

Les unions grammaires

Jean-François Hüe et Jacques-Henri Jayez

Volume 40, numéro 1, mars 1995

URI : <https://id.erudit.org/iderudit/003151ar>

DOI : <https://doi.org/10.7202/003151ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

1492-1421 (numérique)

[Découvrir la revue](#)

Citer cette note

Hüe, J.-F. & Jayez, J.-H. (1995). Les unions grammaires. *Meta*, 40(1), 160-162.
<https://doi.org/10.7202/003151ar>

BLOC-NOTES

LES UNIONS GRAMMAIRES

INTRODUCTION

Dans la nature, le sens s'exprime à partir des formes et de leur langage. La syntaxe de ces formes peut être fort complexe et très irrégulière. Heureusement il est un domaine, celui de nos langues indoeuropéennes, où cette irrégularité et cette complexité sont fortement atténuées. Trois principes sont utilisés à cet effet : l'usage d'un nombre restreint de symboles graphiques, la linéarité de la disposition de ces symboles pour former un texte, l'existence de règles de grammaire en nombre peu élevé et relativement stable au cours du temps.

Cependant, l'immense variété de ce qui doit être décrit par ces langages : les objets, les actions, les aspects temporels, a pour conséquence de rendre difficile l'analyse syntaxico-sémantique des textes en langue naturelle.

Une science particulière, la *linguistique*, s'est développée dans le but d'étudier les langues humaines sous tous leurs aspects. Une sous-branche très riche de cette science générale est la *théorie des langages* (Aho et Ullman 1972 ; Chomsky 1969 ; Ducrot et Todorov 1972 ; Ginsburg 1966 ; Laurière 1986) dont une des ambitions est la réalisation pratique des analyseurs syntaxico-sémantiques que demande la construction des compilateurs des langages artificiels de l'informatique. Cette même théorie des langages, combinée avec d'autres, est à la base des réalisations pratiques des systèmes automatiques de compréhension des langues naturelles (Jayez 1986).

Un grand nombre d'analyseurs syntaxico-sémantiques pour textes en langue naturelle suivent avec rigueur dans leur analyse le principe de la présentation linéaire de nos langues parce que leur conception découle directement de la définition des grammaires par la théorie des langages.

La notion de grammaire introduite par Noam Chomsky (1969), à la vue des concepts qui ont été exposés précédemment, peut être étendue de manière à faciliter la réalisation des analyseurs syntaxico-sémantiques. Nous introduisons pour cela le concept d'**union grammaire**.

DÉFINITION

On appellera *union grammaire* un ensemble $U = \{S, G, F\}$:

- un ensemble S de grammaires au sens de Chomsky :
- une grammaire, au sens de Chomsky, G dont les éléments terminaux sont l'union des éléments de S et des éléments d'un ensemble qui contient au moins les symboles particuliers O (concaténation qui correspond à une situation non ambiguë), O_u (U-concaténation qui correspond à une situation ambiguë) :
- F ensemble de définition des U-concaténations.

L'opération d'**union concaténation** de base est définie sur un **monoïde** $M = E^*$, ($M = UE^n$, n **élément de** \mathbb{N}), d'un ensemble E de symboles par : $xyO_u yz = xyz$, avec x, y, z éléments de M . Cette loi de composition interne sur M est associative et admet un élément neutre noté **nil**.

Exemple

Soit $G_u = \{\{Sa_1, Sa_2\}, Ga, F\}$

avec

$Sa_1 = \{V_{\setminus} Sa_1, V_{\top} Sa_1, ASa_1, PSa_1\}$

où

$V_{\setminus} Sa_1 = \{S_1, T_1\}, V_{\top} Sa_1 = \{a, b, \text{nil}\}, ASa_1 = S_1,$

$PGa = \{S_1 \rightarrow T_1, T_1 \rightarrow aT_1b, T_1 \rightarrow \text{nil}\}$

avec

$Sa_2 = \{V_{\setminus} Sa_2, V_{\top} Sa_2, ASa_2, PSa_2\}$

où

$V_{\setminus} Sa_2 = \{S_2, T_2\}, V_{\top} Sa_2 = \{b, c, \text{nil}\}, ASa_2 = S_2,$

$PGa = \{S_2 \rightarrow T_2, T_2 \rightarrow bT_2c, T_2 \rightarrow \text{nil}\}$

avec

$Ga = \{V_{\setminus} Ga, V_{\top} Ga, AGa, PGa\}$

où

$V_{\setminus} Ga = \{S\}, V_{\top} Ga = \{Sa_1, Sa_2, O\}, AGa = S,$

$PGa = \{S \rightarrow Sa_1 O Sa_2\}$

avec

$F = \{\{O, \text{définition} : x, y, z \text{ éléments du monoïde } \{a, b, c\}^*, xyO_{\setminus} yz \rightarrow xyz, \text{ si } b \text{ inclus dans } x \text{ et } b \text{ inclus dans } z, \text{ sinon } xyO_{\setminus} yz \rightarrow \text{nil}\}\}$

Sa_1 engendre le langage $L(Sa_1) = \{x / x = a^n b^n, n \text{ élément de } \mathbb{N}\}$

Sa_2 engendre le langage $L(Sa_2) = \{x / x = b^n c^n, n \text{ élément de } \mathbb{N}\}$

Ga engendre le langage $L(Ga) = \{Sa_1 O_f Sa_2\}$

Gu engendre le sous-ensemble du langage qui résulte de l'union concaténation O_f du langage engendré par Sa_1 et du langage engendré par Sa_2 , conformément à ce qui est indiqué par le langage $L(Ga)$ et les conditions exprimées par F :

$L(Gu) = L(Sa_1 O_f Sa_2)$

$L(Gu) = L(Sa_1) O_f L(Sa_2)$

$L(Gu) = \{x / x = a^n b^n c^n, n \text{ élément de } \mathbb{N}\}$

La O_f concaténation de deux langages algébriques donne un langage non algébrique.

Pour des langages non algébriques, on déterminera une union grammairale pour laquelle les grammairales de l'ensemble S seront toutes algébriques. L'analyseur syntaxico-sémantique associé à cette union grammairale sera simple à réaliser, et très « parallélisable ».

Analyse syntaxique de «aaabbbccc»

aaabbb — sera analysé comme un élément de $L(Sa_1)$,

— bbbccc sera analysé comme un élément de $L(Sa_2)$,

aaabbbccc sera donc reconnu comme un élément de $L(Gu)$.

ILLUSTRATION SUR DES LANGUES EUROPÉENNES

Nous allons donner quelques exemples simples de règles d'union concaténation tirées de langues européennes : français, espagnol, anglais, allemand (il ne sera pas précisé dans les exemples les catégories syntaxiques ni les règles d'accord).

Nous commencerons par le latin, langue mère des langues romanes.

Illustration en latin

Étant donné deux propositions indépendantes on peut transformer la seconde en subordonnée comme suit :

[Domus][magna][est] et [Domum][video]—>
Domus magna est quam video.

Soit la règle : $x_a y_z$ et $x_b t$ —> $x_a y_z$ *quam t*

[Beatus][es] et [Deus][carus][es]—>
Beatus es qui deus carus es.

Soit la règle : xy et $x_a ty$ —> xy *qui x_b ty*

Illustration en français

Étant donné deux propositions indépendantes on peut transformer la seconde en subordonnée comme suit :

[Le chien][regarde][le lapin] et [Le lapin][mange][un radis]—>

Le chien regarde le lapin qui mange un radis /
Le chien regarde le lapin manger un radis.

Soit la règle : xyz et ztu —> xyz *qui tu / xyz t_{inf} u*

Illustration en espagnol

Étant donné deux propositions indépendantes on peut transformer la seconde en subordonnée comme suit :

[Admiran][el parterre] et [Pués][fueron][a dar la vuelta al estanque grande]—>
Después de admirar el parterre, fueron a dar la vuelta al estanque grande.

Soit la règle $x_a y$ et *Pués zt*—>*Después de x_{inf} y, zt*

Illustration en anglais

Étant donné deux propositions indépendantes on peut transformer la seconde en subordonnée comme suit :

[The mouse][looks at][the cat] et [The cat] is [drinking][milk]—>
The mouse looks at the cat drinking milk.

Soit la règle xyz et *z is tu*—> $xyztu$
ou plus complexe :

[Jubal][was][the son of Lamech and Adah] et [Jubal][is][the father of music]—>
Jubal, the father of music, was the son of Lamech and Adah.

Soit la règle : xyz et xtu —> x, u, y, z

Illustration en allemand

Étant donné deux propositions indépendantes on peut transformer la seconde en subordonnée comme suit :

[Beethoven][trug][zur Bereicherung der musikalischen Kultur][bei] et [Beethoven][komponierte][sein Werk]—>

Indem er sein Werk komponierte, trug Beethoven zur Bereicherung der musikalischen Kultur bei.

Soit la règle : $xy_1 zy_2$ et xtu —> *Indem* genre(x) $ut, y_1 xzy_2$

CONCLUSION

La combinaison du concept **grammaire à contraintes lexicales** (Hüe 1988) et du concept d'**union grammairale** permet de réaliser des analyseurs syntaxiques basés sur le principe de la reconnaissance de certains marqueurs particuliers des enchaînements de phrases pour générer les décompositions du texte en phrases simples. Un traitement sémantique permettra ensuite de réduire les ambiguïtés lorsque cela est possible ou d'indiquer les parties de texte ambiguës.

Deux directions de recherche sont ouvertes au LIANA :

- une première, purement linguistique, est de formaliser précisément les règles constructives et de déterminer les marqueurs d'enchéassement de phrases ;
- une deuxième, purement informatique, est de réaliser un ensemble d'outils logiciels pour implanter les résultats des travaux linguistiques.

JEAN-FRANÇOIS HÛE ET
JACQUES-HENRI JAYEZ

LIANA, Université de Nantes, Nantes, France

RÉFÉRENCES

- AHO, A. et J. D. ULLMAN (1972) : *The Theory of Parsing Translation and Compiling*, New Jersey, Prentice-Hall.
- AUTEBERT, J.-M. (1987) : *Langages algébriques*, Paris, Masson.
- CHOMSKY, N. (1969) : *Structures syntaxiques*, Paris, Seuil.
- COLMERAUER, A. (1986) : «Note sur Prolog III», *Acte du séminaire programmation en logique*, Tregastel.
- DAVALO, E. et A. NAÏM (1989) : *Des réseaux de neurones*, Paris, Eyrolles.
- DUCROT, O. et T. TODOROV (1972) : *Dictionnaire encyclopédique des sciences du langage*, Paris, Seuil.
- FOUQUÈRE, C. (1988) : *Système d'analyse tolérante du langage naturel*, thèse, Université Paris-Nord.
- GINSBURG, S. (1966) : *The Mathematical Theory of Context-free Languages*, New Jersey, McGraw-Hill.
- GOLDBERG, J.-M. (1984) : *Smalltalk-80*, New Jersey, Addison Wesley.
- HAYES, S., HAUPTMANN, S., CARBONELL, S. et M. TOMITA (1986) : «Parsing Spoken Language: a Semantic Case Frame Approach», *Coling*, Bonn.
- HÛE, J.-F. (1984) : *Conception d'un générateur*, DEA, Toulouse.
- HÛE, J.-F. (1988) : «Les grammaires à contraintes lexicales», *Colloque informatique et langue naturelle*, Nantes.
- JAYEZ, J.-H. (1982) : *Compréhension automatique du langage naturel*, Paris, Masson.
- LAURIÈRES, S. (1986) : *Résolution de problèmes par l'homme et la machine*, Paris, Eyrolle.