

Des *constructions* à l'interface entre lexique et grammaire

Marie-Laure Guénot
LPL – CNRS / Université de Provence
mlg@lpl.univ-aix.fr

Proposition de communication orale pour la journée de l'Atala
Interface lexique-grammaire et lexiques syntaxiques et sémantiques
29 janvier 2005

Résumé. — Dans cette communication on tâchera de montrer et d'expliquer l'usage que l'on fait, la distinction et les relations que l'on établit entre *lexique* et *grammaire* dans un ensemble constitué d'un paradigme de description linguistique et d'un modèle formel de représentation. Dans ce but on commencera par observer les emplois de *lexique* et de *grammaire* dans d'autres modèles linguistiques¹, puis on proposera notre approche en l'illustrant à l'aide d'un exemple de description.

1 Lexique et grammaire dans les modèles linguistiques formels

La plupart des modèles de description linguistique courants font partie de ce qu'on appelle les "grammaires lexicalisées". Des grammaires telles que TAG (Joshi (1987)) ou HPSG (Pollard & Sag (1994)) sont caractérisées par l'intégration, au sein même de la représentation d'une unité lexicale (au même titre que celle d'un syntagme), à la fois d'informations *intrinsèques* qui définissent l'objet en tant que partie du *système lexical* de la langue, et d'informations *extrinsèques* qui le définissent en tant que partie du *système grammatical* de la langue. Pour preuve de la distinction nécessaire des deux systèmes pourtant, le recours dans ces modèles à des règles de dérivation lexicale pour obtenir les variantes grammaticales possibles de chaque objet lexical.

On mêle donc dans un seul objet, des informations concernant deux systèmes, certes interdépendants, néanmoins distincts. A tel point que dans des formalismes tels que celui des Grammaires Catégorielles (Wood (1993)) – qui se définit lui-même comme "extrêmement lexicalisé" (*extreme lexicalization*, Kruijff (2004)) – ou que *Dynamic Syntax* (Kempson *et al.* (2001)), ce sont les objets du lexique eux-mêmes, définis sous la forme de *procédures d'analyse*, qui contiennent toutes les informations grammaticales des modèles. En effet dans ce cas il n'existe pas de grammaire distincte, toute ou partie, du lexique lui-même (pas de description d'autres objets), mis à part quelques principes généraux qui régissent les combinaisons possibles, et qui permettent donc de dériver, à partir de ces unités "lexicales", la structure linguistique d'un énoncé sans passer par des descriptions de syntagmes. Les informations proprement lexicales des unités sont exprimées de manière *implicite* à l'intérieur même de leur description procédurale (qui est un ensemble d'informations explicitement grammaticales). Ici le lexique est véritablement la grammaire, et *vice versa*, et les deux sont confondus comme une seule et même entité.

Toutes ces grammaires, comme a commencé de le montrer Kahane (2004), ont en outre ceci de commun que les objets minimaux qu'elles manipulent (disons les "mots") sont interprétables comme étant soit des "ressources" ("*je sais qui je suis*") soit des "besoins" ("*je sais qui je réclame*"), soit des ensembles "ressources-besoins" (chaque trait pouvant être l'un ou l'autre). En revanche il est impossible de ne tenir compte dans ces modèles que *d'une partie* des "ressources" et/ou des "besoins" d'une de ces unités, séparément du reste ; en d'autres termes, il est impossible de manipuler des sous-ensembles variables (d'où la possibilité de les représenter sous forme de systèmes polarisés, dans lesquels chaque unité a une "charge" donnée, définie dans la grammaire). Pour notre part, nous postulons que ce n'est pas toujours l'unité "mot" qui contient en elle-même, considérée dans sa globalité, les informations sur ses relations avec le reste de l'énoncé, et souhaitons marquer clairement cette différence existant entre le système lexical et le système grammatical de la langue, sans avoir à passer par des règles qui ont pour charge de modifier le contenu de base des unités lexicales (i.e., de changer la "grammaire" contenue dans le "lexique"). Selon ce postulat les informations grammaticales (i.e., concernant les relations entre objets d'un énoncé) peuvent être

¹On ne s'intéressera dans cette communication qu'aux modèles formels qui ont pour but d'offrir une représentation des études de linguistique descriptive, et on ne tiendra donc pas compte ici des modèles stochastiques de TALN.

exprimées de manière régulière en fonction d'un ensemble d'informations dont la constitution est variable à volonté, pouvant contenir des caractéristiques provenant de domaines différents (morphologie, syntaxe, sémantique, prosodie, pragmatique). Cela signifie que ces ensembles manipulés dans la grammaire ne correspondent pas forcément toujours à des unités "lexicales", mais peuvent pointer des objets plus larges ou plus fins, en fonction des descriptions effectuées.

2 Les *constructions* à l'interface entre lexique et grammaire

A ce point nous pouvons donner une définition de ce que l'on nommera *lexique* et *grammaire* dans notre paradigme : notre lexique ne contient que des informations intrinsèques, et toutes les informations extrinsèques font l'objet de descriptions grammaticales. Il est évident que dans ce cas comme dans les cas précédents, lexique et grammaire sont intimement liés et fortement interdépendants ; pour autant l'on tâche de distinguer clairement ce qui est propre à l'un de ce qui est propre à l'autre, puisque les deux ne peuvent être confondus.

Dans notre approche, nous nous plaçons dans le paradigme descriptif de la *Construction Grammar* (Kay & Fillmore (1999), ci-après CxG), parce qu'il propose une analyse résolument multi-domaines, non-modulaire et non-compositionnelle, prenant en considération dans des descriptions linguistiques de corpus très fines, des informations relationnelles issues de domaines différents (syntaxe, sémantique, prosodie,...) et fondamentalement interagissantes.

CxG étant uniquement un modèle de description linguistique, et de ce fait ne contraignant pas le formalisme employé pour sa représentation, nous utilisons à cette fin le modèle des Grammaires de Propriétés (Blache (2005), ci-après GP)².

Dans ce paradigme CxG/GP, nous appelons *construction* tous les objets décrits, aussi bien lexicaux que grammaticaux, et tous bénéficient donc du même matériel de description. Ces constructions peuvent être considérées comme des "patrons" plus ou moins complexes, plus ou moins généraux, plus ou moins fins³, faisant référence à des ensembles de traits caractéristiques pouvant être mixtes, et constituant entre eux un réseau complexe de descriptions linguistiques. En d'autres termes, lexique et grammaire en CxG/GP sont deux systèmes, distincts non par leur *forme*, mais par leur *contenu*.

3 Application

Illustrons notre propos par un exemple syntaxique simple⁴.

- (1) a. il pleut
- b. il y a du givre sur le pare-brise
- c. il est arrivé il y a déjà deux ans
- d. il y a les voisins qui arrivent
- e. il y a des gens qui travaillent
- f. il y a des gens ils sont cons

On peut décrire une première construction correspondant à *il*, que l'on représentera formellement comme dans la figure 1, et une autre construction correspondant à *il y a*, que l'on représentera comme dans la figure 2.

Il apparaît dans ces deux figures que le matériel formel utilisé est le même, mais que ce qui distingue une construction décrivant un objet du lexique (figure 1) d'une construction d'objet grammatical (figure 2) c'est que la deuxième contient non seulement des informations lexicales (TRAITS dans la représentation formelle) mais également des informations grammaticales (les PROPRIÉTÉS des GP). Nous classons volontairement *il y a* comme objet *grammatical* (alors qu'il aurait pu être considéré comme une locution et classé parmi les objets du lexique), parce que l'on peut considérer que cette construction est véritablement constituée d'éléments qui sont mis en relation les uns avec les autres, et dont les caractéristiques relationnelles (la *caractérisation* des GP) peuvent être plus ou moins contraintes et impliquent des analyses différentes. Par exemple, le *il y a* des exemples (1b), (1e) et (1f) est moins contraint que celui de (1d) et (1c), puisque dans le premier cas il pourra recevoir un constituant supplémentaire *en* (exemples

²L'une des différences essentielles entre les bases théoriques de GP et de CxG est que GP n'est pas un modèle génératif, en ce sens qu'une grammaire ne doit pas nécessairement décrire *toutes et uniquement* les phrases d'une langue. Cette différence ne pose néanmoins aucun problème, puisqu'elle permet au contraire d'envisager *grammaire* et *grammaticalité* comme deux choses différentes.

³On peut remarquer que ceci, qui est une définition des *constructions* de CxG, correspond parfaitement à la définition de la *granularité* des GP, et ce dès la version de Blache (2001).

⁴Cet exemple est largement inspiré de l'étude de *il y a* proposée dans Cappeau & Deulofeu (2001).

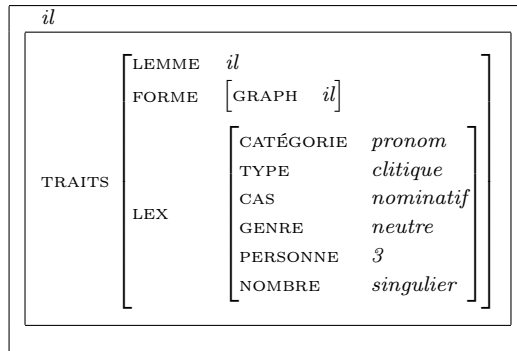


FIG. 1 – Description de *il*.

(2a), (2d) et (2e)) alors que dans le deuxième *il* ne pourra pas (exemples (2c) et (2b)). Cappeau & Deulofeu (2001) posent suite à cette distinction, une opposition entre un emploi “constatif” et un emploi “classifiant”, et Blanche-Benveniste (2000) et Le Goffic (2000) considèrent cela comme la preuve de l’existence de deux *il y a* distincts, dont l’un est une construction verbale (1b) et l’autre une construction prépositionnelle (1c).

- (2) a. *il y en a du givre sur le pare-brise*
 b. * *il est arrivé il y en a déjà deux ans*
 c. * *il y en a les voisins qui arrivent*
 d. *il y en a des gens qui travaillent*
 e. *il y en a des gens ils sont cons*

De la même façon, on peut constater d’autres propriétés qui vont modifier l’analyse de *il y a* (posant une distinction entre une forme microsyntaxique – exemple (1e) – et une forme macrosyntaxique – exemple (1f)), comme par exemple le fait, pour le syntagme dont il fait partie, de pouvoir être déplacé ou non (exemples (3)) :

- (3) a. * *qui travaillent il y a des gens*
 b. *il sont cons il y a des gens / il y en a (des gens)*

Et ainsi, peu à peu l’étude fine de la forme *il y a* en tant que *construction grammaticale*, c’est-à-dire répondant à des caractéristiques de formation, des contraintes portant sur son usage dans la langue (et non sur sa forme lexicale), vont être posées petit à petit et ajoutées à la grammaire, sans que le lexique lui-même ne soit concerné par ces précisions.

Conclusion

Nous avons proposé ici une approche théorique de la distinction entre lexique et grammaire, basée sur une différenciation de contenu dans la description : là où le lexique ne contient que des informations intrinsèques, la grammaire contient aussi des informations extrinsèques. Les deux ne sont pas distinguées par le matériel employé pour les représenter, mais uniquement par leur contenu. Notre approche est la conséquence d’un postulat que nous posons, selon lequel le lexique est un système linguistique à part entière, au même titre que le système grammatical, et que l’on ne peut donc pas les confondre. Chacun des deux systèmes peut contenir un ensemble mixte d’informations issues de domaines différents, qui sont représentés sous une forme homogène permettant leur interaction non-compositionnelle et non-modulaire.

En nous basant sur le modèle descriptif de CxG et formel de GP, nous pouvons représenter à la fois la distinction concrète existant entre lexique et grammaire, ainsi que leur interdépendance, lesquelles permettent la description, la représentation et l’analyse linguistique fines. Les constructions constituent dans ce cadre le matériel de description commun à tous les objets linguistiques (indistinctement qu’ils soient lexicaux ou grammaticaux) ; et ce qui met évidence leur distinction fondamentale (formant deux systèmes distincts) est fourni par le contenu des constructions elles-mêmes.

<i>il y a</i>															
TRAITS	<table style="border: none;"> <tr> <td style="border: none;">FORME</td> <td style="border: none;">[GRAPH <i>il y a</i>]</td> </tr> <tr> <td style="border: none;">LEX</td> <td style="border: none;"> <table style="border: none;"> <tr> <td style="border: none;">MODE</td> <td style="border: none;"><i>indicatif</i></td> </tr> <tr> <td style="border: none;">TEMPS</td> <td style="border: none;"><i>présent</i></td> </tr> <tr> <td style="border: none;">GENRE</td> <td style="border: none;"><i>neutre</i></td> </tr> <tr> <td style="border: none;">PERSONNE</td> <td style="border: none;"><i>3</i></td> </tr> <tr> <td style="border: none;">NOMBRE</td> <td style="border: none;"><i>singulier</i></td> </tr> </table> </td> </tr> </table>	FORME	[GRAPH <i>il y a</i>]	LEX	<table style="border: none;"> <tr> <td style="border: none;">MODE</td> <td style="border: none;"><i>indicatif</i></td> </tr> <tr> <td style="border: none;">TEMPS</td> <td style="border: none;"><i>présent</i></td> </tr> <tr> <td style="border: none;">GENRE</td> <td style="border: none;"><i>neutre</i></td> </tr> <tr> <td style="border: none;">PERSONNE</td> <td style="border: none;"><i>3</i></td> </tr> <tr> <td style="border: none;">NOMBRE</td> <td style="border: none;"><i>singulier</i></td> </tr> </table>	MODE	<i>indicatif</i>	TEMPS	<i>présent</i>	GENRE	<i>neutre</i>	PERSONNE	<i>3</i>	NOMBRE	<i>singulier</i>
FORME	[GRAPH <i>il y a</i>]														
LEX	<table style="border: none;"> <tr> <td style="border: none;">MODE</td> <td style="border: none;"><i>indicatif</i></td> </tr> <tr> <td style="border: none;">TEMPS</td> <td style="border: none;"><i>présent</i></td> </tr> <tr> <td style="border: none;">GENRE</td> <td style="border: none;"><i>neutre</i></td> </tr> <tr> <td style="border: none;">PERSONNE</td> <td style="border: none;"><i>3</i></td> </tr> <tr> <td style="border: none;">NOMBRE</td> <td style="border: none;"><i>singulier</i></td> </tr> </table>	MODE	<i>indicatif</i>	TEMPS	<i>présent</i>	GENRE	<i>neutre</i>	PERSONNE	<i>3</i>	NOMBRE	<i>singulier</i>				
MODE	<i>indicatif</i>														
TEMPS	<i>présent</i>														
GENRE	<i>neutre</i>														
PERSONNE	<i>3</i>														
NOMBRE	<i>singulier</i>														
PROPRIÉTÉS	OBLIGATION	$\left(\begin{array}{l} \text{LEMME } \textit{avoir} \\ \text{TYPE } \textit{principal} \end{array} \right)$													
	CONSTITUANTS	<i>y, il</i>													
	EXIGENCE	$\left(\begin{array}{l} \text{LEMME } \textit{avoir} \\ \text{TYPE } \textit{principal} \end{array} \right) \Rightarrow y \wedge il$													
	LINÉARITÉ	$il \prec y$ $y \prec \left(\begin{array}{l} \text{LEMME } \textit{avoir} \\ \text{TYPE } \textit{principal} \end{array} \right)$													
	ACCORD	$\left(\begin{array}{l} \text{LEMME } \textit{avoir} \\ \text{TYPE } \textit{principal} \end{array} \right). \text{PERSONNE} \approx (\text{FORME } \textit{il}). \text{PERSONNE}$ $\left(\begin{array}{l} \text{LEMME } \textit{avoir} \\ \text{TYPE } \textit{principal} \end{array} \right). \text{NOMBRE} \approx (\text{FORME } \textit{il}). \text{NOMBRE}$													

FIG. 2 – Description de *il y a*.

Références

- Philippe Blache. *Les grammaires de propriétés — des contraintes pour le traitement automatique des langues naturelles*. Hermès Sciences Publications, Paris, 2001.
- Philippe Blache. Property grammars : A fully constraint-based theory. In H Christiansen, P Skadhauge, & J Villadsen, editors, *Constraint Satisfaction and Language Processing*. Springer-Verlag, 2005.
- Claire Blanche-Benveniste. *Approches de la langue parlée en français*. L’essentiel français. Ophrys, Paris, 2000.
- Paul Cappeau & José Deulofeu. Partition et topicalisation : *il y en a* “stabilisateur” de sujets et de topiques indéfinis. *Cahiers de praxématique*, 37, 2001.
- Pierre Le Goffic. *Grammaire de la phrase française*. Langue française. Hachette Supérieur, Paris, troisième édition, 2000.
- Aravind K. Joshi. Introduction to tree adjoining grammar. In Manaster Ramer, editor, *The mathematics of language*, pages 87–114. Benjamins, Amsterdam, 1987.
- Sylvain Kahane. Grammaires d’unification polarisées. In *Actes de TALN 2004*, Fès, Maroc, 2004.
- Paul Kay & Charles J. Fillmore. Grammatical constructions and linguistic generalizations : The *what’s X doing Y?* construction. *Language*, 75(1) :1–33, March 1999.
- Ruth Kempson, Wilfried Meyer-Viol, & Dov Gabbay. *Dynamic Syntax – The flow of language understanding*. Blackwell, Oxford, 2001.
- Geert-Jan Kruijff. Multi-modal combinatory categorial grammar. ESSLLI Lecture Notes, 2004.
- Carl Pollard & Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago, 1994.
- Mary McGee Wood. *Categorial Grammars*. Routledge, London, 1993.