Langue naturelle et langage mathématique; éléments de comparaison

Animateurs : J. PLAZY et J. ROUAULT

I — Introduction

1) *Aspect linguistique*

Le présent travail prend la suite de celui publié dans le Bulletin de l'A.P.M. n° 280 et intitulé "Problèmes posés par les liens possibles entre les enseignements de la grammaire et les mathématiques dans le premier cycle du second degré".

Nous avions alors proposé une structure destinée à rendre compte du fonctionnement de la phrase : cette structure, dite de dépendance, est en fait une arborescence dont les sommets sont étiquetés par les mots de la phrase et dont les arcs sont étiquetés par les relations syntaxiques liant ces mots.

2) *Aspect mathématique*

Il est remarquable de constater que certains ouvrages d'enseignement utilisent un concept très proche du précédent, pour rendre compte de la structure de certaines expressions du langage mathématique : expressions arithmétiques ou algébriques, par exemple.

3) *Conséquence*

L'idée qui nous a guidés pour élaborer ce travail réside dans cette convergence : le même concept formel peut rendre compte de la structure d'une phrase écrite en langue naturelle aussi bien que de la façon dont est construite une expression mathématique.

Nous sommes donc partis de cette structure de dépendance et nous avons cherché à l'utiliser pour cerner les points de convergence, ainsi que les points de divergence, entre langue naturelle et langage mathématique. Cependant, nous n'avons nullement recherché l'exhaustivité : l'outil de comparaison ne permet certainement pas d'y arriver.

4) *Note*

Signalons en passant que les structures de dépendance sont utilisées dans l'analyse automatique des langues naturelles ainsi que dans celle des "expressions arithmétiques" des langages de programmation.

II — Rappels

1) *Structure de dépendance en langue naturelle*

Nous ne donnons pas ici une définition formelle des structures utilisées. Nous nous contentons de rappeler les éléments qui sont indispensables à la compréhension de ce texte.

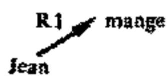
— Une structure de dépendance est une arborescence dont les noeuds et les arcs sont étiquetés.

— Les noeuds sont étiquetés par les unités lexicales présentes dans la phrase dont on veut rendre compte.

— Les arcs sont étiquetés par des relations :
  . R1 à R5
  . CIRC (p) où p est une préposition
  . DET, QUAL, etc...
  . ①, ②, ... ⓝ

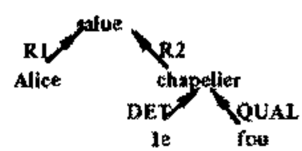
Exemples :

(a1)

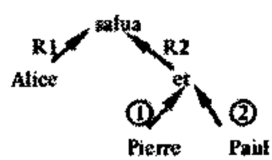


(a2)

(a3)



(a4)



Remarque :

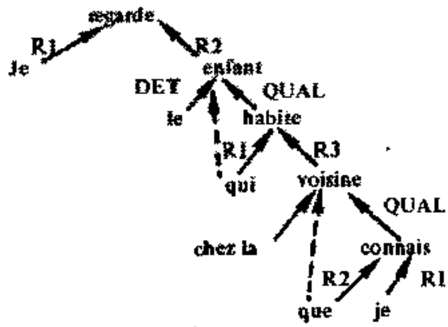
Les exemples (a3) et (a4) sont différents. Ceci est dû à l'étiquette du noeud 1. Dans le cas (a4), en effet, cette étiquette est "et". Cet élément ne peut normalement coordonner que des syntagmes de même fonctionnement. A partir d'eux, il engendre un syntagme de même fonctionnement que les coordonnées.

Exemple :

$$\text{Adj} + \text{“et”} + \text{Adj} \longrightarrow \text{Adj}$$
$$\text{N} + \text{“et”} + \text{N} \longrightarrow \text{N}$$

etc...

(a5)



Remarques :

1) Ces structures obéissent à la règle de projection : on disposera les flèches de telle sorte que l'ordre linéaire de la phrase puisse être restitué par projection sur un axe horizontal.

2) Deux arcs ne peuvent se croiser.

3) Les flèches en pointillé indiquent une relation entre unités lexicales. Dans l'exemple (a5), elles localisent l'antécédent du pronom.

2) *Structure de dépendance en mathématiques*

Pour les besoins de ce travail, nous sommes amenés à modifier les structures utilisées habituellement. Dans les utilisations habituelles des structures de dépendance, on n'étiquette pas les arcs de l'arborescence, et l'on a par exemple :

(a/b)

(b/a)

Nous adoptons une notation qui, étiquetant les arcs, introduit dans le langage mathématique une souplesse analogue à celle des langues naturelles. Nous considérons donc comme distinctes les deux structures :

(a/b)

(b/a) :"a est divisé par b"

Exemples :

(b1)

(−a)

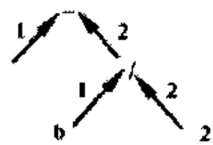
(b2)

(a − b)

(b2')

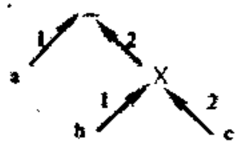
$(a \leqslant b)$

(b3)



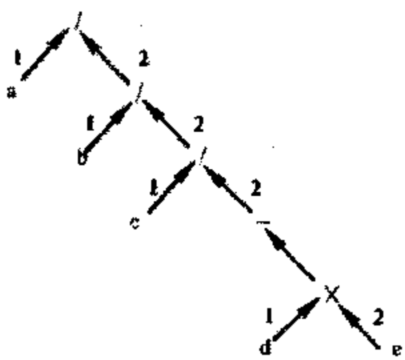
$(a - (b/2))$

(b4)



$(a - (b \times c))$

(b5)



$(a/(b/(c/(-(d \times e)))))$

3) *Notes*

a — Dans ce qui précède, il y a manifestement une parenté de structure entre exemples de même numéro (cf. par exemple a1 et b1). C'est à partir de là que nous procédons à la comparaison.

b — Une difficulté apparaît cependant ici : elle est liée à la distinction opérateur/opérande dans le langage mathématique. Cette distinction se traduit par le fait que seuls les opérandes peuvent apparaître aux feuilles. Il est difficile d'imaginer une telle distinction en langue naturelle, tout au moins au niveau des unités lexicales. Cependant, le fonctionnement des unités lexicales dans une phrase permet de constater que certains mots ne peuvent se trouver aux feuilles de l'arborescence : verbes à un mode personnel, “et”, ...

Remarque sur langue et métalangue :

L'introduction d'unités lexicales de la langue dans un discours écrit dans la métalangue modifie évidemment ce qui vient d'être dit.

Exemples :

"La relation "$\leqslant$" est transitive"

"Le grammairien supprime "et" et "car" dans sa phrase"

c — Pour éviter d'être mal compris, nous voudrions insister sur le fait que les R_i, CIRC(p), etc ... sont effectivement des relations, au sens mathématique du terme. Le seul problème réside dans leur interprétation. Pour nous faire comprendre, établissons le parallèle suivant :

$a \leqslant b$ — Pierre aime Marie

$a \xrightarrow{\leqslant} b$ — Pierre $\xrightarrow{\text{aime}}$ Marie

On met ainsi en évidence le caractère syntaxique des relations introduites. Toute autre interprétation est dangereuse, et notamment l'assimilation automatique d'un verbe à la caractérisation d'une relation.

III — Comparaison ne faisant intervenir que la syntaxe du langage mathématique

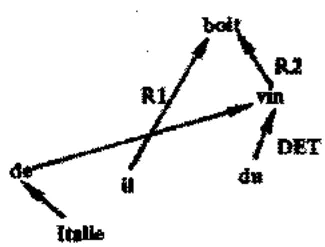
1) *Existence en langue naturelle de phrases qui ne sont pas représentables par une structure de dépendance*

Exemples :

— "L'avion et la voiture sont des moyens de transports aérien et terrestre.

— "Pierre et Jean achètent respectivement un cheval et une alouette"

— "D'Italie, il boit du vin"



Admettre cette représentation revient à s'autoriser le croisement des flèches. Ce qui viole la convention posée au II. 1 (remarque 2).

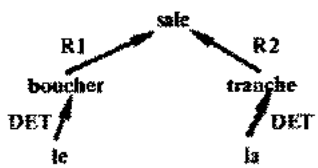
2) *Ambiguïté*

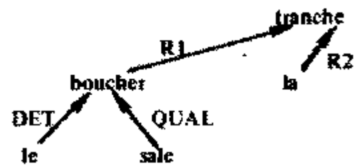
Le langage mathématique est non-ambigu : à chaque expression, on ne peut associer qu'une seule structure.

La langue naturelle est souvent ambiguë, *au niveau syntaxique.* A une phrase donnée, on peut fréquemment associer plusieurs structures.

Exemple :

"Le boucher sale la tranche"





Note : Ne pas confondre ceci avec l'ambiguïté dite sémantique, qui repose sur le fait qu'un mot a plusieurs sens.

Autres exemples :

"La berge plus élargie découvrait les soubassements des murs des jardins qui avaient un escalier de quelques marches descendant à la rivière" (BALZAC — Madame Bovary)

"J'aime la soupe de ma soeur qui est froide"

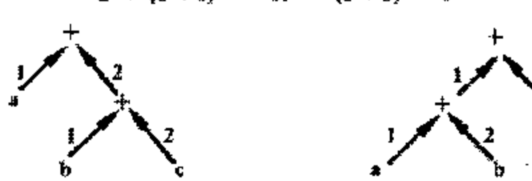
"Le cheval du paysan qui est noir"

3) *Problèmes de valence : nombre de flèches aboutissant à un noeud donné*

La comparaison met ici nettement en évidence la souplesse des langues naturelles ... ou le caractère rigide de la syntaxe du langage mathématique.

Rappelons que nous ne prenons pas en compte ici les propriétés mathématiques des opérations. Ainsi, nous considérons comme distinctes les deux expressions :

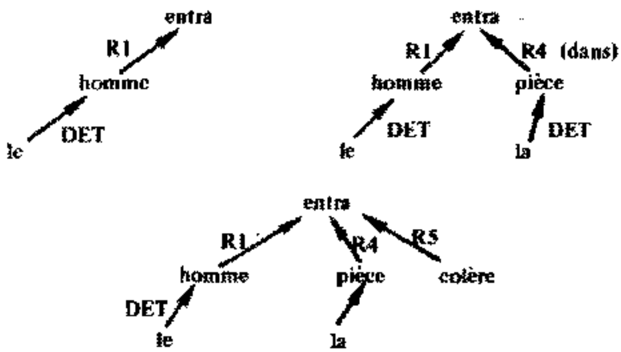
a + (b + c) et (a + b) + c



Dans ces conditions, la valence d'un opérateur mathématique est généralement fixée à la seule valeur 2. Exception : le signe "—" a 1 ou 2 comme valence.

En langue naturelle, la valence d'un verbe est très variable. De plus, on peut lui rattacher un nombre quelconque de circonstanciels.

Exemple :



etc, jusqu'à un cas limite comme :

> "La foudre entra, comme par hasard, avec un hideux éclair, par la fenêtre ouverte, dans l'appartement, à l'instant même".
>
> (VILLIERS DE L'ISLE ADAM)

Remarque : Du point de vue qui nous occupe, il convient de se méfier des simplifications d'écriture qui sont génératrices d'ambiguïtés.

Exemple :

peut s'interpréter soit comme :

$(a - b) - c$

soit comme :

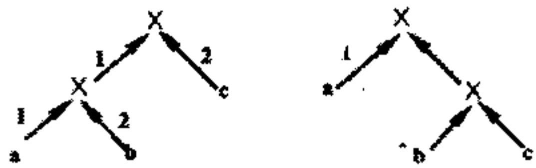
$a - (b - c)$

IV — Comparaison faisant intervenir les propriétés mathématiques des opérateurs

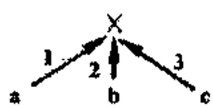
1) *Associativité*

Fréquente en mathématiques, elle paraît n'exister dans la syntaxe des langues naturelles que par l'existence de la structure de monoïde, qui ne nous intéresse pas ici.

En mathématiques, cette propriété permet une plus grande souplesse du langage. En effet, les deux structures suivantes :



étant équivalentes (même interprétation), peuvent être résumées en :



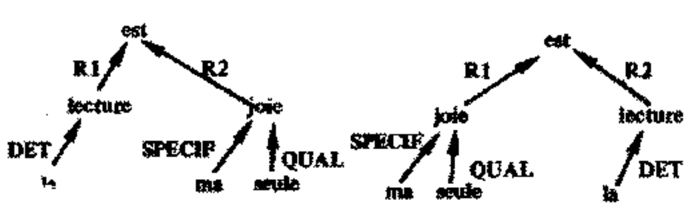
2) *Commutativité*

— En mathématiques :

sont équivalentes (même interprétation), si l'opération est commutative. On peut alors ne pas numéroter les flèches.

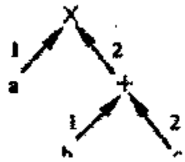
— En langue naturelle, la commutativité est rare, et elle ne joue jamais parfaitement.

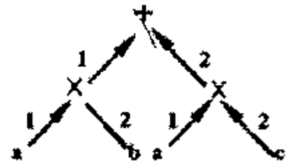
Exemples :



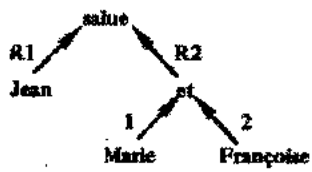
3) *Distributivité*

— En mathématiques, elle permet de conclure à l'équivalence des deux structures :

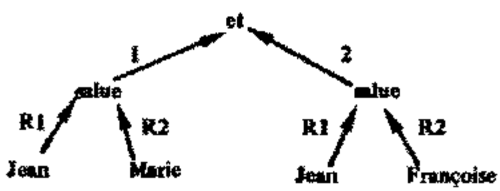




— En langue naturelle, elle joue rarement, et quand elle joue, elle ne conduit pas à une équivalence. Ainsi, la structure :



recouvre, *entre autres*, la suivante :



V — Problème : Chomsky ou Tesnière ?

Nous avions le choix, au départ, entre la théorie de Chomsky ("Grammaire générative") et celle de Tesnière ("Dépendances").

La première suppose l'existence de catégories syntaxiques (Syntagme nominal, Syntagme verbal, ...) ; la seconde se fonde sur les relations entre groupes de mots.

Nous avons préféré cette deuxième démarche : elle met davantage en évidence le fonctionnement dynamique de la phrase en même temps qu'elle donne lieu à une structure plus "intuitive". De plus cette structure nous paraît rendre mieux compte de la réalité linguistique profonde.

Du point de vue pédagogique, notre choix relève plutôt du pari (raisonné ...) que de la science pédagogique. Seules des recherches approfondies en psycholinguistique peuvent apporter une solution à ce problème.