Une grammaire d'interaction du français

Guy Perrier

TALN 2007, Toulouse, 12–15 juin 2007

Une grammaire d'interaction du français

Guy Perrier[1]
(1) LORIA - Université Nancy 2
perrier@loria.fr

Résumé Nous présentons une grammaire du français à relativement large couverture dans le formalisme des grammaires d'interaction. Ce formalisme combine deux idées-forces : la grammaire est vue comme un système de contraintes à travers la notion de description d'arbre, et la sensibilité aux ressources de la langue est utilisée comme principe de composition syntaxique à l'aide de la notion de polarité. Nous donnons un aperçu du pouvoir expressif du formalisme en modélisant quelques phénomènes linguistiques significatifs et nous montrons que l'architecture de la grammaire répond à un souci de réutilisabilité et de faisabilité, crucial quand on cherche à construire des ressources à large couverture : distinction entre une grammaire source modulaire et une grammaire objet obtenue par compilation de la première, indépendance du lexique par rapport à la grammaire. Enfin, nous présentons les résultats d'une évaluation de la grammaire sur une suite de phrases tests, effectuée à l'aide de l'analyseur syntaxique LEOPAR.

Abstract We present a French grammar with a relatively large coverage in the formalism of Interaction Grammars. This formalism combines two key ideas : the grammar is viewed as a constraint system, which is expressed through the notion of tree description, and the resource sensitivity of the language is used as a syntactic composition principle by means of the notion of polarity. We give an outline of the expressivity of the formalim by modelling significative linguistic phenomena and we show that the grammar architecture provides for reusability and tractability, which is crucial for building large coverage resources : a modular source grammar is distinguished from the object grammar which results from the compilation of the first one, the lexicon is independent of the grammar. Finally, we present the results of an evaluation of the grammar with a test suite of sentences achieved with the LEOPAR parser.

Mots-clefs : syntaxe, grammaire formelle, méta-grammaire, grammaire d'interaction

Keywords: syntax, formal grammar, meta-grammar, interaction grammar

1 Introduction

Le travail que nous présentons ici s'inscrit dans une démarche de modélisation des langues à partir de connaissances linguistiques qui fait une place centrale à l'expérimentation. Dans cet objectif, il est nécessaire d'exprimer ces connaissances linguistiques sous forme de grammaires et de lexiques avec la couverture la plus large possible tant en termes de phénomènes linguistiques représentés que de mots auxquels ils s'appliquent. Or, on sait combien il est difficile de construire de telles ressources.

La première difficulté est celle du choix du formalisme pour représenter la grammaire. Actuellement, il n'y a pas vraiment de formalisme qui prévaut dans la communauté scientifique.

Ceux qui sont les plus répandus ont tous leurs points forts et leurs points faibles. Si nous avons conçu un nouveau formalisme, celui des Grammaires d'Interaction (GI), c'est pour faire la synthèse de deux idées importantes exprimées jusqu'ici dans deux types de formalismes différents : l'utilisation de la sensibilité aux ressources des langues comme principe de composition syntaxique qui est un trait caractéristique des grammaires catégorielles (Retoré, 2000) et la vision des grammaires comme systèmes de contraintes qui est celle des grammaires d'unification telles que LFG (Bresnan, 2001) ou HPSG (Sag *et al.*, 2003).

Même si nous utilisons un formalisme original, notre souci est celui de la réutilisabilité, souci qui s'exprime de deux façons :
– Comme pour la conception des langages de programmation, nous distinguons deux niveaux dans la grammaire : la *grammaire source*, qui est écrite par l'humain (le linguiste dans l'idéal) et qui permet d'exprimer les généralisations linguistiques, et la *grammaire objet* qui est directement utilisable par un système de TAL. La première est compilée dans la seconde et nous avons utilisé pour cela XMG (Duchier *et al.*, 2005). XMG fournit un langage de haut niveau pour écrire une grammaire source et un compilateur qui traduit cette grammaire source en une grammaire objet opérationnelle.
– La grammaire est aussi conçue de telle façon qu'elle puisse s'interfacer avec un lexique indépendant du formalisme où les entrées se présentent comme des structures de traits.

C'est de cette manière que nous avons construit une grammaire du français à relativement large couverture dans le formalisme des GI et le but de l'article est de présenter cette grammaire.

2 Les grammaires d'interaction

Les GI[1](Perrier, 2004) sont un formalisme grammatical dédié à la syntaxe et à la sémantique des langues naturelles qui s'appuie sur deux notions, celle de *description d'arbre* et celle de *polarité*.

2.1 Les descriptions d'arbres

Dans une vision dérivationnelle de la syntaxe des langues, les objets syntaxiques manipulés sont en général des arbres qui sont composés de façon plus ou moins sophistiquée (grammaires algébriques, grammaires d'arbres adjoints, grammaires catégorielles ...). Empruntant notre vision à la théorie des modèles (Pullum & Scholz, 2001), nous ne manipulons pas directement des arbres syntaxiques mais des propriétés permettant de les décrire, autrement dit des descriptions d'arbres (Rogers & Vijay-Shanker, 1994). Cette approche est très souple en ce sens qu'elle permet d'exprimer de façon totalement indépendante des propriétés élémentaires d'arbres que l'on peut ensuite combiner librement.

Une description d'arbre peut être vue, soit comme un arbre sous-spécifié, soit comme une spécification d'une famille d'arbres, chaque arbre étant un modèle de cette spécification. La figure 1 donne un exemple de description d'arbre associée au pronom relatif *qui*, lorsqu'il est employé comme complément indirect. Cet emploi donne lieu au phénomène complexe d'une double dépendance non bornée (le *pied piping* en anglais) comme l'illustrent les exemples suivants qui

[1]Pour une présentation complète des GI, le lecteur pourra se reporter à l'article (Perrier, 2004).

sont tous couverts par la description de la figure 1[2].

(a) *Jean [à* **qui***] Pierre a présenté Marie* □ *est ingénieur.*
(b) *Jean [à la femme de* **qui***] Pierre a présenté Marie* □ *est ingénieur.*
(c) *Jean [à la femme de* **qui***] Pierre sait qu'on a présenté Marie* □ *est ingénieur.*

Une description est un ensemble fini de nœuds structurés par deux types de relations : *domination* et *précédence*. Les nœuds, qui représentent des syntagmes, sont étiquetés par des traits décrivant leurs propriétés morpho-syntaxiques. Les valeurs des traits sont des atomes ou des disjonctions d'atomes et elles peuvent être partagées grâce à un mécanisme de co-indexation[3]. Les nœuds peuvent être typés : ils peuvent porter la propriété *Empty* (en fond blanc sur la figure 1) ou *Full*, selon qu'ils ont une forme phonologique vide ou pleine ; ils peuvent porter la propriété *Anchor* (cadre double sur la figure 1), s'ils représentent un nœud ancrant un mot de la langue.

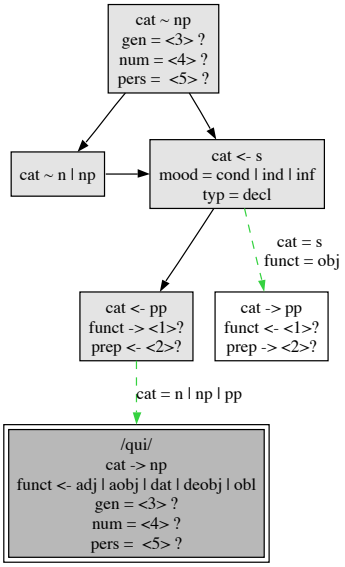

FIG. 1 – Description d'arbre associée au pronom relatif *qui* utilisé dans un complément indirect

Voici les relations entre nœuds qui sont utilisées pour définir les descriptions d'arbres :

1. *Relations de domination* :
   - $A \rightarrow B$ signifie que A est le père de B (flèche vers le bas continue sur la figure 1).
   - $A \rightarrow [B$ signifie que A est le père de B et qu'il n'a pas d'autre fils qui précède B.
   - $A \rightarrow B]$ signifie que A est le père de B et que B ne précède aucun autre fils de A.
   - $A \rightarrow\!*B$ signifie que A domine largement B (clôture réflexive et transitive de la première relation représentée par une flèche discontinue vers le bas sur la figure 1).
   - $A \rightarrow\!*[t_1 = v1, \ldots, t_n = v_n]B$ signifie en plus que tout nœud strictement dominé par A et dominant strictement B doit être étiqueté par une structure de traits subsumée par la contrainte $[t_1 = v_1, \ldots, t_n = v_n]$[4].

[2]Le groupe prépositionnel extrait est placé entre crochets et sa trace dans la proposition relative est représentée par le symbole □.

[3]Lorsque deux traits partagent une même valeur, un indice commun <n> est placé devant leurs valeurs. Lorsqu'un trait a comme valeur la disjonction de tous les atomes de son domaine, cette valeur est notée " ?".

[4]Dans l'implémentation actuelle de la grammaire, le sens de la contrainte est un peu différent dans la mesure où elle s'applique aussi aux deux nœuds reliés par la domination large.

2. *Relations de précédence* :
   - $A >> B$ signifie que A précède immédiatement B (flèche horizontale continue sur la figure 1).
   - $A >* B$ signifie que A précède B (clôture transitive de la précédente relation représentée graphiquement par une flèche horizontale discontinue).

2.2 Les polarités

Les polarités permettent d'exprimer l'état de saturation des arbres syntaxiques. Attachées à des traits qui décorent les nœuds des descriptions, elles ont la signification suivante :
- un trait positif $t \rightarrow v$ exprime une ressource disponible qui doit être consommée ;
- un trait négatif $t \leftarrow v$ exprime une ressource attendue qui doit être fournie ; c'est le dual d'un trait positif ;
- un trait neutre $t = v$ exprime une propriété linguistique qui ne se comporte pas comme une ressource consommable.
- un trait virtuel $t \sim v$ exprime une propriété qui a besoin de se réaliser en se combinant avec un trait réel (positif, négatif ou neutre).

Sur la figure 1, le nœud vide représentant la trace du syntagme prépositionnel extrait de la proposition relative est porteur d'un trait positif $cat \rightarrow pp$ et d'un trait négatif $funct \leftarrow \langle 1 \rangle ?$, qui signifie que ce nœud fournit un groupe prépositionnel qui attend de recevoir une fonction syntaxique. La racine de l'arbre porte un trait virtuel $cat \sim np$ qui signifie que le nœud représente un syntagme nominal virtuel qui doit se combiner avec un syntagme nominal réel.

Les descriptions décorées par des structures de traits polarisés prennent alors la forme de *descriptions d'arbres polarisées (DAP)*.

2.3 La grammaire comme système de contraintes

Une grammaire d'interaction particulière est définie par un ensemble fini de DAP élémentaires qui engendre un langage d'arbres. Un arbre du langage est défini comme un arbre syntaxique modèle d'un ensemble fini d'arbres élémentaires de la grammaire vérifiant deux propriétés particulières : il est à la fois *saturé* et *minimal*.
- Saturé, il réalise une neutralisation complète des polarités présentes ; chaque trait positif $t \rightarrow v$ doit rencontrer dans le modèle son dual $t \leftarrow v$ et vice-versa ; chaque trait virtuel doit rencontrer dans le modèle un trait réel correspondant.
- Minimal, le modèle doit ajouter un minimum d'information à celle présente dans les descriptions initiales (il ne peut ajouter ni relation de domination immédiate, ni trait qui ne sont pas présents dans les descriptions de départ).

L'analyse syntaxique se ramène alors à la résolution d'un système de contraintes. Elle consiste à construire tous les modèles saturés et minimaux d'un ensemble fini de DAP élémentaires. Dans la pratique, la grammaire que nous avons construite est totalement lexicalisée : toute DAP élémentaire possède une ancre unique qui lui permet de se lier à un mot de la langue. La lexicalisation permet, pour l'analyse d'une phrase, de sélectionner uniquement des DAP ancrées par des mots de la phrase. Une fois l'ensemble des DAP sélectionné, la construction d'un modèle saturé et minimal se fait pas à pas à l'aide d'une opération de fusion de nœuds deux par deux, guidée par l'une ou l'autre des contraintes suivantes :

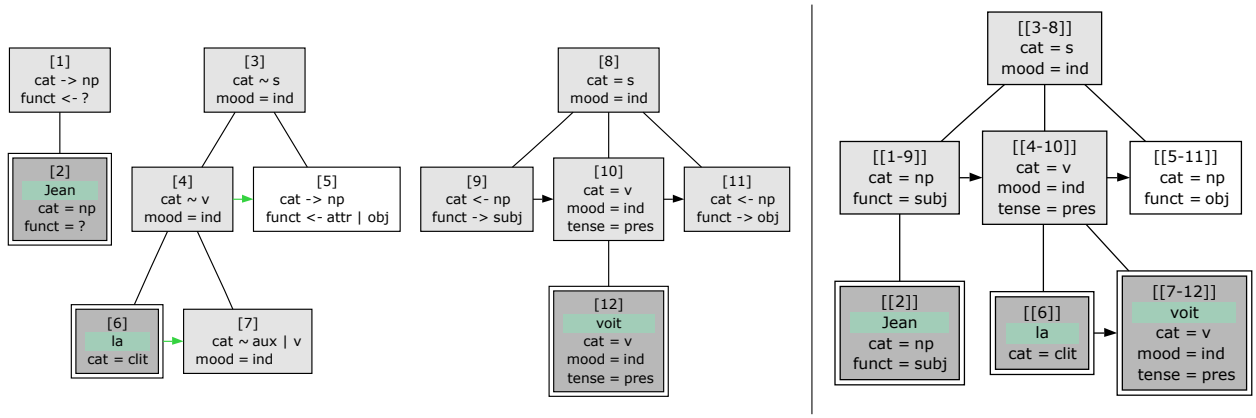

FIG. 2 – DAP associée à la phrase *Jean la voit* et son modèle saturé minimal

– neutraliser un trait positif avec un trait négatif de même nom et porteur d'une valeur qui s'unifie avec celle du premier trait ;
– réaliser un trait virtuel à l'aide d'un trait réel (positif, négatif ou neutre) de même nom et porteur d'une valeur qui s'unifie avec celle du premier trait.

Les contraintes induites par la description font que la fusion de deux nœuds entraîne généralement une superposition partielle de leurs contextes représentés par les fragments d'arbres dans lesquels ils se situent.

Ainsi, les GI combinent les points forts de deux familles de formalismes : la souplesse des *grammaires d'unification* et le contrôle de saturation des *grammaires catégorielles*.

La figure 2 présente un exemple d'analyse syntaxique, celle de la phrase *Jean la voit*[5]. La partie gauche montre l'ensemble des DAP initiales associées par la grammaire à la phrase. La grammaire étant lexicalisée, chacune des DAP est associée à un mot de la phrase et a été extraite d'un lexique. Ces DAP ont été réunies en une seule à laquelle on a ajouté une information de précédence entre les ancres, qui n'apparaît pas sur le schéma, pour prendre en compte l'ordre des mots de la phrase.

Le passage de la description initiale à son modèle donné par la partie droite de la figure 2 est réalisé par une suite de 3 fusions de nœuds[6]. Le simple jeu des contraintes d'arbre fait que ces 3 fusions en entraînent deux autres ainsi qu'une superposition partielle d'arbres.

3 Le pouvoir d'expression des grammaires d'interaction

Dans les limites de cet article, nous ne pouvons étudier de façon exhaustive cette question et nous avons choisi d'en illustrer trois aspects particulièrement significatifs :
– les relations de domination large avec contraintes pour représenter les dépendances non bornées en cascade (pied piping),
– les polarités positives et négatives pour modéliser les paires de mots exprimant la négation,
– les polarités virtuelles pour exprimer la position relativement libre des modificateurs de phrases.

[5]Nous avons simplifié la figure en ne mentionnant pas les traits d'accord.

[6]Dans l'entête de chaque nœud du modèle, on peut retrouver le numéro des nœuds des DAP initiales qui ont été fusionnées.

3.1 Dépendances non bornées et relations de domination larges

Les relations de domination large sont utilisées pour représenter les dépendances non bornées et les structures de traits qu'il est possible d'associer à ces relations permettent d'exprimer des contraintes sur ces dépendances, par exemple les barrières à l'extraction.

Les pronoms relatifs, tels que *qui* ou *lequel*, donnent lieu à des dépendances non bornées en cascade (pied piping) comme dans la phrase : *Jean [dans l'entreprise de* **qui***] Marie sait que l'ingénieur travaille* □ *est malade* :

– Il y a une première dépendance non bornée entre le verbe *travaille* et son complément extrait *dans l'entreprise de qui*. La trace du complément extrait est marquée par le symbole □. Cette dépendance est modélisée dans la DAP associée au pronom relatif *qui* représentée sur la figure 1 par une relation de domination large. La contrainte associée à cette relation de domination exprime que la dépendance du syntagme prépositionnel par rapport au verbe dont il est complément ne peut traverser qu'une suite indéterminée de complétives ou d'infinitives objet. Cela permet de refuser la phrase suivante : ∗ *Jean [dans l'entreprise de* **qui***] Marie qui travaille* □ *le connaît est malade*.
– A l'intérieur du syntagme prépositionnel, il y a une deuxième dépendance non bornée entre la tête du syntagme et le pronom relatif *qui*, qui peut être enchâssée plus ou moins profondément dans ce syntagme. Cette dépendance est aussi représentée sur la figure 1 par une relation de domination large et la contrainte associée exprime que les syntagmes enchâssés sont des noms communs, des syntagmes nominaux ou prépositionnels. Cela permet de refuser la phrase : ∗ *Jean [dans l'entreprise qui appartient à* **qui***] Marie travaille* □ *est malade*.

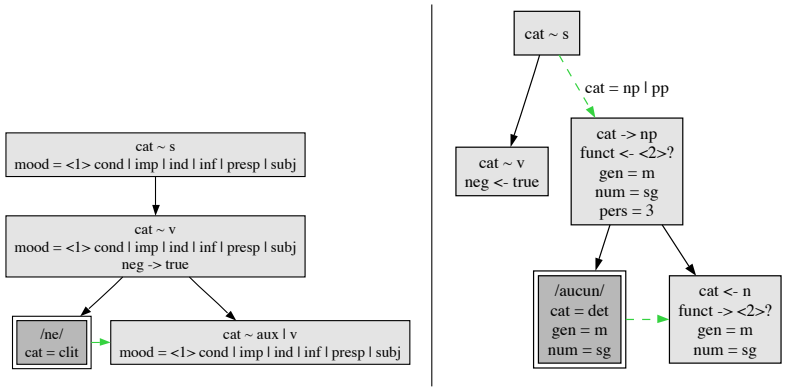

FIG. 3 – DAP associées respectivement à la particule *ne* et au déterminant *aucun*

3.2 L'utilisation des polarités pour modéliser la négation

En français, la négation peut s'exprimer à l'aide de la particule *ne* couplée avec un mot qui peut être un déterminant, un pronom ou un adverbe. La position de la particule *ne* est figée avant un verbe porteur d'une inflexion mais l'autre mot, s'il s'agit d'un déterminant comme *aucun* ou un pronom comme *personne*, peut avoir une position relativement libre dans la phrase, comme le montrent les exemples suivants :

(a) *Jean* **ne** *parle à* **aucun** *collègue.*
(b) *Jean* **ne** *parle à la femme d'***aucun** *collègue.*
(c) **Aucun** *collègue de Jean* **ne** *parle à sa femme.*

Comme le montre la figure 3, le couplage de *ne* avec *aucun* est exprimé par un trait polarisé *neg*

porté par le nœud représentant la projection maximum du noyau verbal : *aucun* est en attente d'un tel trait qui va être fourni par *ne*. La position relativement libre de *aucun* est exprimée par une domination large du nœud représentant la proposition sur le syntagme nominal qu'il introduit. La contrainte associée à cette domination large exprime le fait que *aucun* ne peut introduire que des arguments du verbe tête de la phrase ou leurs compléments. Bien entendu, tous les usages ne sont pas couverts par ces deux descriptions et il est notamment nécessaire de modifier légèrement celle associée à *aucun* pour analyser une phrase comme *Jean* **ne** *voit* **jamais aucun** *responsable*. Il faut ajouter une nouvelle entrée pour *aucun* avec un trait *neg* virtuel au lieu d'être négatif.

3.3 L'adjonction de modificateurs à l'aide de polarités virtuelles

En français, la place des compléments circonstanciels dans la phrase est relativement libre, comme le montrent les exemples suivants :

(a) **Le soir***, Jean va rendre visite à Marie.*
(b) *Jean,* **le soir***, va rendre visite à Marie.*
(c) *Jean va rendre visite* **le soir** *à Marie.*
(d) *Jean va rendre visite à Marie* **le soir**.

Ces variantes expriment des intentions communicatives différentes mais *le soir* est dans tous les cas un complément circonstanciel, modificateur de la phrase.

La polarité *virtuelle* $f \sim v$ n'existait pas dans la version précédente des GI (Perrier, 2004). L'adjonction de modificateurs était effectuée comme dans beaucoup de formalismes (grammaires d'arbres adjoints, grammaires catégorielles ...) par ajout d'un niveau supplémentaire dans l'arbre syntaxique où était le syntagme modifié : à la place d'un nœud de catégorie X était inséré un arbre formé d'une racine de catégorie X et de ses deux fils : le modificateur et le syntagme initial de catégorie X objet de la modification. Si cette introduction d'un niveau supplémentaire est parfois justifiée, le plus souvent elle vient introduire une complexité et une ambiguïté artificielles. Reprenant une idée de (Nasr, 1995) avec son système de polarités noires et blanches, nous avons introduit les polarités virtuelles. Cela nous permet d'ajouter un modificateur comme fils supplémentaire du nœud qu'il modifie sans rien changer au reste de l'arbre syntaxique dans lequel se situe le nœud modifié. La DAP de la figure 1 en donne un exemple car la proposition relative représentée par le fils droit de la racine y apparaît comme un modificateur de groupe nominal. En anglais, on parle alors de *sister adjunction* et elle est utilisée dans certains formalismes (grammaires de dépendance, grammaires de substitution de descriptions (Rambow *et al.*, 2001)). Cette modélisation des modificateurs est beaucoup plus souple que la précédente et nous a permis de traiter les exemples présentés ci-dessus sans difficulté, ainsi que les propositions incises et incidentes, considérées comme des modificateurs de phrases.

4 L'architecture de la grammaire

4.1 L'organisation modulaire de la grammaire

La grammaire a été construite avec l'outil XMG (Duchier *et al.*, 2005) qui permet d'écrire des grammaires d'un haut niveau d'abstraction sous une forme modulaire et de les compiler

ensuite dans des grammaires de plus bas niveau utilisables par des systèmes de TAL. Décrivons brièvement les traits caractéristiques de XMG.

Une grammaire est organisée en une hiérarchie de classes à l'aide de deux opérations de composition : *conjonction* et *disjonction*. Elle est aussi structurée selon plusieurs dimensions qui se retrouvent dans chaque classe. Notre grammaire n'utilise que deux dimensions : la première est la dimension syntaxique où les objets sont des DAP et la seconde est celle de l'interface avec le lexique où les objets sont des structures de traits.

Pour définir la conjonction de deux classes, il est nécessaire de préciser la manière dont les composantes de chaque dimension se combinent :
– pour la dimension syntaxique, c'est l'union des DAP qui est effectuée ;
– pour la dimension des interfaces avec le lexique, c'est l'unification entre structures de traits qui est réalisée.
Pour éviter les collisions d'identificateurs, leur portée est strictement contrôlée ; par défaut, elle est locale à la classe où ils sont déclarés, mais l'on peut exporter un identificateur qui devient alors visible pour l'extérieur. Lorsque l'on combine deux classes, leurs identificateurs exportés doivent être disjoints ; si l'on veut confondre deux identificateurs, il faut le dire explicitement à l'aide d'une équation. A la différence de (Crabbé, 2005) qui utilise un système de couleurs, nous avons choisi de nous servir d'un nombre extrêmement limité de noms de nœuds, pertinents linguistiquement, pour contrôler la conjonction des classes. XMG ne permettant pas de définir des identificateurs globaux, nous avons utilisé des équations entre identificateurs pour contourner le problème.

La grammaire actuelle comprend 448 classes dont 121 classes terminales, qui sont compilées en 2059 DAP. Ces classes sont rangées par famille. Une famille peut être réutilisée par les classes d'une autre. C'est le cas par exemple de la famille *Complement* qui contient les classes définissant les compléments de structures prédicatives. Elle est utilisée par 3 autres familles : *Adjective*, *Noun* et *VerbDiathesis*, qui décrivent respectivement le comportement syntaxique des adjectifs, celui des noms et les différentes diathèses du verbe. La famille *VerbDiathesis* utilise aussi les familles *Verbmorphology* et *Verbfunction* qui décrivent respectivement la morphologie verbale et les fonctions que peut occuper le verbe dans la phrase (y compris lorsqu'il est participe présent et participe passé).

4.2 La liaison avec un lexique indépendant du formalisme

La grammaire, dans sa forme actuelle, est totalement lexicalisée : chaque DAP élémentaire de la grammaire a un unique nœud ancre destiné à être associé à un mot de la langue. Pour cela, chaque DAP est associée à une structure de traits qui décrit de façon indépendante du formalisme un cadre syntaxique correspondant aux mots pouvant ancrer la description. Cette structure de traits constitue l'*interface* de la DAP avec le lexique.

L'ensemble des traits utilisés dans les interfaces sont différents de ceux utilisés dans les descriptions car leur rôle n'est pas le même : ils visent non pas à décrire des syntagmes mais les mots de la langue et ceci d'une façon indépendante du formalisme.

La figure 4 présente dans sa partie gauche une DAP non ancrée correspondant à un verbe transitif à un temps fini dans une configuration canonique. Cette description est accompagnée de son interface et on y a fait figurer les indices associés à certaines valeurs qui montrent que certains traits de l'interface partagent leur valeur avec des traits de la description.

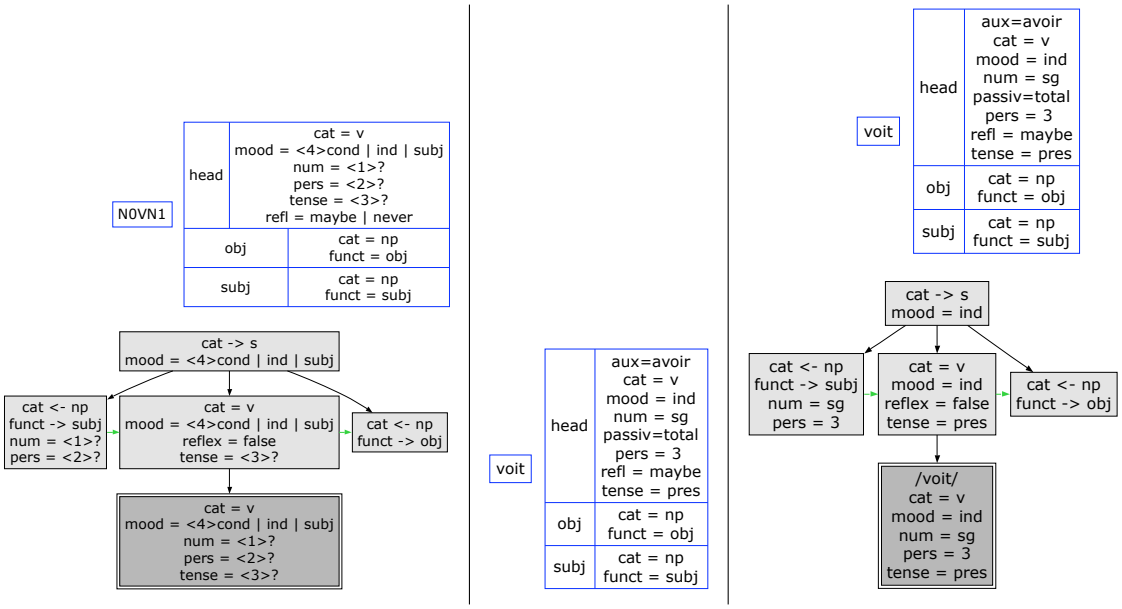

FIG. 4 – DAP non ancrée associée à un verbe transitif, entrée lexicale pour le verbe *voit* et DAP ancrée par le verbe *voit*

Le lexique des mots de la langue associe ceux-ci à des cadres syntaxiques dans un format identique aux interfaces des descriptions. La figure 4 , dans sa partie centrale, montre une entrée d'un tel lexique pour le verbe *voit* : il décrit le cadre syntaxique associé à *voit* lorsque celui-ci est employé comme verbe transitif.

L'ancrage des descriptions de la grammaire se fait ensuite par unification de leurs interfaces avec les cadres syntaxiques compatibles du lexique. Un mécanisme de co-indexation entre valeurs de traits de la description et de l'interface permet un paramétrage de certains traits. La figure 4 montre dans sa partie droite la DAP obtenue par ancrage de la DAP de gauche à l'aide de l'entrée lexicale présentée au centre. Cet ancrage a consisté à unifier l'interface de la DAP non ancrée avec le cadre syntaxique offert par le lexique.

5 Evaluation sur une suite de phrases tests

Notre but est d'évaluer le plus finement et de la façon la moins coûteuse possible la couverture de notre grammaire. Une réponse adaptée est d'utiliser une suite de phrases tests mais il est important que cette suite contienne non seulement des exemples positifs mais aussi des exemples négatifs pour évaluer le pouvoir de surgénération de la grammaire.

Nous avons choisi l'une des rares suites de ce type qui existe pour le français : la TSNLP (Lehmann *et al.*, 1996) qui comprend 1690 phrases positives et 1935 phrases négatives. Elle est loin de couvrir toute la grammaire du français ; notamment, elle contient très peu de phrases complexes et par contre, elle s'attarde beaucoup sur certains phénomènes tels que la coordination ou l'ordre des compléments circonstanciels dans la phrase. On peut même dire que notre grammaire prend en compte des phénomènes ignorés de la TSNLP : voies passive et moyenne, sous-catégorisation des noms et adjectifs prédicatifs, contrôle du sujet des infinitives compléments, propositions relatives, interrogatives . . .

Pour effectuer l'analyse, nous avons utilisé LEOPAR[7], qui est un analyseur syntaxique fondé sur les GI. Avec la grammaire actuelle, il accepte 88% des 1690 phrases positives et rejette 85% des 1935 phrases négatives de la TSNLP. Les 15% de phrases négatives acceptées le sont essentiellement par absence d'intégration de règles phonologiques et de la sémantique dans la grammaire. Les 12% des phrases positives non couvertes le sont pour des raisons très diverses : phrases du langage parlé prenant certaines libertés avec la grammaire, expressions figées ou semi-figées, phénomènes non encore pris en compte (constructions causatives, superlatifs . . .).

6 Perspectives

D'ores et déjà, il est possible d'utiliser LEOPAR avec un lexique à large couverture pour analyser des corpus tout venant. Il est nécessaire d'enrichir la grammaire pour couvrir un certain nombre de phénomènes linguistiques courants non encore pris en compte. Il faudra aussi améliorer les performances de l'analyseur pour faire face à l'explosion potentielle résultant de l'augmentation de la taille de la grammaire se conjuguant avec la longueur des phrases.

Références

BRESNAN J. (2001). *Lexical-Functional Syntax*. Oxford : Blackwell Publishers.

CRABBÉ B. (2005). *Représentation informatique de grammaires fortement lexicalisées : application à la grammaire d'arbres adjoints*. thèse de doctorat, université Nancy2.

DUCHIER D., LE ROUX J. & PARMENTIER Y. (2005). XMG : Un compilateur de métagrammaires extensible. In *TALN 2005, Dourdan, France*.

LEHMANN S., OEPEN S., REGNIER-PROST S., NETTER K., LUX V., KLEIN J., FALKEDAL K., FOUVRY F., ESTIVAL D., DAUPHIN E., COMPAGNION H., BAUR J., BALKAN L. & ARNOLD D. (1996). TSNLP — Test Suites for Natural Language Processing. In *Proceedings of COLING 1996, Kopenhagen*.

NASR A. (1995). A formalism and a parser for lexicalised dependency grammars. In *4th International Workshop on Parsing Technologies (IWPT)*.

PERRIER G. (2004). La sémantique dans les grammaires d'interaction. *Traitement Automatique des Langues*, **45**(3), 123–144.

PULLUM G. K. & SCHOLZ B. C. (2001). On the Distinction between Model-Theoretic and Generative-Enumerative Syntactic Frameworks. In *LACL 2001, Le Croisic, France*, volume 2099 of *Lecture Notes in Computer Science*, p. 17–43.

RAMBOW O., VIJAY-SHANKER K. & WEIR D. (2001). D-tree substitution grammars. *Computational Linguistics*, **27**(1), 87–121.

RETORÉ C. (2000). The Logic of Categorial Grammars. *ESSLI'2000, Birmingham.*

ROGERS J. & VIJAY-SHANKER K. *(1994). Obtaining trees from their descriptions : an application to tree-adjoining grammars.* Computational Intelligence, **10***(4), 401–421.*

SAG I. A., WASOW T. & BENDER E. M. *(2003).* Syntactic Theory : a Formal Introduction. *Center for the Study of Language and INF.*

[7] www.loria.fr/equipes/calligramme/leopar