

# DIALANGUE

Volume 1

Avril 1990

## BULLETIN DE LINGUISTIQUE

Module des lettres / Maîtrise en linguistique  
Université du Québec à Chicoutimi

---

*«[...] En ce qui concerne la langue, toutefois, nous possédons quelques certitudes. La première, dans l'ordre politique, c'est que la question de la langue est inflammable. La deuxième, c'est que qui perd sa langue perd son âme [...] La troisième, c'est que la langue doit être vue. Quand on ne voit plus sa langue, on cesse de la parler. Une ville saute d'abord aux yeux. Le problème de l'affichage est révélateur, c'est le cas de le dire. La quatrième, c'est que la langue doit être portée dans le coeur de ceux qui la parlent.»*

Jean-Paul Desbiens,  
«Sur le bout de la langue»,  
La Presse (Montréal), 22 octobre 1986.

## SOMMAIRE

- Les orientations de recherche des professeurs de linguistique de l'Université du Québec à Chicoutimi
  - Synthèse des mémoires de maîtrise
  - Mémoires et travaux de 1<sup>er</sup> cycle
  - La situation du français à l'Université du Québec à Chicoutimi et dans la région
  - Comptes rendus d'ouvrages récents
-

## DÉVELOPPEMENTS RÉCENTS EN ANALYSE SYNTAXIQUE PAR ORDINATEUR

*François Labelle*

L'introduction du langage de programmation Prolog, en plus d'offrir un cadre adéquat pour le traitement de la sémantique, simplifie grandement la rédaction des grammaires génératives (gg) sur ordinateur. Dans l'exemple suivant, on peut comparer une gg en forme courante et sa traduction en forme Prolog DCG:

	<b>Forme standard</b>	<b>Forme Prolog (DCG)</b>
(1)	P -> SN SV	p --> sn, sv.
(2)	SN -> DET N	sn --> det, n.
(3)	SV -> V SN	sv --> v, sn.
(4)	DET -> le	det --> [le].
(5)	DET -> la	det --> [la]
(6)	N -> chef	n --> [chef].
(7)	N -> soupe	n --> [soupe].
(8)	V -> déguste	v --> [déguste].

Malheureusement, cette forme simple et transparente est insuffisante pour rendre compte de phénomènes courants dans les langues naturelles:

**L'accord grammatical:** Notre petite grammaire produirait et analyserait la phrase incorrecte \***La chef déguste le soupe**. Dans sa forme actuelle, aucun mécanisme ne permet de formuler les contraintes grammaticales d'accord.

**Les restrictions sémantiques:** Notre grammaire produirait et analyserait la phrase incorrecte **La soupe déguste**. Dans sa forme actuelle, aucun mécanisme ne permet de formuler les contraintes syntactico-sémantiques (par exemple, l'obligation d'avoir un objet direct, un sujet de type humain).

**Les phénomènes de dépendance à distance:** Il s'agit généralement de phénomènes tels que la pronominalisation, certains effacements à distance, dans lesquels la nature d'un constituant (l'antécédent) détermine la forme d'un constituant apparaissant à une certaine distance de l'antécédent. Ainsi, par exemple, si nous augmentions notre grammaire de la règle **REL -> que P**, afin de lui permettre de produire des relatives, elle produirait également des structures parasites telles que **La soupe que le chef déguste la soupe**. Il faut en effet

un mécanisme permettant de formuler la contrainte suivante: lorsqu'une phrase, la relative, débute par **que**, il doit y avoir dans cette phrase un constituant manquant et ce constituant doit être un objet direct.

Il est possible de compliquer notre grammaire de façon à pouvoir exprimer ces contraintes. Malheureusement, on y perd rapidement la simplicité et la transparence initiales. Or, deux techniques ont été proposées récemment pour rendre compte des phénomènes mentionnés tout en conservant une grande transparence: les **grammaires d'unification** et les **grammaires contextuelles discontinues**.

### **Les grammaires d'unification (gu):**

Les grammaires d'unification sont des grammaires en forme Prolog DCG, comme celle présentée plus haut, augmentées d'un opérateur d'unification généralement noté **===**. A titre d'exemple, la règle exprimant l'accord en nombre du SN sujet et du SV d'une phrase pourrait avoir l'allure suivante:

```
p --> sn, sv,  
      sn:nombre === sv:nombre.
```

Cette règle, semblable à la règle (1) de notre première grammaire, en plus de combiner le SN sujet et le SV pour former une phrase, ajoute la condition suivante: le nombre du SN doit être le même que le nombre du SV. Techniquement, la tâche de l'opérateur **===** est très riche et tient compte des cas suivants qui peuvent apparaître au cours d'une analyse syntaxique:

- a) Le nombre du SN est déterminé, celui du SV ne l'est pas. Dans ce cas, le SV prendra le nombre du SN.
- b) Le nombre du SV est déterminé, celui du SN ne l'est pas (cas d'inversion, par exemple, où le sujet suit le verbe). Dans ce cas, le SN prendra le nombre du SV.
- c) Le nombre du SN et celui du SV sont tous deux déterminés. Dans ce cas, la règle ne pourra s'appliquer que si le nombre du SN et du SV est identique.
- d) Ni le SN ni le SV n'ont de nombre déterminé. Dans ce cas, il est simplement noté que dès que le nombre de l'un sera déterminé, celui de l'autre le sera aussi.

Donnons un autre exemple. Dans une construction réfléchie, on doit pouvoir exprimer le fait que les traits sémantiques du sujet de la phrase sont identiques aux traits sémantiques de l'objet direct du SV. Dans une grammaire d'unification, cette propriété serait exprimée par la règle suivante:

```
p --> sn, sv,  
      p:sujet:sem === sv:objet:sem.
```

Les grammaires d'unification permettent donc de formuler de manière relativement simple et transparente les contraintes grammaticales et les restrictions syntactico-sémantiques.

### **Les grammaires contextuelles discontinues (gcd):**

Une règle de forme gcd est en fait une **règle complexe composée de deux ou plusieurs sous-règles** séparées par le marqueur /. Sa forme générale est Règle<sub>1</sub> / ... / Règle<sub>n</sub> soit, par exemple:

$$a \rightarrow x / b \rightarrow y / c \rightarrow z.$$

Une règle cdg stipule que si la première sous-règle de gauche s'applique ( dans l'exemple,  $a \rightarrow x$ ), alors il faudra plus tard appliquer les autres sous-règles  $b \rightarrow y$  et  $c \rightarrow z$ , lorsque la situation s'y prêtera. En d'autres mots, ce formalisme permet de représenter un lien de dépendance entre des règles.

C'est exactement ce type de mécanisme qu'il nous faut pour rendre compte des dépendances à distance comme celui qu'on retrouve dans la construction relative. On voudrait pouvoir dire, dans ce cas, que si on applique la règle introduisant un pronom relatif, alors il faudra que, plus loin en cours d'analyse, un des SN de la phrase relative soit absent. Ce qui s'exprimerait, en simplifiant un peu, par la règle cdg suivante (le symbole [] désignant une chaîne nulle):

$$\text{rel} \rightarrow \text{prorel}, p / \text{sn} \rightarrow [].$$

La règle devrait être un peu plus complexe afin de tenir compte de la fonction grammaticale du pronom relatif: par exemple, dans le cas de **que**, c'est le SN objet direct qui devra être nul. En combinant les techniques des grammaires contextuelles discontinues et des grammaires d'unification, cette restriction pourrait être exprimée simplement.

Ces développements récents permettront de réduire de façon significative la complexité des tâches de conception et de rédaction des analyseurs syntaxiques et faciliteront la maîtrise de cet aspect de la linguistique computationnelle.

### **Bibliographie:**

Gazdar, G., C.S. Mellish (1989) Natural language processing in Prolog, Addison-Wesley, Reading, MASS.

Saint-Dizier, P. (198?) Contextual discontinuous grammar, COLING Proceedings.

Sells, P. (1985) Lectures on contemporary syntactic theories, CSLI, Stanford, CA.

Shieber, S. (1986) An introduction to unification-based approaches to grammar, CSLI, Stanford, CA.