout qu'ils n'admettent pas le pied-piping. *Che* est en outre le complémenteur pour une complétive.

Le complémenteur *che* vaut pour le sujet, l'objet et l'attribut. *Cui* vaut pour l'a-objet. Comme on l'a déjà mentionné en effet, l'italien n'a pas l'équivalent de la relative en *dont* pour le gén-objet, en revanche, il présente une relative particulière pour le a-objet, avec complémenteur datif *cui* (381c) (cf. supra Section 4.1).

Cela ne pose pas de problème particulier, et la hiérarchie pour les relatives en italien est modifiée en conséquence.

Il existe également un emploi de *cui* comme génitif, stylistiquement marqué, que l'on ne traite pas. Il s'agit de l'équivalent de la relativisation en français d'un complément de nom en *dont*, mais la syntaxe en est différente en italien : le pronom relatif génitif, *cui*, a une distribution d'adjectif possessif[33] et se place donc après l'article (383). Il peut apparaître dans des constituants complexes (à la différence de *dont*).

(383) a. Aiace, il cui coraggio è noto ... (Cinque, 1988b)
Aiace, le CUI-gén courage est connu...
(Aiace, dont le courage est connu...)

b. Giorgio, alla morte dei cui parenti eravamo già da tempo preparati, ...
Giorgio, à la mort des CUI-gén parents -nous-étions déjà depuis longtemps préparés...
(Giorgio, à la mort des parents duquel nous étions depuis longtemps déjà préparés...)

Une autre différence est que le sujet relativisé peut être à distance de son verbe, comme on l'a déjà mentionné supra, dans la section sur le sujet. On n'a donc pas dans la MGi une différence structurale entre les sujets relativisés et les compléments relativisés. Les arbres pour les sujets relativisés présentent eux aussi deux projections phrastiques, dont l'une (l'inférieure, à droite du complémenteur) peut recevoir l'adjonction d'un verbe pont. On peut donc dès la classe relative introduire un lien de dominance immédiate entre ?Pr et ?P.

En italien, l'objet relativisé (plus généralement extrait) ne s'accorde pas avec le participe. La classe objet-relativisé n'hérite donc pas de objet-extrait, qui ne faisait qu'apporter l'accord.

7.2 Les clivées

Les relatives et les clivées partagent beaucoup de propriétés, mais sont néanmoins bien distinctes, d'une part en sémantique, mais également en syntaxe.

---

[32] Cinque (1988) suggère que dans IL QUALE, QUALE est un adjectif substantivé. Aucun adjectif ne peut intervenir entre IL et QUALE (d'où la contrainte de non adjonction placé sur le N inférieur, dans la Figure 94). En revanche, le tout peut être modifié à droite.

[33] Le possessif en italien est un adjectif toujours prénominal, cooccurrent à l'article :
(i)i suoi occhi
les ses yeux
(ses yeux)

Etant donné la complexité des classes à gérer, et les différences notables entre relatives et clivées nous avons choisi de ne pas faire partager de propriétés dans la hiérarchie.

7.2.1 Les clivées en français

Une phrase comme (384) est ambiguë (Moreau, 1976) entre une interprétation déictique (qui isole un cheval parmi d'autres) et l'interprétation clivée (qui isole un élément qui est parti parmi d'autres qui sont partis).

(384) C'est le cheval qui est parti.

Cette différence sémantique s'accompagne de différences syntaxiques suffisamment importantes pour ne pas analyser la clivée comme une présentative comportant une relative :

- l'antécédent de la relative est plus restreint que le foyer de clivée.

Le foyer de clivée peut être un nom propre, un pronom personnel ou démonstratif, alors qu'ils ne peuvent pas être antécédents de relative restrictive. On peut avoir un foyer de clivée phrastique. On peut également avoir, à l'oral du moins, un foyer de clivée inerrogatif

(385) C'est le fils de qui que tu as rencontré?

- Des ajouts comme des adverbes temporels ou des participiales (gérondives) peuvent être foyers de clivée

(386) C'est en descendant les escaliers qu'il s'est cassé la jambe.

- Un SP peut être foyer de clivé (et c'est le parallèle avec la relativisation d'un SP), mais il existe une clivée concurrente, non ambiguë avec une déictique, qui est souvent meilleure.

(387) a. C'est Marie à qui tu parlé.
b. C'est à Marie que tu as parlé.

Dans la MGf, sont traités les cas où le foyer de clivée est actant du verbe, et dans le cas d'un actant SP, on ne traite que les cas de type (387b). Nous verrons au Chapitre 6 le traitement des clivées dans lesquelles le foyer est un ajout. Nous ne traitons cependant pas les cas où le foyer est un complément de nom d'un actant du verbe (388).

(388) C'est de Pierre que Paul a épousé la soeur.

Les clivées sont représentées par des schèmes avec la copule et le *ce* comme co-têtes, qui s'accordent en nombre et personne. La hiérarchie pour les clivées hérite de la classe extraction.

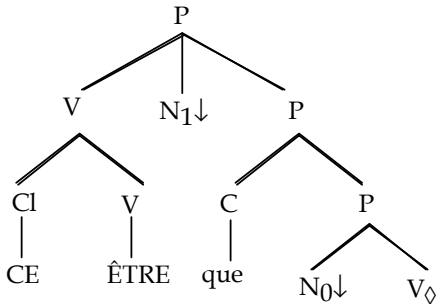


Figure 119. Schème pour une clivée (*C'est N1 que N0 V*)

Seuls les cas où le foyer de clivée est un actant du verbe sont traités actuellement

7.2.2 Les clivées en italien

L'italien dispose comporte également la possibilité de phrases clivées (en italien *frasi scisse*), qui de la même manière qu'en français permettent de mettre en relief un actant, par rapport à un ensemble d'actants pouvant être dans le même rapport au prédicat. Nous avons utilisé la description des clivées en italien de Frison (1988), dont nous reprenons certains exemples ci-dessous.

(389) a. E Giorgio che è arrivato.
Est Giorgio CHE est arrivé
(C'est Georges qui est arrivé.)

b. Sono i tuoi amici che ho incontrato ieri.
Sont tes amis CHE j'ai rencontré hier
(Ce sont tes amis que j'ai rencontrés hier.)

c. E a Giorgio che darò un libro.
Est à Giorgio CHE je-donnerai un livre
(C'est à Georges que je donnerai un livre)

d. E a casa che andrò domani.
Est à maison CHE je-irai demain
(C'est à la maison que j'irai demain)

Comme on peut le voir avec la série (389), la clivée italienne comporte la copule (sans équivalent du pronom français *ce*), qui dans le cas d'un foyer de clivée SN, s'accorde avec ce foyer (que le foyer soit objet du verbe enchâssé, comme en (389b) ou sujet, ou attribut). Dans le cas d'un foyer SP, la copule est au singulier.

Dans le cas d'un foyer de clivée SP, on n'a pas en italien l'alternative entre antéposition de tout le SP ou seulement du SN (cf. l'alternative (387a)/(387b) en français).
Comme pour le français, on peut avoir un foyer de clivée qui est complément de nom (i.e. un SP en *di*) d'un actant direct du verbe (390), ce que nous ne traitons pas dans la MGi. Un foyer de clivée peut être également un ajout, en particulier un adverbe.

(390) E di Liszt che ho sposato la sorella.
(C'est de Liszt che j'ai épousé la soeur.)

L'italien dispose en outre d'un autre type de clivée, dans le cas où le foyer de clivée est sujet de la subordonnée : cette subordonnée peut alors être une infinitive en *a*. Cette construction est plutôt réservée aux foyers de clivée animé. L'infinitif ne peut pas être composé, en effet dans ce cas, c'est la copule qui porte les marques d'aspect (éventuellement accompli). Cette construction n'est pas traitée dans la MGi.

(391) E stato Giorgio a telefonarmi.
Est été Giorgio a me téléphoner.
(C'est Giorgio qui m'a téléphoné.)

De la même manière que pour les relatives, les sujets foyers de clivée admettent un verbe pont (voire plusieurs) entre le foyer et son verbe. En outre, cela est vrai également pour un foyer de clivée qui est sujet de la subordonnée. On donne ci-dessous l'exemple d'un schème pour un intransitif, avec sujet clivé. La copule ESSERE apparaît comme co-tête.

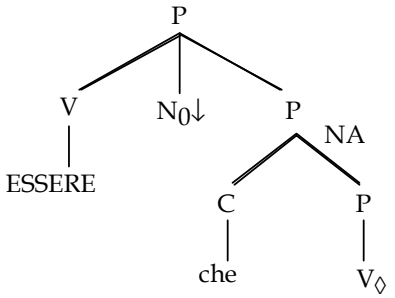


Figure 120. Schème pour une clivée, pour un intransitif (pour dériver (389a))

La représentation des clivées en italien, pour celles traitées dans la MGi, ne pose pas de problèmes particuliers et est parallèle à celle qui est faite pour le français, modulo la différence de structure dans le cas d'un foyer sujet (cf. en français on a une structure plus plate pour bloquer l'adjonction d'un verbe pont). Nous ne détaillons pas les classes ici.

7.3 Les actants grammaticaux "interrogés"

Il s'agit de représenter ici les schèmes pour les questions partielles, avec antéposition (ou extraction) d'un actant grammatical (et non pas interrogatif *in situ*).
On introduit d'abord la représentation des mots interrogatifs dans les TAG sous-jacentes, puis on détaille les classes utilisées pour représenter les actants grammaticaux interrogés.

7.3.1 Pronoms, déterminants, et adverbes interrogatifs

On rappelle que les pronoms interrogatifs en français sont simples (*qui,que,quoi*), composés (LEQUEL) ou renforcés (avec *est-ce que/qui*).
Les déterminants interrogatifs sont QUEL (marqué pour le nombre et le genre) et le déterminant complexe *combien de*. Dans les arbres qui les représentent, ils transmettent le trait wh = + au syntagme nominal.
Les adverbes interrogatifs sont *où, quand, pourquoi, comment* sont représentés par des arbres auxiliaires, modifieurs de phrase, et transmettent le trait wh = +. Les adverbes *où*

et *comment* peuvent cependant représenter des compléments (locatifs, ou complément de manière, comme dans *Comment se porte Jean?*). Ils ancrent donc également des arbres de racine SP (comme cela est général pour les adverbes locatifs et de manière) (Figure 124).
Nous donnons quelques exemples de représentation de ces mots interrogatifs. Ils ancrent des schèmes de manière régulière en fonction de leur catégorie, mais ils ajoutent le trait :wh =+.

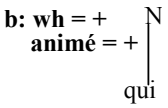


Figure 121

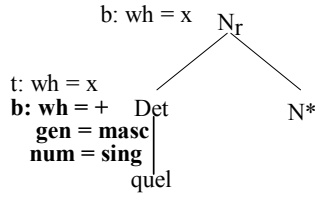


Figure 122

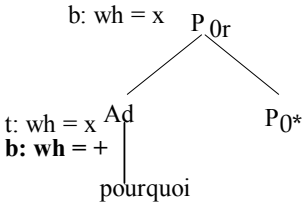


Figure 123

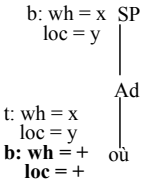


Figure 124

Les formes "renforcées" en *qu/qui est-ce que* (comme dans *qui/qu' est-ce que tu vois?*) sont traitées par adjonction sur ?P de *est-ce que*.
Le comportement particulier de *que/quoi* n'est pas pris en compte actuellement.

L'italien dispose de deux formes simples, *chi* et *che*, respectivement animé+ et animé-.<L'équivalent de LEQUEL est QUALE, marquée pour le nombre, qui ne comporte pas de déterminant (à la différence de l'équivalent de LEQUEL relatif).

(392) a. **Chi/quali** vedi?
(Qui/lesquels vois-tu?)

b. **Che/quale** vuoi?
(Que/lequel vois-tu?)

c. A **chi/quale** pensi?
(A qui/lequel penses-tu?)

d. A **che** pensi?
(A quoi penses-tu?)

Les formes renforcées n'existent pas en italien.
Les déterminants interrogatifs sont *che*, invariable, QUALE (QUEL), variable en nombre, et QUANTO (combien de), variable en genre et nombre. On donne Figure 126 un exemple d'arbre élémentaire pour ces déterminants.

(393) a. Che musica senti?
(Quelle musique écoutes-tu?)

b. In quale città ti hanno trasferito?
(Dans quelle ville t'ont-ils transférés?)

c. Quanta pasta vuoi?
COMBIEN.fs pâte tu-veux
(Combien de pâtes veux-tu?)

Dans le cas animé-, on trouve souvent, en particulier à l'oral, à la place du pronom *che*, le syntagme *che cosa* voire *cosa* tout court. On représente ce dernier cas par un arbre

élémentaire spécifique (Figure 125) ancré par *cosa*, qui porte les traits wh+ et det+ (ce qui bloque l'adjonction d'un déterminant).

(394) a. (che) cosa vuoi?
(CHE) chose tu-vois
(Que vois-tu?)

b. Di (che) cosa ti preoccupi?
De (CHE) chose te préoccupes
(De quoi te préoccupes-tu?)

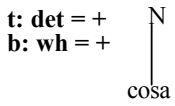


Figure 125

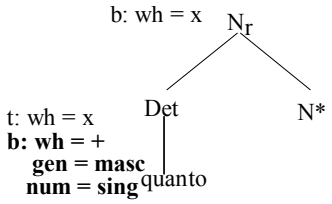


Figure 126

Les adverbes interrogatifs sont *come* (comment)*, dove* (où)*, quando* (quand)*, perché* (pourquoi), et sont traités comme pour le français. *Come* et *Dove* ancrent des arbres initiaux de projection SP.

7.3.2 Les arbres élémentaires pour les actants syntaxiques interrogés en français

On a vu comme propriété du sujet, qu'en français, le sujet extrait se distingue des autres actants en cela qu'il n'admet pas d'extraction à distance, où en tous cas pas selon le schéma général (cf. supra, les séries (277) et (278)).
Pour cette raison, la TAG du français utilise les mêmes arbres que le sujet soit interrogé ou pas (Abeillé, 1991) ce qui bloque l'adjonction de verbe pont entre le sujet et son verbe. Un trait :wh est partagé par le sujet et la phrase, pour distinguer au niveau phrastique *Une femme est venue / Quelle femme est venue?* (Figure 127).
Pour les autres actants interrogés, on définit des arbres spécifiques, avec l'actant antéposé portant le trait wh=+ (Figure 128 et Figure 129), et une projection phrastique supplémentaire, au dessus de ?P. Les dépendances non bornées sont représentées par adjonction au noeud ?P.

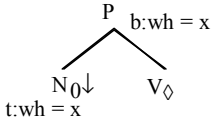


Figure 127. Sujet interrogé ou pas (français)

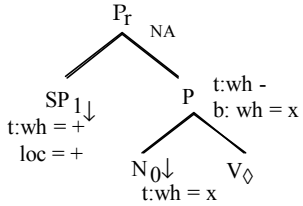


Figure 128. locatif interrogé

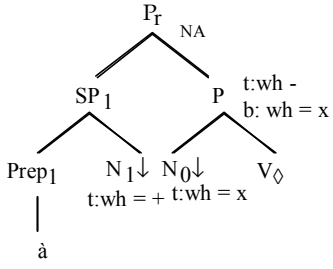


Figure 129. a-objet interrogé

Les schèmes pour les compléments interrogés (antéposés) valent aussi bien pour les interrogatives indirectes que pour les directes. Les verbes à interrogative indirecte reçoivent l'interrogative par substitution (cf. infra Section 6.3.5).
Dans le cas où le sujet n'est pas inversé, l'interrogative directe (395c) peut comprendre un clitique sujet inversé (ce qui est représenté par un trait :clS:dir = -/inv pour la réalisation de ce clitique, cf. supra, la section sur le sujet). Cette possibilité est bloquée dans le cas d'une interrogative indirecte (395d) : le verbe à interrogative impose l'absence de clitique sujet inversé (via le trait :inv, qui est corrélé à la présence d'un sujet nominal inversé, et à la présence d'un sujet clitique inversé (trait :clS:dir)).

(395) a. Quelle voiture a acheté Paul?
b. Je ne sais pas [quelle voiture a acheté Paul].
c. Quelle voiture Paul a-t-il acheté?
d. Je ne sais pas [quelle voiture Paul a acheté].

Dans le cas d'actants phrastiques, l'antéposition est en général mauvaise (396a,c). On préfère une autre tournure, avec extraction de l'élément interrogé hors de l'actant phrastique. Si cet actant est sujet (396c), on utilise l'impersonnel pour pouvoir extraire (396d).

(396) a. ??Où aller propose Paul?
b. Où Paul propose-t-il d'aller?
c. *Quoi faire te plaît en particulier?
d. Qu'est-ce qu'il te plaît de faire?

7.3.3 Les arbres élémentaires pour les actants grammaticaux interrogés en italien

La syntaxe des actants interrogés en italien est relativement plus simple puisque :
- le sujet n'est pas à différencier des autres actants
- il n'y a pas de formes renforcées

Pour le sujet interrogé, on utilise des schèmes avec une position sujet extraite, comprenant un niveau phrastique supplémentaire. Les autres schèmes ne diffèrent pas du français.

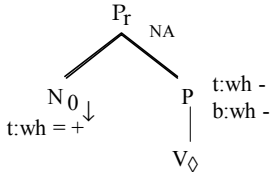


Figure 130. Schème pour l'interrogation du sujet en italien

7.3.4 Représentation dans la MGf et la MGi

On donne Figure 131 le schéma des classes de la MGf pour représenter l'interrogation des actants grammaticaux. La classe racine complément-interrogé hérite d'extraction (on obtient ainsi l'incompatibilité avec une autre extraction). Elle place en position extraite une constante ?quest, de catégorie N ou SP (pour un obl-obj interrogé) mais pas P (cf. on extrait pas un complément phrastique). La constante ?quest est dominée (au sens large) par une constante phrastique supplémentaire ?Pr. Dans le cas d'un SP interrogé, un noeud SP intervient sur le chemin entre ?quest et ?Pr, ce qui est spécifié par l'héritage de la topologie des compléments prépositionnels. Dans le cas d'un objet interrogé (non clitique *que*), le lien de dominance sera traduit par un lien de dominance immédiate lors de la traduction en arbre (étant donné que les deux extrémités ne sont pas unifiables, car de catégories différentes).
Les sous-classes ne font que spécifier la fonction associée à ?quest, ce qui fera unifier cette constante avec l'argument adéquat.
Le cas du clitique *que* est traité à part, puisqu'il ne fait pas intervenir de niveau phrastique supplémentaire (et n'apparaît pas dans la figure).

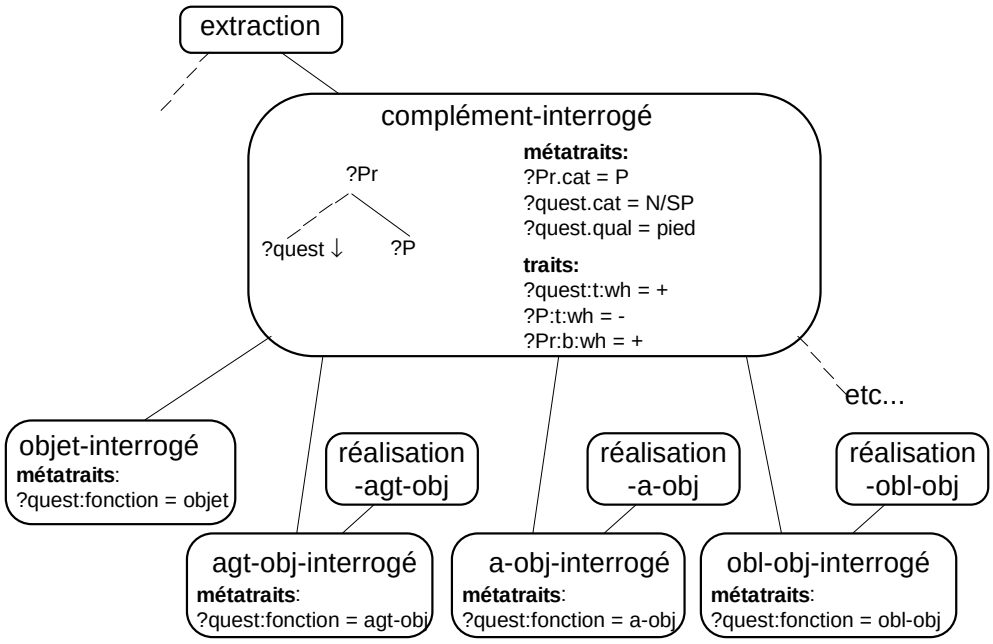


Figure 131. Les compléments interrogés dans la MGf

Pour l'italien, on considère une classe supplémentaire pour le sujet interrogé (et donc la classe générale s'appelle actant-interrogé).

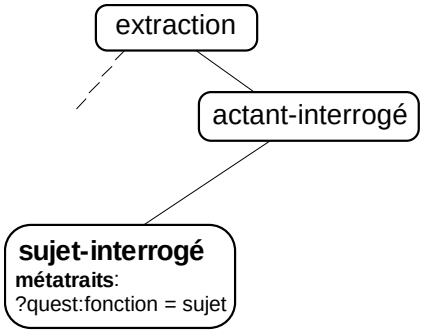


Figure 132. Les compléments interrogés dans la MGi (différences par rapport au français)

Chapitre 6 :
La représentation des modifieurs

Outre la représentation des schèmes à ancre verbale, nous avons également conçu la représentation des schèmes pour les éléments qui sont, soit toujours, soit généralement, des modifieurs syntaxiques. Il peut s'agir d'une préposition, d'un adverbe, d'un complémenteur, d'un adjectif ou d'un nom. Le travail a été fait pour le français principalement, avec une adaptation sommaire pour l'italien. L'analyse fine des distinctions français / italien reste à faire en ce qui concerne les modifieurs. Le travail pour les noms dans leur emploi d'actants sous-catégorisés (non modifieurs) et le contraste pour l'italien reste à faire, ainsi que la représentation des conjonctions de coordination[1].

La particularité d'un modifieur syntaxique, c'est qu'il dépend syntaxiquement d'un élément qu'il gouverne sémantiquement (cf. Chapitre 1, la distinction des arbres élémentaires selon l'interface syntaxe/sémantique en jeu; et pour plus de détails, Chapitre 7).
La représentation en TAG des modifieurs, où le modifieur constitue l'ancre de l'arbre élémentaire, et où l'élément modifié est représenté par un noeud pied, constitue un traitement original, qui permet d'identifier ancre en TAG et foncteur sémantique. Suivant ce principe, la MG considére que la sous-catégorisation d'un lexème est le reflet des arguments sémantiques (plus les actants syntaxiques non référentiels), parmi lesquels un argument peut se réaliser en syntaxe comme le gouverneur syntaxique du lexème (cf. également Chapitre 2).

Par exemple, pour *Paul joue dans le jardin*, on considère que la préposition locative a deux arguments, sémantiques arg0 (le localisé) 'Paul joue' et arg1 (la localisation) 'le jardin', mais en syntaxe, arg0 est réalisé comme le gouverneur syntaxique de *dans*.
Pour représenter cela dans la MGf, nous considérons que la préposition locative a une sous-catégorisation à deux actants (actant0 correspondant à arg0, et actant1 correspondant à arg1), sauf que actant1 porte la fonction "complément-de-préposition" par rapport à la préposition, et la préposition porte la fonction "modifieur" par rapport à l'actant0 (ou inversement, l'actant0 porte la fonction modifieur-inverse par rapport à la préposition).
On donne l'arbre élémentaire ancré par *dans*, à utiliser pour dériver *Paul joue dans le jardin.* Il contient bien deux positions argumentales pour les arguments sémantiques de la préposition, mais l'arbre dans son ensemble est une projection de P0 et pas de la préposition.

[1] C'est-à-dire qu'actuellement les schèmes pour les noms ne sont pas intégrés à la MGf et la MGi.

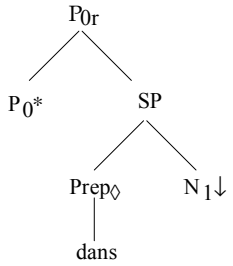


Figure 133. Arbre élémentaire pour *dans*

Pour décrire la représentation des modifieurs, nous suivons l'organisation en trois dimensions. La dimension 1 spécifie les différentes sous-catégorisations initiales retenues (qui doivent correspondre à autant de familles de schèmes), la dimension 2 les redistributions (dont l'utilisation est légèrement détournée par rapport à l'utilisation qui en est faite pour les verbes), et la dimension 3 les différentes réalisations pour les fonctions grammaticales retenues.

1 Les différentes sous-catégorisations initiales distinguées

Les différentes familles de modifieurs sont distinguées selon :
la catégorie de l'ancre (qui apparaît en majuscule dans le nom de famille)
les catégories possibles pour les arguments de l'ancre

Les fonctions initiales distinguées sont :
- modifieur
- complément de préposition
- complément de préposition locatif (spatial ou temporel)
- complément de complémenteur (plein)
- complément d'adjectif
- complément d'adjectif en *de*

Nous avons choisi de n'utiliser qu'une seule fonction modifieur, pour un modifieur de phrase, de verbe, de nom ... Cette fonction est systématiquement portée par l'ancre, et est portée par rapport à l'argument 0 (cf. la convention de numéroter 0 l'argument qui se réalise comme tête syntaxique, s'il existe). La fonction modifieur peut être portée par des ancres de catégories différentes. En (397a) l'adjectif modifie le nom, en (397b,c,d,e), la phrase est modifiée respectivement par une préposition, un complémenteur plein, un nom, un adverbe.

(397) a. ballon **bleu**
b. **Dans** le jardin, Paul joue.
c. **Quand** Paul est arrivé, Pierre est parti.
d. Ce **matin**, Paul est arrivé.
e. **Hier**, Paul est arrivé.

Inversement, chaque catégorie de modifieur ne peut modifier que certaines catégories. Donc la fonction modifieur est systématiquement précisée, en fonction de la catégorie du modifieur, et de la catégorie du modifié. Mais nous utilisons pour cela simplement les métatraits :cat de constantes qui représentent le modifieur et le modifié, sans nommer par des noms de fonctions différents les différents sous-types de modifieur (par exemple modifieur prépositionnel de phrase; modifieur adverbial d'adjectif ...).

Nous donnons ci-dessous la liste des familles (= des sous-catégorisations initiales, sous-classifiées selon la catégorie du lexème pouvant sélectionner la famille et les catégories possibles des arguments).

1.1 Sous-catégorisations initiales pour les prépositions

Les différentes familles de modifieur à ancre prépositionnelle (il s'agit des prépositions pleines) sont distinguées selon la catégorie du complément de la préposition : il peut s'agir d'un SN (398a), d'une complétive (398b), d'une infinitive. Dans le cas d'une infinitive, il faut distinguer le cas (rare) sans complémenteur (398c), et celui avec complémenteur *de* (398d). Les deux cas sont déterminés lexicalement par la préposition. Le complément de la préposition peut également être un syntagme locatif (au sens large de locatif temporel ou locatif spatial), dont la catégorie peut varier (cf. (398e), (398f), où l'on a un SP, un adverbe, un SN comme complément de préposition).

(398) a. **A** la campagne, Paul dort bien.
b. **Après** que Paul est arrivé, Pierre est parti.
c. **Après** être arrivé, Paul a pris un bain.
d. **Avant** de partir, Paul a pris un bain.
e. **De** chez nous / là / la maison, on aperçoit la tour Eiffel.
f. **Depuis** longtemps / avant la guerre / l'armistice.

L'argument modifié peut être phrastique (comme dans tous les exemples de la série (398)), et il peut alors systématiquement être verbal également (399).

(399) Paul prononce **avec** emphase un discours émouvant.

L'argument modifié peut être également nominal (400), auquel cas, on ne considère que le cas où le complément de la préposition est nominal.

(400) Les bottes **avec** des éperons sont interdites.

Les autre cas de syntagmes prépositionnels sont considérés comme des cas de complémentation (cf. un SP peut être complément d'un adjectif par exemple).

Nous pouvons donc donner la liste suivante de familles à ancre prépositionnelle, avec la sous-catégorisation initiale correspondante (exprimée en utilisant la notation modifieur-1, pour la fonction modifieur inverse) :

s0Pn1 • Préposition + SN,
• Modifieur de phrase ou de verbe (exemple *dans*)
Sous-cat. initiale : (modifieur-1[P/V], complt-de-prep[N])

s0Pcs1 • Préposition + *que* Pind/subj ou *de* Pinf,
• Modifieur de phrase ou de verbe (exemple *avant*)
Sous-cat. initiale : (modifieur-1[P/V], complt-de-prep[Psubj/ind/*de*inf])

s0Ps1 • Préposition + *que* Pind/subj ou ∅ Pinf,
• Modifieur de phrase ou de verbe (exemple *après*)
Sous-cat. initiale : (modifieur-1[P/V], complt-de-prep[Psubj/ind/*de*inf])

s0Ppp1 • Préposition + locatif,
• Modifieur de phrase ou de verbe (exemple *de, jusque, depuis*)
Sous-cat. initiale : (modifieur-1[P/V], complt-de-prep-loc)

n0Ppn1 • Préposition + SN,
• Modifieur de N (exemple *avec*)
Sous-cat. initiale : (modifieur-1[N] , complt-de-prep[N])

Nous verrons à la Section 3 les différentes réalisations pour les fonctions intervenant dans ces sous-catégorisations.

1.2 Sous-catégorisations initiales pour les complémenteurs (pleins)

Nous avons mentionné au Chapitre 5, que nous considérons les traditionnelles conjonctions de subordination comme des complémenteurs, dans le cas où l'on ne peut les décomposer en préposition + complémenteur *que* en syntaxe[2] (cf. supra (398b) où l'on sépare la préposition *après* du complémenteur *que*). Il s'agit donc par exemple de *quand, si, lorsque* .... L'argument modifié est toujours une phrase, et le complément du complémenteur est une phrase finie simple (sans complémenteur régi).

(401) Pierre est parti **quand** Paul est arrivé.

Nous donnons ci-dessous un exemple de schème qui peut être ancré par un complémenteur, et qui est dans son emploi de post-modifieur (comme en (401)).

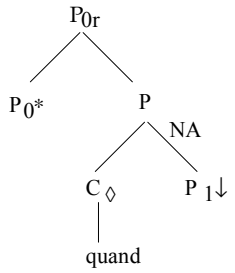


Figure 134. Arbre élémentaire pour dériver (401)

On a donc la famille supplémentaire :

s0Cs1 • Complémenteur plein + Pind/subj

[2] Il s'agit également des cas comme *bien que* où *que* apparaît séparé, mais où l'autre partie n'est pas utilisée indépendamment. On considère alors un mot composé *bien-que*.

• Modifieur de phrase ou de verbe (exemple *quand*)
Sous-catégorisation initiale : (modifieur-1[P/V], complt-de-complémenteur)

1.3 Sous-catégorisations initiales pour les adverbes

Un adverbe peut être modifieur de phrase, de verbe, d'adjectif, d'adverbe, et même pour certains de noms et de SP.
Au niveau sémantique, il s'agit d'un foncteur qui prend un argument. Dans le détail, l'adverbe montre souvent une interface syntaxe/sémantique complexe. On peut citer les adverbes "orientés agent", qui modifie syntaxiquement une phrase ou un verbe, mais sont liés sémantiquement au sujet de la phrase (comme dans *Paul a gentilment cédé sa place*), ou les adverbes énonciatifs, qui modifie en syntaxe une phrase, mais qui en sémantique caractérisent l'acte d'énonciation (comme dans *Sincèrement, Paul n'aurait pas dû faire ça*[3]).

Nous opérons la classification des adverbes suivante (partiellement reprise de Abeillé et Godard (1997) et les références citées) :

- Adverbes qui modifient un adjectif ou un adverbe (comme *très, trop, si, aussi*), ou un adjectif seulement (comme *intellectuellement*).
  D'où la famille **a0Ad**, dont la sous-catégorisation initiale est (modifieur-1[A/Ad])
- Adverbes (peu nombreux) qui modifient un nom, un SP ou un adverbes (comme *juste, même, aussi, seulement*).

(402) a. J'ai lu même ce livre.
b. Lisez-le même rapidement.
c. Ils étaient là même avant le match.

D'où la famille **x0Ad**, dont la sous-catégorisation initiale est (modifieur-1[N/Ad/SP])

Parmi les adverbes pouvant modifier un verbe ou une phrase, on peut distinguer ceux qui sont plutôt orienté phrase, et ceux plutôt orientés verbe (mais l'on ne peut pas formellement restreindre la catégorie de l'élément modifié à seulement P ou seulement V).
Le test pour distinguer les deux types d'adverbe est la possibilité pour l'adverbe d'être modifieur de phrase antéposé à la phrase. Si cela est possible on parlera d'adverbe de phrase, sinon d'adverbe de verbe.

- Les adverbes de verbe sont par exemple *bien, mal, trop, beaucoup, peu*, qui effectivement ne peuvent pas être antéposés à la phrase. On ne peut cependant pas exclure qu'il modifie une phrase (et soient alors postposés). Cela tient en partie au fait que l'on ne dispose pas d'un niveau SV entre le verbe et la phrase.
  D'où la famille **v0Ad**, dont la sous-catégorisation initiale est (modifieur-1[V/P], :type=verbe). On ajoute le trait :type=verbe, pour distinguer cette sous-catégorisation initiale de la suivante, utilisée pour les adverbes de phrase.

[3] Il s'agit d'un cas d'ellipse d'un verbe énonciatif. Dans *Sincèrement, je crois que Paul n'aurait pas dû faire ça*, l'adverbe modifie bien en syntaxe une phrase dont le prédicat ('croire') est précisé sémantiquement par l'adverbe.

- Les adverbes de phrase regroupent les modaux (*peut-être, probablement*), les locatifs (temporels comme *demain* ou spatiaux comme *là, ici*), les adverbes de manière.

(403) a. Demain/ Probablement, Paul partira.
b. Ici, Jaurès a été assassiné.
c. Habilement, Paul a répondu à la question sans y répondre.

D'où la famille **s0Ad**, dont la sous-catégorisation initiale est (modifieur-1[V/P], :type=phrase).

Pour les adverbes négatifs (*pas, jamais*...), nous utilisons une famille spécifique (v0Ad-neg), comprenant simplement deux schèmes, pour l'emploi post-verbal, après un verbe fini, et pour l'emploi préverbal devant un infinitif.
Avant de passer aux autres catégories de modifieurs, nous voyons brièvement la gestion du placement des adverbes, lorsqu'ils s'adjoignent à un noeud verbal.

1.3.1 Placement des adverbes

Les adverbes peuvent s'adjoindre aussi bien sur ?Vanc que sur ?Vmarq. Avec un traitement des clitiques en morphologie, on n'a pas besoin de bloquer par des traits l'éventuelle adjonction d'adverbes entre les clitiques et leur hôte, ou entre le clitique sujet et les clitiques compléments.
Dans le cas d'un complexe causatif, on admet aussi bien un adverbe entre le verbe causatif et l'infinitif, qu'après l'infinitif. Pour ce deuxième cas, l'adjonction se fait sur l'infinitif.

(404) a. Paul fait souvent sortir ses élèves.
b. Paul fait sortir souvent ses élèves.

Nous ne détaillons pas l'analyse précise du placement des adverbes, qui nécessite encore du travail dans la MGf et la MGi.
On a mentionné que la particule négative *ne* devait recevoir deux traitements, un en morphologie, devant un verbe fini, et un en syntaxe, dans le contexte d'un verbe infinitif (Recourcé, 1995). L'analyse précise reste à fournir.

Pour l'italien, nous renvoyons également à un travail futur l'étude du placement de l'adverbe *non*. Du point de vue sémantique, *non* a un emploi sémantiquement "vide", lorsqu'il accompagne un forclusif (406b), mais il a également un emploi négatif[4](406a), (405), généralisé du fait de l'absence du forclusif *pas* en italien. A noter que dans le cas d'un forclusif nominal (*nessuno (personne)*, *niente (rien)*), l'ordre verbe/forclusif influe sur la présence du *non* (cf. le contraste (406a)/(406b).

(405) Non vengo.
Je ne viens pas.

(406) a. Nessuno è venuto
Personne est venu
(Personne n'est venu)

b. Non è venuto nessuno
NEG est venu personne.

[4] En français *ne* a encore quelques emplois comme négatif, comme dans *Je ne peux tolérer cela.*

(Il n'est venu personne)

1.4 Sous-catégorisations initiales pour les adjectifs

Pour les adjectifs, nous considérons comme sous-catégorisation initiale le cas où ils sont employés comme épithète, et sont donc modifieurs d'un de leurs arguments sémantiques (qui peut être le seul argument)[5].
L'argument modifié est nominal. L'adjectif peut en outre avoir un complément avec préposition régie, nominal (407) ou phrastique. Dans le cas phrastique, on distingue le cas où le complément est un de-obj des autres cas. Le de-obj est réalisé comme une infinitive avec préposition *de* (408a), ou comme une complétive directe (408b), qui peut alterner avec une tournure en *de ce que* (408c). Le de-obj est pronominalisable en *en* (408d). On retrouve donc les caractéristiques de la réalisation phrastique de la fonction de-obj pour les verbes (cf. Chapitre 5). Nous classons ensemble les autres cas de compléments phrastiques (principalement infinitivaux), comme par exemple (409). Dans le cas nominal, on ne distingue pas non plus plusieurs fonctions, mais il s'agit d'une approximation, qui tient au fait que l'implémentation actuelle ne couvre pas la pronominalisation du complément (comme dans *Paul y est fidèle*).

(407) Paul est fidèle à ses convictions.

(408) a. Paul est heureux de partir.
b. Paul est heureux que Marco vienne.
c. Paul est heureux de ce que Marco vienne.
d. Paul en est heureux.

(409) Paul est lent à se mettre au travail.

L'implémentation actuelle comprend donc les familles :

n0A	Adjectif, sans complément
n0A(pn1)	Adjectif, avec complément nominal Sous-catégorisation initiale : modifieur-1[N], complément d'adj.[N]
n0A(ps1)	Adjectif, avec complément infinitival Sous-catégorisation initiale : modifieur-1[N], complément d'adj.[Pinf]
n0A(ds1)	Adjectif, avec de-objet Sous-catégorisation initiale : modifieur-1[N], de-objet[Pinf/Psubj]

Pour les familles n0A(pn1) et n0A(ps1), la préposition n'est pas précisée dans les schèmes, mais doit l'être dans le lexique syntaxique.

[5] Nous avons insisté au Chapitre 2, Section 1.6.2, que nous ne considérons pas l'argument modifié comme lélément qui porte une fonction par rapport à lancre (en l'occurrence il s'agirait de considérer le nom comme sujet de l'adjectif). Nous gardons la définition traditionnelle de fonction grammaticale, définition syntaxique, qui permet de rendre bien compte du fait qu'un élément a un seul gouverneur syntaxique (et ne peut porter qu'une seule fonction).
On peut cependant noter que nous avons fait une entorse à ce principe, lorsque nous avons pris en compte pour les verbes, les emplois de modifieurs (comme les participiales ou les relatives).

La représentation des adjectifs est incomplète. En particulier, il faut ajouter la représentation des adjectifs comme *difficile* (pour les tours en *Adj à Vinf*).

1.5 Sous-catégorisations initiales pour les noms (emplois de modifieur uniquement)

Nous considérons les emplois de nom comme modifieur de phrase ou de verbe, comme en (410).

(410) a. **Ce matin**, Paul s'est levé de bonne heure.
b. Marie avait terminé **la veille** les papardelle à la sauce du chasseur.

Du point de vue sémantique, il faut considérer que c'est la syntaxe qui exprime le prédicat sémantique "est-localisé-dans-le-temps", et non pas un élément lexical. L'argument modifié est sémantiquement l'élément localisé temporellement. Le modifieur constitue un argument de "est-localisé-dans-le-temps", i.e. la localisation spatiale.

Il faut considérer la famille suivante :

s0N • Nom, modifieur de phrase ou verbe
Sous-catégorisation initiale : (modifieur-1[P/V])

2 Les redistributions pour les modifieurs

Jusqu'à présent, les seuls cas de redistributions que nous avons cités concernaient des fonctions sous-catégorisées par un verbe. Effectivement la notion de redistribution est surtout pertinente pour les verbes. On peut tout de même citer les cas des noms prédicatifs, pour lesquels on peut considérer un passif. Par exemple (411b) serait la varainte passive de (411a). L'intégration des familles de schèmes à ancre nominale (autre que modifieur) reste cependant à faire, et nous ne pouvons donc la décrire ici.

(411) a. La déclaration des droits de la personne (par + *de) Amnesty International
b. La déclaration d'Amnesty International

Pour l'adjectif, on peut également considérer le cas de l'attribut de l'objet, comme une variante causative du cas avec attribut du sujet. Ainsi (412b) comporte un argument causateur supplémentaire par rapport à (412a). La différence avec le causatif vu pour les verbes est qu'ici le verbe RENDRE est lexicalement causatif, et n'est pas analytiquement formé à partir d'un verbe à attribut du sujet. Nous ne traitons donc pas ces cas par redistribution.

(412) a. Paul semble fou.
b. Marie a rendu Paul fou.

Pour les modifieurs, nous utilisons la notion de redistribution de façon un peu détournée : nous avons défini des "redistributions" pour spécifier la catégorie de l'argument modifié, par un élément portant la fonction modifieur. Cela revient à préciser la fonction modifieur comme modifieur de phrase, modifieur de nom, modifieur de verbe ...

En effet nous avons vu Section 1, que certaines sous-catégorisations initiales laissent plusieurs catégories possibles pour l'argument modifié. Or cet argument n'apparaît pas en dimension 3, dans la mesure où il ne porte pas de fonction (c'est le modifieur qui porte la fonction, et en dimension 3, on fait varier la catégorie du modifieur). Distinguer les différentes catégories pour l'argument modifié en dimension 3 est une solution possible, mais lourdes : il faudrait gérer manuellement toutes les possibilités de couples (catégorie argument modifié/catégorie modifieur). Pour cette raison nous avons choisi de représenter la catégorie de l'argument modifié en dimension 2. Le compilateur MG→TAG opère tous ces croisements, que l'on peut contrôler en ajoutant les contraintes sur les catégories sous la forme de méta-traits appropriés.

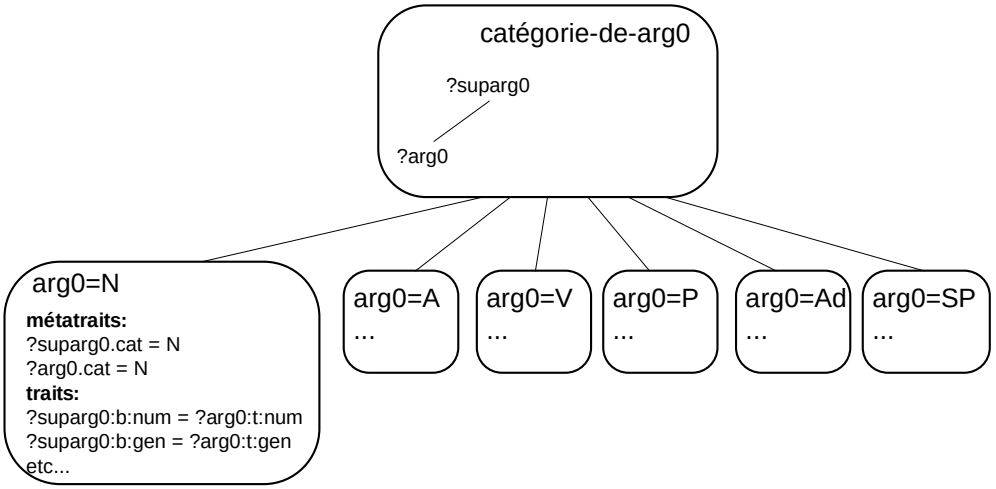


Figure 135. Classes de la dimension 2, pour les modifieurs : spécification de la catégorie de l'argument modifié

Nous avons également considéré comme une redistribution le cas de non réalisation locale de l'argument 0. Par exemple, pour les prépositions, nous avons vu dans les chapitres sur les schèmes à ancre verbale la possibilité pour un verbe de sous-catégoriser un SP locatif. Ainsi en (413a), on considère la préposition comme modifieur de la phrase *Paul joue*, mais en (413b), le SP est complément du verbe. L'argument 0 de la préposition (le "localisé") apparaît comme sujet du verbe de mouvement : il s'agit d'un cas de contrôle exactement comparable au cas *Paul veut partir*. L'argument 0 de la préposition est partagé sémantiquement avec le verbe, et en syntaxe il est réalisé comme sujet du verbe.

(413) a. Paul joue dans le jardin.
b. Paul va dans le jardin.

Comme pour le cas du contrôle du sujet d'un infinitif, nous considérons un arbre élémentaire pour la préposition, avec une trace pour son argument zéro. La préposition projette un SP qui peut être utilisé comme complément (en étant substitué dans l'arbre pour un verbe à complément locatif).

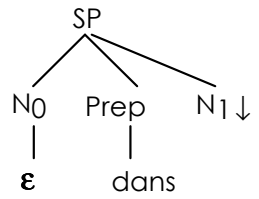


Figure 136. Arbre élémentaire pour une préposition locative, avec argument 0 contrôlé

Mais à la différence du cas de contrôle du sujet d'un infinitif, nous traitons ce cas par redistribution. En effet, en (413a) la préposition porte la fonction modifieur par rapport à la phrase. En (413b), elle n'est plus modifieur de la réalisation de l'argument 0 (*Paul*), mais projette une structure (un SP) dont elle est la tête syntaxique. Nous représentons cela par la redistribution modifieur-1→vide(contrôlé). Il s'agit d'un cas particulier puisque le modifieur perd sa fonction, sans en acquérir une localement (i.e. au sein de son arbre élémentaire) une autre.

On a le même cas de figure, pour un adverbe locatif ou de manière, qui peut lui aussi servir de complément pour un verbe (comme dans *Paul revient ici*, *Paul va bien*). On a également le même cas pour un adjectif employé comme attribut (cf. Chapitre 3, la MGf et la MGi comprennent des familles verbales avec attribut du sujet et de l'objet).

Avec l'implémentation actuelle, pour dériver *Paul est beau*, on utilise l'arbre élémentaire de la Figure 137 pour l'adjectif.

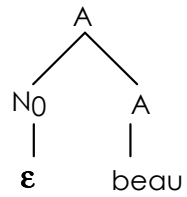


Figure 137. Arbre élémentaire pour un adjectif avec argument 0 contrôlé

Enfin nous utilisons une redistribution pour représenter le cas d'ellipse de l'argument 0, ce qui revient à employer le modifieur comme tête d'une structure normalement projetée par l'argument 0. On a le cas d'ellipse du nom devant l'adjectif (414a), et l'adjectif projette alors un SN. On peut également considérer le cas d'ellipse d'une phrase, pour un emploi de complémenteur plein. Il s'agit par exemple de phrase comme (414b), où c'est l'on considère le complémenteur plein *pourvu que* qui est considéré comme la tête.

(414) a. Le rouge est à moi.
b. Pourvu que Paul arrive à temps!

Nous utilisons pour ces cas une redistribution modifieur-1→vide(ellipse).

3 Les réalisations des fonctions modifieur et dépendants de modifieurs

Nous passons en revue les différentes réalisations possibles pour le modifieur, en fonction de sa catégorie et de la catégorie de l'argument modifié. Puis les différentes réalisations pour les fonctions sous-catégorisables par un modifieur (complément de préposition, complément d'adjectif ...).

3.1 Le modifieur

Le modifieur peut être antéposé à l'argument qu'il modifie, ou bien postposé, mais cela n'est possible que pour un modifieur de catégorie P, N, A, ou V (les modifieurs de SP ou d'adverbe sont antéposés). Dans les deux cas, l'argument 0 (l'argument modifié) eset représenté par un noeud pied (représenté dans la MGf par la constante ?arg0), et est dominé immédiatement par un noeud de même catégorie (représenté par la constante ?suparg0).
Les schèmes pour un modifieur antéposé de phrase valent systématiquement pour un modifieur "assertif" et pour un modifieur portant le trait :wh = +, i.e. comportant un interrogatif, et ce de manière régulière pour toutes les catégories de modifieurs pouvant modifier une phrase (modifieur adverbial (415a); modifieur prépositionnel (415b); modifieur nominal (415c)).

(415) a. Quand as-tu dormi?
b. Pour quelle raison as-tu dormi?
c. Quel jour es-tu arrivé?

On représente cela par un partage du trait :wh entre le modifieur et la projection de l'argument modifié.

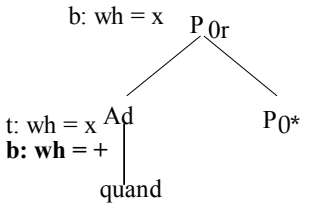


Figure 138. Arbre élémentaire ancré par un adverbe interrogatif (partage du trait :wh)

En pratique on doit distinguer plusieurs classes pour le cas de modifieur antéposé, suivant la catégorie de l'argument modifié, et parfois selon la catégorie de l'ancre, de manière à pouvoir ajouter les traits appropriés. En particulier un modifieur adverbial, de type "adverbe de verbe" (cf. supra, la famille v0Ad), ne peut être antéposé au verbe que si celui-ci est infinitif ou participial (cf. *Paul a bien dormi, Paul veut bien dormir*).
Nous ne détaillons pas tous les différents sous-cas d'antéposition d'un modifieur.

Les autres types de réalisation d'un modifieur sont comme syntagme relatif relatif, ou comme foyer de clivée, valant pour des modifieurs de phrase uniquement.

On peut avoir un foyer de clivée modifieur, pour un modifieur adverbial (416a), prépositionnel (que la préposition ait comme complément un SN (416b), une infinitive (416c), ou une complétive (416d)), complémenteur plein (416e).

(416) a. C'est demain que Paul part.
b. C'est pour Marie qu'il a fait ça.
c. C'est après être parti qu'il a téléphoné.
d. C'est après que Marie soit partie qu'il a téléphoné.
e. C'est quand Paul est parti que Marie a téléphoné.

Nous donnons Figure 139 l'exemple du schème pouvant être ancré par *pour* pour dériver (416b).

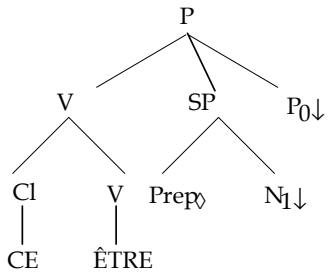


Figure 139. Schème pour un adverbe foyer de clivée

Dans le cas d'un adverbe, seul un sous-ensemble des adverbes de phrase (sélectionnant la famille s0Ad) peuvent être foyer de clivée. Les adverbes modaux (comme *peut-être*) ne l'admettent pas. La sélection du schème pour la clivée est gérée par un trait spécifique (:cq).
Pour représenter les modifieurs foyer de clivée dans la MGf, nous utilisons la classe complément-clivé, superclasse introduite pour les clivées avec foyer complément de verbe (cf. Chapitre 5) (il s'agir de complément par opposition à sujet clivé, cf. les structures pour l'extraction du sujet sont différentes des autres, pour le français).

Un modifieur prépositionnel, avec complément nominal (i.e. une préposition sélectionnant la famille s0Pn1), peut constituer un syntagme relatif (417a). On peut également avoir un adverbe relatif, il s'agit en français de *où* (417b,c).

(417) a. La maison dans laquelle Paul dort
b. Là où Paul dort.
c. Le lit où Paul dort

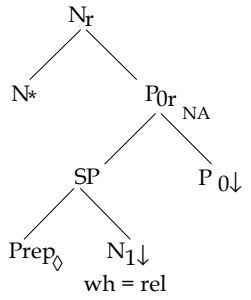


Figure 140. Schème pour une préposition, comme tête d'un syntagme relatif (pour dériver (417a))

Pour représenter ce schème, ainsi que les schèmes pouvant être ancrés par *où*, pour les relatives illustrées en (417b,c), nous définissons des classes qui héritent de classes définies pour les relatives au Chapitre 5 (sp-relativisé et complément-relativisé).

3.2 Les fonctions dépendant d'un modifieur

Les réalisations des fonctions dépendant d'un modifieur sont très simples, puique dans tous les cas, il y a une réalisation possible (étant donnée que l'on ne traite pas l'extraction du complément d'un adjectif, ou sa pronominalisation).
Par exemple, pour une préposition, le complément de la préposition est toujours localisé à la droite de la préposition et dominé par un SP. On utilise cepedant plusieurs classes selon la catégorie du complément (SN, infinitive sans complémenteur, infinitive avec complémenteur *de*, complétive).

Ceci conclut notre description des MGf et MGi. Comme nous l'avons dit au début de ce chapitre, l'adaptation de la représentation des modifieurs à l'italien n'a été faite que sommairement.
Plus généralement, le travail reste à faire pour les familles à ancre nominale, et les phénomènes de coordination, qui n'ont pas du tout été abordés ici, sont à représenter également au sein de la MGf et la MGi.

Partie III

Après avoir décrit la métagrammaire et son instanciation pour le français et l'italien, nous développons dans cette partie différents prolongements.
Au Chapitre 7, nous étudions l'interface syntaxe/sémantique telle qu'elle est réalisée en TAG, en prenant en compte les principes linguistiques que nous avons retenu pour les arbres élémentaires d'une TAG. Cela nous amène à étudier une procédure d'interprétation de l'arbre de dérivation comme un graphe de dépendances sémantique.
Au Chapitre 8, nous décrivons les applications de la MGf et la MGi, et donnons quelques pistes pour des applications envisageables.

Chapitre 7
TAG et l'interface sémantique/syntaxe

Nous avons introduit au Chapitre 1 un principe de bonne formation des arbres élémentaires en termes de "complétude" sur la réalisation des arguments sémantiques d'un lexème, comme positions à étendre au sein d'un arbre élémentaire ancré par ce lexème. Ainsi, les arbres élémentaires d'une grammaire TAG réalisent-ils l'interface syntaxe/sémantique lexicale, en cela qu'ils doivent fournir la description de la réalisation syntaxique d'argument sémantique.

Nous avons cité au Chapitre 1 les deux approches principales pour ajouter un calcul sémantique à une analyse TAG : il s'agit soit d'utiliser des TAG synchrones, soit de fonder l'analyse sémantique sur la structure qui encode la dérivation en TAG — l'arbre de dérivation.
Notre but dans ce chapitre est d'étudier la seconde voie. Nous entendons étudier en profondeur le statut linguistique des arbres élémentaires, ainsi que celui de l'arbre de dérivation, qui en découle. Nous allons voir que parce qu'une grammaire TAG est lexicalisée, et étant donnés les principes de minimalité sémantique et de cooccurrence prédicat-arguments, l'arbre de dérivation peut être interprété comme la représentation de liens entre des unités sémantiques et leurs arguments, i.e. des dépendances sémantiques.
L'idée n'est pas nouvelle. L'arbre de dérivation est souvent présenté dans la littérature TAG comme présentant des propriétés intéressantes pour l'analyse sémantique (cf. par exemple Abeillé, 1993, Shieber et Schabes, 1994). Une étude prenant en compte les principes de bonne formation (PMS et PCPA) est cependant manquante, et nous proposons de la fournir. Nous précisons qu'il s'agit d'un travail initié conjointement avec Sylvain Kahane (Candito et Kahane, 1998a, 1998b).
Considérer l'arbre de dérivation comme base d'un calcul sémantique n'est pas suffisant en soi, si l'on veut prendre en compte des phénomènes typiques d'interface syntaxe/sémantique tels que la portée des quantificateurs. C'est une des motivations pour la définition des TAG synchrones. Cependant, nous pensons que d'une part il est important de connaître le statut exact de l'arbre de dérivation en TAG et qu'il est dommage de ne pas en tirer parti. Une procédure d'interprétation des arcs de l'arbre de dérivation comme dépendances sémantiques (telle que nous la décrivons Section 2.1) peut être appliquée avec une ambition plus modeste que l'analyse sémantique. Par exemple on peut simplement collecter des données de fréquences d'apparition de dépendances lexicalisées (cf. Section 4), ce qui peut être utilisé pour la désambiguisation syntaxique par exemple.

D'autre part, les TAG synchrones ont le désavantage d'augmenter la capacité générative du modèle, à moins de se restreindre à des dérivations isomorphes entre la TAG syntaxique et la TAG sémantique (cf. Shieber, 1994)[1], ce qui diminue fortement l'attrait des TAG synchrones, étant donné que la plupart des exemples d'utilisation utilise des dérivation non synchrones (qu'il s'agisse de l'utilisation des TAG synchrones pour coupler TAG syntaxique et sémantique (cf. Shieber et Schabes, 1990; Abeillé, 1994b) ou pour coupler deux TAG syntaxiques pour la traduction automatique (Abeillé et al., 1990)).

Nous nous intéressons au rapport entre dépendances sémantiques et dépendances syntaxiques (et nous utilisons pour cela les travaux de Mel'cuk au sein de la Théorie Sens-Texte), ce qui ne constitue qu'une partie des questions liées à l'interface syntaxe/sémantique.
Nous détaillons dans une première section les notions de dépendances sémantiques et syntaxiques, afin de pouvoir caractériser les différents types d'arbres élémentaires, selon le rapport sémantique/syntaxe qu'ils mettent en jeu, et d'étudier les critères qui déterminent le typage des arbres élémentaires en arbres initiaux et arbres auxiliaires.
Nous fournissons ensuite une procédure d'interprétation de l'arbre de dérivation comme un graphe de dépendances sémantiques. Puis nous étudions certains cas problématiques. Enfin, nous citons des applications possibles, existantes ou envisagées de cette lecture de l'arbre de dérivation.

1 Dépendances et têtes

Les grammaires syntagmatiques contemporaines ont toutes introduit la notion de tête syntaxique. La théorie X-barre est intégrée à GB; la notion de tête syntaxique est un pas décisif pour l'émergence de GPSG, puis de HPSG, et la notion de tête est comprise dans la notion de fonction grammaticale, utilisée par LFG. Or la notion de tête syntaxique, est proche de celle de dépendance syntaxique. Ainsi la plupart des grammaires syntagmatiques utilisent à la fois la notion de constituance et celle de dépendance.
Elles se rapprochent en cela des théories linguistiques qui n'utilisent que la notion de dépendance. Les grammaires de dépendances, depuis Tesnière (1959), utilisent la notion de dépendance syntaxique, comme une relation entre un élément lexical et un autre, pour représenter la structure syntaxique d'un énoncé.
La Théorie Sens-Texte (ci-après TST) (Mel'cuk, 1988), utilise la notion de dépendance seulement, et dépasse le cadre des travaux de Tesnière, en cela qu'elle comprend la formalisation nécessaire aux représentations non seulement syntaxique mais également morphologique et sémantique. La TST distingue en particulier trois types de dépendances, les dépendances sémantiques, les dépendances syntaxiques[2] et les dépendances morphologiques.

[1] Rambow et Satta (1996) proposent une solution alternative de système de dérivations synchrones, qui permet de traiter en particulier la "montée" des quantificateurs, en conservant la capacité faible du modèle utilisé de manière synchrone. Cependant, cela ne permet pas de résoudre tous les cas de divergences syntaxe/sémantique qui nécessitent des dérivations non isomorphes.
[2] Plus précisément sont distinguées des dépendances syntaxiques profondes et des dépendances syntaxiques de surface. Les premières sont des abstractions sur les secondes, c'est-à-dire qu'une dépendance syntaxique profonde peut être réalisée par différentes dépendances syntaxiques de surface. Nous ne pensons pas que cette

Etant donné que l'arbre de dérivation relie (par des arcs de substitution ou des arcs d'adjonction) des arbres élémentaires lexicalisés, il peut être considéré comme une structure de relations entre items lexicaux, et donc une structure de dépendances (les grammaires de dépendances ont la particularité de considérer des relations entre items lexicaux). Nous allons donc nous intéresser particulièrement aux différentes notions de dépendances sémantiques et syntaxiques, cette deuxième notion, moins directement perceptible, étant sujette à de nombreuses controverses. Comme nous nous plaçons cependant dans le cadre d'une grammaire syntagmatique (TAG), nous intégrons également la notion de tête syntaxique d'un constituant.

1.1 Dépendances sémantiques

Une **dépendance sémantique** relie de manière orientée un sémantème à un de ses arguments (sémantiques[3]). On dira que le sémantème en question est un **gouverneur sémantique** de l'argument, et l'argument est un **dépendant sémantique** du sémantème. Dans un énoncé, un même élément peut être gouverné sémantiquement par plusieurs gouverneurs sémantiques. Par exemple dans *Je vois un petit chat noir*. Le sémantème 'chat' est gouverné sémantiquement pas 'voir', 'petit' et 'noir'. La représentation des dépendances sémantiques d'un énoncé n'est donc généralement pas un arbre.

Nous précisons que nous suivons la Théorie Sens-Texte en prenant le terme argument sémantique au sens large. Un sémantème est considéré comme ayant des arguments sémantiques dés lors qu'il fonctionne comme un foncteur sur ces arguments, au niveau logique. Une dépendance sémantique est établie entre une unité sémantique incomplète (en ce sens qu'elle doit être complétée d'arguments) et la ou les expressions qui la rendent complète sémantiquement. A ce niveau, on met sur le même plan des relations entre ce qui est généralement appelé un prédicat et ses arguments, et des relations par exemple entre une unité exprimant le temps ou l'aspect et son argument. Un prédicat exprime un état ou un événement, des marques de temps ou d'aspect constituent des informations supplémentaires sur un état ou un événement. Si cette distinction est cruciale pour l'interprétation sémantique, il reste que dans les deux cas, le sémantème (prédicat ou marque d'information) "attend" un argument au niveau logique. Une marque de temps s'applique à un état ou événement, comme un prédicat unaire (exprimé par un intransitif par exemple) s'applique à un argument sujet.
Nous excluons cependant du domaine des dépendances sémantiques la quantification. La relation syntaxique entre un quantificateur et l'élément quantifié a une portée en sémantique qui peut être large, et le simple outil des dépendances sémantiques entre sémantèmes ne nous semble pas approprié pour représenter cela (cf. Polguère, 1990 pour une étude dans le cadre de la TST).

---

distinction soit nécessaire ici. Ce qui nous intéresse, c'est la distinction entre dépendance sémantique et dépendance syntaxique.

[3] Nous utilisons le terme argument pour désigner un argument sémantique, et le terme actant pour désigner un actant grammatical, qu'il s'agisse de la réalisation grammaticale d'un argument ou pas (cf. Chapitre 2).

1.2 Dépendances syntaxiques et tête syntaxique

Nous définissons maintenant la notion de **dépendance syntaxique**, en nous inspirant directement de la définition de Mel'cuk (1988, à paraître). Une dépendance syntaxique orientée entre f1 et f2 est établie, de telle sorte que f1 est *le* **gouverneur syntaxique** de f2, et f2 est un **dépendant syntaxique** de f1, si :

(A1) les constructions syntaxiques dans lesquelles peuvent se trouver le couple f1 f2, sont déterminées pour la plupart par la présence de f1 (définition librement adaptée et simplifiée de la notion de valence passive de Mel'cuk, 1988:113)

(A2) dans le cas où le critère A1 est inopérant, ce qui est systématiquement le cas pour f1 et f2 de même catégorie, f1 doit être l'élément qui transmet le plus de traits morphologiques au contexte extérieur à f1 f2. C'est-à-dire que le nombre d'inflections de f1 qui sont contrôlées par ou qui contrôlent une inflection d'un élément extérieur est supérieur à celui pour f2.

(A3) dans le cas où A1 et A2 sont inopérants (en particulier en l'absence de morphologie suffisamment riche), Mel'cuk s'en remet à la sémantique pour déterminer le gouverneur syntaxique : il considère que c'est l'élément qui détermine le type de référent du groupe f1 f2, i.e. f1 f2 est une sorte de f1, comme "cheval blanc" est une sorte de cheval. C'est ce que nous définissons infra comme la tête sémantique.

Un élément dans un énoncé a au plus un gouverneur syntaxique (sauf dans le cas de représentations de la coordination par têtes multiples). Donc la représentation des dépendances syntaxiques d'un énoncé est un arbre.

La définition du gouverneur syntaxique peut être mise en parallèle avec celle que l'on pourrait donner pour la **tête syntaxique** dans le cadre d'une grammaire syntagmatique. Nous faisons référence aux travaux de Zwicky (1985, 1993) qui distingue différents critères de définition de la tête syntaxique.

Le critère A1 ci dessus est parallèle à la définition de la tête comme l'élément qui détermine la catégorie du constituant supérieur, étant donné que la catégorie d'un constituant est établie sur la base de sa distribution (cela correspond à ce que Zwicky appelle le *category determinant*). Il est cependant à noter que nous retenons le critère A1 de Mel'cuk, comme faisant intervenir les constructions syntaxiques, c'est-à-dire une notion de la distribution qui s'affranchit de considérations sémantiques (les restrictions de sélection par exemple). Ainsi, pour *est malade*, *malade* impose des restrictions de sélection (cf. **la table est malade/ Jean est malade*), mais les constructions permises sont déterminées par *est* (le tout peut comprendre un sujet et constituer une phrase finie, qui peut être enchâssée ou être indépendante, etc...).

Le critère A2 correspond à la définition de la tête comme l'élément qui transmet au constituant auquel il appartient la majorité des traits morphologiques (et c'est ce qui est appelé le *morphological locus* par Zwicky (1985)). La conjonction des deux critères (Zwicky, 1993) revient à considérer que la catégorie est également un trait transmis par la tête au constituant supérieur, et l'on retrouve alors la notion de tête au sens de GPSG ou HPSG (précisément définie en termes de transmissions de traits de tête).

Enfin le critère A3 correspond à ce que Zwicky appelle le *semantically characterizing participant*, et que nous appellerons tête sémantique (cf. infra).

Une différence entre tête syntaxique, au sens de GPSG ou HPSG, et gouverneur syntaxique, c'est que la tête d'un constituant, en tant que constituant immédiat de celui-ci,

peut elle-même être syntagmatique. Mais on peut par récurrence définir la tête lexicale d'un syntagme comme la première tête (au sens ci-dessus) qui soit lexicale, lorsqu'on isole récursivement la tête d'un syntagme, puis la tête de la tête, etc... La notion parallèle de celle de gouverneur syntaxique est en grammaire syntagmatique la notion de tête lexicale syntaxique.

Les auteurs que nous avons cités (Mel'cuk et Zwicky) notent les divergences possibles entre les différents critères. Mel'cuk indique clairement un ordre de priorité dans les critères (A1 > A2 > A3) ce qui ne laisse pas d'indétermination possible. Un gouverneur syntaxique (comme un auxiliaire de temps par exemple) peut satisfaire le critère A2 (il est le locus morphologique) mais pas le critère A3 (la sémantique de *être parti* est un sous-type de 'partir').
Zwicky donne conjointement les différents critères (*category determinant, morphological locus, et semantically characterizing participant*) comme conjoints, tout en citant explicitement des cas de divergences, qu'il classe comme "spécifieurs". Il s'agit des auxiliaires, des noms de type déterminant complexe (i.e. *une pile d'assiettes, un verre de vin*), ainsi que des adpositions (prépositions en français) régies. La divergence principale dans ces cas concerne le critère sémantique et les deux autres critères. Nous suivons Mel'cuk en considérant une hiérarchie de critères, et en particulier, nous donnons la priorité aux critères A1 et A2 sur le critère A3, c'est-à-dire que nous avons entendons le terme tête syntaxique au sens de HPSG.

Dans le cas de TAG, nous avons déjà fourni au Chapitre 1 les principes régissant la bonne formation des arbres élémentaires que nous retenons (principe de minimalité sémantique et principe de cooccurrence prédicat-arguments). L'ancre d'un arbre élémentaire est l'élément qui sémantiquement gouverne des arguments, représentés en syntaxe par les positions à étendre de l'arbre élémentaire. L'ancre peut donc ne pas être la tête syntaxique (au sens pris ci-dessus). Une autre définition de la tête syntaxique, comme l'élément qui détermine la sous-catégorisation, ne nous paraît pas s'appliquer à tous coups à l'ancre d'un arbre élémentaire. Plus précisément, dans le cas d'un arbre élémentaire représentant un modifieur syntaxique, l'arbre élémentaire comprend une position noeud pied (pour l'élément modifié), qui n'est pas sous-catégorisé par l'ancre (avec la définition traditionnelle de sous-catégorisation). Ainsi dans *Dans le jardin, Paul joue*, la préposition sous-catégorise seulement un SN (*le jardin*). Or en TAG, l'arbre élémentaire ancré par cette préposition comprend une position à substituer pou le SN sous-catégorisé, mais aussi une position de noeud pied pour la phrase modifiée. L'arbre élémentaire constitue une structure qui représente une projection de l'argument sémantique modifié (la phrase *Paul joue*).
Donc nous ne retenons pas que l'ancre d'un arbre élémentaire est la tête au sens de l'élément définissant la sous-catégorisation. De manière à donner une caractérisation uniforme des arbres élémentaires, nous retenons le PCPA, dans sa version sémantique : l'arbre élémentaire est la structure minimale telle que sont comprises des positions à étendre pour tous les arguments exprimés de l'ancre.

Cependant le statut syntaxique des arguments sémantiques d'un lexème peut différer selon les cas. Nous étudions ci-dessous les divergences syntaxe/sémantique, pour ce qui est des dépendances.

1.3 Divergences entre dépendances syntaxiques et sémantiques

Les notions de dépendances sémantique et syntaxique ne sont bien sûr pas équivalentes. Mel'cuk propose (Mel'cuk, 1988:119s) les différentes possibilités de dépendances sémantique, syntaxique, et morphologique pouvant exister entre deux mots, exemples à l'appui dans différentes langues. Nous ne prenons pas en compte les dépendances morphologiques ici[4]. On a alors outre le cas d'aucune dépendance syntaxique ou sémantique entre deux mots, quatre configurations possibles.

	dépendance sémantique			dépendance syntaxique		
1	f1	—sém→	f2	f1	—synt→	f2
2	f1	—sém→	f2	f1	←synt—	f2
3	f1	—sém→	f2	f1		f2
4	f1		f2	f1	—synt→	f2

Table 14 : Les différentes possibilités de dépendances sémantique et syntaxique entre deux mots (outre aucune dépendances)

Le cas 1 peut être illustré avec f1 comme verbe, et f2 comme complément du verbe (au sens large incluant le sujet), par exemple dans *Paul dort*.
Le cas 2 peut être illustré avec pour f1 un adjectif en emploi épithète, et f2 le nom qu'il modifie. En sémantique dans *cheval blanc*, 'blanc' gouverne le nom, mais en syntaxe, il est gouverné par lui.
Les cas 3 et 4, sont beaucoup moins fréquents dans la langue, et toujours discutables. Le cas 3 peut être illustré par un adjectif en emploi attribut. En sémantique, dans *le cheval est blanc*, 'blanc' gouverne 'cheval', mais il n'y a pas de dépendance syntaxique (directe) entre les deux, avec la définition de la dépendance que nous avons retenue[5]. Un autre exemple moins controversé peut être fourni avec un cas de contrôle anaphorique comme dans *Pour honorer le cortège, les drapeaux ont été déployés par les villageois*, où 'villageois' peut être considéré comme un argument sémantique de 'honorer' (les villageois honorent le cortège), sans qu'il y ait aucune dépendance syntaxique entre les deux.
Un exemple à la fois pour le cas 3 et le cas 4 peut être celui d'un adjectif et d'un nom prédicatif, où l'adjectif est sémantiquement lié à l'argument du nom comme dans *un silence confus de l'enfant* (Mel'cuk, 1988:126) : 'confus' gouverne sémantiquement 'enfant' sans qu'il y ait une dépendance syntaxique entre les deux (cas 3), et *silence* gouverne syntaxiquement *confus* sans qu'il y ait de dépendance sémantique entre les deux (quoique la construction syntaxique tend à établir un lien sémantique entre les deux). Un autre exemple pour le cas 4, peut être celui d'un SP extraposé, comme dans *La différence est grande entre A et B* (Abeillé, 1994b), où *entre* est sémantiquement lié à *différence*, sans qu'il y ait de lien syntaxique entre les deux[6].

[4] Mais Mel'cuk montre, exemples à l'appui, la même indépendance des dépendances morphologiques par rapport aux dépendances sémantiques et syntaxiques, que celle existant entre dépendances syntaxiques et sémantiques.
[5] Dans de nombreuses théories, le sujet est considéré comme le sujet de l'adjectif. Cependant, avec les critères de dépendance syntaxique retenus, l'attribut est dépendant syntaxiquement de la copule dans notre exemple. Ce qui fait considérer le sujet comme le sujet de l'adjectif est un critère sémantique pour nous (le sujet est un argument sémantique de l'adjectif).
[6] Abeillé (1994b) relate une distinction faite par Gunnarson entre différents cas d'extraposition de SP. Dans l'exemple fourni avec *différence/entre*, plusieurs tests indiquent une absence de lien syntaxique entre le SP et

Nous reviendrons sur les cas 3 et 4, non prototypiques, infra. La différence entre les cas 1 et 2 permet de définir la notion de **modifieur syntaxique**, comme un lexème dont le sémantème gouverne sémantiquement un ou plusieurs arguments, mais dont un de ces arguments se réalise en syntaxe comme le gouverneur syntaxique (cas 2). C'est la notion classique de modifieur, tel qu'un adjectif dans son emploi épithète, une relative, un adverbe...
Cependant, il existe à notre avis une notion sémantique de la modification, qui se ramène souvent à la notion syntaxique mais pas toujours.

1.3.1 Modifieurs syntaxiques et modifieurs sémantiques

Si l'on considère les deux expressions *verre de vin* et *verre à vin*, les dépendances syntaxiques et sémantiques suivent la Figure 141 ci-dessous. Dans les deux expressions, *verre* est la tête syntaxique[7], *vin* est un dépendant syntaxique de *verre*. Et dans *verre de vin*, le sémantème 'verre1' (unité de mesure de capacité) a pour argument sémantique 'vin'; dans *verre à vin*, le sémantème 'verre2' (récipient) a pour argument sémantique 'vin'. Les dépendances syntaxiques et sémantiques sont les mêmes dans les deux cas.

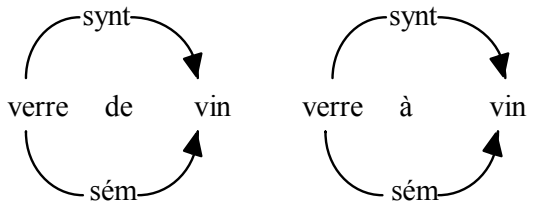


Figure 141. Dépendances syntaxiques et sémantiques

Or il existe bien sûr une différence cruciale entre *verre de vin* et *verre à vin*, c'est que dans le premier cas, on parle de vin, dans le second cas on parle de verre. Nous introduisons pour cela le terme de **tête sémantique**[8] : la tête sémantique d'un ensemble de dépendances sémantiques est l'élément tel que la sémantique de l'ensemble est un sous-type de la sémantique de cet élément. 'ensemble peut être considéré comme le sémantème qui "résume" ou encore qui est une sous-spécification d'un ensemble de sémantèmes liés par des dépendances sémantiques (voir par exemple Candito et Kahane, 1998b). Dans nos exemples, un *verre à vin* est une spécification de verre (par la fonction), *un verre de vin* est une spécification de vin (par la quantité). Polguère (1990), dans le cadre de la TST utilise le terme de "noeud dominant" mais cela interfère avec la notion de tête syntaxique, c'est pourquoi nous avons préféré introduire un nouveau terme[9]. Zwicky (1993) parle de

---

le nom : la pronominalisation possible du SN (*La différence est mince entre A et B, mais elle est grande entre C et B*) et la possibilité d'antéposer le SP (*Entre A et B, la différence est grande*).

[7] Dans *verre de vin*, le locus morphologique est *verre*, cf. par exemple *les verres de vin se vidèrent/*vida*.

[8] Zwicky mentionne le terme de *semantic head* en l'attribuant à Jespersen.

[9] Plus précisément Polguère fournit deux caractérisations de la notion de noeud dominant, une qui correspond à ce que nous appelons tête sémantique et l'autre qui définit le noeud dominant comme celui qui fournit la tête syntaxique dans la structure (ou les) structures syntaxiques qui réalisent la structure sémantique. Il s'agit dans la très grande majorité des cas de la même notion, mais il existe quelques exceptions comme *verre de vin*. Une solution alternative serait de considérer qu'au niveau de la syntaxe profonde (la TST distingue en effet syntaxe profonde et de surface), la réalisation de 'vin' est tête syntaxique. Mais cela ne fait que reculer la divergence au niveau de la correspondance avec la syntaxe de surface.

l'élément sémantiquement caractérisant, et l'associe au concept de tête syntaxique (en isolant cependant des cas typiques de divergence, dont en particulier les déterminants complexes comme *verre de*).
Etudions précisément la différence existant entre *verre à* et *verre de*. Dans *verre de vin*, la tête sémantique est 'vin'. Nous donnons Figure 142 les dépendances sémantiques pour *verre de vin* et *verre à vin*, en notant d'un •sém l'extrémité de la dépendance sémantique où se trouve la tête sémantique.

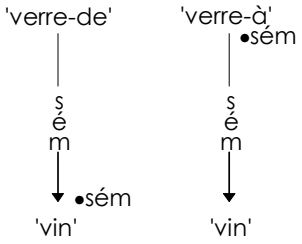


Figure 142. Dépendances sémantiques et tête sémantique

Cette notion va bien représenter le fait que les deux portions de graphe sémantique vont être reliées à un gouverneur sémantique par la tête sémantique. Ainsi pour *boire un verre de vin*, 'boire' gouverne 'vin, alors que pour *casser un verre de vin*, 'casser' gouverne 'verre-à' (Figure 143).

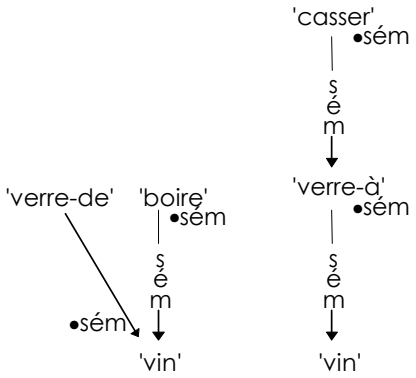


Figure 143

On peut alors définir un **modifieur sémantique**, comme un sémantème qui gouverne un argument, mais dont cet argument est la tête sémantique du tout[10]. En général, la réalisation de cet argument est aussi la tête syntaxique du tout, cela rejoint alors la notion de modifieur syntaxique. C'est le cas d'un adjectif épithète. Mais dans des cas comme *verre de vin*, la tête syntaxique exprime un modifieur sémantique (Figure 144). Nous ne voyons pas de cas de modifieur syntaxique qui ne soit pas un modifieur sémantique.

[10] Cette notion est comparable à ce qui est appelé spécifieur en GB (terme repris par Zwicky (1993) dans son étude sur les caractéristiques de la tête syntaxique).

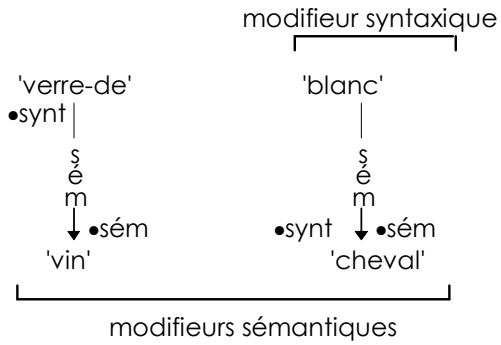


Figure 144

Un autre exemple de modifieur sémantique qui ne soit pas un modifieur syntaxique peut être celui d'un verbe à montée aspectuel, comme /VENIR-DE/. Dans *Chiara dit que Paul vient de partir*, l'argument sémantique de 'dire' c'est 'partir' (à moins d'insister sur *vient de*), qui constitue la tête sémantique de la complétive. Le verbe à montée est donc un modifieur sémantique. Pourtant la tête syntaxique de la complétive est le verbe à montée (d'après le critère A2 supra). Celui-ci n'est donc pas un modifieur syntaxique.

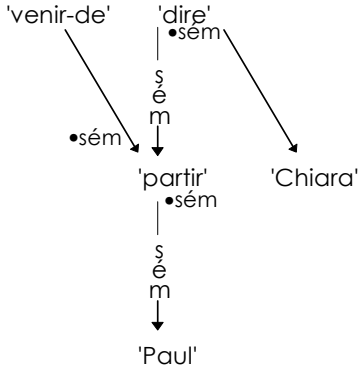


Figure 145. Dépendances sémantiques pour *Chiara dit que Paul vient de partir*.

La notion de tête sémantique doit nécessairement accompagner celle de dépendance sémantique, de manière à rendre compte de la différence entre *verre de/verre à*, *vouloir partir/ venir de partir*...

La détermination de la tête sémantique est aisée dans le cas d'entités, dans le cas de prédicats dont les arguments sont des entités également. Ainsi *Paul a couru trois heures* dénote une course[11]. Dans le cas de phrases complexes (comme coordination de propositions ou comme phrase avec subordonnée), la détermination de la tête sémantique peut être plus délicate, voire ambiguë. Nous verrons des exemples dans les sections suivantes. Nous commençons par étudier la caractérisation des arbres élémentaires d'une TAG, selon l'interface syntaxe/sémantique en jeu.

[11] La notion de tête sémantique est à distinguer de notions discursives, comme le focus par exemple. Dans *Paul a couru trois heures*, l'élément nouveau de la phrase peut être *trois heures*, mais la sémantique de la phrase ne se ramène pas à une durée.

1.4 Caractérisation des arbres élémentaires sous l'angle de l'interface sémantique/syntaxe

Nous avons donné au Chapitre 1 une version purement sémantique du principe de cooccurrence prédicat-arguments. Nous pensons qu'une telle formulation permet une caractérisation uniforme des arbres élémentaires d'une TAG, alors que ceux-ci sont hétérogènes du point de vue syntaxique.
Pour les arbres initiaux, dans la plupart des cas la racine de l'arbre initial représente une projection de l'ancre (comme l'arbre de la Figure 160, pour un nom prédicatif et ses arguments réalisés comme compléments du nom). Mais cela n'est pas toujours le cas. On peut donner l'exemple d'un arbre élémentaire pour une clivée (Abeillé, 1991) où la tête syntaxique est la copule, représentée comme co-tête, dont la racine de l'arbre est une projection.

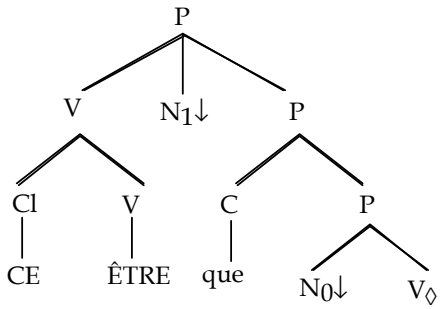
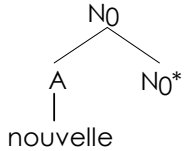


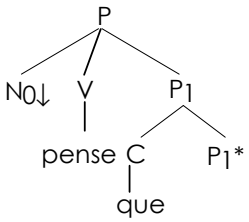
Figure 146. Schème pour une clivée (*c'est N1 que N0 V*)

Il faut plutôt dire pour les arbres initiaux, qu'ils sont la structure minimale telle qu'elle contienne les positions pour les arguments sémantiques de l'ancre. Un des arguments peut se réaliser en dehors du domaine dont l'ancre est la tête syntaxique, comme par exemple un foyer de clivée; cet argument sera tout de même localisé dans l'arbre de son gouverneur sémantique.

Pour les arbres auxiliaires on a donné une première distinction au Chapitre 1, selon la localisation de la tête syntaxique du constituant représenté par l'arbre (cf. Kroch, 1989; Shieber et Schabes, 1994). Pour un arbre auxiliaire enchâssant, la tête syntaxique du constituant représenté par l'arbre est l'ancre; pour un arbre auxiliaire modifieur, la tête syntaxique du constituant représenté par l'arbre est représentée par le noeud pied. C'est le cas par exemple pour un arbre élémentaire ancré par un adjectif dans son emploi épithète.



Arbre auxiliaire modifieur



Arbre auxiliaire enchâssant

Nous venons de voir qu'aussi bien pour les arbres initiaux que pour les arbres auxiliaires, l'ancre n'est pas systématiquement la tête syntaxique du constituant représenté par l'arbre. Une solution alternative pour une caractérisation syntaxique des arbres élémentaires serait

de considérer qu'un arbre élémentaire comprend des positions pour tous les éléments sous-catégorisés par l'ancre (et dans le cas d'extraction, comme dans une clivée, un actant sous-catégorisé se réalise en dehors de la projection de l'ancre). Cependant, on doit alors généraliser la notion de sous-catégorisation, pour y intégrer le gouverneur syntaxique dans le cas d'un modifieur syntaxique. Nous préférons situer la généralisation au niveau sémantique, en imposant qu'un arbre élémentaire contienne les positions pour l'expression des arguments sémantiques, que ceux-ci se réalisent comme dépendant ou comme gouverneur syntaxique (Candito et Kahane, 1998b).

1.5 Critères de typage des noeuds argumentaux en noeuds pied et noeuds à substituer

Avec cette caractérisation des arbres élémentaires, le noeud pied a le même statut que les noeuds à substituer. Ces noeuds sont censés représenter un argument sémantique de l'ancre, que l'arbre soit initial ou auxiliaire, enchâssant ou modifieur. On peut vérifier que c'est en général un principe implicite dans les TAG existantes, c'est-à-dire que, en dehors des cas de divergences fortes entre syntaxe et sémantique, le principe est suivi.

Si un noeud pied et un noeud à substituer ont le même statut en sémantique, qui est de représenter un argument sémantique de l'ancre, il reste à voir selon quels critères se détermine la distribution des noeuds pied et des noeuds à substituer.

Il existe une contrainte formelle qui est qu'un arbre a au plus un noeud pied.
Du point de vue linguistique, en général en TAG, les arbres auxiliaires sont utilisés pour les modifieurs syntaxiques, ainsi que pour les auxiliaires, les verbes à montée, les verbes à contrôle, les verbes à complétive. Abeillé (1991) propose une caractérisation linguistique des éléments à représenter comme des arbres auxiliaires en TAG, comme étant ceux correspondant aux modifieurs au sens harrissien du terme (Harris, 1976). D'où une justification théorique pour l'adjonction des auxiliaires, des verbes à montée, des verbes à complétives (bien que pour ces derniers, il y ait une justification d'ordre plus pratique, qui est de pouvoir traiter la syntaxe des dépendances non bornées).
Nous pensons que la notion de modifieur au sens de Harris n'est pas syntaxique mais sémantique, et qu'elle est à rapprocher de la notion de modifieur sémantique vue supra. Le pari en TAG, c'est de représenter systématiquement tout modifieur *sémantique* par un arbre auxiliaire (voir Abeillé (1994a) pour le traitement de *verre à* comme arbre initial, et *verre de* comme arbre auxiliaire). La généralisation sur le typage des arbres en arbres initiaux/auxiliaires est la suivante :

Les arbres élémentaires auxiliaires sont ceux qui représentent un modifieur sémantique.

Passons en revue les différents cas d'utilisation de l'adjonction en TAG. Les modifieurs syntaxiques sont adjoints, or ils sont aussi des modifieurs sémantiques. Voyons le cas de modifieurs sémantiques qui ne sont pas des modifieurs syntaxiques.
Les auxiliaires sont adjoints, or, avec la définition de tête syntaxique que nous avons retenue, un auxiliaire est tête syntaxique par rapport au participe, mais il n'est pas tête sémantique. Il est donc bine un modifieur sémantique.

Les verbes à complétive sont en TAG représentés par des arbres auxiliaires, du moins dans le cas d'extraction hors de la complétive. Pour ces verbes, la détermination de la tête sémantique est plus délicate. Il existe certains arguments pour considérer les verbes à complétives comme des modifieurs sémantiques. Dans une phrase comme (418a), il y a ambiguïté sur la tête sémantique, entre 'penser' et 'malade'. Mais (418a) peut être, au moins dans certains cas, quasiment synonyme de (418b,c). C'est un argument théorique en TAG en faveur de l'analyse de verbes à complétive par des arbres auxiliaires s'adjoignant sur la complétive.

(418) a. Paul pense que Pierre est malade.
b. Pierre est malade, pense Paul.
c. Selon Paul, Pierre est malade.

Mais il faut bien reconnaître qu'il y a également des motivations plus "techniques" pour adjoindre un verbe à complétive sur sa complétive, qui sont de pouvoir ainsi représenter les dépendances non bornées (et qui précèdent historiquement la justification théorique). Par exemple, les verbes à complétive "non pont", c'est-à-dire qui n'admettent pas l'extraction hors de leur complétive (comme /S'ECRIER/), sont représentés par des arbres initiaux, ce qui bloque bien l'extraction (Abeillé, 1991).
De plus, un verbe à complétive "pont" peut dans certaines constructions ne pas admettre d'extraction. C'est le cas par exemple si un des arguments du verbe à complétive est relativisé. Le verbe pont reçoit alors sa complétive par substitution.

(419) La personne à qui Paul a promis que tout serait prêt.

Ainsi pour dériver (419), on utilise un arbre auxiliaire de racine et noeud pied nominal, ancré par *promis*, qui s'adjoint sur l'antécédent (*personne*) (Figure 147). Formellement, la complétive *tout serait prêt*, ne peut qu'être substituée, puisqu'un même arbre ne peut pas avoir plusieurs noeuds pieds[12]. Cela permet dans le même temps de représenter le fait qu'on ne peut pas extraire hors de cette complétive, du fait de l'îlot relatif.

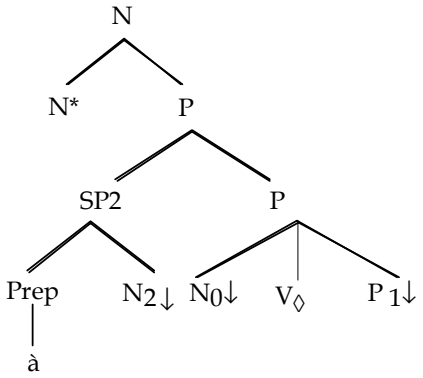


Figure 147. Schème pour un verbe pont, avec complément relativisé

Mais il est intéressant de remarquer[13] que dans ce cas, la tête sémantique de la relative semble bien être le verbe pont (/PROMETTRE/). Il y aurait un parallèle à faire entre des constructions syntaxiques et position de la tête sémantique. En (419), un argument du verbe pont est extrait, et la tête sémantique est le verbe pont. En (420), c'est un argument du

[12] Permettre plusieurs noeuds pieds pour un même arbre augmente la capacité générative du formalisme.
[13] Il s'agit d'une remarque de Sylvain Kahane.

verbe enchâssé qui est extrait, et la tête sémantique de la relative semble plutôt être le verbe enchâssé. Pour ces deux types de situation, TAG traite la complémentation du verbe pont respectivement par substitution et par adjonction.

(420) La personne à qui Paul croit que Pierre parlait.

L'exemple de l'utilisation de l'adjonction pour un verbe à montée est intéressant. Traiter systématiquement les verbes à montée (du sujet) par adjonction à un noeud verbal représente bien le fait qu'au niveau sémantique il s'agit de prédicats unaires, qui au niveau de l'ordre des mots apparaît au milieu de la réalisation de son argument (cf. le verbe à montée apparaît entre le sujet et le verbe infinitif). On est donc forcé d'adjoindre, pour une question d'ordre des mots. La question intéressante est de savoir si ces cas d'adjonction sont effectivement utilisés pour des modifieurs sémantiques, c'est-à-dire si l'adjonction est également motivée par la théorie.
Or il nous semble que l'on a un continuum de situations entre des verbes à montée modifieurs sémantiques (i.e. où la tête sémantique de la combinaison Vmontée + infinitif est réalisée par l'infinitif) et d'autres qui peuvent alterner, entre fournir une information subalterne, ou constituer le centre de la prédication. Les deux options peuvent être testées en considérant un verbe à montée lui-même enchâssé. Par exemple en (420a), il nous semble que la tête sémantique de la complétive est réalisée par le verbe à montée. L'argument sémantique de 'affirmer' est 'finir' (au moins dans un certain contexte où l'on présuppose que Chiara peignait la cathédrale). En (420b) le verbe à montée peut être considéré comme une information subalterne, la tête sémantique de la complétive étant alors 'vouloir'.

(421) a. Paul affirme que Chiara a fini de peindre la cathédrale.
b. Paul disait que Chiara semblait ne plus vouloir finir.

Donc, face à un traitement uniforme des verbes à montée par adjonction, indépendamment motivé, le critère sémantique d'adjonction de modifieur sémantique n'est que partiellement suivi. Malgré une tendance à retrouver pour les éléments généralement adjoints en TAG des modifieurs sémantiques (également modifieurs syntaxiques ou pas), on ne peut pas s'en remettre à cet unique critère pour déterminer quel argument sémantique doit être représenté comme un noeud pied. La notion de modifieur sémantique nous semble appropriée. Mais elle ne fixe pas complètement la distribution des noeuds pieds dans la grammaire, étant donné que pour certains éléments lexicaux, le fait qu'ils soient tête sémantique ou plutôt qu'un de leurs arguments sémantiques le soit peut dépendre du contexte et une même phrase peut être ambiguë pour la tête sémantique. Dans d'autres cas, des questions de syntaxe ont priorité pour le typage des noeuds argumentaux en noeud pied/noeud à substituer (par exemple l'ordre des mots imposé par un verbe à montée).

2 Arbre de dérivation et graphe de dépendances sémantiques

Parce que l'arbre de dérivation est constitué de relations entre arbres élémentaires lexicalisés, correspondant éventuellement à des unités poly-lexicales (pour les idiomes notamment), il a été comparé à un arbre de dépendances (Rambow et Joshi, 1992).

Plus précisément, Rambow et Joshi (1992) relie l'arbre de dérivation (dorénavant l'**AD**) à un arbre de dépendances syntaxiques profondes au sens de la Théorie Sens-Texte (Mel'cuk, 1988). Mais ils sont confrontés au fait que les arcs d'adjonction ne sont pas uniformes pour ce qui est de la dépendance syntaxique qu'ils dénotent (puisque comme nous le verrons, ils sont uniformes pour ce qui est de la dépendance sémantique, et on a vu les divergences entre dépendances sémantiques et dépendances syntaxiques).
Nous avons proposé d'interpréter les arcs de l'arbre de dérivation comme des dépendances sémantiques (Candito et Kahane, 1998a; 1998b), ce qui est une conséquence naturelle de considérer les arbres élémentaires comme des unités sémantiques (Abeillé, 1991) et de considérer, pour tout type d'arbre élémentaire, que les positions à étendre (noeuds pied et à substituer) représentent un argument sémantique du lexème ancrant l'arbre.
Il est d'ailleurs à noter que le fait que l'arbre de dérivation doive représenter la sémantique, et s'affranchir de la syntaxe n'est pas nouveau. Abeillé (1994a) par exemple, distingue la représentation de *verre de vin* et *verre à vin* sur des bases sémantiques uniquement (cf. supra). Kroch (1989) mentionne que la dérivation TAG "devrait préserver la compositionnalité sémantique ("should preserve compositional semantics"). On peut citer également Shieber et Schabes (1994), qui motivent leur définition alternative de la dérivation, par des arguments qui se ramènent à des considérations sémantiques. Enfin Danlos, dans le système G-TAG (Danlos, 1998b) utilise l'arbre de dérivation comme représentation intermédiaire entre le niveau conceptuel et le niveau syntagmatique.

Nous fournissons ci-dessous l'ébauche d'une procédure pour l'interprétation de l'AD comme un graphe de dépendances sémantiques, et de la procédure inverse. Nous sommes consciente qu'il s'agit d'une ébauche, et qu'un travail beaucoup plus approfondi est nécessaire pour que ces procédures soient opérationnelles.

2.1 Procédure d'interprétation de l'arbre de dérivation

Nous prenons la définition de la dérivation de Schieber et Schabes (1994) (cf. Chapitre 1), où plusieurs adjonctions modificatrices (i.e. d'arbres auxiliaires modifieurs) peuvent être opérées à un même noeud.
Alors, pour interpréter un arbre de dérivation comme un graphe de dépendances sémantiques, il faut (Figure 148):

- traduire les noeuds (des arbres élémentaires) par le sémantème signifié par le lexème qui ancre l'arbre élémentaire; sauf les noeuds correspondants à des flexions (comme les auxiliaires de temps), qui peuvent être incorporés comme traits.
- lire un arc de substitution comme une dépendance sémantique où le gouverneur sémantique est l'arbre recevant la substitution;
- lire un arc d'adjonction comme une dépendance sémantique où le gouverneur sémantique est l'arbre auxiliaire adjoint.

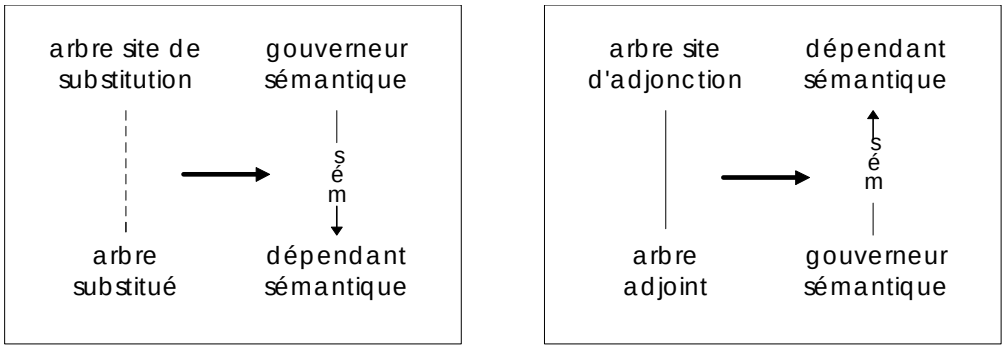


Figure 148. Interprétation des arcs de l'AD comme des dépendances sémantiques

Un aspect important est que cela vaut pour tout type d'arc d'adjonction, qu'il s'agisse d'une adjonction enchâssante, ou d'une adjonction modificatrice, à condition que l'adjonction ait lieu sur l'artère de l'arbre de l'arbre recevant l'adjonction (c'est-à-dire le chemin de l'ancre principale à la racine)[14].

Nous fournissons
Figure 149, l'AD et l'interprétation correspondante pour la phrase (422), volontairement compliquée (!). Nous ne faisons pas figurer les arcs de l'AD correspondant à l'ajout de déterminants (cf. supra, ils ne sont pas considérés comme réalisation de dépendances sémantiques).

(422) La nouvelle bibliothèque possède le livre que Marie pense que Jean cherche.

Pour calculer le numéro d'argument (cf. les arguments sémantiques sont distingués par une numérotation) à partir de l'adresse de substitution ou d'adjonction, il faut supposer que les positions argumentales (noeuds pied et à substituer) portent leur numéro. Le calcul est alors fait de la manière suivante :

- pour un arc de substitution, le numéro d'argument est celui porté par le noeud à substituer (dans l'arbre site de substitution, i.e. à l'extrémité *supérieure* de l'arc de l'AD)
- pour un arc d'adjonction, le numéro d'argument est celui porté par le noeud pied (dans l'arbre adjoint, i.e. à l'extrémité *inférieure* de l'arc de l'AD).

[14] L'adjonction a effectivement lieu sur l'artère dans la plupart des cas. On peut cependant citer le cas (qui nous a été fourni par Martine Smets) d'un arbre avec co-tête prépositionnelle, où le SP peut recevoir un modifieur. Par exemple pour dériver *Paul donne des fleurs seulement à Marie*, l'arbre de *seulement* s'adjoint sur le SP contenu dans l'arbre de *donne*, l'arc correspondant est difficilement interprétable comme dépendance sémantique entre *donne* et *seulement*. Par ailleurs nous excluons les arcs d'adjonction correspondant à une quantification de l'interprétation (cf. infra, la représentation des déterminants).

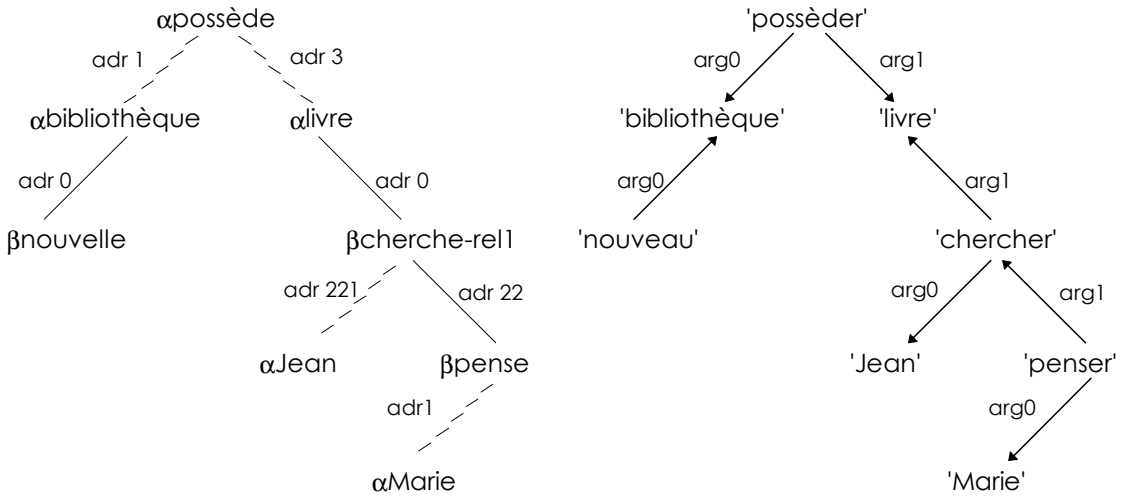


Figure 149. Arbre de dérivation pour (422), et interprétation en termes de dépendances sémantiques

Nous donnons Figure 150 le graphe de dépendances sémantiques obtenu, réagencé. En "retournant" les arcs d'adjonction, on peut effectivement obtenir un graphe qui n'est pas un arbre. L'AD de départ et le graphe de dépendances obtenu sont cependant isomorphes si l'on ne tient pas compte de la direction des arcs.

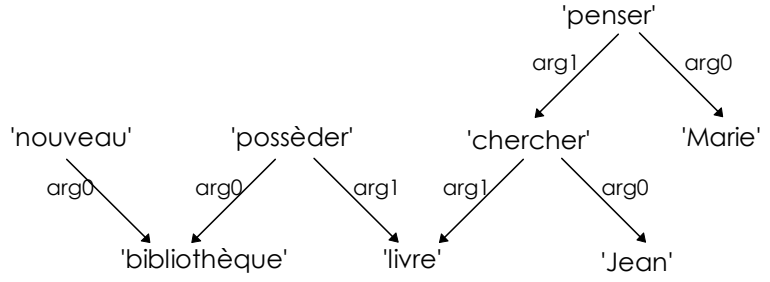


Figure 150. Dépendances sémantiques obtenues pour (422), réagencées

L'AD peut donc être lu comme un graphe de dépendances sémantiques, proche de la structure sémantique de la Théorie Sens-Texte (TST). Cette structure (dorénavant SSém) est en effet définie comme un graphe, dont les noeuds sont des sémantèmes, et les arcs des dépendances sémantiques, étiquetés par un numéro d'argument, qui sert à distinguer les différents arguments d'un sémantème[15].

On peut cependant noter trois différences entre l'AD et la SSém :

- Les arcs de l'AD sont typés en arcs d'adjonction et arcs de substitution, alors qu'aucun typage équivalent n'est présent dans une SSém. En outre l'AD est un arbre, alors que cette contrainte n'existe pas pour la SSém. Nous verrons infra qu'en "retournant" la direction des arcs d'adjonction on obtient également une structure non arborescente pour l'AD. Cependant, il reste que l'on ne peut pas obtenir de cycle (au sens non orienté),

[15] Dans la TST, la structure sémantique ne constitue pas à elle seule la représentation sémantique d'une phrase (ou plutôt d'un ensemble de phrases en relation de paraphrase), elle est augmentée d'une structure sémantico-communicative, qui fournit des informations de type thème/rhème.

alors que cela est pertinent en linguistique, et possible pour la SSém (cf. infra Figure 151).

- La TST considère des sémantèmes qui peuvent correspondre à des unités inférieures au mot : les sémantèmes lexicaux sont dissociés des sémantèmes flectionnels. En TAG, nous avons vu que les arbres élémentaires constituent des unités sémantiques qui correspondent au minimum à un mot et pas à un morphème. Cela a pour conséquence que les marques de temps et d'aspect par exemple, constituent systématiquement des noeuds de la représentation sémantique de la TST, ce qui n'est généralement pas le cas pour l'AD (sauf cas particulier d'adjonction de la copule par exemple).
- Cette dernière différence est la plus importante : dans une représentation sémantique à la TST (ou plus généralement, dans une représentation sémantique tout court), un même noeud (sémantème) peut être réalisé linguistiquement par plusieurs expressions, dans un énoncé qui réalise la représentation sémantique. Ces expressions sont alors dans une relation de coréférence. En TAG, les expressions coréférentes vont constituer des noeuds distincts de l'AD (i.e. des arbres élémentaires distincts), ou, dans le cas de contrôle, aucun noeud (cf. Le sujet contrôlé n'est pas réalisé phonologiquement).

Prenons l'exemple de la phrase (423), dans le sens où le pronom réfère à *Paul*. La représentation sémantique à la TST est fournie à la Figure 151, l'AD correspondant à la phrase (et son interprétation en termes de dépendances sémantiques) est fourni Figure 152, où l'on peut noter que le pronom *il* et *Paul* sont représentés par des noeuds distincts. Pour le pronom, la procédure d'interprétation ne peut pas telle quelle assigner un sémantème : il faut augmenter cette procédure d'un module de résolution de la coréférence, aussi bien pour les expressions anaphoriques que pour récupérer, dans le cas de contrôle, l'argument « sujet contrôlé ».

(423) Paul_i dit qu'il_i part.

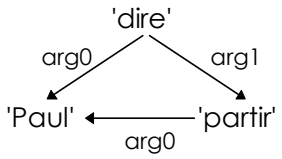


Figure 151. Structure sémantique à la TST, pour (423) (hors sémantèmes flexionnels)

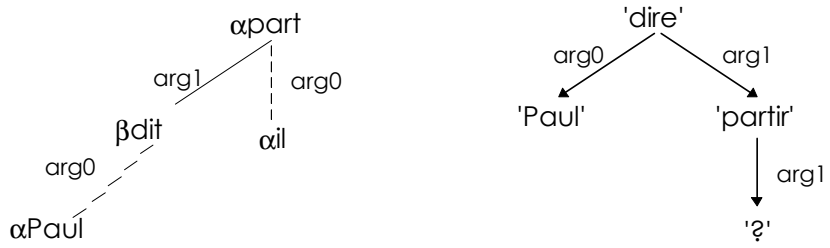


Figure 152. Arbre de dérivation pour (423), et son interprétation en dépendances sémantiques

2.2 Procédure de traduction d'un graphe de dépendances sémantiques en un arbre de dérivation

Nous avons présenté la procédure d'interprétation de l'AD dans le sens de l'analyse : disposant de la dérivation d'une phrase, il s'agit de l'interpréter (mais comme nous venons de le voir, la procédure est à augmenter d'un module pour traiter la coréférence). Dans le sens de la génération, on peut également définir une procédure pour fournir un arbre de dérivation à partir d'un graphe de dépendances sémantiques.

Avant de détailler cela, nous pouvons citer le système de génération G-TAG (Danlos, 1998b), qui utilise comme niveau intermédiaire entre un niveau conceptuel et la phrase, un arbre de dérivation TAG. Le système construit un arbre de dérivation à partir d'une représentation conceptuelle, ce qui est donc comparable à la procédure que nous allons décrire ci-dessus. Cependant la procédure que nous proposons (purement théorique) prend en compte de manière unifiée tous les types d'adjonction (modificatrice et enchâssante) et construit un arbre de dérivation isomorphe au graphe d'entrée. La seule opération utilisée est celle qui consiste à retourner un arc. Cela diffère de la procédure implémentée au sein de G-TAG, avec laquelle l'arbre de dérivation construit n'est pas isomorphe à l'entrée (cf. Meunier, 1997).

Nous avons vu les trois différences existantes entre AD et SSém graphe de dépendances sémantiques de la TST). Faisons abstraction de la différence concernant les sémantèmes flectionnels : admettons que l'on ne traite pas les inflections. Faisons également abstraction dans un premier temps des cas nécessitant une pronominalisation.

Il reste que pour passer d'une SSém à un arbre de dérivation, il faut typer les arcs de dépendance sémantique en arc d'adjonction et de substitution, et de telle sorte que les contraintes formelles sur l'AD soient respectées : un AD est un arbre. Donc cette procédure de traduction SSém→AD suppose de définir un noeud d'entrée du graphe de dépendances sémantiques, et un parcours de toutes les dépendances, tel que le parcours est continu. Les dépendances parcourues dans le sens positif sont traduites en un arc de substitution. Les dépendances parcourues dans le sens négatif sont traduites en un arc d'adjonction, dans le sens indiqué Figure 153. Par ailleurs, étape non triviale (...), les sémantèmes doivent recevoir une lexicalisation.

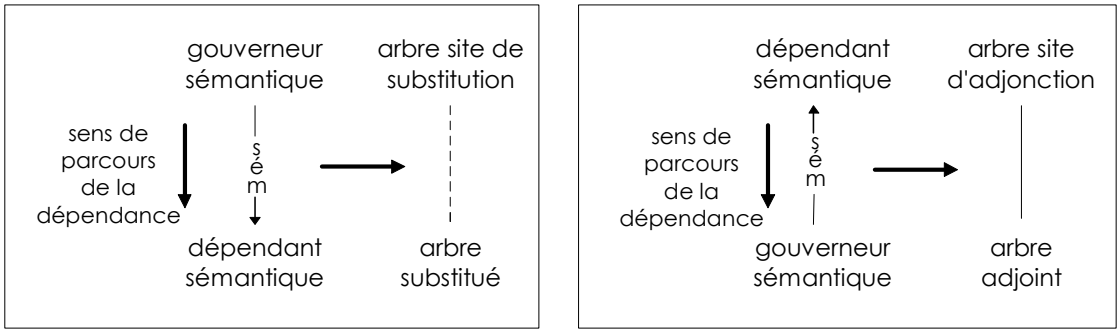


Figure 153. Calcul de l'arbre de dérivation à partir d'un graphe de dépendances sémantiques

Prenons l'exemple de la phrase (424), dont la représentation en termes de dépendances sémantiques est donnée Figure 154.

(424) Paul est arrivé, puis Marie a téléphoné.

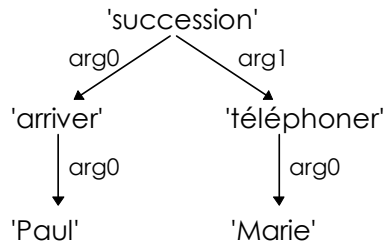


Figure 154. Dépendances sémantiques pour (424)

Supposons que l'on dispose des lexicalisations pour les lexèmes en jeu (illustrées par la phrase (424)), et que l'on choisisse pour noeud d'entrée 'succession', alors toutes les dépendances sémantiques de la Figure 154 sont parcourues positivement, elles sont donc toutes traduites comme des arcs de substitution. Cela signifie en particulier que la lexicalisation de 'succession' doit recevoir ses deux arguments par substitution. L'arbre de dérivation obtenu est fourni Figure 155 (on y néglige les auxiliaires de temps, qui apportent les marques de temps et aspect).

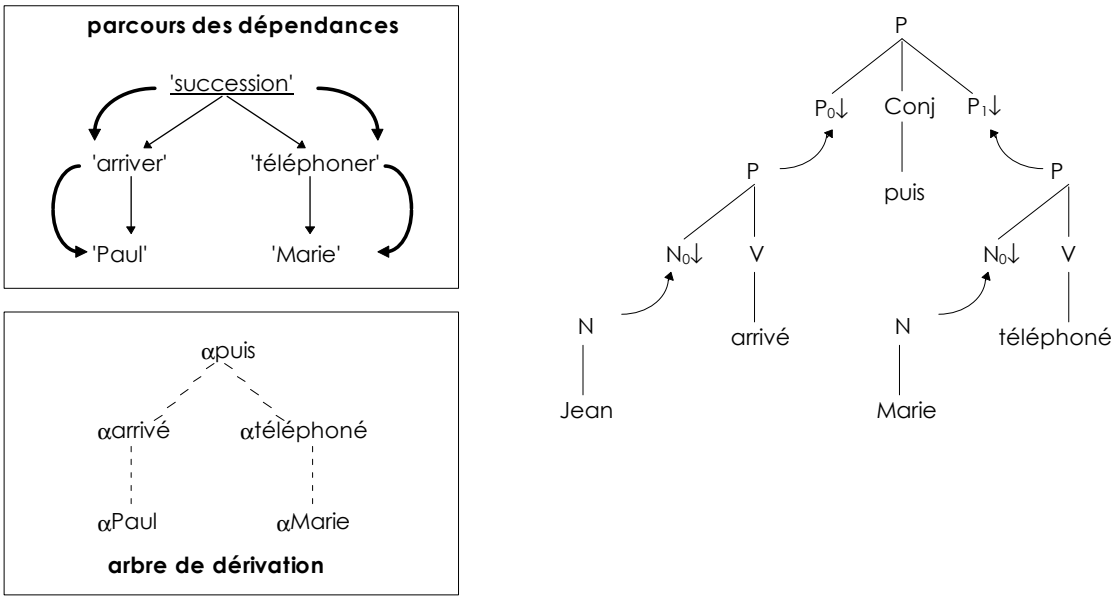


Figure 155. Dérivation pour (424), obtenue en parcourant le graphe de la Figure 154 à partir de 'succession'

Si l'on choisit au contraire pour noeud d'entrée le noeud 'arriver', alors la dépendance entre 'arriver' et 'succession' doit être parcourue négativement, et donne lieu à un arc d'adjonction. L'arbre de dérivation obtenu est fourni Figure 156.

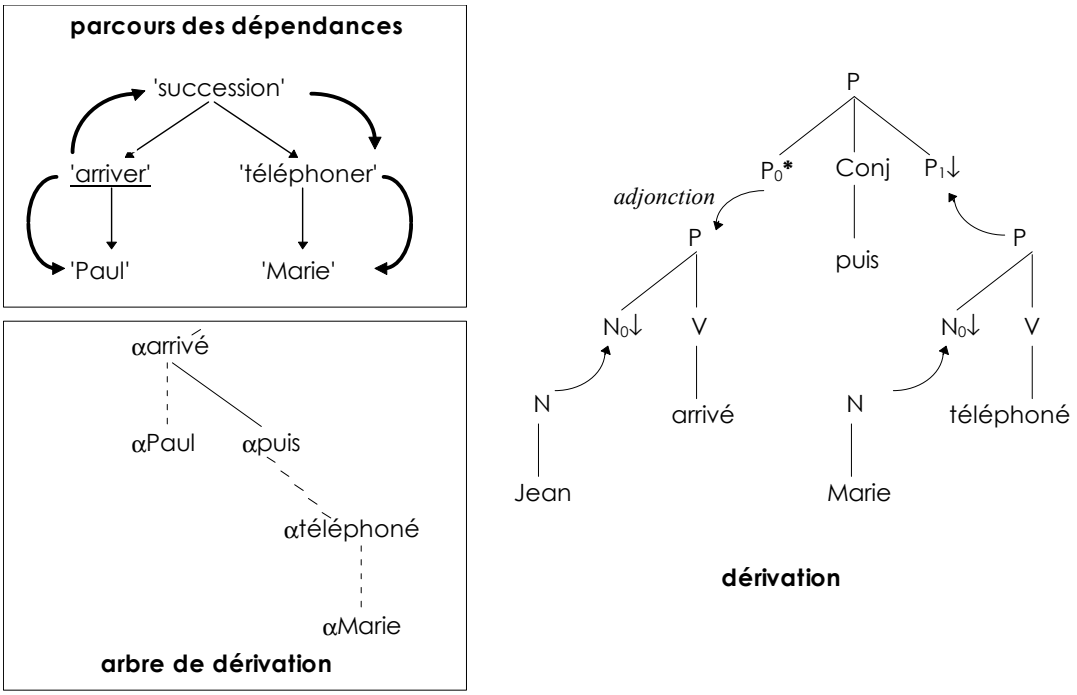


Figure 156. Dérivation pour (424), obtenue en parcourant le graphe de la Figure 154 à partir de 'arriver'

Avec cette procédure, le choix du noeud d'entrée détermine totalement la distribution des noeuds argumentaux en noeuds pied et noeuds à substituer. Dans l'exemple que l'on a donné le noeud P0 de l'arbre associé à *puis* est alternativement un noeud à substituer et un noeud pied, sans changer de topologie.

Le noeud d'entrée choisi est traduit comme la racine de l'AD construit. Cette racine ne peut que recevoir des adjonctions ou des substitutions, ce qui implique, si l'on considère que toutes les adjonctions sont utilisées pour des modifieurs sémantiques, que le noeud racine représente la tête sémantique de l'énoncé. Nous pouvons alors donner un sens à la différence existant entre les deux parcours proposés pour traduire la SSém associée à (424) : il s'agit d'une différence de tête sémantique. Ainsi, les deux parcours distincts illustrés à la Figure 155 et à la Figure 156 peuvent-ils être interprétés comme une ambiguïté sur la tête sémantique associée à la phrase (424). Si la tête sémantique est la succession[16], alors il n'y a aucun modifieur sémantique, et donc aucune adjonction. Si la tête sémantique est 'arriver' (c'est-à-dire si le coup de téléphone est précisé comme une information subalterne) alors 'succession' apparaît comme un modifieur sémantique : elle gouverne sémantiquement un élément qui en est la tête sémantique. La réalisation linguistique de 'succession' est alors traitée par adjonction.

Ce résultat est directement à relier à la procédure de passage de la représentation sémantique à la représentation syntaxique profonde dans le cadre de la TST. Polguère (1990) y définit précisément le noeud d'entrée[17] du graphe de dépendances sémantiques

[16] La succession peut être interprétée comme la tête sémantique dans le cas où l'on considère les deux événements coordonnés sur un pied d'égalité en termes d'importance, ou bien éventuellement si l'on considère que la succession est l'information saillante (pour ce deuxième cas, il faudrait plutôt considérer une réalisation linguistique comme *Paul est arrivé* ***et ensuite seulement*** *Marie a téléphoné*.

[17] La détermination du noeud d'entrée est faite sur la base de la structure sémantico-communicative. Son automatisation complète est encore à l'étude.

comme le noeud qui fournit la tête syntaxique (et donc la racine de l'arbre de dépendances syntaxiques profonde).

Reste à étudier le lien, fort, entre un parcours approprié d'une représentation sémantique et les contraintes syntaxiques liées aux lexicalisations des sémantèmes. La distribution des noeuds pied et à substituer peut être contrainte par des questions d'ordre des mots, ou de syntaxe (on ne veut pas par exemple adjoindre un nom sur son verbe).

Nous avons volontairement donné un cas ci-dessus, où un même élément peut être considéré comme tête sémantique ou pas et être de ce fait adjoint ou pas, avec comme conséquence au niveau des arbres élémentaires en jeu, une simple différence de typage d'un noeud en noeud à substituer et noeud pied. Mais il existe d'autres cas, où la syntaxe impose une construction différente selon l'élément qui est tête sémantique. Nous pensons par exemple à la relation existant entre une phrase assertive et un syntagme nominal correspondant à la relativisation d'un actant nominal de la phrase assertive (c'est la relation existant entre (425a) et (425b)).

(425) a. Jean cherche un livre
b. un livre que Jean cherche

En ignorant les questions de portée du quantificateur *un* (et donc en ne le faisant pas apparaître dans les dépendances sémantiques), ces deux énoncés partagent les mêmes dépendances sémantiques, données Figure 157 (qui ne représentent qu'une partie de la sémantique des deux énoncés).

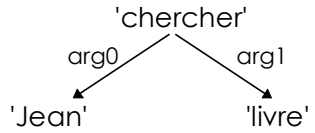


Figure 157. Dépendances sémantiques en jeu pour (425a) et (425b)

La TST représente la différence sémantique des deux énoncés au niveau de la structure sémantico-communicative (que nous détaillons pas ici), qui décompose la SSém en domaines de thème rhème (et récursivement), en distinguant pour chaque sous-domaine un noeud dominant, i.e. dans nos termes la tête sémantique. La différence sémantique entre (425a) et (425b) peut effectivement être bien représentée par une différence de tête sémantique : en (425b), la tête sémantique est 'livre' (le tout réfère à un livre), en (425a) la tête sémantique est 'chercher'. Pour la procédure de traduction de la SSém en un arbre de dérivation, cela est crucial : pour dériver (425a), en partant de 'chercher', toutes les dépendances sémantiques sont parcourues positivement et sont traduites en arcs de substitution. Pour dériver (425b), en partant de 'livre', la dépendance entre 'livre' et 'chercher' est parcourue négativement, et est traduite en un arc d'adjonction.
C'est là que la grammaire peut jouer son rôle, par un contrôle sur la distribution des noeuds pied et à substituer pour les positions argumentales : le seul moyen de réaliser par un arc d'adjonction la dépendance entre un nom (lexicalisation de 'livre') et un verbe (lexicalisation de 'chercher') est de faire du verbe un modifieur, i.e. une relative ou participiale. De manière plus technique, une grammaire TAG, parmi les arbres élémentaires associés au verbe CHERCHER ne va comprendre qu'un certain nombre d'arbres où la position argumentale pour l'argument 1 est un noeud pied : l'arbre pour la relativisation de cet

argument (cf. Chapitre 5), nécessaire pour dériver (425b) et l'arbre pour les participiales présentes et passées (*le livre (étant) cherché par Jean*, cf. Chapitre 5).
La différence de tête sémantique entre (425a) et (425b) va de pair avec une différence de structure syntaxique, ce qui n'était pas le cas pour l'ambiguïté de tête sémantique dont on a parlé pour la phrase (424).

Nous avons vu avec l'interprétation de l'AD pour la phrase (422) que l'on peut obtenir à partir d'un AD un graphe de dépendances qui n'est pas un arbre. Il est cependant impossible d'obtenir un graphe comportant un cycle (au sens non orienté). Inversement, à partir d'un graphe de dépendances sémantiques comportant un cycle (au sens non orienté), comme par exemple dans la Figure 158 ci-dessous, on ne peut pas obtenir directement un AD valable. Dans ce cas, le cycle doit être coupé à un noeud pouvant être exprimé par un pronom. Pour le graphe de la Figure 158, en supposant disposer d'une méthode pour déterminer où couper le cycle, on peut en coupant au noeud 'Paul' (i.e. en dupliquant son noeud), construire un AD pour la phrase *Paul dit qu'il part*. Le travail pour automatiser cette gestion des cycles reste à faire (ainsi que celui pour la détermination de quel noeud dupliqué doit être pronominalisé, cf. **Il_i dit que $Paul_i$ part*).

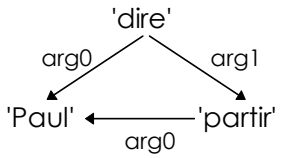


Figure 158. Graphe de dépendances sémantiques avec cycle (au sens non orienté)

Il est à noter que pour cette phrase, selon la tête sémantique choisie ('dire' ou 'partir', cf. supra, la discussion sur les verbes ponts comme modifieurs sémantiques), le verbe pont sera représenté par un arbre élémentaire avec pour la complétive un noeud à substituer ou un noeud pied. La procédure suppose donc que dans le cas d'extraction hors d'une complétive — ce qui nécessite que le verbe pont soit adjoint — la tête sémantique est le verbe enchâssé (cf. supra).

3 Cas problématiques

Le PCPA tel qu'il a été formulé est bien sûr contraignant en syntaxe. Il peut obliger à distinguer différents cas, là où la syntaxe est uniforme (Section 3.1). De plus, le PCPA ne garantit pas à lui seul que les positions prévues pour les arguments sémantiques d'une ancre soient effectivement remplies par les "bons" arguments. Nous étudions quelques exemples Section 3.2).

3.1 Différentes sémantiques pour une syntaxe uniforme

Le PCPA oblige à faire une distinction là où la syntaxe est uniforme. Par exemple, la plupart des adjectifs peuvent être considérés comme des sémantèmes avec un argument, réalisés par le nom que l'adjectif modifie (dans son emploi épithète). Mais il existe des adjectifs qui représentent un argument du nom prédicatif avec lequel ils se combinent, comme dans *le voyage présidentiel* ou *le refus soviétique* (il s'agit de "sujets-adjectifs"). Un

strict respect du PCPA imposerait que les arbres élémentaires pour les noms prédicatifs prévoient des positions à substituer adjectivales pour ces cas, ce qui n'est pas fait actuellement, d'autant qu'une étude sur la sélection mutuelle nom/sujet-adjectif doit être faite pour ne pas surgénérer.
Un autre exemple concerne les déterminants, qui, pour une syntaxe à peu près uniforme, ont plusieurs statuts possibles en sémantique. Un déterminant peut être considéré parfois comme un foncteur qui prédique une propriété (défini, indéfini, démonstratif ...) (426a,b), mais il a dans le même temps un statut de quantifieur. En sémantique le possessif peut exprimer un sémantème à deux arguments 'posséder', dont un des arguments, le possesseur, est incorporé (426c). Mais le possessif peut également correspondre à un argument, il est alors gouverné sémantiquement par le nom qu'il modifie. Ainsi en (426d), *sa* représente un argument pronominalisé de *disparition*. Enfin, parfois la présence du déterminant est plutôt contrainte par le contexte syntaxique et sa sémantique est moins claire.

(426) a. J'ai acheté ce livre.
b. J'ai acheté le livre rouge.
c. J'ai pris son livre.
d. J'ai appris sa disparition.

Pour les déterminants, on peut distinguer deux traitements pour les différents cas cités. Si le déterminant ne correspond pas à un argument, il s'agit alors d'un quantifieur. Nous considérons que le déterminant n'a pas à proprement parler d'argument sémantique, son apparition pouvant être totalement déterminée par d'autres facteurs (par exemple, dans le cas d'une relative restrictive, on aura systématiquement un défini). Il est cependant représenté par un arbre auxiliaire (Figure 159). Pour marquer qu'il ne s'agit pas d'une relation sémantème-argument, mais d'une relation d'un autre type (la quantification), on peut ne pas numéroter le noeud pied. Le lien entre le déterminant et l'élément sur lequel il s'adjoint est en prendre en compte par un mécanisme supplémentaire pour le traitement de la quantification.

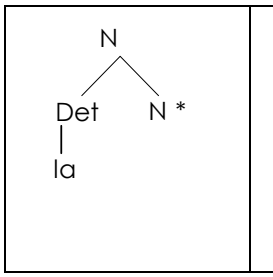


Figure 159

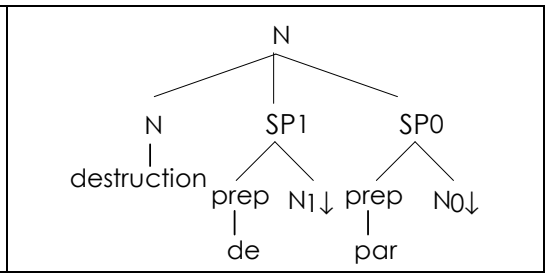


Figure 160

Au contraire, si le déterminant correspond à un argument, il doit être substitué (Figure 161), et les arbres pour les noms correspondants doivent comprendre une position à substituer appropriée (Figure 162).

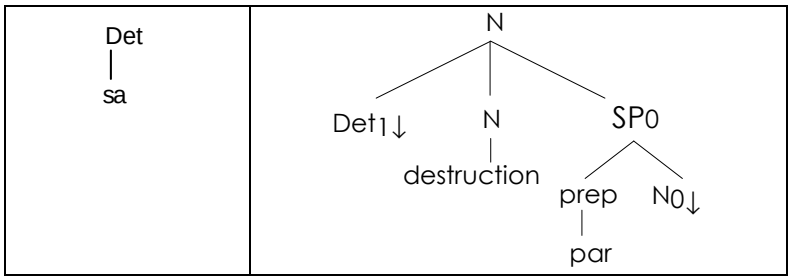


Figure 161 Figure 162

3.2 Quelques cas où le PCPA n'est pas respecté

Le PCPA requiert que les positions noeuds pied ou à substituer représentent systématiquement un argument sémantique de l'ancre. En pratique, nous avons cité des cas où certains éléments sont substitués, sans qu'ils constituent des arguments sémantiques de l'ancre. C'est le cas par exemple de la copule ou d'un verbe d'état, utilisés comme "auxiliaire du passif".
Pour représenter la variation possible pour l'auxiliaire du passif (cf. Chapitre 4, on peut avoir aussi bien la copule, que tout verbe d'état, et l'on a alors l'interprétation stative du passif), et pour représenter le caractère unique et obligatoire de cet auxiliaire du passif, le moyen le plus simple en pratique est de substituer cet auxiliaire.
On peut donc penser à affaiblir le PCPA en requérant que tous les arguments aient une position argumentale prévue (comme noeud pied ou à substituer), sans requérir que toutes les positions à étendre soient argumentales.

Un autre cas où cet affaiblissement est nécessaire concerne les relatives avec "pied-piping", comme en (427) par exemple.

(427) La personne au père de laquelle Van Gogh avait donné un tableau

Nous donnons Figure 163 la combinaison d'arbres nécessaire pour dériver (427). L'arbre élémentaire pour le verbe comportant un complément SP relativisé (cf. Chapitre 5) comporte quatre positions à étendre, pour seulement trois arguments. Dans le cas sans pied-piping (par exemple *la personne$_i$ à laquelle$_i$ VanGogh avait donné* ...), la position N* et la position N2↓ sont étendues par des éléments qui sont coréférents. Mais dans le cas avec pied-piping comme en (427), ces deux positions ne sont pas coréférentes (*personne* n'est pas coréférente à *le père de laquelle*, mais à *laquelle*). On peut voir que c'est la position N* qui n'est pas systématiquement un argument du verbe de la relative. D'où l'intérêt d'affaiblir le PCPA, en ne considérant comme argumentales que certaines positions.

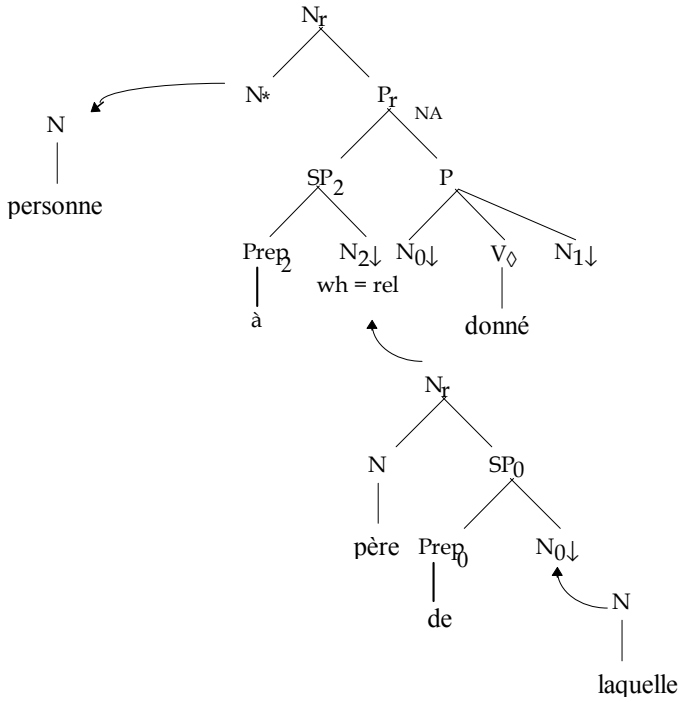


Figure 163. Combinaisons d'arbres pour dériver (427)

Dans l'interprétation de l'AD, les arcs correspondants à l'extension de positions non argumentales ne doivent pas être pris en compte[18]. Ainsi pour l'AD pour le SN donné en (427), les dépendances obtenues sont-elles correctes, à ceci près qu'il faut un mécanisme supplémentaire pour récupérer le sémantème à associer au pronom relatif Figure 164.

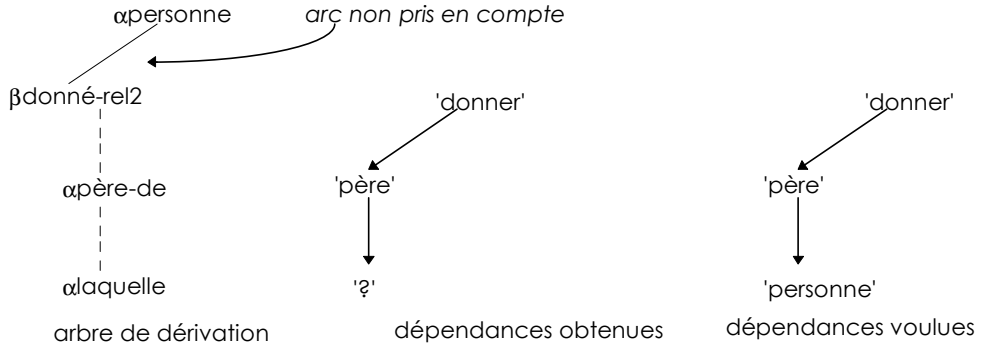


Figure 164. Arbre de dérivation pour (427), dépendances obtenues et dépendances attendues

3.3 Divergences dans l'interface syntaxe/sémantique

Nous n'avons pas considéré tous les cas d'interfaces syntaxe/sémantique. En effet si l'on reprend les différentes configurations possibles quant à la dépendance syntaxique pouvant exister entre un gouverneur sémantique et un de ses arguments (Table 14), on a trois

[18] Dans un précédent travail (Candito et Kahane, 1998b), nous considérions les dépendances sémantiques fournies par l'AD comme erronées, mais il suffit en fait de ne pas prendre en compte un des arcs de l'AD.

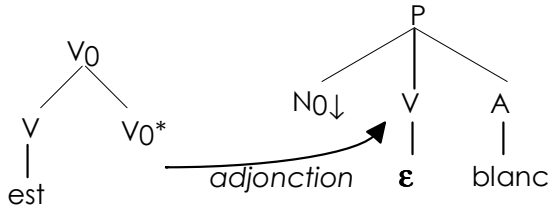
possibilités : 1. il y a dépendance syntaxique dans le même sens; 2. il y a dépendance syntaxique en sens inverse; 3. il n'y a pas dépendance syntaxique[19]. Pour chaque argument apparaissant dans un arbre élémentaire, on a a priori ces trois possibilités. Mais on peut remarquer qu'un seul argument peut entrer dans le cas 2 (i.e. se réaliser comme le gouverneur syntaxique de l'ancre), puisqu'un dépendant syntaxique a un seul gouverneur syntaxique. Pour un arbre élémentaire dans son ensemble (pouvant avoir plusieurs arguments sémantiques), on ne retient comme possibilités que les cas suivants, où un seul des arguments peut ne pas être dépendant syntaxique de l'ancre[20] :

I. Tous les arguments de l'ancre se réalisent en syntaxe comme des dépendants syntaxiques de l'ancre.
II. Un des arguments de l'ancre se réalise en syntaxe comme le gouverneur syntaxique de l'ancre.
III. Un des arguments de l'ancre n'est pas relié à l'ancre par une dépendance syntaxique dans un sens ou dans l'autre.

Le cas I est celui d'un arbre initial ou d'un arbre auxiliaire enchâssant. Le cas II est celui d'un arbre auxiliaire modifieur. Reste le cas III. TAG a plusieurs solutions pour ce cas. Nous prenons deux exemples, en (428a) un adjectif attribut, en (428b) une infinitive sous un verbe à montée, en (428c) une infinitive sous un verbe à contrôle.

(428) a. Le cheval est blanc.
b. Pierre semble regretter son acte.
c. Marie veut manger une polenta.

En (428a), 'cheval' est sémantiquement gouverné par 'blanc', mais en syntaxe le gouverneur syntaxique est la copule, qui n'apporte sémantiquement que les marques de temps et aspect (mais un autre verbe d'état aurait un apport sémantique plus important). Le PCPA impose pour représenter ces cas, d'utiliser un arbre élémentaire ancré par l'adjectif, avec une position pour le nom. La copule peut être adjointe, mais techniquement, il faut prévoir pour cela des arbres élémentaires qui comportent un noeud verbal dominant la chaîne vide (cf. l'analyse pour l'anglais, XTAG group, 1998). Cela n'est pas très satisfaisant d'un point de vue syntaxique[21]. Mais il est clair qu'il est difficile à la fois de ménager les principes liés à la sémantique (PMS et PCPA) et de conserver des arbres syntagmatiques uniquement motivés en syntaxe.



[19] Le cas 4 (dépendance syntaxique sans dépendance sémantique) n'est pas à traiter ici, puisqu'on part d'une dépendance sémantique.

[20] On peut bien sûr imaginer des cas complexes, combinant deux types de configurations entre l'ancre et un premier argument et l'ancre et un second argument. Par exemple la paire (ancre, argument1) pourrait relever du cas 2, et la paire (ancre, argument2) pourrait relever du cas 3. Mais il semble difficile de trouver des exemples.

[21] Du point de vue computationnel, l'utilisation d'une trace comme co-tête pour le noeud V attendant l'adjonction n'a aucune incidence, étant donné que l'arbre comporte une ancre réalisée phonologiquement (*blanc*). L'arbre ne sera sélectionné en analyse que si la phrase comprend le mot *blanc*.

Figure 165. Adjonction de la copule sur l'arbre d'un adjectif en emploi attribut (adapté pour le français, de l'analyse du XTAG group (1998))

Pour un verbe d'état autre que la copule, la contribution sémantique est plus importante, mais reste un prédicat unaire. Pour *Le cheval semble éreinté*, on obtient l'arbre de dérivation de la Figure 166 (en négligeant le déterminant), qui est satisfaisant du point de vue des dépendances sémantiques que l'on peut en tirer.

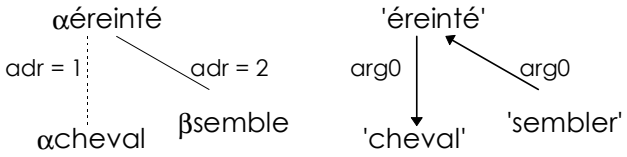


Figure 166. Arbre de dérivation et interprétation en dépendances sémantiques pour *Le cheval semble éreinté*

En (428b), le sujet est gouverné sémantiquement par l'infinitif, sans qu'il y ait dépendance syntaxique directe entre les deux (même si on postule l'existence d'un sujet non réalisé pour l'infinitif). En syntaxe, le sujet dépend du verbe à montée, sans en dépendre sémantiquement. Cette divergence sémantique/syntaxe est traitée par adjonction (cf. Chapitre 3, la section sur les verbes déficients pour l'analyse précise). Pour dériver (428b), *Pierre* est substitué dans l'arbre de l'infinitif, qui comprend bien une position pour chacun de ses arguments sémantiques. Puis *semble*, dont le sémantème est un prédicat unaire, est adjoint, ce qui est schématisé en (428b').

(428) b'. Pierre [semble] regretter son acte.

En (428c), la situation est différente puisque le sujet est gouverné sémantiquement à la fois par l'infinitif et par le verbe à contrôle. Ce "partage" ne peut pas être représenté directement en TAG, c'est-à-dire qu'on ne peut pas prévoir une position à étendre pour le sujet à la fois dans l'arbre du verbe à contrôle et dans l'arbre de l'infinitive. Il s'agit d'un cas évident où le PCPA ne peut pas être complètement tenu. L'analyse retenue dans la TAG du français est de prévoir une position à substituer pour le sujet dans l'arbre du verbe à contrôle, mais pas dans l'arbre de l'infinitive. Le "partage" d'argument sémantique est représenté par la coindexation de traits (voir Chapitre 5, la section sur la réalisation des arguments phrastiques, pour l'analyse précise).

(428) c'. [Marie veut] manger une polenta.

Enfin en ce qui concerne le cas n°4 prévu Table 14, c'est-à-dire le cas d'une dépendance syntaxique entre deux éléments sans qu'il y ait dépendance sémantique, il y a également plusieurs cas de figure en TAG. Un premier cas est celui où l'un des deux éléments en jeu (le gouverneur syntaxique ou le dépendant syntaxique) est sémantiquement vide. Le PMS impose que dans ce cas l'élément sémantiquement vide soit représenté comme une co-tête. Par exemple en français, le sujet non référentiel *il* est directement présent dans des arbres élémentaires spécifiques pour la construction impersonnelle (qu'elle alterne ou pas avec un correspondant personnel).
Un autre cas est fourni en (428b), où l'on a vu que le sujet dépend syntaxiquement du verbe à montée, sans qu'il y ait de dépendance sémantique entre les deux. L'analyse par adjonction du verbe à montée permet de résoudre la divergence entre syntaxe et sémantique.

Cette analyse repose crucialement sur le fait de pouvoir adjoindre un élément qui vienne "prendre" la tête syntaxique de la structure.
Cette analyse est utilisée également pour les verbes à montée de l'objet en anglais, par le XTAG group, qui suivent en cela GB. Ainsi, pour dériver (429), l'arbre de *believe* est adjoint sur une infinitive avec sujet.

(429) [I believe] John to be the most stupid teacher in the world.
?Je crois Jean être le professeur le plus stupide du monde.

Mais comme nous l'avons déjà mentionné au Chapitre 3, l'adéquation de cette analyse aux exigences d'interface avec la sémantique se fait cependant au détriment de l'analyse syntaxique, dans la mesure où les régularités syntaxiques liées au statut d'objet de l'argument "monté" ne sont pas capturées. Donc par exemple, on doit considérer deux analyses, selon que /BELIEVE/ est à l'actif (analysé comme verbe à montée de l'objet, adjoint à une phrase) ou au passif (analysé comme verbe à montée du sujet, adjoint à un noeud verbal). L'assignation du cas accusatif au sujet de l'infinitive doit petre faite par un mécanisme supplémentaire, alors qu'elle serait naturelle si l'argument monté était considéré comme un objet. Pour le français on a le même problème, avec en outre la nécessité de faire monter le clitique sur le verbe enchâssant.
Ces cas de verbes à montée de l'objet posent un dilemme si l'on veut à la fois ménager les principes sémantiques et la syntaxe. On peut cependant remarquer qu'ils sont relativement rares, que l'objet de ces constructions n'a généralement pas toutes les propriétés d'objet, et que dans certains cas, l'objet acquiert le statut d'argument sémantique du verbe enchâssant (cf. Chapitre 3). Dans la MGf et la MGi, nous avons considéré pour les verbes à montée de l'objet une analyse comme verbes à contrôle (cf. Chapitre 3). Les éventuelles déficiences syntaxiques (comme l'absence de passif par exemple) sont bloquées au cas par cas, via le lexique.

Enfin, pour des exemples comme *un silence confus de l'enfant*, ou *franchement, il n'aurait pas du dire cela*, on voit bien que l'on ne peut pas prétendre récupérer les dépendances sémantiques appropriées. Tout ce que l'on peut imposer pour les arbres élémentaires correspondants, c'est qu'ils contiennent le bon nombre de positions à étendre pour représenter les arguments. Mais on ne peut pas assurer que ces positions correspondront toujours à un argument effectif. Pour *confus*, on utilise un arbre élémentaire classique pour un adjectif en emploi épithète. Il peut ou pas se combiner avec un arbre qui représente l'argument de l'adjectif (c'est le cas dans *enfant confus*, mais pas *dans silence confus*). Pour *franchement*, on utilise un arbre classique pour un adverbe de phrase, qui s'adjoint sur un noeud phrastique (cf. Chapitre 6, la représentation des adverbes). Mais le noeud pied phrastique ne représente jamais l'argument sémantique de cet adverbe, cet argument étant le locuteur.

3.3.1 Le PCPA ne garantit pas à lui seul une interprétation correcte de l'arbre de dérivation

Plus généralement, on doit conclure que l'interprétation de l'arbre de dérivation comme un graphe de dépendances sémantiques n'est pas garantie par le seul respect du PCPA : le PCPA impose qu'un arbre élémentaire comprenne une position à étendre pour chaque

argument sémantique exprimé de l'ancre, mais cela ne garantit pas que ces positions vont être effectivement remplies par le bon argument sémantique.
Plusieurs exemples peuvent être fournis pour lesquels on peut prévoir une divergence systématique de ce type, i.e. où une position argumentale (noeud pied ou noeud à substituer), n'est pas remplie par le "bon" argument sémantique.

(430) a. Paul pense que la pluie a cessé de tomber.
b. Que Paul veuille bien rester a surpris Marie.

Pour (430a), la divergence provient du fait que le verbe à montée est la tête sémantique de la complétive, en effet, "l'objet de pensée" est la terminaison de la pluie (cf. supra la discussion sur le statut de modifieur sémantique ou pas des verbes à montée du sujet). L'arbre pour le verbe à complétive *pense que* comporte bien une position argumentale pour "l'objet de pensée", mais il s'agit d'un noeud pied phrastique. Donc il ne pourra jamais être étendu par un verbe à montée, qui est représenté par un arbre de racine verbale (cf. Figure 167, les adjonctions nécessaires pour dériver (430a)).

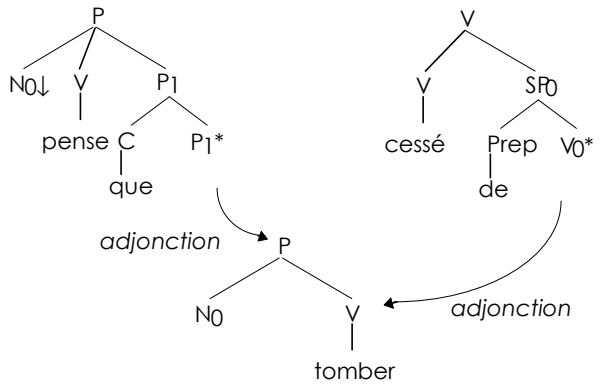


Figure 167. Adjonctions pour dériver (430a).

Nous donnons Figure 168 l'arbre de dérivation pour (430a), les dépendances que l'on peut y lire, qui diffèrent des dépendances attendues.

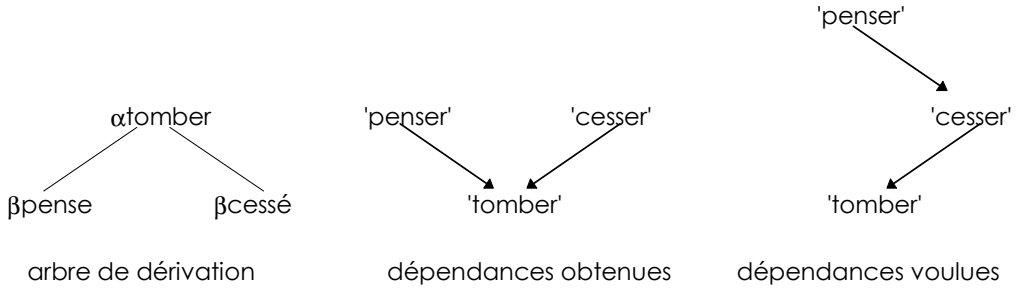


Figure 168. Arbre de dérivation pour (430a), dépendances obtenues et dépendances attendues

L'exemple (430b) ne constitue pas véritablement un problème, mais conforte plutôt la généralisation qui veut que soient adjoints des modifieurs sémantiques. En (430b), les dépendances sémantiques en jeu sont données Figure 169. Nous considérons que la tête sémantique de la complétive sujet c'est 'vouloir' (le sujet de la surprise c'est la volonté de Paul), donc le verbe pont *veuille* n'est pas employé comme modifieur sémantique ici, il n'est donc pas sensé être adjoint, mais doit plutôt recevoir sa complétive par substitution

(ce qui permet de dériver le bon ordre des mots, l'adjonction n'est nécessaire du point de vue syntaxique que dans le cas d'extraction hors de la complétive). Le verbe *surprend* reçoit lui sa complétive sujet par substitution, ce qui permet de représenter l'impossibilité d'extraire hors de la complétive sujet (cf. Abeillé, 1991). Du point de vue sémantique, 'surprendre' constitue la tête sémantique du tout, il est donc valable de ne pas représenter ce verbe par adjonction. On peut vérifier qu'alors les dépendances obtenues à partir de l'AD sont correctes.

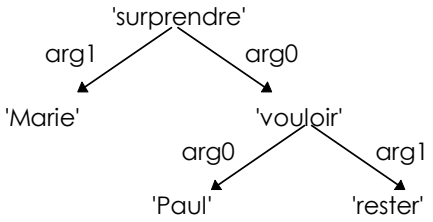


Figure 169. Dépendances sémantiques pour (430b)

Cependant, si l'on choisit de traiter dans tous les cas les verbes pont par adjonction (pour avoir un traitement uniforme de ceux-ci, qu'il y ait extraction ou pas hors de la complétive), les dépendances obtenues ne sont pas correctes (cf. Figure 170 et Figure 171). Cela confirme que la généralisation sur le critère de typage des arbres élémentaires en arbres initiaux et arbres auxiliaires est pertinente : dans les cas où elle n'est pas respectée, les dépendances sémantiques obtenues peuvent ne pas être correctes.

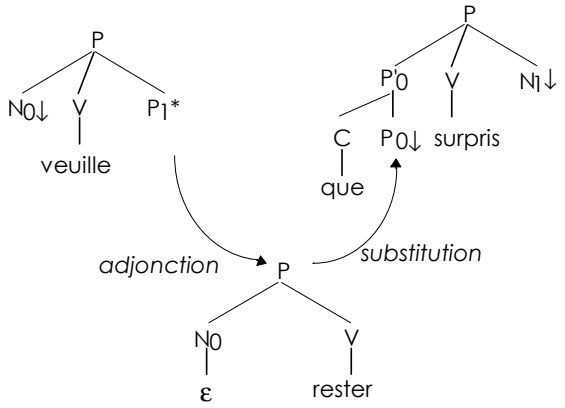


Figure 170. Combinaisons d'arbres pour dériver (430b)

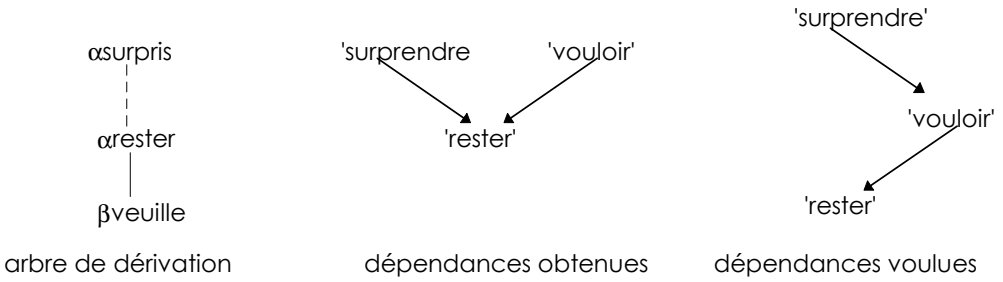


Figure 171. Arbre de dérivation pour (430b) (simplifié), si l'on veut adjoindre le verbe pont (*veuille*) dépendances obtenues et dépendances attendues

Ces exemples donnent un aperçu minime de la complexité de l'interface syntaxe/sémantique. On ne peut pas prétendre qu'une grammaire TAG réalise complètement cette interface, sans calcul supplémentaire. Cependant nous pensons que les principes sémantiques cités peuvent être relativement bien tenus pour ce qui est des catégories majeures, et permettent de faciliter un calcul supplémentaire, de toutes façons nécessaire pour les questions de quantification (pour lesquelles les déterminants jouent un rôle) et de portée.

4 Application à l'analyse ou la génération

Malgré les problèmes cités, il reste que dans bon nombre de cas, l'arbre de dérivation représente assez bien les dépendances sémantiques, mises à part les relations de quantification, et la pronominalisation. Cela a bien sûr des applications directes pour l'utilisation des TAG, à la fois en analyse et en génération.

Pour l'analyse, on peut envisager un calcul sémantique à partir de l'arbre de dérivation. On peut également envisager de rassembler des données statistiques sur des dépendances sémantiques lexicalisées. C'est une des motivations, en germes, pour la définition de la dérivation proposée par Shieber et Schabes (1994) (cf. Chapitre 1). Retenant leur définition de la dérivation, Schuler (1998) propose explicitement un modèle statistique de dépendances sémantiques, permettant d'estimer les probabilités pour un analyseur TAG stochastique (Resnik, 1992).

Les TAG probabilistes ont été proposées comme amélioration des CFG probabilistes (Jelinek et al., 1990) et comme modèle probabiliste lexicalisé permettant de prendre en compte les dépendances structurales entre mots.
Les CFG probabilistes sont des grammaires hors-contexte, où une probabilité est associée à chaque règle CF de la grammaire. On peut alors calculer la probabilité d'une analyse comme le produit des probabilités des règles qui ont servi à la dérivation. Cette définition utilise l'hypothèse d'indépendance d'application des différentes règles.
Le postulat pour permettre à une CFG probabiliste de fournir l'analyse correcte parmi l'ensemble d'analyses pour une phrase donnée, est de considérer que l'analyse correcte est celle ayant la plus forte probabilité (définie comme ci-dessus). Or ce postulat est faux (cf. Resnik, 1992). Un simple exemple permet de s'en convaincre. Supposons que le corpus d'entraînement contienne plus d'emplois de verbes intransitifs que de verbe avec complément SP, alors la règles VP $\rightarrow$ V sera plus probable que VP $\rightarrow$ V SP, et ce indépendamment des exigences de sous-catégorisation du verbe effectif.
D'autres types de modèles probabilistes pour la désambiguisation ont été proposés, qui prennent en compte la probabilité de dépendances syntaxiques entre deux mots (voir par exemple Collins, 1997). Les TAG stochastiques apparaissent comme un moyen d'allier lexicalisation et informations structurales, indépendamment de la distance effective entre deux mots en relation de dépendance (cf. la factorisation de la récursion par l'opération d'adjonction).
Dans une TAG probabiliste, une probabilité est associée à tout arc (d'adjonction ou de substitution) susceptible d'apparaître dans une dérivation. La probabilité d'une dérivation est alors calculée en fonction de la probabilité de tous les arcs en jeu. La dérivation la plus

probable pour une phrase est celle qui a la probabilité la plus élevée parmi toutes les dérivations possibles de cette phrase.

Ce modèle intéressant au niveau théorique se heurte cependant à une difficulté majeure : tel quel, il définit un espace de probabilités trop important (toutes les adjonctions et les substitutions possibles). On doit donc d'une manière ou d'une autre assigner des probabilités non pas aux adjonctions ou substitutions possibles, mais à des classes de substitutions ou d'adjonctions. Un type de classement peut consister à "délexicaliser", i.e. compter dans une même classe toutes les substitutions (à la même adresse) non pas d'un arbre a[lexème fléchi1] dans un arbre b[lexème fléchi2], mais d'un schème a dans un schème b. Le modèle perd cependant l'avantage majeur d'être sensible aux ancres lexicales en jeu, dans son assignation d'une probabilité à une combinaison entre deux schèmes.
L'alternative proposée par Schuler (1998) est de classer les arcs de substitution et adjonctions non pas selon les schèmes en jeu, mais selon les dépendances sémantiques en jeu[22], entre deux lexèmes. L'idée est de faire abstraction des différents moyens syntaxiques d'exprimer une dépendance sémantique similaire.

(431) a. Des enfants fabriquent ces vêtements.
b. Des vêtements que fabriquent des enfants
c. Des vêtements fabriqués par des enfants

La Figure 172 donne les trois arcs d'arbre de dérivation qui représentent le lien entre *vêtements* et une forme de /FABRIQUER/. Ces trois arcs sont lus comme une même dépendance sémantique.

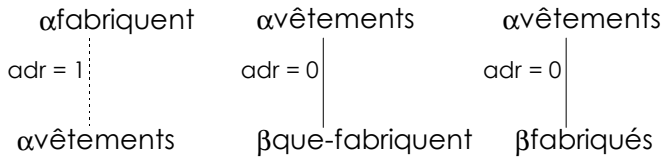


Figure 172. Arcs d'arbre de dérivation pour les trois liens syntaxiques entre /VETEMENT/ et /FABRIQUER/

On peut en outre imaginer un niveau d'abstraction supplémentaire, qui serait de classer les dépendances entre sémantèmes plus généraux que ceux définis pour chaque lexème, c'est-à-dire considérer des sémantèmes valant pour plusieurs lexèmes[23], en particulier un lexème verbal et sa nominalisation (c'est ce qui est considéré au sein de la TST). Cela revient par exemple à considérer pour un sémantème comme 'fabriquer', le lexème nominal FABRICATION, comme réalisation possible.

(431) d. La fabrication des vêtements par des enfants est interdite.

[22] Schuler (1998) a proposé en parallèle de nos travaux la lecture explicite des arcs d'adjonction d'arbres auxiliaires modifieurs comme des dépendances sémantiques à lire "dans l'autre sens", et propose d'utiliser cette lecture aux TAG probabilistes. Il n'a cependant pas réalisé l'étude précise de l'interprétation de l'arbre de dérivation, et de sa généralisation à tous les types d'adjonction.
[23] La Théorie Sens-Texte de Mel'cuk fait un grand usage de cette notion de sémantème pouvant ête réalisés par des lexèmes différents, éventuellement de catégorie syntaxique différente. Voir également le système G-TAG (Danlos, 1998b), où sont la base lexicale associée à un concept peut comprendre des réalisations verbales ou nominales.

Plus généralement, on peut également envisager la définition de classes d'équivalence de sémantèmes, pour grouper par exemple des verbes converses (comme /ACHETER/ et /VENDRE/)[24]. La dépendance sémantique entre *Le garçon vend ...* et celle entre *achète au garçon* peuvent être comptabilisées comme instance d'une même dépendance, légèrement plus abstraite.

[24] Par exemple, le système G-TAG utilise un niveau conceptuel dont les concepts peuvent être réalisés par des converses.

Chapitre 8
Applications et perspectives

Nous décrivons dans ce chapitre l'évaluation de la TAG compilée de la MGf, ainsi que les applications réalisées ou envisagées, du formalisme de la MG, des MG pour le français et l'italien, et de la procédure d'interprétation de l'arbre de dérivation.

1 Grammaires de référence du français et de l'italien

Les grammaires compilées à partir de la MGf et la MGi peuvent constituer des grammaires de référence pour ces deux langues, dans la mesure où elles comprennent la représentation de contraintes fines, ainsi que certains phénomènes relativement marginaux (comme par exemple certaines diathèses comme le passif impersonnel, relativement courant pour des verbes à objet phrastique mais soutenu pour des verbes à objet nominal).
Leur représentation au sein de la métagrammaire fournit de plus un outil performant pour la maintenance de ces grammaires, et leur adaptation pour différentes applications, pourvu que l'on dispose d'un analyseur ou générateur implémentant la dérivation au sens des grammaires d'arbres adjoints. En particulier, elles peuvent être utilisées pour l'analyse ou la génération. On peut dans ce cas chercher à utiliser des grammaires réversibles, mais cela n'est pas obligatoire. Pour une application, certaines contraintes codées dans la MGf et la MGi peuvent être considérées comme inutiles et être ignorées (c'est en particulier le cas pour la génération, cf. Section 2, l'utilisation de la MGf pour une grammaire utilisée dans le système de génération G-TAG). La représentation des grammaires au sein de la métagrammaire permet justement de compiler des grammaires simplifiées et/ou variant sur certains points par rapport à la grammaire "de référence".

1.1 Evaluation de la TAG du français, sur le corpus de test TSNLP

La TAG du français formée des schèmes compilés de la MGf et des lexiques morphologique et syntaxique existants à TALANA a été évaluée en utilisant le corpus de test TSNLP (*Test Suite for Natural Language Processing*, Estival et Lehmann, 1997), et en utilisant l'analyseur syntaxique pour les TAG, intégré au système XTAG (Paroubek et al., 1992).

La procédure d'évaluation (extraction des phrases françaises du corpus, écriture automatique des scripts de lancement d'analyse) a été mise en place par Benoît Robichaud. Nous avons analysé tous les "items" du corpus, classés par catégorie (phrase, SN, SA, SP), par grammaticalité[1] et par phénomènes (selon la classification TSNLP).
Les résultats obtenus pour les items grammaticaux sont fournis Table 15[2].

Les lexiques ont été complétés de façon à ce qu'aucun mot ne soit inconnu. Un prétraitement morphologique effectue quelques "désamalgames" (les prépositions + articles[3], prépositions + pronom relatif ...).
Les principales sources d'échecs sont :

- Une sous-catégorisation initiale manquante pour un élément lexical (par exemple, un verbe transitif sans objet, une utilisation transitive de PARLER (*parler hindi*);
- Schèmes manquants (arbres causatifs, impératifs avec clitique post-verbaux), qui n'avaient pas été encore définis lors de l'évaluation;
- Non couverture de certains phénomènes d'accord (pour la forme de politesse en français, ou la coordination "distribuée", comme dans *deux bandes bleue et jaune*);
- Non couverture de certaines constructions (comme les relatives en *à* ou le "gapping")

Les ambiguïtés fréquemment relevées concernent principalement des ambiguïtés d'attachement, qui se traduisent en TAG par des ambiguïtés sur le site d'une adjonction. En particulier, une ambiguïté qui peut être considéré comme artificielle est celle d'un modifieur adjoint soit au noeud ?P soit au noeud ?Vmarq, pour un verbe intransitif (où dont l'unique complément est extrait). D'autres ambiguïtés de site d'adjonction sont linguistiquement pertinentes. Par exemple, en ne contraignant pas le trait :det pour l'adjonction d'une relative, on permet deux possibilités d'analyse ((det N) relative) ou (det (N relative)), ce qui est interprétable comme la distinction relative restrictive / relative appositive.
On relève également systématiquement une ambiguïté pour un syntagme prépositionnel entre une analyse comme complément et une analyse comme modifieur, par exemple pour le complément d'agent.

[1] TSNLP utilise 4 niveaux d'acceptabilité. Outre des items acceptables (acceptabilité 1) et des items totalement inacceptables (acceptabilité 0), sont également considérées des items plutôt inacceptables, dont l'inacceptabilité porte sur un phénomène mineur. Nous n'avons pris en compte que les deux premiers cas (acceptabilités 0 et 1).
[2] Nous ne fournissons pas d'évaluation en termes d'efficacité de traitement, c'est-à-dire sous forme de temps de calcul, car l'objet de l'évaluation est la grammaire TAG du français, et non pas l'analyseur du système XTAG, qui est aujourd'hui dépassé en termes de rapidité d'exécution et en termes de fonctionnalités. En particulier, il ne fournit aucune analyse partielle, ce qui fait que l'évaluation que nous fournissons est stricte, au sens où ne sont considérées comme analysées que les phrases totalement analysées.
[3] *Des* et *du* sont systématiquement considérés comme *de* + *les*, *de* + *le*.

Nombre de mots	**Nombre d'items**	**Nombre d'items** ayant reçu au moins 1 analyse	**% d'items** ayant reçu au moins 1 analyse	**Nbre moyen d'analyses / item** ayant reçu au moins 1 analyse
1 mot	9	8	88.88	1
2 mots	114	98	85.96	1.42
3 mots	231	201	87.01	1.85
4 mots	168	134	79.76	1.40
5 mots	258	215	83.33	1.87
6 mots	183	158	86.33	2.24
7 mots	190	157	82.63	2.76
8 mots	148	112	75.67	2.25
9 mots	90	72	80	4.40
10 mots	87	59	67.81	3.69
11 mots	37	27	72.97	8.96
12 mots	36	25	69.44	10.27
13 mots	31	18	58.06	11.16
14 mots	14	7	50	12.28
15 mots	22	14	63.63	23.21
16 mots	13	8	61.53	10.75
17 mots	8	1	12.5	16
19 mots	2	1	50	4
22 mots	2	0	0	-
TOTAL	1643	1315	80.03	2.96

Table 15. Evaluation pour les phrases grammaticales TSNLP (par longueur)

Phénomène[4]	Nbre d'items	Nombre d'items ayant reçu au moins 1 analyse	% d'items ayant reçu au moins 1 analyse
negation	106	56	52.83
c-agreement	75	60	80
np-agreement	143	135	94.40
tense	83	82	98.79
modification	291	277	95.18
passive	162	148	91.35
question	189	159	84.12
imperative	110	64	58.18
conditional	13	7	53.84
c-complementation	180	166	92.22
ap-complementation	9	7	77.77
np-complementation	12	10	83.33
np-coordination	167	118	70.65
pp-coordination	84	38	45.23
pron-coordination	4	4	100
det-n-ap-coord	10	7	70
det-ap-coord	4	4	100

Table 16. Evaluation pour les phrases grammaticales TSNLP (par phénomène)

On peut comparer ces résultats avec une évaluation antérieure réalisée en 1996, avec une grammaire de taille réduite, comprenant environ 830 schèmes (il s'agissait d'une grammaire écrite entièrement manuellement). Les lexiques étaient identiques (mis à part les changement dans la classification en familles, opérés entre la version 1996 de la grammaire et la version actuelle). En 1996, 65% des phrases grammaticales étaient analysées, mais les phrases avec coordination n'avaient pas été prises en compte[5]. Cependant, l'augmentation de la couverture observée avec la grammaire actuelle va de pair avec une augmentation de l'ambiguïté, puisqu'en 1996, on obtenait 1,5 analyses par phrase.

La grammaire a été également évaluée dans sa capacité de rejeter des phrases agrammaticales. La Table 17 fournit les résultats pour les items agrammaticaux de TSNLP, classés par phénomènes. 82% sont rejetées (contre 91% en 1996, mais là encore, mis à part les items comprenant une coordination). Les domaines de surgénération sont principalement la concordance de temps, la négation et la coordination, qui requièrent encore du travail. Il est de plus à noter que certains jugements d'inacceptabilité du corpus TSNLP sont discutables, en particulier, les

[4] Nous gardons la terminologie de la classification par phénomènes TSNLP.

[5] Avec la grammaire actuelle, si l'on met à part les phrases avec coordination, 86% des phrases sont analysées.

contraintes sur la concordance de temps entre adverbe et temps grammatical sont trop restrictives[6].

Phénomène	Nombre d'items	Nombre d'items ayant reçu au moins 1 analyse	% d'items rejetés
tense	279	42	84.94
conditional	18	11	38.88
imperative	175	21	88
question	157	65	58.59
modification	50	37	26
negation	218	24	88.99
passive	112	33	70.53
c-agreement	144	16	88.88
np-agreement	514	38	92.60
c-complementation	514	67	86.96
ap-complementation	22	0	100
np-complementation	19	5	73.68
np-coordination	159	38	75.64
pp-coordination	40	2	95
pron-coordination	13	13	0
det-n-ap-coord	30	0	100
det-ap-coord	29	0	100
TOTAL	2300	402	82.52%

Table 17. Evaluation pour les phrases **a**grammaticales TSNLP (par phénomène)

Pour l'italien, nous n'avons pas procédé à une évaluation de cette ampleur, faute de lexiques adéquats (en outre TSNLP ne comprend pas de phrases italiennes). Nous avons seulement réalisé quelques tests d'analyses, en utilisant un lexique morphologique compilé du lexique MULTEXT, comprenant 150000 formes fléchies, et un lexique syntaxique de test, écrit manuellement.

Nous fournissons à l'annexe C quelques exemples d'analyses (sous la forme d'arbres dérivés et d'arbres de dérivation), pour le français et l'italien.

L'évaluation de la TAG du français que nous venons de décrire a permis d'évaluer la couverture de la grammaire, et de mettre en évidence certaines lacunes. La MGf doit être modifiée/augmentée en conséquence (et justement une des motivations principales de la représentation sous forme de MG est de faciliter la maintenance).
Pour autant, un réel système d'analyse syntaxique du français, performant et robuste, n'a pas été encore mis en place, faute d'analyseurs performants. Par chance, récemment, des analyseurs basés sur les TAG ont été réalisés ou sont en cours de réalisation, qui sont

[6] Par exemple *Demain l'ingénieur vient* est classé comme inacceptable.

nettement plus performants que l'analyseur du système XTAG (voir par exemple Issac, 1997; Kinyon, 1997; Lopez, 1998, Kercadio, 1998; Halber, 1998). Une collaboration devrait avoir lieu entre TALANA et le LoRIA à Nancy, pour coupler l'analyseur de P. Lopez aux TAG développées à TALANA. Une collaboration est également prévue avec A. Halber, pour le test de son analyseur probabiliste[7].

2 Application en génération de textes : Compilation de TAG pour le système G-TAG

Nous avons plusieurs fois cité le générateur de textes G-TAG, qui prend en entrée une représentation conceptuelle et fournit en sortie un texte exprimant cette représentation. La réalisation linguistique est faite en s'appuyant sur une grammaire TAG : le générateur construit à partir de la représentation conceptuelle un arbre de dérivation, par spécification progressive d'un arbre de "g-dérivation", dont certains noeuds sont des varibales qui divent être instanciées par des fragments d'arbre de dérivation.
G-TAG a été utilisé pour générer des textes de divers domaines techniques (logiciel, chimie et aéronautique), en français et en anglais. Les schèmes de la TAG du français utilisée ont été compilés à partir de la MGf.
Utiliser la MGf pour compiler la grammaire utilisée par G-TAG présente deux avantages majeurs.

- G-TAG s'appuie sur une représentation des schèmes comme la combinaison d'un nom de famille et d'une structure de traits atomiques (appelés t-traits). Par exemple, l'arbre pour un transitif strict, à arguments nominaux, représentant un passif sans agent est représenté par l'information n0Vn1[T-passif, T-sans-agent][8]. Cette représentation est obtenue de manère immédiate avec une représentation de la grammaire qu sein de la MG : il suffit d'associer aux classes pertinentes de la MGf l'instanciation des t-traits appropriés (ainsi par exemple, le t-trait T-passif est-il associé à la classe passif-personnel, et le t-trait T-sans-agent est-il associé à la classe agt-objet-omis.
- Un second avantage d'utiliser la représentation au sein de la MG est qu'il a été possible d'opérer des coupes dans la grammaire : le générateur pour une application donnée n'utilise pas toutes les constructions prévues (comme par exemple le moyen impersonnel, ou le passif impersonnel dans le cas d'arguments nominaux). Par souci d'efficacité, il vaut mieux ne pas charger au sein du générateur des schèmes inutiles. Les coupes sont obtenues de manière immédiate par une compilation à partir de la MGf (il suffit de ne pas prendre en compte certaines classes terminales des dimensions 1, 2, ou 3, ou bien d'ajouter des contraintes de cooccurrence négatives, lors de la compilation.

[7] Une comparaison pourra alors être réalisée avec un autre système complet d'analyse du français, tel que le système FIPS (Laenzingler et Wehrli, 1991; Wehrli, 1997:214-233), qui est basé sur une grammaire de type GB, pour le français.
[8] Les t-traits donnés explicitement valent +, les t-traits non mentionnés valant par défaut -.

La compilation d'une TAG italienne pouvant servir de grammaire à G-TAG est en cours (les schèmes sont disponibles, il faut cependant construire les lexiques adéquats).

Une autre application qui pourrait mettre à profit la représentation des connaissances syntaxiques au sein de la métagrammaire est la génération en langue contrôlée. La représentation en MG permet de définir aisément un sous-ensemble de grammaire, correspondant aux spécifications de la langue contrôlée, cette grammaire pouvant alors être utilisée pour de la génération en langue contrôlée (voir par exemple Lux (1998), pour le français, avec une partie applicative s'insérant dans le système G-TAG, et donc basée sur une grammaire TAG).

3 Supertagging

Parce qu'une TAG est entièrement lexicalisée, on peut concevoir les schèmes comme des "supercatégories" associées aux lexèmes fléchis d'une phrase. Parmi tous les schèmes que peut ancrer un lexème fléchi, un seul est pertinent pour un emploi donné du lexème fléchi, dans le cas où la phrase le comprenant n'est pas ambiguë.
De là, Srinivas (1997) propose d'appliquer les techniques d'étiquetage automatique (le tagging) pour un "super-étiquetage" (le supertagging), qui consiste à essayer d'assigner la super-étiquette correcte aux mots d'une phrase (i.e. le schème que ce mot doit ancrer dans l'analyse syntaxique correcte de la phrase). Ces techniques sont principalement l'utilisation de modèles statistiques (trigrammes).

Mais bien sûr, la difficulté du super-étiquetage par rapport à l'étiquetage est sans comparaison. Alors qu'un mot peut être étiqueté par, disons, une dizaine au grand maximum de catégories, un mot, en particulier un verbe, peut être super-étiqueté par des centaines de schèmes (i.e. peut ancrer des centaines de schèmes)[9].

C'est là où une représentation hiérarchisée d'une TAG a son rôle à jouer : elle permet de définir une caractérisation des schèmes comme des structures de traits (a priori atomiques). Une structure de traits sous-spécifiée subsume un ensemble de schèmes. Par exemple la structure de traits a-un-sujet = Ninv, a-un-gén-objet = relativisé, subsume tous les schèmes qui comportent un sujet nominal inversé et un gén-objet en position relativisée (qu'il s'agisse de schèmes à l'actif ou au passif).
Au lieu d'essayer d'associer une super-étiquette, tâche qui s'avère trop difficile si la grammaire a une couverture assez large, on peut n'assigner que des traits (ce qui est déjà cité comme perspective par Srinivas (1997)). Il faut pour cela définir les attributs pertinents, dont les valeurs sont de fait mutuellement exclusives, tels que les schèmes doivent, pour être bien formés, fixer une valeur unique pour chaque attribut. On peut alors appliquer les techniques d'étiquetage automatique, en étiquetant non pas avec des catégories ou des super-catégories, mais en étiquetant avec des valeurs pour un attribut donné. Par exemple, pour tout verbe, il faut assigner une valeur (la plus probable) au

[9] De plus, une super-étiquette s'applique en théorie à un ensemble de mots (cf. un schème peut comprendre plusieurs ancres), pas forcément contigus.

trait a-un-objet, cette valeur pouvant être - (pour aucun objet direct), relativisé (pour un objet relativisé) etc...

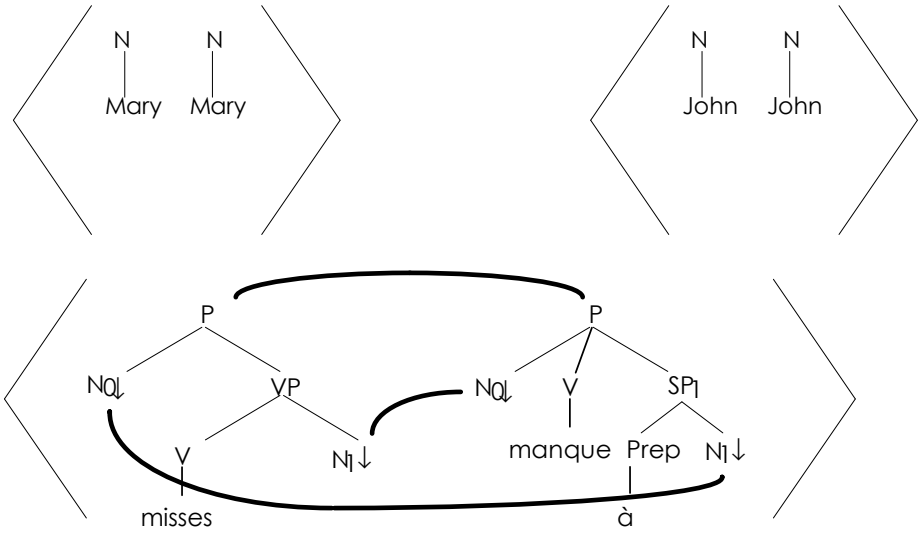
On peut envisager diverses techniques pour étiqueter (étiquetage probabiliste par tri-grammes, étiquetage à l'aide d'heuristiques ...). Des heuristiques peuvent être définies, qui diffèrent selon l'attribut considéré (des heuristiques différentes vont permettre de conclure à la présence d'un élément relativisé, ou à la présence d'un sujet inversé).

La difficulté réside dans la définition des attributs pertinents, c'est-à-dire de définir la partition la plus appropriée pour représenter les différents schèmes. On peut envisager de réaliser de multiples essais, en modifiant cette partition. Une autre difficulté réside dans le fait que les attributs diffèrent selon la catégorie du mot à étiqueter. On ne peut donc pas simplement appliquer un étiquetage par tri-grammes, en considérant les deux valeurs étiquetées pour les deux mots précédents le mot courant, puisque pour les deux mots précédents l'attribut considéré ne sera vraisemblablement pas pertinent.
Bien que nous n'ayons pas à l'heure actuelle de réponses précises à apporter à ces questions, nous envisageons de travailler dans cette direction à l'avenir. Ce type d'application permet à notre avis de marier des traitements probabilistes avec des descriptions linguistiques riches.

4 Traduction automatique

Enfin, nous pouvons envisager une application de la représentation d'une TAG au sein de la MG, en traduction automatique.
Abeillé et al. (1990) ont proposé d'utiliser les TAG synchrones pour la traduction automatique. Nous avons déjà parlé des TAG synchrones appliqués à la dérivation synchrone d'une analyse syntaxique et d'une analyse sémantique. En traduction automatique, l'idée est de coupler une TAG syntaxique pour la langue source avec une TAG syntaxique pour la langue cible. La traduction consiste en trois étapes : une phrase source est analysée, l'arbre de dérivation obtenu est associé à un arbre de dérivation pour la langue cible, puis la phrase de la langue cible est générée en opérant les combinaisons définies par l'arbre de dérivation cible.
L'étape d'association d'un arbre de dérivation source (Ad source) à un arbre de dérivation cible (AD cible) passe par l'association à chaque arbre élémentaire apparaissant comme noeud de l'AD source, d'un arbre élémentaire correspondant pour la langue cible. Ces paires d'arbres élémentaires doivent être fournies dans le lexique de transfert. En outre, au sein d'une paire d'arbres élémentaires, des liens doivent être établis entre certains noeuds, tels que les combinaisons ayant lieu à un noeud ont lieu au noeud qui lui est relié (c'est la base de la définition des dérivations synchrones). Plus précisément, chaque fois que dans un arbre élémentaire source (Asource), un noeud lié (Nsource, lié à Ncible) reçoit une adjonction ou une substitution d'une arbre Bsource, le noeud Ncible, dans l'arbre apparié à Asource, reçoit l'adjonction ou la substitution d'un arbre Bcible, tel que Bsource et Bcible sont appariés.

Nous reprenons ci-dessous l'exemple de paires d'arbres élémentaires anglais/français fournies par Abeillé et al. (1990), utilisées pour traduire *John misses Mary/Mary manque à John.*



Les auteurs citent divers exemples de divergences classiques apparaissant entre langue source et langue cible, qui peuvent être simplement résolues avec une approche à la TAG synchrone. L'exemple ci-dessus montre comment gérer la différence de sous-catégorisation pour des prédicats. TAG permet également de coupler un arbre élémentaire dont les ancres sont Vsupport + nom avec un verbe (pour traduire par exemple *give a cough* par *tousser*).

Si la proposition est attractive, elle repose sur un lexique de transfer fait de paires d'arbres élémentaires (lexicalisés), pour lesquels des liens entre noeuds sont établis, en particulier pour les positions argumentales à étendre. Il est évident que la construction d'un tel lexique doit être en partie automatisée. Nous pensons que la représentation des TAG source et cible au sein de la MG peut en fournir le moyen.
L'idée est de ne requérir pour le lexique de tranfert que la déclaration pour les deux lexèmes de leur sous-catégorisation initiale (i.e. la famille de schèmes qu'ils sélectionnent) et la correspondance entre arguments sémantiques. Un programme doit être capable de dériver automatiquement les paires d'arbres élémentaires, avec les liens entre les positions argumentales, grâce à une représentation des schèmes au sein de la MG (qui fournit l'information concernant les positions argumentales, ainsi que l'information sur le type de construction qu'illustre un schème). Donc par exemple, pour la paire de lexèmes /MISS/ /MANQUER/, il faudrait déclarer entre les sous-catégorisations initiales (arg0/sujet; arg1/objet) et (arg0/sujet; arg1/a-objet) la correspondance suivante : arg0→arg1; arg1→arg0.
Il faut distinguer différents cas de transfert entre lexèmes :

1. Le cas où les familles source et cible sélectionnées sont les mêmes (par exemple entre /EAT/ /MANGER/)

2. Le cas où les familles diffèrent (comme entre /MISS/ /MANQUER/). On peut en outre distinguer dans ce dernier cas
    2a. Le cas où les deux familles existent pour les deux langues, mais que pour une paire de lexèmes la sélection de la famille diffère dans les deux langues,
    2b. Le cas où une famille (une sous-catégorisation initiale) n'existe pas dans l'autre langue (ou du moins n'a pas de parallèle exact). C'est le cas par exemple pour les ditransitifs en anglais, comme /GIVE/, dont la sous-catégorisation initiale peut être (arg0/sujet; arg1/objet; arg2/to-objet), avec la fonction to-objet définie comme pouvant subir le dative-shift.

Pour le cas 1, les sous-catégorisations des lexèmes cible et source sont les mêmes, ce qui entraîne que les familles comprennent à peu de chose près les mêmes schèmes, modulo les différences de propriétés des fonctions pour les deux langues (un objet en français n'a pas les mêmes propriétés qu'un objet en anglais).
On peut dans ce cas fournir un appariement de schèmes assez simplement, avec un appariement systématique des arguments sémantiques de même indice, et un appariement des constructions syntaxiques.
Nous pensons que cet appariement peut être réalisé en représentant (via une représentation dans la MG) les schèmes comme des structures de traits, et en appariant des traits. Dans le cas 1, les différents traits définissant un schème source peuvent être simplement reportés pour la définition du schème cible.
Par exemple, pour apparier un schème pour un sujet relativisé d'un intransitif (comme dans *(l'homme) qui partit*) avec le correspondant anglais, en supposant une correspondance arg0→arg0 entre les deux lexèmes, on peut garder pour le schème cible, les mêmes traits que le schème source (à savoir le trait n0V (pour la sous-catégorisation initiale) et le trait arg0-relativisé).

Il est cependant à noter que l'appariement peut ne pas être bijectif, même dans ce cas. Par exemple, pour le français, la réalisation des actants comme clitique est distinguée de la réalisation nominale (ce qui donne lieu à des schèmes différents). Cette distinction n'est pas faite en anglais, qui comprend des pronoms forts uniquement.
Donc par exemple le schème anglais pour un transitif à arguments nominaux, avec arguments en position canonique) doit être apparié à quatre schèmes en français (celui avec sujet et objet en position canonique, celui avec sujet clitique, celui avec objet clitique, et celui avec sujet et objet clitiques). Pour une traduction dans le sens français→anglais, cela ne pose pas de problèmes, mais en sens inverse, il faut fournir un mécanisme pour pouvoir choisir entre les quatre possibilités[10].

Pour le cas 2a) comme pour le cas 2b), un même argument sémantique au moins, est associé dans les deux langues à deux fonctions grammaticales différentes. Il est donc important de réaliser un appariement de traits entre arguments et pas entre fonctions finales.

[10] Le choix entre les différentes possibilités doit être guidé par la réalisation des actants. Pour *She eats fish*, le mécanisme de traduction de *she* doit soit pointer vers un arbre de racine Cl dans le cas d'un traitement syntaxique de clitiques, soit remonter un trait clitique sur l'arbre élémentaire verbal, dans le cas d'un traitement des clitiques en morphologie. Dans les deux cas, cela impliquera la sélection d'un arbre élémentaire verbal avec sujet clitique (en faisant échouer les autres sélections).

Supposons que l'on veuille calculer l'arbre à apparier à un arbre élémentaire ancré par *misses*, avec sujet relativisé en anglais, et objet en position canonique (pour dériver par exemple la relative *(the man who misses Mary*). L'arbre cible à calculer peut être illustré par la relative *(l'homme) à qui Mary manque*.
Le schème source doit être représenté comme la combinaison des traits suivants :

[n0Vn1, sans-redistribution, arg0-relativisé, arg1-pos-can]

Voyons comment sont traduits ces différents traits, pour construire une structure de traits représentant le schème cible. Le trait représentant la famille (la sous-catégorisation initiale), doi être traduit en fonction des sous-catégorisations initiales fournies dans le lexique. Donc dans notre exemple, n0Vn1 doit être traduit en n0Van1. Le trait sans-redistribution est simplement reporté pour le schème cible.
Pour "traduire" les deux autres traits (arg0-relativisé, arg1-pos-can), il faut utiliser la correspondance entre arguments fournies dans le lexique (dans notre exemple arg0→arg1; arg1→arg0), et remplacer les indices argumentaux source par les indices argumentaux cible. Ainsi arg0-relativisé doit devenir arg1-relativisé et arg1-pos-can doit devenir arg0-pos-can. C'est la grammaire du français qui impose que la combinaison de traits obtenue [n0Van1, sans-redistribution, arg1-relativisé, arg0-pos-can] corresponde à un schème cible de structure assez différente du schème source[11] (où en particulier les arguments 0 et 1 ont les fonctions sujet et a-objet respectivement).
Avec cet exemple, nous avons montré l'importance d'exprimer les traits représentant les schèmes non pas en faisant référence aux fonctions finales (par exemple sujet-relativisé) mais en faisant référence aux indices argumentaux. Ces traits peuvent être calculés grâce à une représentation des schèmes au sein de la MG (on dispose à la fois de l'information sur la fonction finale associée à un argument, et du type de réalisation pour chaque fonction finale).

Cependant, une difficulté supplémentaire réside dans le fait que dans le cas 2 (a ou b), il faut s'attendre à ce que certaines réalisations pour un argument ne soit pas disponible pour l'argument dans la langue cible, du fait que les deux arguments portent une fonction différente au départ[12]. De plus certains les redistributions applicables aux sous-catégorisations initiales peuvent différer. Par exemple, pour la paire /GIVE/ /DONNER/, la sous-catégorisation initiale de /GIVE/ permet le dative-shift, qui n'existe pas en français. Dans ce cas on doit se rabattre sur un arbre "sans dative-shift" en français.
Ainsi pour traduire *Paul gave Mary a flower*, il faut avoir apparié la combinaison de schèmes [n0Vn1ton2, dative-shift, arg0-pos-can, arg1-pos-can, arg2-pos-can] avec la combinaison [n0Vn1an2, sans-redistribution, arg0-pos-can, arg1-pos-can, arg2-pos-can], qui fournit bien le schème adéquat pour dériver *Paul donna une fleur à Mary*.

[11] Pour un trait donné argi-pos-can, la langue cible peut faire correspondre différents sous-cas. Nous avons déjà donné supra l'exemple de l'éventuelle réalisation clitique des actants en français, qui est appariée en anglais à une réalisation en position canonique. Donc on doit obtenir pour notre exemple plusieurs schèmes cible [n0Van1, sans-redistribution, arg1-relativisé, arg0-pos-can] et [n0Van1, sans-redistribution, arg1-relativisé, arg0-clitique].
[12] Et même si l'argument porte la même fonction dans les deux langues, les fonctions peuvent différer dans le détail. Par exemple certaines langues ne vont pas permettre la relativisation d'un complément oblique.

Prenons un cas un peu plus compliqué d'un dative-shift au passif, avec en outre omission de l'agent. Pour dériver en anglais *Mary was given a flower*, on utilise un schème que l'on peut représenter sous la forme de la combinaison de traits suivante :

[n0Vn1ton2, dative-shift-puis-passif, arg0-omis, arg1-pos-can, arg2-pos-can].

Une traduction adéquate serait *On a donné une fleur à Mary*. Pour calculer les traits définissant le schème cible, on obtient via le lexique la "traduction" du trait n0Vn1ton2 par n0Vn1an2. Pour obtenir l'arbre cible adéquat, il faut faire correspondre à la séquence de redistributions dative-shift-puis-passif, la séquence sans-redistribution. Comme tous les argi correspondent à un argument de même indice en français. les traits restants sont repris tels quels. On obtient alors la combinaison :

[n0Vn1an2, sans-redistribution, arg0-omis, arg1-pos-can, arg2-pos-can]

Cependant, la combinaison des traits (n0Vn1an2, sans-redistribution) fournit pour arg0 la fonction finale sujet, qui ne compte pas parmi ses réalisations l'omission. Il faut donc prévoir une réalisation de sujet à faire correspondre à une omission en anglais. Cette réalisation en français peut être, comme nous l'avons donné dans la traduction, un sujet indéfini *on*.

Nous avons pris des exemples de traduction anglais/français, car les deux langues présentent des disparités plus grandes que pour le couple italien/français. Dans le détail cependant, le même type de problème se pose. En particulier, nous avons vu la correspondance systématique que l'on peut établir entre une construction impersonnelle en français et une construction avec sujet inversé en italien (qu'il s'agisse du sujet d'un inaccusatif ou d'un inergatif). On a ainsi la paire *Il est arrivé trois personnes/Sono arrivate tre persone*.

Le mécanisme à prévoir est légèrement différent de celui illustré supra. En effet, il faut faire correspondre non pas un trait à un autre, mais une paire de traits (impersonnel, arg0-pos-can) à une paire de traits, en distinguant en outre le cas d'un inergatif, pour lequel la paire de traits cible est (sans-redistribution, arg0-inversé) du cas d'un inergatif, pour lequel la paire de traits cible est (objet→sujet, arg0-inversé) (en considérant l'implémentation simplifiée de l'inaccusativité fournie au Chapitre 4).

Nous avons fourni une piste pour le développement d'un algorithme d'appariement automatique d'arbres élémentaires et de positions argumentales, pour la construction semi-automatique d'un lexique de transfert, utilisable par des TAG synchrones. Cette construction s'appuie sur la représentation des schèmes comme des combinaisons de traits, permise par une représentation via la MG. Une telle procédure est de toutes façons indispensable pour espérer utiliser réellement les TAG synchrones pour la traduction. Il ne s'agit bien sûr que d'une ébauche succincte, et dans le détail de nombreux points sont à travailler. Nous pensons cependant qu'il s'agit d'une application prometteuse.

Conclusion

Nous avons défini la métagrammaire, comme un formalisme permettant une représentation hiérarchisée de la syntaxe d'une langue; nous avons défini et implémenté une procédure de compilation de cette représentation en un ensemble de schèmes pour une grammaire d'arbres adjoints. Cette procédure suit des principes de bonne formation pour les schèmes construits, principes que nous avons énoncés en utilisant les notions de fonctions grammaticales, initiale et finale, empruntées à et adaptées de la Grammaire Relationnelle. Nous avons décrit un fragment conséquent de syntaxe française et italienne, comme deux métagrammaires parallèles.
Enfin, nous avons étudié le statut linguistique de l'arbre de dérivation en TAG. Il apparaît qu'en faisant certaines hypothèses sur les arbres élémentaires, l'arbre de dérivation peut être lu comme un graphe de dépendances sémantiques, malgré quelques cas problématiques. Inversement, à partir d'un graphe de dépendances sémantiques, un arbre de dérivation peut être calculé, modulo la gestion de la coréférence.

L'architecture de description syntaxique proposée, à deux niveaux (la métagrammaire et un formalisme lexicalisé tel que les grammaires d'arbres adjoints) comporte plusieurs avantages :

- Disposer d'un niveau de description plus abstrait et hiérarchisé (qui utilise l'héritage de propriétés entre types généraux et sous-types plus spécifiques) permet de s'affranchir des choix d'un formalisme implémenté particulier (en l'occurrence les grammaires d'arbres adjoints).
- Disposer d'un niveau de description hiérarchisé rend possible à la fois une description fine des contraintes syntaxiques, et l'expression de généralisations qui rendent compte des régularités de la langue. Du point de vue pratique, la redondance étant limitée grâce au partage d'informations, les métagrammaires du français et de l'italien peuvent ainsi être manipulées et maintenues avec souplesse. Leur réutilisation peut être envisagée (ce qui est l'exception en matière de grammaires électroniques) : elles peuvent constituer des grammaires de référence, en cela qu'elles reprennent des analyses linguistiques modernes, et relativement consensuelles, et être adaptées via la représentation au sein de la MG, pour différentes applications, d'analyse et/ou de génération.
- Par la définition au niveau général une organisation des informations, à vocation multilingue (voire universelle), et le respect de ces contraintes par le compilateur MG→TAG, le formalisme prend en charge des contraintes linguistiques (avec par exemple un principe contrôlant la satisfaction des exigences de sous-catégorisation); il constitue un cadre paramétrable pour différentes langues, comprenant des contraintes à vocation universelle, que l'on peut envisager d'enrichir dans le futur, en permettant d'intégrer des paramètres représentant la variation entre langues. Nous avons choisi d'utiliser à la fois les concepts de dépendance (typées en fonctions

grammaticales) et de constituance, ce qui constitue la tendance générale de ces dernières années, où l'on reconnaît la nécessité de ne pas réduire les fonctions à des positions structurales. Nous avons gardé la notion de diathèse de base, ce qui revient à considérer des fonctions initiales (ou profonde), qui peuvent différer des fonctions finales (ou surfaciques), malgré la tendance actuelle à définir des modèles déclaratifs. Ce choix a été guidé principalement par les données italiennes, avec l'existence en particulier d'un passif causatif, qu'il nous semble peu justifié de décrire de manière déclarative.

- La définition et l'implémentation du compilateur MG→TAG permet une utilisation effective dans un système implémenté, en bénéficiant des avantages non négligeables de TAG, qui sont en particulier un pouvoir génératif contraint, donnant lieu à des algorithmes de reconnaissance/dérivation en temps polynomial, ainsi que la lexicalisation intégrale de la grammaire.

Outre la possibilité d'application en ingénierie linguistique, ce système à deux niveaux permet de tester par des moyens informatiques les hypothèses retenues pour le modèle linguistique en général et la description d'une langue en particulier, tant pour leur cohérence que pour leur adéquation à décrire.

- Pour le formalisme des TAG, l'ajout d'une représentation au sein de la métagrammaire constitue un enrichissement du formalisme, puisque celui-ci prend en charge certaines contraintes de bonne formation sur les structures élémentaires construites (les schèmes construits satisfont les exigences de sous-catégorisation supposées) et sur la complétude de la grammaire (les croisement valides de phénomènes sont opérés de manière systématique).

  Cela constitue également un enrichissement des structures élémentaires, avec possibilité de référer explicitement aux notions de fonctions grammaticales, finales et initiales, et d'indice argumental. Cette possibilité est mise à profit pour des applications telle que la génération d'une TAG pour le système de génération G-TAG. Les applications envisagées au Chapitre 8 reposent elles aussi sur cet enrichissement des structures élémentaires d'une TAG. Elles constituent des axes de recherche intéressants.

Les phénomènes syntaxiques abordés pour le français et l'italien couvrent une bonne partie de la syntaxe de ces langues, pour le verbe et la phrase. La représentation des modifieurs est plus succincte, elle est cependant originale dans la mesure où elle utilise le principe de TAG d'une sous-catégorisation "sémantique", i.e. comprenant les arguments sémantiques, l'un d'entre eux pouvant être la tête syntaxique.
Nous avons fourni un contraste systématique pour les deux langues, qui bien que très proches, diffèrent dans le détail. Nous nous sommes appuyée sur une grammaire TAG du français existante (Abeillé, 1991). Sa "redescription" au sein de la métagrammaire a permis d'en vérifier la cohérence et d'en augmenter la couverture. Le travail pour l'italien a été réalisé par contraste systématique avec les analyses pour le français.
La description fournie couvre les trois domaines d'information retenus dans la métagrammaire : la sous-catégorisation initiale, les redistributions, les réalisations des fonctions grammaticales.
Le premier domaine a supposé de dégager les fonctions grammaticales retenues pour les deux langues, et il s'agit de notions traditionnelles.

Le domaine des redistributions a fait l'objet d'une étude assez poussée des données. L'originalité du travail tient au fait que nous couvrons les changements de diathèse majeurs (passif, impersonnel, moyen, causatif, réfléchis), en fournissant systématiquement les contraintes d'interaction entre ces différents phénomènes. Pour l'italien, nous avons retenu l'hypothèse inaccusative, en considérant des sous-catégorisations finales sans sujet possibles. Outre les différences syntaxiques classiques en inaccusatifs et inergatifs, cela nous a permis de rendre compte de distinctions fines dans le contexte du *si* nominatif, mises en évidence par Cinque (1988a).
Le domaine des réalisations de fonctions couvre les phénomènes principaux, pour les positions canoniques, inversées, extraites et clitiques. Nous avons proposé un mécanisme intégrant le représentation des clitiques pronominaux en morphologie (Miller, 1991), ce qui permet de résoudre plusieurs problèmes inévitables avec un traitement des clitiques en syntaxe en TAG.

Nous avons également réalisé une recherche, en collaboration avec Sylvain Kahane, sur le statut linguistique des unités en TAG, avec en particulier la proposition d'un principe de bonne formation, qui permet une caractérisation uniforme des arbres élémentaires. Si ce travail n'est pas radicalement nouveau, mais s'appuie sur une idée couramment avancée que le résultat de l'analyse TAG est « proche de la sémantique », une étude complète n'était cependant pas disponible. Nous pensons que cette étude peut fournir des bases théoriques solides pour des applications de traitement automatique utilisant les TAG, non seulement pour la dérivation syntaxique mais également sémantique.

Références bibliographiques

Abeillé A. 1991 : *Une grammaire lexicalisée d'Arbres Adjoints pour le français*. Thèse de Doctorat, Université Paris 7.

Abeillé A. 1992 : Synchronous TAGs and French pronominal clitics. Actes de COLING'92, Nantes, pp. 60-66.

Abeillé A. 1993 : *Les nouvelles syntaxes*, Armand Colin.

Abeillé A. 1994a : *Syntaxe et sémantique : interactions dans une grammaire d'unification*. In Rastier et al. 1994, pp. 141-165.

Abeillé A. 1994b : Syntax or semantics? Handling nonlocal dependendies with MCTAGs or Synchronous TAGs. Computational Intelligence, vol. 10, n°4, pp. 471-485.

Abeillé A. à paraître : Extraction out of NP and clitic-noun dependencies in French. In Abeillé et Rambow (éds.) à paraître.

Abeillé A. 1998 : Verbes "à montée" et auxiliaires dans une grammaire d'arbres adjoints. *Lynx*.

Abeillé A. et Schabes Y. 1989 : Parsing idioms with a lexicalized Tree Adjoining Grammar. Actes du 4ème European ACL, Manchester.

Abeillé A., Schabes Y. et Joshi A. 1990. Using Lexicalized TAG for Machine Translation. Actes de COLING'90. Helsinki

Abeillé A., Godard D. 1996a : La complémentation des auxiliaires en français, *Langages* 122, pp. 32-61.

Abeillé A., Godard D., Miller P., Sag I. 1996b : French unbounded dependencies. In Salari S. et Dini L. (éds), *Romance in HPSG*, CSLI, Chicago Press.

Abeillé A., Daille B., Candito M.-H et Robichaud B. 1996c : FTAG un analyseur de phrases françaises basé sur les grammaires d'arbres adjoints. Actes de ILN'96, Nantes.

Abeillé A. et Godard D. 1997 : The syntax of French negative adverbs, in Forget, Hirschbuhler, Martineau et Rivero (éds.), *Negation and polarity syntax ans semantics*. J. Benjamins, Amsterdam, pp. 1-27.

Abeillé A., Godard D. et Miller P. 1997 : Les causatives en français : un cas de compétition syntaxique. *Langue française*, n°115, pp. 62-74.

Abeillé A., Godard D., Sag I. 1998 : Two kinds of composition in French complex predicates. *Syntax and Semantics*, 30, *Complex Predicates in Nonderivational Syntax* Hinrichs, Kathol et Nakasawa (éds.), pp. 1-42.

Abeillé A. et Rambow O. à paraître. *Tree Adjoining Grammars*. CSLI publications, Stanford.

Abeillé A. et Candito M.-H. à paraître : FTAG : A Lexicalized Tree Adjoining Grammar for French. In Abeillé et Rambow (éds.) à paraître.

Acquaviva P. 1991 : Funzioni delle frasi subordinate : Frasi argomentali : Completive e soggettive. In Renzi et Salvi (éds.), 1991, pp. 633-674.

Agard F. 1984 : *A course in Romance Linguistics, volume 1 : A synchronic view*, Georgetown University Press, Washington.

Alsina A. 1996 : *The role of Argument Structure in Grammar: Evidence from Romance*. Lecture Note Series. CSLI publications, Stanford.

Barbero C. Lesmo L. Lombardo V. et Merlo P. 1998 : Integration of syntactic and lexical information in a hierarchical dependency grammar. Actes de l'atelier *Processing of Dependency-based Grammars*. ACL-COLING'98, Montréal.

Bashung K. 1992 : *Grammaires d'unification à traits et contrôle des infinitives en français*. Adosa, Clermont Ferrand.

Becker T. 1993 : *HyTAG: A new Type of Tree Adjoining Grammars for Hybrid Syntactic Representation of Free Word Order Languages*. Thèse de Doctorat, Université de Saarbrücken.

Becker T. 1994 : Patterns in Metarules. Actes 3ème colloque international sur les Grammaires d'Arbres Adjoints (TAG+3), Paris.

Belletti A. et Rizzi L. 1988 : Psych-verbs and theta theory. *Natural Language and Linguistic Theory,* vol. 6, pp. 291-352.

Benincà P. et Salvi G. 1988 : L'ordine degli elementi della frase e le costruzioni marcate : l'ordine normale degli elementi nella frase semplice. In Renzi et Salvi (éds.), pp. 119-128.

Benincà P. et Cinque G. 1991 : Frasi subordinate al participio: participio presente. In Renzi (éd.), pp. 604-610.

Bertuccelli M. 1991 : Frasi subordinate al participio: participio passato. In Renzi (éd.), pp. 593-603.

Bleam T. 1994. Clitic Climbing and the Power of Tree Adjoining Grammar. Actes de TAG+3, Paris.

Bouillon P. (éd.) 1994 : Traitement automatique des langues naturelles. Editions Duculot / AUPELF/UREF.

Bresnan J. (éd.). 1982 : *The mental representation of grammatical relations*, MIT Press.

Bresnan J. 1982a : The passive in lexical theory. In J. Bresnan (éd.), 1982, pp. 1-86.

Bresnan J. 1982b : Polyadicity. In J. Bresnan (éd.), 1982, pp. 149-172.

Bresnan J. et Kanerva J. 1989 : Locative inversion in Chichewa: A case study of factorization in Grammar. *Linguistic Inquiry*, vol. 20, n°1, pp. 1-50.

Bresnan J. et Zaenen A. 1990 : Deep Unaccusativity in LFG. In Dziwirek, Farell et Mejias-Bikandi (éds.) *Grammatical Relations: A Cross-Theoretical Perspective.* CSLI Publications, Stanford.

Burzio L.1986 : *Italian syntax*, Reidel, Dordrecht.

Candito M.-H. et Kahane S. 1998a: Une grammaire TAG vue comme une grammaire Sens-Texte précompilée. Actes de TALN'98, Paris.

Candito M.-H. et Kahane S. 1998b : Can the derivation tree represent a semantic graph? An answer in the light of Meaning-Text Theory. Actes du colloque TAG+4, Philadelphia.

Candito M.-H. et Kahane S. 1998c : Defining DTG derivations to get semantic graphs. Actes du colloque TAG+4, Philadelphia.

Carpenter R. 1991 : The generative power of Categorial Grammars and Head-driven Phrase Structure Grammar with lexical rules. *Computational Linguistics*, vol. 17, n°3, pp. 301-314.

Carpenter R. 1992 : *The logic of typed feature structures*. Cambridge, Cambridge University Press.

Cinque G. 1988a : On *Si* constructions and the Theory of *Arb*. *Linguistic Inquiry*, vol. 19, n° 4, pp. 521-581.

Cinque G. 1998b : La frase relativa. In Renzi (éd.) 1988, pp.443-506.

Chomsky N. 1957 : Structures syntaxiques. Trad. française, Seuil, 1969.

Collins M. 1997 : Three generative, lexicalised models for statistical parsing. Actes de ACL'97.

Comrie B. 1977 : In defense of spontaneous demotion : the impersonal passive. *Syntax and Semantics*, 8, *Grammatical Relations*, Cole & Saddock (éds.).

Cordin P. et Calabrese A. 1988. I pronomi personali. In Renzi (éd.) 1988, pp. 535-592.

Daelemans W., De Smedt K. et Gazdar G. 1992 : Inheritance in Natural Language Processing. *Computational Linguistics*, vol. 18, n°2, pp. 205-218.

Danlos L. et Meunier F. 1996 : G-TAG, un formalisme pour la génération de textes : présentation et applications industrielles. Actes de ILN'96, Nantes.

Danlos L. 1998a : Linguistic ways for expressing a discourse relation in a lexicalized text generation system. Actes de l'atelier *Discourse Relations and discourse markers*, ACL-COLING'98, Montréal.

Danlos L. 1998b : G-TAG : un formalisme lexicalisé pour la génération de textes inspiré de TAG. *Traitement automatique des langues (T.A.L.)*, vol. 39, à paraître.

Davis A. 1995 : *Linking Theory and the Hierarchical Lexicon*. Thèse de PhD., Université de Stanford.

Dini L. 1994 : Italian absolute phrases: an interaction between syntax and semantics. In *Quaderni del laboratorio di linguistica*'9/94.

Dini L. 1995 : Unaccusative behaviors. In *Quaderni del laboratorio di linguistica*'9/95.

Ducrot O. et Schaeffer J.-M. 1995 : *Nouveau dictionnaire encyclopédique des sciences du langage*. Seuil, Paris.

Elia A. 1984 : *Le verbe italien : les complétives dans les phrases à un complément*. Biblioteca della ricerca. Schena-Nizet, Paris/Fasano di Puglia.

Estival D. et Lehmann S. 1997 : TSNLP : des jeux de phrases test pour le TALN, *Traitement Automatique des Langues (T.A.L.)*, vol. 38, n°1, pp. 155-172.

Evans R. et Gazdar G. 1989 : Inference in DATR. Actes 4° EACL, Manchester, pp. 66-71.

Evans R., Gazdar G. et Moser L. 1993 : Prioritised Multiple Inheritance in DATR. In *Inheritance, Defaults and the Lexicon*, Briscoe, de Paiva & Copestake (éds.) pp. 137-147. Cambridge University Press.

Evans R., Gazdar G. et Weir D. 1994 : Using Inheritance to describe LTAG. Actes 3° colloque international sur les Grammaires d'Arbres Adjoints (TAG+3), Paris.

Evans R., Gazdar G. et Weir D. 1995 : Encoding Lexicalized Tree Adjoining Grammar with a Nonmonotonic Inheritance Hierarchy. ACL-95.

Fava E. 1991 : Funzioni delle frasi subordinate : Frasi argomentali : Interrogative indirette. In Renzi et Salvi (éds.) , 1991, pp. 675-720.

Flickinger D. 1987 : Lexical Rules in the Hierarchical Lexicon. Thèse de Doctorat. Université de Stanford.

Frank R. 1992 : *Syntactic locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. Thèse de PhD, Université de Pennsylvania.

Fraser N. et Hudson R. 1992 : Inheritance in word grammar. Computational Linguistics, vol. 18, n°2, pp. 133-138.

Frison L. 1988 : L'ordine degli elementi della frase e le costruzioni marcate : Le frasi scisse. In Renzi (éd.) pp. 194-225.

Gaatone D. 1993 : Les locutions verbales et les deux passifs en français. Langages, 109, Larousse, pp. 37-52.

Gazdar G., Klein E., Pullum G. et Sag I. 1985 : *Generalized Phrase Structure Grammar*, Harvard University Press.

Godard D., Sag I., Abeillé A., mns : Reflexive verbs, impersonal and causative constructions in French. Manuscript non publié (1996).

Godard D. 1988 : *Les relatives en français*. Editions CNRS.

Grévisse M. 1969 : *Le bon usage*. Duculot, Liège.

Grimshaw J. 1982 : On the Lexical Representation of Romance reflexive clitics. In Bresnan (éd.), 1982, pp. 87-148.

Grimshaw J. 1990 : *Argument structure*. Linguistic Inquiry monograph 18. MIT Press.

Gross M. 1968 : *Grammaire transformationnelle du français : Syntaxe du verbe*. Larousse. Paris.

Gross M. 1969 : La notion d'objet en français. *Langages*, vol. 1.

Gross M. 1975 : *Méthodes en syntaxe*. Hermann, Paris.

Habert B. 1995 : TALN : la programmation à objets sans objet?. Actes colloque Traitement Automatique du Langage Naturel, Marseille, juin 95.

Halber A. 1998 : Tree-Grammar linear typing for unified super-tagging/probabilistic parsing models. Actes du colloque TAG+4, Philadelphia, pp. 54-57.

Harris Z. 1976 : *Notes du cours de syntaxe*. Seuil, Paris.

Hyams N. 1986 : The interpretation and acquisition of Italian impersonal *Si*. *Syntax and Semantics,* 19, *The syntax of pronominal clitics*, Borer (éd.), pp. 219-240.

Issac F. 1997 : *Analyse Syntaxique et Apprentissage des Langues*. Thèse de Doctorat de l'Université Paris 13.

Jackendoff R. 1990 : *Semantic Structures*. MIT Press.

Jelinek F., Lafferty J. et Mercer R. 1990 : Basic methods of probabilistic context-free grammars. Technical Report RC 16374, IBM, New York.

Jouannaud M-.P. 1994 : *L'inversion du sujet nominal en français*. Mémoire de DEA de linguistique théorique et formelle. Université Paris 7.

Joshi A., Levy L. et Takahashi M.. 1975 : Tree Adjunct Grammars. Journal of the Computer and System Sciences, 10:1, pp. 136-163.

Joshi A., Vijay-Shanker K. et Weir D. 1991 : The convergence of mildly context-sensitive grammar formalisms. In *Foundational issues in Natural Language Processing*, Sells et al. (éds.), MIT Press, pp. 31-81.

Kahane S. et Mel'cuk I. 1997 : Synthèse des phrases à extraction en français contemporain. Actes de la Conférence de Syntaxe et Sémantique de Paris (CSSP).

Kahane S., de Kercadio Y. et Candito M.-H. en préparation : Une analyse des relatives en TAG qui permet l'interprétation sémantique directe.

Kaplan R. et Bresnan J. 1982 : Lexical-Functional Grammar: A Formal System for Grammatical Representation. In Bresnan (éd.), pp.173-281.

Kasper R. et Rounds W. 1986 : A logical semantics for feature structure. Actes du 24ème ACL, pp. 257-266.

Kasper R., Kiefer B., Netter K. et Vijay-Shanker K. 1995 : Compilation of HPSG to TAG. Actes d'ACL'95, Boston.

Kayne R. (1975) : *French Syntax : the transformational cycle*. MIT Press, Cambridge.

Kayne R. (1976) : French relative *que*. In Hensey et Lujan (éds.), *Current Studies in Romance Linguistics*. Georgetown University Press, pp. 255-299.

Keenan E. et Comrie B. 1977 : Noun-Phrase Accessibility and Universal Grammar. Linguistic Inquiry, 8:1, pp. 63-99.

de Kercadio Y. 1998 : An improved Earley parser for LTAG. Actes du colloque TAG+4, Philadelphia, pp. 84-87.

Kinyon A. 1997 : Un algorithme d'analyse LR(0) pour les grammaires d'arbres adjoints lexicalisées. Actes de TALN'97, Grenoble.

Kulick, S. 1997 : Generalized transformations and restructuring in Romance, in Actes de la Eastern States Conference on Linguistics (ESCOL'97), Yale.

Kulick, S. 1998 : Clitic climbing in Romance: "restructuring", causatives, and object-control verbs, in Actes de TAG+4, Philadelphia, août 98.

Koenig, J.P. 1994 : *Lexical underspecification and the syntax/semantics interface*. Thèse de PhD., University of California at Berkeley, 1994 (révisée 1996).

Koenig J.P. et Jurafsky D. 1994 : Type Underspecification and On-Line type Construction in the Lexicon. Actes West Coast Conference in Linguitics 13 (WCCFL).

Kroch A. et Joshi A. 1985 : Some aspects of the linguistic relevance of Tree Adjoining Grammar. Rapport technique, CIS Dept, Université de Pennsylvanie.

Kroch A. 1987 : Unbounded Dependencies and Subjacency in a Tree Adjoining Grammar, in Manaster-Ramer (éd.), *The Mathematics of Language*, J. Benjamins.

Kroch A. 1989 : Assymetries in long distance extraction in a TAG grammar. In Baltin et Kroch (éds.) : *Alternative conceptions of Phrase Structure*, Chicago Press.

Laenzingler C. et Wehrli E. 1991 : FIPS : Un analyseur interactif pour le français. *TA Informations*, vol. 32, n°2, pp.35-49.

Lamiroy B. 1993 : Pourquoi il y a deux passifs. Langages, 109, Larousse, pp. 53-72.

Lazard G., 1994a : L'actance, puf, Paris.

Lazard G., 1994b : L'actant H : sujet ou objet?, BSL 89/1.

Levin B. et Rappaport M., 1995 : *Unaccusativity. At the Syntax-Lexical Semantics Interface*. Linguistic Inquiry Monograph 26, MIT Press, Cambridge et London.

Lopez P. 1998 : Analyse guidée par connexité de TAG lexicalisées. Actes de TALN'98, Paris, pp. 202-211.

Lux V. 1998 : *Elaboration d'un Français Rationalisé Modulaire (FREM) : Test du FREM en génération automatique.* Thèse de Doctorat, Université Paris 7.

Manzini R. 1986 : On Italian *Si*. *Syntax and Semantics,* 19, *The syntax of pronominal clitics*, Borer (éd.), pp. 241-262.

Marandin J.-M. 1997 : *Dans le Titre se Trouve le Sujet : L'inversion locative en Français*. Manuscript non publié, Université Paris 7.

Mel'cuk I., 1988 : *Dependency Syntax : Theory and Practice*, State University of new York Press, Albany.

Mel'cuk I., 1997 : Vers une Linguistique Sens-Texte, Leçon inaugurale, Collège de France.

Mel'cuk I., à paraître : Dependency in Linguistic Description. In *Dependency and Valency*. Handbook of Linguistics and Communication Science, De Gruyter, Berlin/New York.

Mandelblit N. 1997 : *Grammatical Blending: Creative and Schematic Aspects in Sentence Processing and Translation*. Thèse de PhD, University of California in San Diego.

Meunier F. 1997 : *Implémentation de G-TAG, formalisme pour la génération inspiré des grammaires d'arbres adjoints*. Thèse de Doctorat, Université Paris 7.

Miller P. 1991 : *Clitics and Constituents in Phrase Structure Grammar.* Thèse de doctorat, Université d'Utrecht.

Miller P. et Sag I. 1997 : French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, vol. 15, pp. 573-639.

Monachesi P. 1995 : A grammar of Italian Clitics. Thèse de Doctorat, Tilburg, ITK.

Moreau M.L. 1976 : *C'EST Etude de syntaxe transformationnelle*. Editions universitaires de Mons.

Muller C. 1995 : J'entends l'homme qui bégaie qui bégaie: une autre analyse des relatives de perception. In Bat-Zeev et L. Kupferman (éds.). Mélanges Gaatone, Gent.

Napoli D.J. 1981 : Semantic interpretation vs. lexical governance : clitic climbing in Italian. *Language*, vol. 57, n°4, pp. 841-887.

Paroubek P., Schabes Y. et Joshi A. 1992 : XTAG : a graphical Workbench for developing TAGs. Actes 4° Conference on Applied NL Processing, Trento, pp. 223-227.

Perlmutter, D. 1978 : Impersonal passives and the unaccusative Hypothesis. Actes du Fourth Annual meeting of the Berkeley Linguistics Society.

Perlmutter, D. 1983 : Personal versus impersonal constructions. Natural Language & Linguistic Theory, 1983, vol 1, pp. 141-200.

Perlmutter, D. et Postal, P. (éds.), 1983 : Studies in Relational Grammar 1. University of Chicago Press, Chicago.

Perlmutter, D. et Rosen, C. (éds.) 1984 : Studies in Relational Grammar 2. University of Chicago Press, Chicago.

Polguère A. 1990 : *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*. Thèse de Doctorat, Université de Montréal.

Polguère A. 1992 : Remarques sur les réseaux Sens <=>Texte. In Clas (éd.) *Le mot, les mots, les bons mots*, P.U.M, Montréal.

Pollard C. et Sag I. 1987 : *Information-based Syntax and Semantics*. CSLI Series, University of Chicago Press.

Pollard C. et Sag I. 1994 : *Head-driven Phrase Structure Grammar*. CSLI Series, University of Chicago Press.

Pollock J.Y. 1984 : On case and the syntax of infinitives in French, in Guéron et al. (éds.) *Grammatical Representation*, Foris, Dordrecht.

Pollock. J.Y. 1986 : Sur la syntaxe de *en* et le paramètre du sujet nul. In *La Grammaire modulaire*. M. Ronat et D. Couquaux (éd).

Pustejowski J. 1995 : *The generative Lexicon*. Cambridge, MIT Press.

Rambow O., 1994 . *Natural language syntax and formal systems.* Thèse de Doctorat. Université de Pennsylvania.

Rambow O. et Joshi A. 1992 : A formal look at dependency grammars and phrase-structure grammars, with special consideration of word-order phenomena. In Wanner (éd.) *Current Issues in Meaning-Text Theory*. Pinter, London.

Rambow O., Vijay-Shanker K. et Weir D. 1995 : D-Tree Grammars. Actes d'ACL'95.

Rambow O. et Satta G. 1996 : Synchronous models of language. Actes de ACL'96. Santa Cruz.

Rambow O. et Vijay-Shanker K. 1998 : *Wh*-islands in TAG and Related Formalisms. Actes du colloque TAG+4. Philadelphia.

Rastier F., Cavazza M. et Abeillé A. 1994 : *Sémantique pour l'analyse. De la linguistique à l'informatique*. Masson, Paris.

Recourcé G. 1995 : *L'association négative en français; étude linguistique et formelle de la particule* ne. Thèse de Doctorat. Université Paris 7.

Renzi L. 1988 : *Grande Grammatica di consultazione,* Vol. I. Il Mulino, Bologna.

Renzi L. et Salvi G. 1991 : *Grande Grammatica di consultazione,* Vol. II. Il Mulino, Bologna.

Resnik P. 1992 : Probabilistic tree-adjoining grammar as a framework for statistical natural language processing. Actes de COLING'92, Nantes, pp. 418-424.

Rivière N. 1981 : *La construction impersonnelle en français contemporain*, Editions Jean-Fayard, Paris.

Rizzi L. 1978 : A restructuring rule in Italian Syntax. In Keyser (éd.) *Recent transformational studies in European languages*. MIT Press, Cambridge, pp. 113-158.

Robichaud B. 1995 : La cliticisation en français. Mémoire de DEA, Université Paris 7.

Rogers J. et Vijay-Shanker K. 1992 : Reasoning with descriptions of trees. Actes d'ACL'92, pp. 72-80.

Rogers J. et Vijay-Shanker K. 1994 : Obtaining trees from their descriptions : an application to Tree-Adjoining Grammars. *Computational Intelligence*, vol. 10, n° 4, pp. 401-421.

Rosen C. 1982 : The unaccusative hypothesis and the "inherent clitic" phenomenon in Italian. Chicago Linguistic Society, vol. 18.

Rosen C. 1984 : The interface between Semantic Roles and Initial Grammatical Relations. In Perlmutter et Rosen (éds.) 1984.

Rouveret A. et Vergnaud J.R. 1980 : Specifying reference to the subject: French causatives and conditions on representations. *Linguistic Inquiry*, vol. 11, pp. 97-102.

Russel G., Ballim A., Caroll J. et Warwick-Armstrong S. 1993 : A Practical Approach to Multiple Default Inheritance for Unification-based Lexicons. In *Inheritance, Defaults and the Lexicon*, Briscoe, de Paiva & Copestake (éds.) pp. 137-147. Cambridge University Press.

Ruwet N. 1972 : Les constructions pronominales neutres et moyennes, in *Théorie syntaxique et syntaxe du français*. Seuil, Paris, pp. 87-124.

Sag I. 1987 : Grammatical Hierarchy and Linear Precedence. *Syntax and Semantics,* 20, *Discontinuous Constituency. Syntax and Semantics*, Huck, Ojeda et Almerindo (éds.), pp. 303-340.

Salvi G. 1988 : La frase semplice. In Renzi (éd.), pp. 29-114.

Salvi G. 1991 : L'accordo. In Renzi et Salvi (éds.), pp. 227-244.

Schabes Y. 1990 : *Mathematical and computational aspects of lexicalized grammars*. Thèse de doctorat, Université de Pennsylvania, Philadelphia.

Schabes Y., Abeillé A. et Joshi A.. 1988 : Parsing strategies with lexicalized grammars : Tree Adjoining Grammars. Actes COLING'88, Budapest, vol. 2, pp. 578-583.

Schuler W. 1998 : Exploiting semantic dependencies in parsing. Actes du colloque TAG+4, Philadelphia.

Sells P., Zaenen A. et Zec D. 1987. Reflexivization Variation: Relations between Syntax, Semantics, and Lexical Structure. In *Working papers in grammatical theory and discourse structure*, Iida, Wechsler et Zec (éds.), CSLI.

Shieber S. 1986 : An introduction to unification-based theories of grammar. CSLI, University of Chicago Press.

Shieber S. et Schabes Y. 1990 : Synchronous Tree Adjoining Grammars. Actes de COLING'90, Helsinki.

Shieber S. et Schabes Y. 1994 : An alternative conception of tree-adjoining derivation. *Computational Linguistics*, vol. 20, n°1, pp. 91-124.

Shieber S. 1994 : Restricting the weak generative capacity of Synchronous TAGs. *Computational Intelligence*, vol.10, n°4, pp.371-385.

Skytte G., Salvi G. et Manzini R. 1991 : Frasi subordinate all'infinito. In Renzi et Salvi (éds.), pp. 483-570.

Smets M., Evans R. 1998 : A compact encoding of a DTG grammar. Actes du colloque TAG+4. Philadelphia.

Sornicola R. 1994 : On word-order variablility: a study from a corpus of Italian. *Lingua e Stile* / a. XXIX, n°1, mars 1994.

Srinivas B. 1997 : *Complexity of lexical descriptions and its relevance for partial parsing*. Thèse de PhD., Université de Pennsylvania, Philadelphia.

Steele G.L. 1990 : *Common Lisp*. Digital Press. Bedford. Massachusetts. 2ème édition.

Tasmowski L. 1984 : ?**Lui faire téléphoner quelqu'un d'autre*: une stratégie? *Linguisticae Investigationes* VIII:2, pp. 403-427.

Tesnière L. 1959 : *Eléments de syntaxe structurale*. Klincksieck, Paris.

Touretsky D.S. 1986 : *The Mathematics of Inheritance Systems*, London/Los Altos : Pitman/Morgan Kaufman.

Vijay-Shanker K. 1987 : A Study of Tree Adjoining Grammars. Thèse de Doctorat. Université de Pennsylvania.

Vijay-Shanker K. et Joshi A. 1989 : LFG functional uncertainty is a corollary in TAG. Actes de ACL'89.

Vijay-Shanker K. 1992 : Using descriptions of trees in a Tree Adjoining Grammars. *Computational Linguistics*, vol. 18, n°4, pp. 481-517, MIT Press.

Vijay-Shanker K. et Schabes Y. 1992 : Structure Sharing in Lexicalized Tree Adjoining Grammar. Actes COLING'92, Nantes, pp. 205-211.

Webelhuth G. 1998 : Causatives and the nature of argument structure. *Syntax and Semantics*, 30, *Complex Predicates in Nonderivational Syntax* Hinrichs, Kathol et Nakasawa (éds.), pp.369-422.

Wehrli E. 1980 : Constructions infinitives: Compléments VP et leurs implications théoriques. Thèse de Doctorat, Université Mc Gill, Montréal.

Wehrli E. 1986 : On some properties of French Clitic *Se*. *Syntax and Semantics*, 19, *The syntax of pronominal clitics*, Borer (éd.), pp. 263-282.

Wehrli E. 1997 : *L'analyse syntaxique des langues naturelles. Problèmes et méthodes*. Masson, Paris.

Xia F., Palmer M., Vijay-Shanker K. et Rosenzweig J. : Consistent grammar development using partial tree-descriptions for Lexicalized Tree-Adjoining Grammars. Actes du colloque TAG+4, Philadelphia, pp. 180-184.

XTAG research group (1998) : A lexicalized Tree Adjoining Grammar for English, Technical Report IRCS 98-03, University of Pennsylvania.

Zaenen A. et Dalrymple M. 1996 : Les verbes causatifs "polymorphiques" : les prédicats complexes en français. In *Nouveaux raisonnements syntaxiques*, Abeillé & Godard (éds), Langages, #122, pp. 79-95.

Zribi-Hertz A. 1982 : La construction "se-moyen" en français et son statut dans le triangle: moyen/passive/réfléchi. *Linguisticae Investigationes* 6, pp. 345-401.

Zwicky A. 1985 : Heads. *Journal of Linguistics.*. n°21, pp. 1-29.

Zwicky A. 1993 : Heads, bases and functors. In Corbett et al. (éds.) *Heads in Grammatical Theory*, Cambridge.

Zwicky A. et Pullum G. 1983 : Cliticisation vs. Inflection: English *n't*. *Language*, vol. 59, n°1, pp. 502-513.

Annexe A
MGtoTAG Compiler
A system for representing hierarchically a TAG & Compiling it into a flat grammar

This document is intended to be a user guide for writing a metagrammar, in a format that can serve as input to the MGtoTAG compiler, that outputs a set of tree sketches, to be coupled with a lexicon to form a TAG.
MGtoTAG compiler has been implemented in lisp (Lucid Common Lisp, release 4.1). The metagrammar for a given language is stated as an inheritance network in Common Lisp Object System.

1 Input format to MGtoTAG compiler

To serve as input to MGtoTAG compiler, a metagrammar has to be written as an ASCII file, containing two parts : the definition of specific parameters for tree sketches generation , and a sequence of class definitions.

1.1 Parameters for tree sketches generation

The following parameters have to be defined :

- `*families*` for the list of initial subcat frames (they correspond to terminal classes of dimension 1)
- `*redistributions*` for the list of sequences of redistributions, including the case of no redistribution (they correspond to terminal classes of dimension 2)
- `*realizations*` for the list of realizations of syntactic functions

  The realizations are grouped by function.

  Syntax : `( <realization-of-a-function>* )`
  where `<realization-of-a-function> := ((<class-name>+) :function <symbol>)`

- `*disjunctive-nodes*` contains the list of disjunctive nodes

- `*inheritance-constraints*` contains the list of (positive) cooccurrence constraints
- `*neg-inheritance-constraints*` contains the list of negative cooccurrence constraints
- `*constrained-crossings*` contains the list of constrained crossings

To define a parameter, use the following syntax :

```
(defparameter <parameter> <value>)
```

1.2 Class definitions

Classes of the inheritance network are identified by a <class-name>. The properties associated with a them are distributed over the following slots :

- **superclasses** : list of direct superclasses
- **contants** : constants' specification, allowing to add meta-feature structures to constants, and to declare statements of equality of constants.
- **equations** : equations for the target tree sketches
- **description** : partial description of tree
- **name** : string contributing to the name of the target tree sketches

The value of the slots are computed when an instance is created, as a combination of the local values and the inherited values. For instance, the value of the slot description is the conjunction of the local value and the inherited values.

To define a class, use the following syntax :

```
(<class-name>
     <list-of-direct-superclasses>
     <call-to-slot-specification-function>*
     )
```

where

```
<list-of-direct-superclasses> :=   (<direct-superclasses>*)|
                                   ( TOP )
```

and `<call-to-slot-specification-function>` are function calls to specify slot values. The different functions available for each slot are listed after the next section. The following example is a class definition with one function call only (init-name) :

```
(PASSIVE
     (PASSIVE-MORPH OBJET->SUBJECT SUBJECT->BY-OBJECT)
     (init-name "-pass" :direction :right))
```

List of direct superclasses

TOP stands for the top class of the hierarchy, and has to be directly inherited if no other class of the defined hierarchy is directly inherited.

An important feature of the class definition is that the list of direct superclasses is an *ordered list, from the most specific to the most general*. Indeed CLOS uses prioriterized inheritance for multiple inheritance networks. That is, it always defines a total order on the superclasses of a given class. This entails that the direct supeclasses of a class be ordered by the network writer.

Though in most cases of class definitions, this order is irrelevant (because of a very restricted use of non monotonic inheritance), it is crucial in the classes which specify a redistribution of functions, relying on a certain order of execution. For instance, the subject has to be demoted *before* promoting an object in subject position.

In our implementation, inheritance of properties is performed *from the most general superclass to the most specific*. Si if OBJECT->SUBJECT describes promotion of object to subject, and SUBJECT->BY-OBJECT describes demotion of subject to by-object, and if PASSIVE inherits both classes, then OBJET->SUBJECT has to be more specific than SUBJECT->BY-OBJECT, thus in the direct superclasses list of the definition of PASSIVE, OBJET->SUBJECT has to come before SUBJECT->BY-OBJECT, as shown in the class definition example above.

Slot 'constants'

This slot is used to add specifications of constants, which associate meta-feature structures to constants and which can also state equality of constants. Note that for a given crossing class, it is not mandatory that all the inherited constants (namely the constants specified in the 'constants' slot of some supeclass) appear in the description inherited by the crossing class.

Meta-features are features used by the metagrammar (such as part of speech, index, initial grammatical function...), distinct from the features that are associated with the tree sketches nodes of the represented TAG (mode, agr...)[1]. Meta-features induce an additional constraint on the trees satisfying a description Φ : the meta-features of a constant must be compatible, and the meta-features of constants naming the same node must also be compatible.

Distinguished meta-features are :

Meta-feature name	Meaning	Value Type
`:cat`	List of possible parts of speech	List of strings
`:index`	Index of corresponding node[2]	String
`:initial-function`	Initial grammatical function	Symbol
`:function`	Current grammatical function[3]	Symbol

[1] Note that it can be chosen to decorate the tree sketches with the meta-features as well.

[2] If the index is an integer, it means that the constant names a node representing a semantic argument, otherwise, the index is used to distinguish between nodes of a same tree sketch, that have the same part of speech.

[3] The final grammatical function associated with the constant is the function value, after all redistributions have taken place.

:qual	Type of the represented leaf node	Element in subst, foot, head, co_head[4] (defaults to subst)
:position		
:constraints	Used to add an NA constraint on a node	Boolean (nil/t) (defaults to nil)
:unique	Used to specify that the constant must not unify with any other constant	Boolean (nil/t) (defaults to nil)

The meta-feature :function can be revised (and this represents a redistribution). The :cat value is the intersection of all the inherited :cat values. The :qual value for a leaf node defaults to subst. Namely, in a crossing class, after the unification of all inherited descriptions is computed, any leaf node whose :qual value is unspecified is assumed to be a substitution node. The :constraint and :unique meta-feature default to nil. All other meta-features values are monotonically inherited.
All meta-features used by the MGtoTAG compiler are declared in a list value of the *keys* variable (currently in file constantes.lisp). New meta-features have to be added to that list.

- init-constants is a function used to add a list of constants' specification.
  A constant specification is made of a set of constant (which are as a result declared to name the same node) plus a meta-feature structure associated with these constants.

```
Syntax :  (init-constants '<constants-spec-list>)
where     <constants-spec-list>  :=   ( <constant-spec>* )
          <constant-spec>        :=   ( <constants-list>
                                        <meta-feature-spec>* )
          <constants-list>       :=   ( <symbol>+ )
          <meta-feature-spec>    :=   <boolean-feat> <boolean>
                                 |    :cat <cat-list>
                                 |    :ind <string>
                                 |    :function <symbol>
                                 |    :initial-function<symbol>
                                 |    :qual <node-quality>
          <boolean-feature>      :=   :unique
                                 |    :constraint
          <boolean>              :=   t | nil
          <node-quality>         :=   subst  |  foot  |  head  |
co_head
          <cat-list>             :=   ( <string>* )
```

Example 1:

[4] :subst stands for substitution node, :foot for a foot node, head for an uninstantiated anchor and co_head for an anchor directly instantiated in the tree sketch, such as the preposition *by* for passive in English.

```
(init-constants  '(((arg1) :cat ("N" "Cl") :ind "1"
                :initial-function objet :function objet)))
```

Example 2:

```
(init-constants  '(((objet quest))))
```

In example 1, a constant named `arg1`, is said to be of category either N or Cl, of index 1. It bears the initial function `objet`, which is also its current function. In example 2, the constants `objet` and `quest` are declared equal.

- `init-constants-with-index` is similar to `init-constants`, Except that it takes as argument a grammatical function, and systematically adds to the constants introduced, the index of the constant bearing that function. This is useful for instance to add an index to a PP node dominating an argumental node, in dimension 3.

Syntax : `(init-constants-with-index '<function>`
`<constants-spec-list>)`

Example:

```
(init-constants-with-index  'by-obj
                            '(((by-obj-mother)        :cat
("PP"))))
```

- `revise-function` is used to revise the function of an argument, and optionnally to add a constant specification on the constant involved (the one whose function is revised). The function revise-function fails if no argument with specified function to revise is not found. This insures that when a redistribution changes a function into another, the first one has to be present in the subcategorization of the predicate.

Slot 'description'

This slot is used to state a partial description of tree, to be satisfied by the target tree sketches. Partial descriptions of trees are here stated as a list of quasi-trees, interpreted conjunctively[5]. The value of the slot 'description' of a given class is the conjunction of all the inherited values and the local value.

Partial descriptions express statements on constants. For a given crossing class, any constant appearing in the inherited partial description have to be declared in the slot 'constant' of some superclass. But this constraint holds for crossing classes, but not for core classes of a metagrammar : a description in a certain class can make reference to a constant that is specified (i.e. to which meta-features are added) in any other core class.

Two lisp functions are available to add a quasi-tree to the value of the 'description' slot.

[5] We take the formal definition of quasi-trees of (Rogers and Vijay-Shanker, 1994). It is a partial description whose minimal referent is unique. Intuitively, it has everything of a tree, except that some pairs of constants, instead of being related either by precedence of parenthood, are related by dominance. The minimal referent is then the tree computed by collapsing these dominance links.
Representing partial descriptions as conjunction of quasi-trees allows to represent any partial description (a conjunction of quasi-trees is not necessarily a quasi-tree), except statements of linear precedence without reference to a dominant node. However, this has no impact given the linguistic needs of our application.

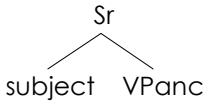
- `init-simple-desc` can be used if the quasi-tree has the properties of a tree, namely if no path are underspecified.

```
Syntax :  (init-simple-desc '<quasi-tree-simple> )
where     <quasi-tree-simple>  := ( <root-simple>
                                    <quasi-tree-simple>* )
          <root-simple>        := symbol
```

Example:

```
  (init-simple-desc '(Sr (subject) (VPanc)))
```

This example adds the following quasi-tree:



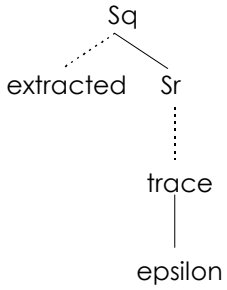
- `init-description` is used otherwise.

```
Syntax :  (init-description '<quasi-tree> )
where     <quasi-tree>         := ( <root> <quasi-tree>* )
          <root>               := ( symbol boolean )
          <boolean>            :=     t | nil
```

Example:

```
  (init-description '((Sq nil)
                      ((extracted t))
                      ((Sr   nil)   ((trace   t)   ((epsilon
  nil))))))
```

This example adds the following quasi-tree:



Slot 'equation'

This slot is used to express the features equations associated to the target tree sketches[6]. Equations make reference to constants, taht are eventually translated into the concatenation of the part of speech and the index, if it exists[7].

[6] Currently, we make a strong distinction between meta-features and features. In a future implementation we intend to relax this distinction.

[7] This is required by the XTAG tree sketch format, that MGtoTAG compiler currently generates. Another format could easily be generated.

- `init-equations` is used to add a list of equations.

```
Syntax : (init-equation `( <equation>+ ))
where   <equation>      := ,(cr-eq   <feature-path>   <feature-
  path)
                        | ,(cr-eq <feature-path> <value>)
        <feature-path>  := <symbol> <top-or-bottom> <feature>
        <top-or-bottom> := "t" | "b" | ()
        <feature>       := <string>
        <value>         := <string>
```

Example:

```
  (init-equations  `(,(cr-eq Sq "b" "extracted" "+")
                     ,(cr-eq Sq "b" "inv" Sr "t" "inv")))
```

This function call adds the following equations:

```
Sq.b:<extracted> = +
Sq.b:<inv> = Sr.t:<inv>
```

where Sq and Sr are two constants, that have to be translated.

Slot 'name'

This slot is used to add a "piece of tree sketch name". Its value type is string. For a crossing calls, the value of the slot is the concatenation of all inherited values, and the local value. The concatenation is performed from the most general to the most specific superclass. In each class, the direction of the concatenation (either right or left) is specified.

- `init-name` is used to add a substring to the slot 'name' (either left- or right-concatenated).

```
Syntax : (init-name <string> :direction <direction>)
where   <direction>     := :right | :left
```

Example:

```
  (init-name "-pass" :direction :right)
```

- `init-name+index` is used to add a substring plus the index of a constant bearing a specified function. This is useful if one wants to recors in the name of a tree, the index of an argument in marked position, and not its function.

```
Syntax : (init-name+index      <grammatical-function>
                               <string>:direction <direction>)
```

Example:

```
  (init-name+index "W" 'subject :direction :left)
```

This example is called in class EXTRACTED-SUBJECT, in the English demo metagrammar (Annexe B). It concatenates to the left the string "W" plus the index of the constant

bearing the function subject. So in the case of active voice, the tree sketch name will start with "W0", in the case of passive voice, it will start with "W1".

2 Installation and execution

MGtoTAG compiler is a lisp program that uses functionnalities from the Common Lisp Object System module. It has been tested to run with Lucid 4.0 and 4.1, on a sun machine (SparcII, SparcV, ...). To run the program:

1. Set the parameters in file load-gener.lisp

- `*ltag-generator-dir*` for the directory where the source files are located;
- `*hierarchy*` for the (full) path of the file containing the hierarchy to load;
- `*families-output-dir*` for the directory where you want the tree sketches families to be written.

2. Set the parameter `*families*` in the file describing the hierarchy, depending on the families you want to generate

3. Let `<clos>` be the unix command to call a lisp image including the Common Lisp Object System module; let `<path>` be the location of the file load-gener.lisp; and let `<ext>` be the extension of the files to be generated; then MGtoTAG compiler can be run with the following command (which can be inserted in script):

```
<clos> -l <path>/load-gener.lisp -eval "(genere-familles
       *families* '<ext>)"
```