

Une MétaGrammaire pour les adjectifs du français

Nicolas BARRIER
Laboratoire LLF - Université Paris VII
UFR de Linguistique, case 7003
2, place Jussieu
75251 Paris Cedex 05 - FRANCE
nbarrier@linguist.jussieu.fr

Mots-clefs – Keywords

Grammaires d’arbres adjoints, MétaGrammaire, Développement, Constructions adjectivales
Tree-Adjoining Grammars, MetaGrammar, Development, French Adjectival constructions

Résumé - Abstract

Initialement développée au sein de l’université Paris VII depuis maintenant près de quinze ans, la grammaire FTAG, une implémentation du modèle des grammaires d’arbres adjoints pour le français, a connu ses dernières années, diverses évolutions majeures. (Candito, 1996) a ainsi réalisé l’intégration d’un modèle de représentation compact et hiérarchique d’informations redondantes que peut contenir une grammaire, au sein d’un système déjà existant. Ce modèle, que nous appelons MétaGrammaire (MG) nous a permis, en pratique, de générer semi-automatiquement des arbres élémentaires, et par là même, d’augmenter de façon considérable les différents phénomènes syntaxiques couverts par notre grammaire. Un soin tout particulier a donc été apporté pour traiter les prédicats verbaux, en laissant cependant (partiellement) de côté le prédicat adjectival. Nous présentons donc ici une nouvelle implémentation de ce prédicat dans le cadre d’une extension de la grammaire FTAG existante.

We present here a new implementation of the adjectival predicate for the FTAG grammar, a french implementation of the Tree Adjoining Grammar model. FTAG has known over the years many improvements. (Candito, 1996) hence integrated an additional layer of syntactic representation within the system. The layer, we called MetaGrammar (MG) let us improve the syntactic coverage of our grammar by generating semi-automatically dozens of new elementary trees.

1 Principes généraux

La grammaire FTAG s’inscrit dans le cadre du développement de grammaires d’arbres adjoints lexicalisés à large couverture pour le français. Elle a comme particularité d’être générée semi-automatiquement à partir d’une MétaGrammaire indépendamment de toutes applications visées ou domaines particuliers. Comme toute grammaire, elle est liée aux compétences du locuteur. Nous présentons dans cette section les bases du formalisme, et le concept même de MG.

1.1 Le formalisme LTAG

Le modèle des grammaires d’arbres adjoints tire son nom de l’utilisation d’arbres syntagmatiques, et non de règles de réécriture classiques, comme c’est le cas avec les grammaires *hors-contexte*. Ces arbres élémentaires, *auxiliaires* ou *initiaux*, sont utilisés comme éléments atomiques combinés à l’aide de deux opérations : l’une spécifique au formalisme TAG nommée *adjonction*, l’autre plus classique nommée *substitution*.

Ces arbres sont tous lexicalisés, c'est-à-dire qu'au moins une de leurs feuilles est occupée par un élément du vocabulaire terminal, différent de la chaîne vide. Cette contrainte permet de mieux rendre compte du fait que d'une part, les propriétés associées aux mots dépendent du contexte et que d'autre part, les propriétés associées aux syntagmes dépendent du type d'éléments lexicaux mis en jeu. Tout arbre élémentaire lexicalisé doit avoir au niveau de ses feuilles au moins un nœud servant d'*ancree* lexicale.

Chaque nœud d'un arbre élémentaire est également renseigné par deux *structures de traits* top et bottom, utilisées comme contraintes. On trouve en partie *top*, les traits indiquant les relations du nœud avec les nœuds qui le dominent, ou du même niveau ; en partie *bottom*, les traits indiquant les relations du nœud avec ceux qu'il domine. Ces structures de traits sont mises à jour par *unification* lors de la combinaison des arbres entre eux. Nous renvoyons à (Joshi et al., 1975) pour davantage de détails.

1.2 Le concept de MétaGrammaire

La MétaGrammaire, mise au point par M.H Candito¹, suivant la proposition de (Vijay-Shanker et Schabes, 1992), est un réseau d'héritage multiple, un mécanisme de partage de propriétés syntaxiques entre des unités structurées dans une hiérarchie à trois dimensions qui induit un raisonnement quasi-monotone. Chaque propriété syntaxique de cette hiérarchie est déclarée comme un ensemble de descriptions partielles d'arbres - des "bouts" d'arbres -, telles qu'elles ont été exposées dans (Rogers et Vijay-Shanker, 1994)². Ces définitions peuvent laisser sous-spécifiées certaines relations entre nœuds - chaque sous-classe du réseau venant alors enrichir ces contraintes, en spécifiant certaines de ces relations.

Afin de construire des structures d'arbres pré-lexicalisées respectant le principe de co-occurrence prédicat/argument et de grouper les structures appartenant à la même famille d'arbres, la MG utilise, en plus des descriptions partielles, des fonctions syntaxiques. La sous-catégorisation est exprimée comme une liste de parties du discours possible à laquelle est associée une liste de fonctions. Cette sous-catégorisation initiale est celle du cas non marqué, qu'une redistribution peut venir modifier, rendant ainsi l'héritage non monotone³. Les arbres élémentaires partageant la même sous-catégorisation initiale ne diffèrent que par la réalisation de surface de leurs fonctions syntaxiques, et leurs redistributions. Chaque classe relève donc nécessairement d'une des trois dimensions suivantes :

- Dimension 1 : Sous-catégorisation initiale
- Dimension 2 : Redistribution des fonctions syntaxiques
- Dimension 3 : Réalisation de surface des fonctions syntaxiques

A partir de cette hiérarchie à trois dimensions écrite à la main, un compilateur génère automatiquement les arbres élémentaires associés aux descriptions partielles dans un processus à deux étapes. Le compilateur commence par engendrer toutes les classes croisées pour les traduire ensuite en arbre(s) élémentaire(s) en spécifiant totalement les relations de dominance et de précédenace linéaire laissées sous-spécifiées dans les descriptions partielles. Chaque classe croisée hérite précisément d'une classe terminale de dimension 1, puis d'une classe terminale de dimension 2, puis d'autant de classes terminales de dimension 3 qu'il n'y a de fonctions syntaxiques à réaliser issues de dimension 1⁴.

Par exemple, à partir des informations issues des classes représentées en FIG. 1, le compilateur génère automatiquement l'arbre élémentaire de la figure⁵. La logique de description de ces classes utilise un langage de type déclaratif. Concrètement, les informations s'organisent autour de variables (*globales*) auxquelles est associée une liste possible de catégories du discours, de fonctions (syntaxiques), et éventuellement d'index - chaque variable désignant un nœud de l'arbre. Ces variables sont utilisées comme éléments atomiques permettant une description partielle d'arbres, dans une notation *père-fils*. Ces variables peuvent également se voir renseignées d'une liste d'équations, que l'on peut venir réviser.

¹La MétaGrammaire a d'abord été implémentée sous *Lucid Common Lisp 4.1* par M.H Candito, mais le portage CLISP, tout comme l'intégration d'un module de génération "XML", est de notre fait. Le programme est donc maintenant disponible pour un grand nombre d'architectures (y compris les plus récentes).

²*Dominance immédiate, parenté, et précédenace.*

³La topologie des descriptions partielles est en revanche monotone.

⁴Le concepteur de la MG est donc ainsi amené à déclarer explicitement une classe intermédiaire de dimension 2 pour toute composition de redistributions, comme c'est par exemple le cas, dans le cadre d'une construction *causative/passive*. Mais d'autres implémentations n'ont pas cette limitation (Xia et al., 1998).

⁵Les indices associés aux arguments représentent les fonctions initiales des arguments avant redistribution, et non la position de l'argument pour sa fonction finale.

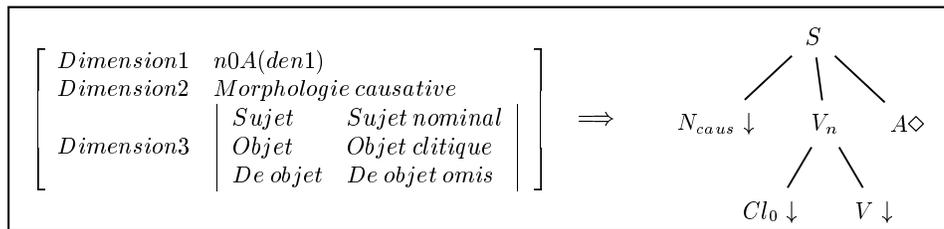


Figure 1: Exemple

Chaque classe comporte également une liste (*ordonnée*) de super-classes directes. Les relations entre ces classes peuvent être contraintes à l'aide de quatre opérations différentes. Les deux premières imposent des contraintes d'héritage positif et négatif ; les dernières sont plus spécifiques : l'une permet d'interdire tout croisement entre classes héritant d'une même super-classe (cf. *extraction*, argument relativisé vs clivé), tandis que l'autre autorise des croisements du type "si une classe hérite de B et de C alors elle doit également hériter de A".

Pour autant, des problèmes spécifiques liés à l'implémentation actuelle de notre système viennent se greffer à la difficulté de décrire correctement la grammaire d'une langue. De par l'utilisation de variables globales, et non locales, organiser un ensemble d'informations syntaxiques au sein d'un réseau d'héritage multiple rend parfois l'écriture d'une classe délicate, l'aspect de la hiérarchie se voyant modifié par l'introduction de classes intermédiaires qui n'ont pas nécessairement lieu d'être, et qui remettent ainsi en cause la notion même d'héritage. Ceci dit, il existe également des problèmes d'ordonnement des différents arguments entre eux et de partage des informations, mais ceci dépasse le cadre de cet article.

2 Une MétaGrammaire pour les adjectifs

Il paraît important de noter ici qu'il ne s'agit pas de réaliser une description du prédicat adjectival en particulier, mais plutôt d'exposer une nouvelle implémentation de l'adjectif dans le cadre d'une extension de la grammaire FTAG existante. En ce sens, le lecteur devra se tourner vers d'autres références pour une étude précise sur le sujet. Citons par exemple (Taln Atelier, 1999), (Noailly, 1999) et (Boonen, 2000). Nous tenons également à signaler que les choix linguistiques généraux imposés par cette nouvelle implémentation suivent ceux déjà définis par (Abeillé, 1991) et (Candito, 1999), pour leur propre implémentation d'une grammaire TAG. Nous renvoyons le lecteur à (Abeillé et Candito, 2000) pour tout renseignement concernant la grammaire FTAG ⁶.

Suivant la proposition de (Boonen, 2001), qui a également réalisé l'intégration d'un lexique syntaxique adjectival pour la grammaire FTAG et étendu le système de gabarit existant, nous avons choisi de ne plus dissocier les constructions à attribut du sujet des autres constructions adjectivales. Nous pouvons ainsi maintenant encoder des relatives, ou des clitiques pour les différents types de compléments de l'adjectif, chose délicate auparavant avec les choix opérés⁷ (ces constructions auraient nécessité l'utilisation des multicomponent TAG). Nous considérons l'adjectif comme la tête du prédicat adjectival, qu'il se trouve en position d'épithète ou dans une construction avec copule, et les constructions à attribut de l'objet, comme une variante des constructions causatives. Nous définissons également la notion de *sujet* comme la catégorie modifiée par l'adjectif.

2.1 Sous catégorisation initiale

L'organisation précédente de la grammaire, réalisée par (Candito, 1999), faisait référence à cinq familles de schèmes différentes pour encoder le prédicat adjectival. Quatre d'entre elles étaient dédiées à l'encodage des adjectifs employés comme épithètes⁸, alors que la dernière seulement, réalisait l'intégration des constructions à attribut du sujet⁹.

L'extension proposée définit seize schémas de sous-catégorisation différents, sur le modèle de ceux déjà définis pour les verbes. Le sujet peut y être de type nominal ou phrastique, de même que ses compléments, introduits par

⁶Une page web en cours de développement, est également disponible [<http://ftag.linguist.jussieu.fr>]

⁷L'adjectif était substitué dans l'arbre du verbe attributif

⁸n0A [4 arbres], n0Apn1 [2 arbres], n0Aps1 [2 arbres], et n0Ades1 [7 arbres].

⁹n0Va1 [32 arbres].

la préposition *à* ou *de*. Nous avons également ajouté des familles pour des adjectifs qui prennent un complément, phrastique ou nominal, introduit par une préposition sous-spécifiée¹⁰. Cette préposition, désignée par la lettre *p*, est alors renseignée au cas par cas pour chaque adjectif, dans les lexiques syntaxiques. La raison d'une telle distinction est fonction du fait qu'un syntagme prépositionnel introduit par une préposition autre que *à* ou *de*, ne peut être cliticisé.

Le tableau 1 donne pour chaque sous-catégorisation initiale un exemple de réalisation possible. Les noms des familles sont normalement transparents et suivent le schéma d'oblicité des arguments par rapport à la tête. La catégorie de l'ancre est indiquée en majuscule. On note le sujet avec l'indice 0 et les compléments de 1 à 2. Les syntagmes nominaux sont renseignés par *n*, et les compléments phrastiques par *s*. Sont entre parenthèses, les arguments optionnels.

Famille	Exemple
n0A	Jean est barbu
n0A(pn1)	Jean est fort <i>en</i> histoire
n0A(àn1)	Jean est sourd <i>à</i> cette proposition
n0A(den1)	Jean est amoureux <i>de</i> Marie
n0A(àn1)(pn2)	Jean est supérieur <i>à</i> Paul <i>en</i> mathématiques
n0A(àn1)(den2)	Jean est redevable <i>à</i> Paul <i>de</i> 100 francs
n0A(den1)(pn2)	Jean est quite <i>de</i> ses dettes <i>envers</i> la société
n0A(ps1)	Le café est bon <i>pour</i> rester éveillé
n0A(às1)	Jean est attentif <i>à</i> ce que tout se passe bien
n0A(des1)	Jean est certain <i>de</i> pouvoir assister à la réunion
n0A(àn1)(des2)	Jean est reconnaisant <i>à</i> Paul d'être venu
s0A	Faire des erreurs est inacceptable
s0A(pn1)	Arrêter de fumer est bon <i>pour</i> la santé
s0A(ps1)	Faire du sport est bon <i>pour</i> éviter les crises cardiaques
s0A(an1)	Faire du sport est nécessaire <i>au</i> corps
s0A(den1)	Faire du sport est indépendant <i>de</i> vos autres activités

Table 1: *Sous catégorisation initiale des adjectifs du français : exemples*

Nous distinguons pour chaque argument de l'adjectif, les fonctions initiales suivantes : *sujet*, *prep-objet*, *à-objet*, *de-objet*, *prep-objet-second*, et *de-objet-second*. Ces fonctions pourront se voir modifiées dans le cadre d'une redistribution syntaxique intervenant en dimension 2. Comparativement à la hiérarchie verbale déjà définie, la sous-catégorisation initiale pour les constructions adjectivales se trouve donc plus réduite, et fait appel à un jeu de fonctions plus restreint (pas d'objet direct, ni de locatif).

2.2 Redistribution

La MétaGrammaire des adjectifs pour le français couvre les phénomènes suivants de redistribution : *épithète*, *attribut du sujet*, *causatif* et *impersonnel*. Parmi les constructions causatives, on distingue le causatif *actif*, le causatif *passif* et le causatif *passif impersonnel sans agent*.

1. un homme heureux [*Epithète*]
2. Jean est barbu [*Attribut du sujet*]
3. Il est inacceptable de commettre des erreurs [*Impersonnel*]
4. A. Poulain rend les gens heureux [*Causatif actif*]
5. Des gens sont rendus heureux par A. Poulain [*Causatif passif*]
6. Il est rendu impossible de faire cela [*Causatif passif impersonnel*]

Le Tableau 2 montre les différentes règles lexicales mises en jeu dans chaque redistribution pour le passage d'une fonction syntaxique dite initiale à une fonction syntaxique dite finale. Par exemple, dans le cadre d'une construction

¹⁰Autre que *à* ou *de*.

impersonnelle, le sujet phrastique initial devient complément phrastique indirect final de l'adjectif, et le clitique *il* est promu sujet. Une contrainte est également ajoutée, puisqu'il est impossible d'opérer cette construction, lorsque le sujet est de type nominal (ou clitique) :

1. Faire du sport est bon pour le corps / Il est bon pour le corps de faire du sport
2. Le sport est bon pour le corps / * Il est bon pour le corps le sport

Construction	Sujet initial			Redistribution
	N	Cl	S	
Attribut	+	+	+	Aucune
Causatif actif	+	+	-	Sujet initial devient objet Introduction d'un sujet causateur
Causatif passif	+	+	+	Sujet causateur devient agent causateur Objet devient sujet
Causatif passif Impersonnel sans Agent	+	+	+	Agent causateur omis Introduction d'un clitique sujet figé impersonnel
Épithète	+	-	-	Sujet initial devient Sujet épithète
Impersonnel	-	-	+	Sujet initial devient Complément phrastique indirect Introduction d'un clitique sujet figé impersonnel

Table 2: Redistributions avec contraintes

2.3 Réalisation de surface des fonctions syntaxiques

Les différentes fonctions grammaticales que nous distinguons pour représenter la sous-catégorisation initiale ont été exhibées lors de la description de la dimension 1 en section 2.1. Pour la sous-catégorisation finale, nous avons introduit en section 2.2, les fonctions *objet*, pour les constructions causatives, *par-objet*, pour les constructions causatives passives, et *cpl-phr-indirect*, pour les constructions impersonnelles. Les classes terminales de la dimension 3 sont obtenues à partir de deux sources d'information : les fonctions syntaxiques finales de l'argument sous-catégorisé et les constructions qui peuvent lui être associées. Nous donnons en TAB. 3, la liste des principales réalisations de surface que nous avons implémentées¹¹, pour les quelques fonctions syntaxiques définies.

	Réalisation de surface					
	Nominal	Clitique	Clivé	Phrastique	Relativisé	Non réalisé
SUJET	Canonique Inversé	X	Nominal Phrastique	X	<i>qui</i>	
PREP-OBJET	X		Nominal Phrastique	X	<i>qui, lequel quoi</i>	X
À-OBJET	X	X	Nominal Phrastique	X	<i>qui, lequel quoi</i>	X
DE-OBJET	X	X	Nominal Phrastique	X	<i>dont</i>	X
PREP-OBJET-SECOND	X		Nominal		<i>qui, lequel quoi</i>	X
DE-OBJET-SECOND	X	X	Nominal	X	<i>dont</i>	X
CPL-PHR-INDIRECT				X		
OBJET	Antéposé Post-posé	X				
PAR-OBJET	X					X

Table 3: Réalisation de surface des fonctions syntaxiques

La FIG. 2 montre deux exemples de réalisation de schémas d'arbre avec argument extrait pour le prédicat adjectival. Dans le premier arbre, le premier complément N_1 de l'adjectif est relativisé ; dans le second, il est clivé. Ces arguments sont réalisés à l'intérieur de l'arbre régi par l'ancre adjectivale. Nous donnons pour information d'autres schémas d'arbre disponibles en ligne à l'adresse suivante http://www.linguist.jussieu.fr/nbarrier/n0A_den1_

¹¹Les arguments questionnés seront prochainement intégrés à la hiérarchie.

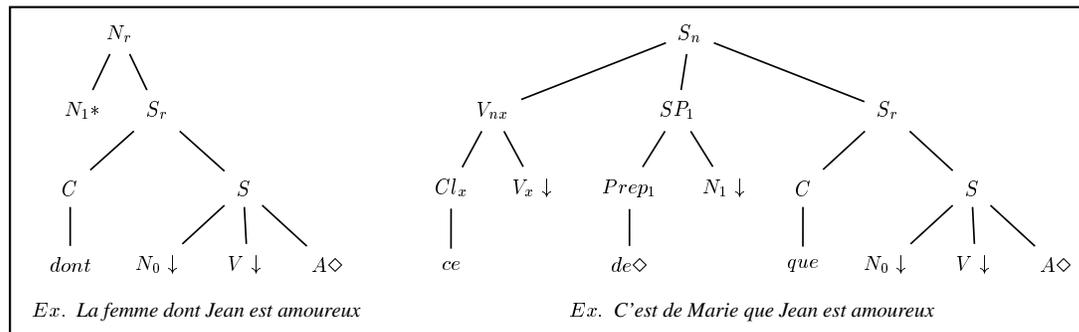


Figure 2: Arbres élémentaires avec argument extrait

Conclusion

L'implémentation d'une hiérarchie pour le prédicat adjectival en FTAG n'a pas posé spécifiquement de problèmes, si ce n'est qu'il a fallu réaliser ce travail au sein d'un système existant, et par là même, respecter les conventions précédemment adoptées. Le principal écueil se situait dans la réalisation des différents arguments, ou plus précisément de leur position entre eux. Le point de vue adopté pour FTAG est de disposer de structures relativement plates, ce qui n'est pas pour faciliter l'implémentation. Le développement de cette MétaGrammaire s'est fait de façon incrémentale, avec révision partielle des données, ie que l'on a commencé à s'intéresser à la réalisation des premiers arguments de l'adjectif (sujet vs SP1), pour ensuite venir compléter ces premières descriptions par d'autres compléments. Peu à peu, d'autres phénomènes ont été intégrés.

En résumé, cette hiérarchie adjectivale contient approximativement 80 classes terminales pour 16 schémas de sous catégorisation initiale et 5 redistributions. Au total, ce ne sont pas moins de 2690 nouveaux schémas d'arbre qui sont venus s'ajouter aux 5250 arbres existants. Ce travail reste cependant incomplet, dans la mesure où celui-ci n'a pas encore fait l'objet d'une évaluation sur corpus. Il faudra plus particulièrement s'attacher aux structures de trait associées aux nœuds et veiller à les enrichir. Il conviendra enfin d'effectuer un travail similaire pour les noms en position d'adjectif.

References

- Taln, Atelier (1999). Description des adjectifs pour les traitements informatiques. *Organisé par Pierrette Bouillon et Evelyne Viegas*.
- Abeillé, A. (1991). « Une grammaire lexicalisée d'arbres adjoints pour le français ». *PhD thesis, Université Paris VII*.
- Abeillé, A. et Candito, M.-H. (2000). « FTAG: A Lexicalized Tree Adjoining Grammar for French », pages 305–329. *CSLI Publications. Edited by Anne Abeillé and Owen Rambow*.
- Boonen, D. (2000). « Typologie syntaxique des constructions adjectivales en français ». *Master's thesis, Katholieke Universiteit Leuven*.
- Boonen, D. (2001). « Le prédicat adjectival en FTAG ». *DEA Université Paris VII*.
- Candito, M.-H. (1996). « A principle-based hierarchical representation of LTAGs ». *Dans Proceedings of COLING-96*.
- Candito, M.-H. (1999). « Représentation modulaire et paramétrable de grammaires électroniques lexicalisées ». *PhD thesis, Université Paris VII. Application au français et à l'italien*.
- Joshi, A., Levi, L., et Takahashi, M. (1975). « Tree adjunct grammars ». *Dans Journal of the Computer and System Sciences*.
- Noailly, M. (1999). L'adjectif en français. *Ophrys*.
- Rogers, J. et Vijay-Shanker, K. (1994). « Obtaining Trees from their Descriptions: An application to Tree Adjoining Grammars ».

Vijay-Shanker, K. et Schabes, Y. (1992). « Structure sharing in lexicalized tree adjoining grammar ». Dans Proceedings of COLING-92.

Xia, F., Palmer, M., Vijay-Shanker, K., et Rosenzweig, J. (1998). « Consistent Grammar Development Using Partial-Tree Descriptions for Lexicalized Tree-Adjoining Grammars ». Dans TAG+4 Proceedings.