



**HAL**  
open science

## Filtres dans une grammaire catégorielle

Karine Kray-Baschung, Gabriel G. Bès

► **To cite this version:**

Karine Kray-Baschung, Gabriel G. Bès. Filtres dans une grammaire catégorielle. L'ordre des mots dans une grammaire catégorielle, 1990, Clermont-Ferrand, France. pp.93-103. halshs-00372938

**HAL Id: halshs-00372938**

**<https://shs.hal.science/halshs-00372938>**

Submitted on 5 Aug 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gabriel G. Bès / Karine Baschung

---

## Filtres dans une grammaire catégorielle

## RÉSUMÉ

La concaténation par adjacence à partir d'un prédicat pose problème aux grammaires qui l'adoptent, ce qui est souvent le cas des grammaires catégorielles. L'ordre des mots en français, en particulier certains cas de négation, exige l'introduction d'un mécanisme plus complexe comportant un filtrage des structures grammaticales.

## Filters in Categorical Grammar

### ABSTRACT

The concatenation of adjacent signs starting from a distinguished predicate is a real difficulty for grammars which adopt it. This is often the case for categorial grammars. French word order, particularly with respect to negation, requires a more complex mechanism including an operation of filtering on grammatical structures.

Gabriel G. Bès, Karine Baschung  
Université Blaise-Pascal (UFR LAC)  
GRIL  
34, avenue Carnot  
F-63037 Clermont-Ferrand Cedex (France).

On peut concevoir une phrase comme une suite  $u = u_1, \dots, u_n$ . Si la longueur de  $u > 1$ , on admet que  $u = u_1 v = w u_n$  où  $v = u_2, \dots, u_n$  et  $w = u_1, \dots, u_{n-1}$ . Un langage  $\mathcal{L}$  est alors un ensemble de suites  $u$ . Une grammaire  $\mathcal{G}$  spécifie ces suites en leur assignant une représentation.

Une grammaire catégorielle de type UCG (Cf. Zeevat et al. 87) est composée de règles constructivistes (l'application fonctionnelle à gauche et à droite) qui opèrent par concaténation d'éléments adjacents à partir d'une tête ou pivot ( $= P_i$ ).  $\mathcal{G}$  assigne à  $u$  de  $\mathcal{L}$  une représentation  $S$ . En reprenant la définition de Jurie & Bès 91, on dira que  $\mathcal{G}$  possède la propriété CKP ("Central Kernel Property"), car pour tout  $u$  spécifié par  $\mathcal{G}$  ( $u$  auquel  $\mathcal{G}$  assigne par ses règles une représentation  $S$ ),  $\mathcal{G}$  doit également spécifier  $v$  et/ou  $w$ .

Il n'est pas vrai que tout  $u$  grammatical spécifié par  $\mathcal{G}$  est construit à partir de  $v$  ou de  $w$  qui soient des suites grammaticales. A preuve l'exemple (1), où (1a) =  $u$ , (1b) =  $v$  et (1c) =  $w$  (dans tous les exemples les SN sont traités comme des éléments uniques).

- |     |     |                           |
|-----|-----|---------------------------|
| (1) | (a) | Jean construit une chaise |
|     | (b) | construit une chaise      |
|     | (c) | Jean construit            |

Cet exemple est un cas relativement simple où la non grammaticalité de (1b) et de (1c) résulte de la non consommation de toutes les valences de *construit*. Il révèle cependant une question de fond: les suites spécifiées par les règles de concaténation de  $\mathcal{G}$  ne sont pas toutes grammaticales. Une contrainte supplémentaire aux règles doit donc être ajoutée à  $\mathcal{G}$ , contrainte qui, à notre connaissance, n'a pas été explicitée dans les présentations des grammaires

catégorielles. Dans le cas d'une grammaire de type UCG, cette contrainte est relativement facile à spécifier et à implémenter pour les situations illustrées par (1): les phrases grammaticales seront celles où la valeur *Cat* (pour *Category*) dans *S* est *sent*. Cette contrainte est un filtre qui doit vérifier que les valences obligatoires de toute entrée prédicative ont effectivement été consommées.

Mais un filtre n'est pas exigé seulement pour rendre compte des problèmes que l'on vient d'évoquer. Il est possible de caractériser en général les situations où un filtre est requis par une grammaire qui, telle les grammaires UCG, concatène des éléments adjacents à partir d'un élément *P*.

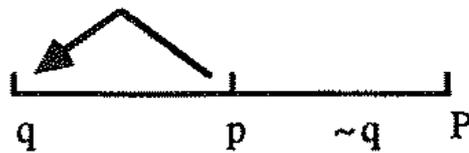
Pour aboutir à cette caractérisation, on peut utiliser un métalangage descriptif ou ensemble d'axiomes  $\Theta$ , qui, de manière purement déclarative, définira formellement les propriétés qui doivent être satisfaites par les suites de  $\mathcal{L}$ , conçues comme des modèles de  $\Theta$  (cf. Jurie & Bès 89). Par ailleurs, il est possible de typifier les situations où  $\Theta$  ne possède pas la propriété CKP, ce qui implique que pour tout  $u$  modèle de  $\Theta$ , il n'est pas vrai que  $v$  et/ou  $w$  soient des modèles de  $\Theta$ . Dans Jurie & Bès 91, il a été démontré que si les conditions dans (2) sont réunies,  $\Theta$  ne possède pas la propriété CKP.

- (2)
- (a)  $p$  et  $q$  sont disjoints
  - (b)  $\exists P$
  - (c)  $\exists p \longrightarrow \exists q$
  - (d) Il existe un modèle de  $\Theta$  dans lequel il est vrai que  $\exists ij (p(i) \wedge F(i) \wedge q(j) \wedge F(j) \wedge \forall k (i < k \longrightarrow \sim q(k)))$

Dans (2),  $p$  et  $q$  sont des éléments de  $u$ ,  $P$  est l'élément pivot (dans les exemples qui suivent, le verbe) à partir duquel on concatène à gauche, par  $F$  ("Forward application"), ou à droite par  $B$  ("Backward application"),  $i$ ,  $j$ ,  $k$  sont des positions occupées par les éléments de  $u$  et ' $<$ ' est le symbole de précédence (Il est possible de donner une formulation analogue à (2) pour les concaténations à droite en substituant  $B$  à  $F$ ).

On peut représenter graphiquement (2) par (3).

(3)



Il existe en français des situations où les conditions de (2) sont satisfaites ; à noter que l'élément  $q$  exigé par (2c) n'est pas simplement une valence de  $P$  comme dans (1). Ces situations vont exiger de  $\mathcal{G}$  une vérification par filtrage des suites spécifiés par  $\mathcal{G}$ . Il s'agira en particulier de vérifier (2c) alors que l'on est dans la situation (2d).

Par exemple, dans (4), (4a) peut être complété à gauche par (4b) mais non par (4c), alors que (4d) est agrammatical.

(4) Particule de négation *ne* et *SN nég(atif)*.

- (a) ne regarde Marie
- (b) Personne
- (c) \*Jean
- (d) \*ne personne regarde Marie

L'axiome à l'oeuvre dans (4) est  $\exists ne \rightarrow \exists(SNneg \vee pas)$  qui, conjointement avec les axiomes sur la linéarité provoque une situation définie par (2).

On a par ailleurs les situations (5) à (8) qui rentrent toutes dans le cadre de (2). Elles sont présentées selon le schéma de (4) en évoquant par ailleurs, de manière informelle, l'axiome qui les détermine.

(5) Particules de négation et infinitif

- (a) pas donner ce livre aux enfants
- (b) Il souhaite ne
- (c) \*Il souhaite
- (d) \*Il souhaite pas ne donner ce livre aux enfants

(pas concaténé à gauche d'un infinitif exige un *ne* à sa gauche)

## (6) Clitiques avec impératif et négation

- (a) le prépare
- (b) ne le prépare pas
- (c) \* $\emptyset$  le prépare pas
- (d) \*le ne prépare pas

(un clitique concaténé à gauche d'un impératif exige *ne* à sa gauche)

(7) Participe non accordé avec l'auxiliaire *être*

- (a) est préparé les examens pendant les vacances
- (b) Marie s'
- (c) \*Marie
- (d) \*Marie est se préparé les examens pendant les vacances

(l'auxiliaire *être* concaténé à gauche d'un participe non accordé qui admet par ailleurs *avoir*, exige *se* à sa gauche)

(8) Participe accordé avec *avoir*

- (a) a préparée pour l'examen
- (b) Quelle amie Pierre  
Pierre l'
- (c) \*Pierre
- (d) \*Pierre a quelle amie préparée pour l'examen  
\*Pierre a la préparée pour l'examen

(l'auxiliaire *avoir* concaténé à gauche d'un participe accordé exige un objet direct à sa gauche)

L'inversion du sujet combinée avec l'extraction des compléments des enchâssées détermine une autre situation où un filtre est exigé. En général, un sujet non clitique à droite du verbe exige un constituant interrogatif à sa gauche. On a donc dans  $\Theta$  l'axiome de (9) (où  $XP$  note un constituant quelconque,  $nom$  : un constituant nominatif,  $wh$  : un constituant interrogatif,  $xB$  et  $xF$  :  $x$  concaténé par  $B$  et  $F$  respectivement.)

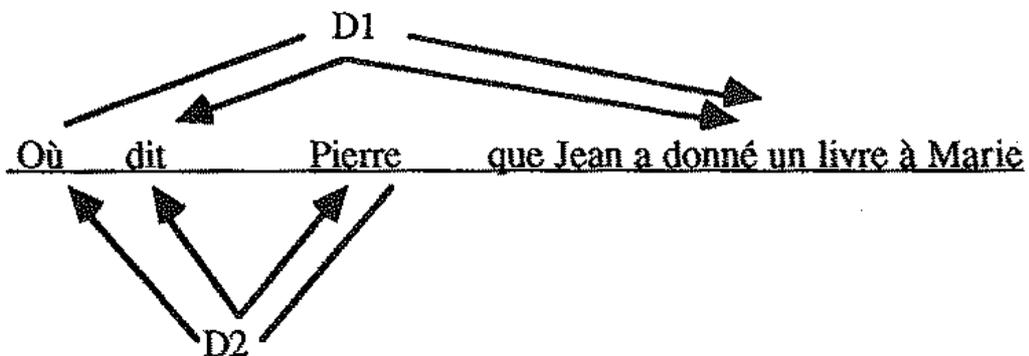
$$(9) \quad \exists \text{SN}[\text{nomB}] \longrightarrow \exists \text{XP}[\text{whF}]$$

En français, (9) peut être satisfait par (10), où *XP* est instancié par un complément de l'enchâssée (*où* porte sur *a donné*).

(10) Où dit Pierre que Jean a donné un livre à Marie ?

On admet que (10) est formé par les constituants indiqués dans (11) sur lesquels  $\Theta$  définit les dépendances notées par les fléchés ( $\leftrightarrow$  note les dépendances entre les valences et  $P_i$ ).

(11)



Etant donné les contraintes sur  $\mathcal{G}$  (concaténation par adjacence à partir de  $P_i = \text{verbe}$ ), la construction du  $S$  assigné à (10) ne peut être que (12a) ou (12b).

- (12) (a) Où dit Pierre  
 (b) Où dit Pierre

Or, dans les deux cas, on crée une situation de filtrage. Dans (12a), pour vérifier D1 et dans (12b) pour vérifier D2.

Aux observations précédentes, il faut ajouter les suivantes même si, à première vue, elles ont peu de rapport. Comme le montre (13), certains verbes exigent des modifieurs obligatoires, et, comme l'illustre (14), les modifieurs négatif et interrogatif ont, par rapport à la linéarité, un comportement différent des autres.

- (13) (a) Jacques {habite, se trouve} à Paris  
 (b) \*Jacques {habite, se trouve}

- (14) (a) A Paris, Jacques a visité les musées  
 (b) Où Jacques a visité les musées  
 (c) \*Où, Jacques a visité les musées  
 (d) Nullepart Jacques n'a visité les musées  
 (e) \*Nullepart, Jacques n'a visité les musées

(15) rsume le faisceau d'observations prcdentes.

(15)

- (a) Une procédure de filtrage est exigée pour vérifier la consommation des valences obligatoires.  
 (b) Les verbes doivent être sous-catégorisés par rapport à l'optionnalité des modifieurs.  
 (c) Les modifieurs interrogatif et négatif ont un comportement différencié par rapport à la linéarité.  
 (d) Ce sont les constructions interrogatives et négatives qui déterminent, conjointement avec la linéarité, les conditions d'un filtrage autre que celui exigé par (a).

Dans FG (= French Grammar) on propose une solution qui tient compte de (15) (a) à (d). Les éléments de cette solution sont indiqués ci-après. Le signe FG possède sept champs. (16) présente un signe argument et (17) un signe foncteur.

- (16) :C1 (Catégorie syntaxique)  
 :Varg (Ensemble de valences de l'argument)  
 :T1 (Traits)  
 :SR1 (Sensibilité aux règles)  
 :S1 (Sémantique)  
 :O1 (Optionnalité)  
 :W1 (Phonologie)
- (17) :C /(C1 : Vsa : T1 : SR1 : S1 : O1 : W1)  
 :Vf (Ensemble des valences du foncteur)  
 :T  
 :SR  
 :S  
 :O  
 :W

FG utilise deux règles de concaténation ( $F$  et  $B$ ) qui opèrent, à l'exception du champ 2 (valences), par unification directe de termes Prolog, et une règle d'effacement, appelée à la fin du processus de concaténation. Les valences sont traitées comme des éléments appartenant non à des listes mais à des ensembles; sur la procédure *mbu* qui, moyennant des opérations ensemblistes, permet de calculer les valences du  $Sr$  (= signe résultant de la concaténation), on consultera Baschung 90 et, dans ce même volume, la communication de Martine Emorine sur les verbes supports. Les règles  $F$  et  $B$  concatènent les phonologies du foncteur et de l'argument et permettent d'obtenir respectivement  $WW1$  et  $W1W$ . Sur le calcul des valeurs des champs 1, 3, 4, 5 et 6 dans  $Sr$ , on consultera Bès et Gardent 88, Gardent & al 89. La règle d'effacement est une règle unaire dont le rôle est double. Elle élimine les valences dont  $O1 = op$  et elle filtre les suites auxquelles les règles de  $\mathcal{G}$  ont assigné un  $S$  qui ne satisfait pas aux axiomes d'implication de  $\Theta$  du type  $\exists p \rightarrow \exists q$  (cf. (2c) et (9)). Le mécanisme (analogue à celui incorporé dans FUG, cf. Shieber 86) en est le suivant.

Dans le champ  $Varg$  de toute entrée verbale, on incorpore une valence dont  $C1 = mod$  et, selon le verbe,  $O1 = op$  (*tionnelle*) ou bien *nop* (*non optionnelle*). Les autres valences de  $Varg$  correspondent à des arguments verbaux, leur  $C1 = np$ . Elles sont *op* ou bien *nop*. Dans  $T1$  de *mod* on introduit les traits qui vont permettre d'implémenter les vérifications exigées par la règle d'effacement.

Lors de la concaténation avec  $P_i$ , le constituant qui instancie  $p$  de (2c) introduit dans  $T1$  de l'argument verbal un trait de blocage qui n'unifie pas avec le trait correspondant, spécifié dès le lexique, dans la valence *mod*. Ainsi, suite à la concaténation de (18a) on obtient le  $Sr$  correspondant de (18b) :

- (18) (a) ...regarde Marie ...  
 (b)  $Varg : \{(mod : [..., \sim inv2, ...] : ...), \dots\}$   
 $\alpha 1$   
 $T1 : [..., inv2, ...]$   
 $\alpha 2$

Le constituant qui instancie  $q$  de (2c) efface, au moment de concaténer, le trait de blocage introduit par  $p$ . Ainsi, suite à la concaténation à gauche de (18a) d'un élément interrogatif on obtient (19a) et (19b) dans son  $Sr$  associé.

- (19) (a) ...quel livre regarde Marie ...  
 (b)  $Varg : \{(\text{mod} : [\dots, \sim \text{inv}2, \dots] : \dots), \dots\}$   
 $\alpha 1$   
 $T1 : [\dots, -, \dots]$   
 $\alpha 2$

Par unification de  $\alpha 1$  avec  $\alpha 2$  la règle d'effacement vérifie que l'axiome (9) de  $\Theta$  a été respecté. Elle pourra donc supprimer la valence  $mod$ , de même que toutes les valences optionnelles des  $Sr$  dont  $Cat = sent$ . On définit ainsi formellement l'ensemble des phrases grammaticales spécifiées par  $\mathcal{G}$  : c'est l'ensemble  $S$  dont  $Cat = sent$  et  $\{Varg\} = \emptyset$ .

## Références

**Baschung K. (1990).** *Grammaires d'unification à traits et contrôle des infinitives en français*. Clermont-Ferrand, Adosa, 1992.

**Bès G.G. & C. Gardent (1989).** "French order without order". *Proceedings of the 4th Conference of the European Chapter of the Association for computational Linguistics*. Manchester, 1989, p. 249-255.

**Gardent C., G.G. Bès, P.-F. Jurie & K. Baschung (1989).** "Efficient parsing for French". *Proceedings of the Conference of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, 1989, p. 280-287.

**Jurie P.-F. & G.G. Bès (1989).** "The control of grammars". *Contribution to ACORD Deliverable T 7.3*, Université de Clermont-Ferrand II, 1989.

**Jurie P.-F. & G.G. Bès (1991).** "The control of UCG grammars". Dans G.G. Bès et Th. Guillotin (eds). *The construction of a Natural Language and Graphic Interface : Results and Perspectives from the ACORD Project*. A paraître, Springer, 1992.

**Shieber (1986).** *An introduction to unification - Based Approaches to Grammar*. CSLI, 1986.

**Zeevat H., E. Klein & J. Calder (1987).** "Unification Categorical Grammar". Dans N.J. Haddock, E. Klein & G. Morrill (eds). *Categorical Grammar, Unification Grammar and Parsing*. Edinburgh, Centre for Cognitive Science, 1987, p.195-222.