

Une métagrammaire pour les noms prédicatifs du français

BARRIER Sébastien, BARRIER Nicolas
Laboratoire LLF - Université Paris 7
UFR de Linguistique
2, Place Jussieu - 75251 Paris Cedex 05
{sbarrier;nbarrier}@linguist.jussieu.fr

Mots-clefs – Keywords

Grammaires d'arbres adjoints, FTAG, noms prédicatifs, verbes supports, métagrammaire, ordre des constituants
Tree-Adjoining Grammars, FTAG, predicative noun, support verb constructions, metagrammar, constituency order

Résumé - Abstract

La grammaire FTAG du français a vu ces dernières années ses données s'accroître très fortement. D'abord écrits manuellement, les arbres qui la composent, ont ensuite été générés semi-automatiquement grâce à une Métagrammaire, développée tout spécialement. Après la description des verbes en 1999, puis celle des adjectifs en 2001-2002, c'est maintenant au tour des verbes supports et des noms prédicatifs de venir enrichir les descriptions syntaxiques de la grammaire. Après un rappel linguistique et technique des notions de verbe support et de métagrammaire, cet article présente les choix qui ont été entrepris en vue de la description de ces nouvelles données.

We present here a new implementation of support verbs for the FTAG grammar, a french implementation of the Tree Adjoining Grammar model. FTAG has know over the years many improvements. (Candito, 1999) hence integrated an additional layer of syntactic representation within the system. The layer, we called MetaGrammar let us improve the syntactic coverage of our grammar by genrating semi-automatically thousands of new elementary trees.

1 Les verbes supports

Nous rappelons ici brièvement les principaux intérêts linguistiques des constructions à verbe support. Pour davantage de renseignements, le lecteur pourra par exemple se tourner vers (Giry-Schneider, 1978), (Giry-Schneider, 1987) ou encore (Danlos, 1992). Les travaux du LADL dans le domaine restent une riche source d'informations, avec ceux de (Harris, 1968) sur les verbes opérateurs.

Le rôle du verbe support est essentiellement de porter des informations concernant le temps et l'aspect (Gross, 1981). Il introduit avec lui un nom, appelé nom prédicatif, qui donne la valeur sémantique au prédicat de la phrase. Ce nom peut être employé au sein d'un syntagme nominal¹ (que nous notons SNvsupp) ou d'une construction à verbe support², tout en conservant un lien paraphrasetique entre ces deux usages : *Max commet un crime contre Luc* vs. *Le crime de Max contre Luc*.

Comme le nom prédicatif a une valeur lexicale, il ne peut être ni cliticisé, ni interrogé. Il peut en revanche être relativisé³ ou extrait. Mais alors qu'un verbe ordinaire accepte seulement un seul type d'extraction :

¹De la forme *Det N_{pred} de N₀ (Prep_X N_X)*. Un verbe support ne peut se construire avec SNvsupp. Une phrase comme *Max a commis le crime de Luc contre Léa, n'est pas acceptable, sauf si elle a le sens spécial de Max a commis le même crime que Luc contre Léa.

²De la forme *N₀ V_{supp} Det N_{pred} (Prep_X N_X)*. Dans cette structure, *N₀*, sujet du *V_{supp}*, est aussi argument de *N_{pred}*

³La relative est une paraphrase du syntagme nominal SNvsupp : *Le crime de Max contre Luc* = *Le crime que Max commet contre Luc*.

- *Max raconte un crime contre Luc*
- Extraction : *C'est un crime contre Luc que Max a raconté,*
mais pas : **C'est contre Luc que Max a raconté un crime.*

il faut pouvoir rendre compte d'une double analyse avec verbe support comme :

- *Max commet un crime contre Luc*
- Extraction : *C'est un crime contre Luc que Max commet,*
et *C'est contre Luc que Max commet un crime*

La séquence *Det N_{pred} Prep₁ N₁* est donc analysée de 2 façons concurrentes : l'une comme simple constituant (i.e. un syntagme nominal), l'autre comme deux constituants (un syntagme nominal et un syntagme prépositionnel)⁴

Chaque nom prédicatif peut accepter un verbe support qui peut fournir diverses variantes aspectuelles :

- neutre : *Max a l'espoir de retrouver son livre*
- inchoative : *Max prend l'espoir de retrouver son livre*
- durative : *Max garde l'espoir de retrouver son livre*
- terminative : *Max a perdu l'espoir de retrouver son livre*

Mais ces variations ne sont pas toujours si nettes, et on a alors à faire à une variante stylistique : *Max (casse/nourrit) l'espoir de retrouver son livre.*

2 Les Métagrammaires

Les Métagrammaires sont au sein de nombreux projets de développement. Elles nous permettent ici d'encoder les verbes supports pour la grammaire FTAG. Nous présentons donc le fonctionnement de la Métagrammaire mise au point par (Candito, 1999) et repris par (Barrier, 2002), qui est utilisée pour le français, l'italien et le coréen⁵.

2.1 Principes et fonctionnement

La Métagrammaire mise au point par (Candito, 1999), suivant la proposition de (Vijay-Shanker & 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 - intuitivement des "bouts" d'arbres. Ces définitions peuvent laisser sous-spécifiées certaines relations entre noeuds - chaque sous-classe du réseau venant enrichir ces contraintes, en spécifiant certaines de ces relations.

Afin de construire des structures 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. 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 définies, à savoir, la sous-catégorisation initiale pour la dimension 1, la redistribution des fonctions syntaxiques pour la dimension 2, et la réalisation de surface des fonctions syntaxiques pour la dimension 3. 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 autant d'éléments à unifier permettant une description partielle d'arbres, dans une notation *père-fils*.

⁴Le même phénomène se produit lors de la passivation : avec un verbe ordinaire, on ne peut avoir **Un crime a été raconté par Max contre Luc*, mais avec un verbe support il est tout à fait possible d'obtenir *Un crime a été commis par Max contre Luc*.

⁵D'autres compilateurs alternatifs existent cependant, et on pourra se tourner vers (Gaiffe *et al.*, 2002) pour davantage de détails.

3 Réalisation en FTAG

Les choix linguistiques 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 & Rambow, 2000) pour tout renseignement complémentaire concernant les grammaires TAG et FTAG.

3.1 Les noms prédicatifs et FTAG

Différentes descriptions pour les noms prédicatifs avaient été proposées pour la grammaire FTAG, par (Abeillé, 1991), mais elles comportent selon nous deux inconvénients majeurs. D'une part, ces descriptions n'ont pas été généralisées par l'utilisation d'une méta-grammaire, et n'ont pas été incluses dans la version actuelle de la grammaire, d'autre part, elles placent sur pied d'égalité, verbe support et nom prédicatif, puisque les phrases à verbe support y sont représentées comme des arbres élémentaires à têtes multiples – la tête lexicale comprenant le verbe, le nom prédicatif (et éventuellement le déterminant et des propositions). De telles descriptions correspondent davantage selon nous à des expressions figées.

Nous proposons donc de renseigner la tête lexicale par le nom prédicatif seul ; le verbe support sera quant à lui substitué. Le choix du verbe support est déterminé par le nom prédicatif lui-même, de par l'utilisation d'un trait ou d'un métatrait (si plusieurs supports peuvent être utilisés). Les arbres concernant les syntagmes nominaux sont des arbres initiaux et non auxiliaires, comme c'est le cas généralement, puisque le nom est tête. Nous rendons compte de la double analyse, même si celle-ci s'avère peu satisfaisante dans la mesure où elle ne fait pas référence à une ambiguïté de la langue. Elle nous permet cependant de placer le syntagme prépositionnel avant le nom prédicatif. L'annexe présenté en fin d'article, fournit quelques détails liés à l'implémentation.

3.2 Ordre des constituants

Avec l'essor des MG, des problèmes d'un ordre nouveau sont apparus. Alors que l'on s'intéressait autrefois au pur phénomène syntaxique, puis à la façon adéquate de le représenter, nous sommes maintenant confrontés à un problème d'une autre envergure. Puisque la MG se veut généralisante, il nous faut non seulement comprendre ce que les différentes descriptions syntaxiques ont en commun et ce qu'elles partagent comme informations, mais aussi savoir comment encoder l'ordre entre constituants, car chaque élément est libre de se déplacer au sein d'un arbre tant que les contraintes qui le définissent sont satisfaites. Nous considérons donc le problème de l'ordre entre constituants en termes de dominance et de précédece, ie que tout syntagme qui ne serait pas à la place à laquelle il serait effectivement attendu serait considéré comme problématique.

La première des solutions que l'on pourrait envisager serait d'encoder l'ordre en dimension 1 ou 2, mais la propagation de ces contraintes serait réalisée trop tôt, alors que d'autres choix pourraient s'appliquer et seraient guidés par la réalisation des arguments en dimension 3⁷.

La seconde idée serait de construire des niveaux syntaxiques intermédiaires pour permettre le rattachement de nœuds deux à deux. En pratique, cette solution n'est cependant guère applicable car trop contraignante et directement liée à la théorie syntaxique dans laquelle elle veut s'appliquer.

Une troisième solution serait de se reposer sur le calcul du modèle minimal⁸ et d'introduire des contraintes d'ordre en dimension 3, comme suggéré par (Gerdes, 2002). Si nous considérons l'exemple des verbes bi-transitifs à 3 arguments en anglais comme mentionné, où il s'agit d'encoder la place de l'objet direct D par rapport à l'objet indirect I avec la préposition *to*⁹, nous pouvons introduire indifféremment une contrainte d'ordre linéaire $d < i$ en dimension 3 dans la classe de l'objet direct D , ou dans la classe de l'objet indirect I . Puisque chacun des éléments à ordonner doit figurer dans la structure résultante, la contrainte d'ordre peut donc s'appliquer du fait d'un partage correct de l'information entre les deux classes.

On peut cependant noter que quand la classe des verbes bi-transitifs exige la réalisation de ces deux objets, les classes D et I sont appelées indépendamment l'une de l'autre sans pour autant pouvoir porter d'hypothèse sur le contenu de la classe concurrente, si bien que rien n'impose la présence de l'autre classe. La position de (Gerdes,

⁷Il faut encoder un ordre entre réalisation de fonctions et non un ordre entre fonctions.

⁸Le terme de modèle minimal apparaît pour la première fois dans (Gerdes, 2002), qui le définit comme "modèle d'une description avec un nombre minimal de nœuds", ie qu'il s'agit des référents minimaux de la description où le nombre de nœuds serait minimal. Ces derniers ont d'ailleurs été considérés à tort par l'auteur comme clairement équivalents au modèle référent de (Candito, 1999).

⁹*Mary gave a book to Peter* vs **Mary gave to Peter a book*.

2002) présente donc, selon nous, certains désavantages. En particulier, elle engendre un déséquilibre entre les différentes classes mises en jeu, et alors que l'application d'une contrainte d'ordre devrait se faire de façon partagée, c'est uniquement la classe qui porte la contrainte qui dicte son comportement à l'autre. Bien évidemment, nous pouvons pour "l'ergonomie de la représentation (vouloir) réutiliser la classe dans un autre contexte", avec cependant un double risque potentiel, puisque rien n'impose, d'une part, que la constante "inconnue" ait à s'unifier avec une autre lors du calcul du modèle minimal, et qu'il nous faille d'autre part gérer une accumulation à priori non quantifiable des contraintes d'ordre à définir pour les autres placements à réaliser. Ceci nous amène donc tout naturellement au problème de la lisibilité, puisque nos contraintes ne s'appliqueraient finalement qu'au cas par cas, et qu'elles sembleraient imposer la présence d'autres classes facultatives.

Jusqu'ici aucune des solutions que nous avons étudiées ne semble satisfaisante tant d'un point de vue linguistique qu'implémental. La dernière proposition que nous avons abordée ne partageait pas correctement l'information et imposait certains choix non anodins et en particulier la position de la contrainte d'ordre n'était pas sans conséquences. A défaut de vouloir systématiser la réalisation de l'ordre entre les différents constituants de la phrase en faisant rentrer des contraintes dans des modules de dimension 1, 2 ou 3, nous proposons donc une approche moins généraliste en choisissant d'encoder l'ordre linéaire au sein de nouvelles classes de dimension 4, "indépendamment" des classes de dimension 1, 2 ou 3, à l'aide d'un système de règles explicites et conditionnelles. (Candito, 1999) avait déjà défini de tels modules sans pour autant leur accorder de statut particulier. Nous avons donc repris et étendu son implémentation qui permettait d'ajouter aux croisements déjà définis de nouvelles classes dans la liste de précedence¹⁰.

Conclusion

Cette nouvelle réalisation permet d'enrichir les Métagrammaires du français déjà encodées, et présente la MGC comme un outil mature et robuste, particulièrement adapté au français. Les corrections que nous y avons apportées la rendent beaucoup plus souple et simple à l'usage¹¹. 5 des 18 familles que nous avons définies sont d'ores et déjà finalisées, et totalisent à elles seules, plus de 2050 arbres. Ce travail demeure cependant incomplet, dans la mesure où une évaluation sur corpus n'a pas encore été entamée.

Références

- ABEILLÉ A. (1991). *Une grammaire lexicalisée d'arbres adjoints pour le français*. PhD thesis, Université Paris 7.
- ABEILLÉ A., BARRIER N. & BARRIER S. (2001). La grammaire FTAG. Documentation interne.
- ABEILLÉ A. & RAMBOW O. (2000). *Tree Adjoining Grammars*. USA: CSLI Publications.
- BARRIER N. (2002). Une métagrammaire pour les adjectifs du français. In *TALN 2002*.
- BOONEN D. (2001). Le prédicat adjectival en ftag. Master's thesis, Université Paris 7.
- CANDITO M.-H. (1999). *Représentation modulaire et paramétrable de grammaires électroniques lexicalisées. Application au français et à l'italien*. PhD thesis, Université Paris 7.
- DANLOS L. (1992). Support verb constructions. In *Journal of French Linguistic Study*.
- DANLOS L. (1998). GTAG: un formalisme lexicalisé pour la génération inspiré de TAG. *TAL*, **39-2**.
- GAIFFE B., CRABBÉ B. & ROUSSANALY A. (2002). A new metagrammar compiler. In *Proceedings of TAG+6*.
- GERDES K. (2002). *Topologie et grammaires formelles de l'allemand*. PhD thesis, Université Paris 7.
- GIRY-SCHNEIDER J. (1978). *Les nominalisations en français*. Genève-Paris: Droz.
- GIRY-SCHNEIDER J. (1987). *Les prédicats nominaux en français*. Genève-Paris: Droz.
- GROSS M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, **63**.
- HARRIS Z. S. (1968). *Mathematical Structure of Languages*. Wiley-Interscience.
- VIJAY-SHANKER K. & SCHABES Y. (1992). Structure sharing in lexicalized tree adjoining grammar. In *Proceedings of COLING-92*.
- VIVÈS R. (1983). *Avoir, prendre, perdre : constructions à verbe support et extensions aspectuelles*. PhD thesis, Université Paris 8.

¹⁰On pourrait également envisager un système à la (Gaiffe *et al.*, 2002) où les croisements avec la dimension 4 seraient obtenus à l'aide d'un système à base de traits polaires. Mais pour les raisons que nous avons exposées précédemment, nous avons écarté cette solution.

¹¹Pour faciliter le développement, nous avons revu les messages de *debugging* pour les étendre : chaque schème se voit également maintenant renseigné d'une description en structure de traits, similaire dans l'esprit à celle de (Danlos, 1998) ou (Gaiffe *et al.*, 2002), qui permet d'une part d'établir un pont entre lexique et syntaxe, et d'autre part, d'obtenir une "vue" de l'entrée appréhendable par un non expert TAG.

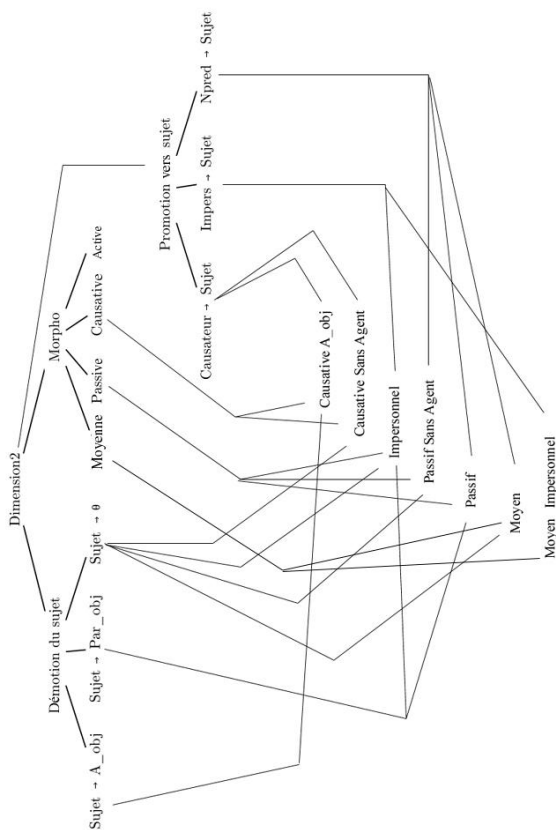
ANNEXE : Brève description des dimensions

Les nouvelles familles que nous définissons reprennent les conventions de notation déjà adoptées par (Candito, 1999) pour les verbes et par (Boonen, 2001) et (Barrier, 2002) pour les adjectifs :

Famille	Exemple
n0vN	Max prend une douche
n0vN(an1)	Max fait du chantage à Luc
n0vN(den1)	Max fait la censure de cette page
n0vN(pn1)	Max a commis un crime contre Luc
n0vN(loc1)	Max fait un pèlerinage à Lourdes
n0vN(des1)	Max garde l'espoir de réussir
n0vN(ps1)	Max fait des efforts pour réussir
n0vPN(pn1)	Max est en colère contre Luc
n0vPN(den1)	Max est dans l'ignorance de cet événement
n0vPN(as1)	Max a de la peine à se décider
s0vN	Que Max doit partir fait scandale
s0vN(den1)	Que Max doit partir fait la joie de Luc
s0vPN(den1)	Que Max doit partir est à l'avantage de Luc
n0vN(den1)(an2)	Max fait le récit de ses aventures à Luc
n0vN(den1)(pn2)	Max fait la division de ce nombre par 6
n0vN(den1)(loc2)	Max fait une expédition de colis en Amérique
n0vN(pn1)(pn2)	Max fait une plaisanterie sur Marc avec Luc
n0vN(an1)(des2)	Max donne l'ordre à Luc de partir

Nous ne décrivons pas ici la dimension 1, dont l'intérêt reste limité en comparaison des dimensions 2 et 3. La dimension 2 prend en charge les actifs, les passifs (avec et sans agent), les impersonnels, les causatifs, ainsi que les constructions moyennes et nominales :

- *Max commet un crime (contre Luc) - Actif*
- *Un crime est commis (par Max) (contre Luc) - Passif*
- *Il est commis un crime (par Max) (contre Luc) - Impersonnel passif*
- *Léa fait commettre un crime (à Max) (contre Luc) - Causative A-obj*
- *Un crime se commet (contre Luc) en 5 minutes - Moyenne*
- *Il se commet un crime toutes les 5 minutes - Moyenne Impersonnel*
- *Le crime (de Max) (contre Luc) - Syntagme nominal*



La dimension 3, quant à elle, s'avère beaucoup trop riche pour pouvoir être représentée ici graphiquement. Nous fournissons donc ci-après un tableau récapitulant les différentes réalisations de surface des principales fonctions syntaxiques.

	Réalisation de surface						
	Nominal Canonique Inversé	Clitique	Clivé Nominal	Phrasique	Relativisé ¹²	Interrogé	Non réalisé
Sujet	X	X	X	X	qui	X	
Nom prédicatif	X		Nominal		que		
Prep Obj	X		Nominal	X	X	X	X
A Obj	X	X	Nominal		X	X	X
De Obj	X	X	Nominal		X	X	X
Prep Obj2	X		Nominal		X	X	X
A Obj2	X	X	Nominal		X	X	X
CpI Phr Ind				X			
Par Obj	X		Nominal			X	X

¹²La grammaire FTAG établit une distinction entre complémentateur et pronom relatif. Cette distinction tient au fait que les compléments pour les relatives introduisent aussi des complétives