



COLLECTION «PROGRAMMATION»

Publiée sous la direction de L. NOLIN

*Publications de l'Institut de Programmation
de la Faculté des Sciences de Paris*

Notions sur les grammaires formelles

par

Maurice GROSS et André LENTIN

Institut Blaise-Pascal

Préface de Noam CHOMSKY

2^{ème} édition revue et corrigée

GAUTHIER-VILLARS ÉDITEUR
55, quai des Grands-Augustins - Paris 6^e

1970

BIBLIOTHEQUE DU CERIST

157 239

OUVRAGES DE LA COLLECTION

« PROGRAMMATION »

Parus :

ARSAC J., LENTIN A., NIVAT M. et NOLIN L. — Algol. Théorie et pratique

GROSS M. et LENTIN A. — Notions sur les grammaires formelles.

FAURE R. — Eléments de la recherche opérationnelle. Avec la collaboration de *N.S. Guillot Le Garf* et de *M. Bloch*.

NOLIN L. — Formalisation des notions de machine et de programme.

LERMAN I. C. — Les bases de la classification automatique.

Sous presse :

FAURE R. et HEURGON E. — Structures ordonnées et algèbres de Boole.

© GAUTHIER-VILLARS, 1967

Toute reproduction, même partielle, de cet ouvrage est interdite. La copie ou reproduction, par quelque procédé que ce soit : photographie, microfilm, bande magnétique, disque ou autre, constitue une contrefaçon passible des peines prévues par la loi du 11 mars 1957 sur la protection des droits d'auteurs.

AVERTISSEMENT

Le présent ouvrage procède de l'enseignement que les Auteurs ont été appelés à donner, ces dernières années, dans le cadre de diverses Chaires ou Institutions, parmi lesquelles il convient de citer : le Centre de Linguistique Quantitative, de la Faculté des Sciences de Paris, créé à l'instigation du regretté Professeur Favard; la Chaire d'Analyse numérique de la Faculté des Sciences de Paris (Professeur : René de Possel) : option de Troisième Cycle; la Chaire de Physique Mathématique de l'Université de Toulouse (Professeur M. Laudet) : diplôme d'études approfondies, mention « Traitement de l'Information »; le Département de Linguistique de l'Université de Pennsylvanie ⁽¹⁾ (Professeur Z. S. Harris); l'Institut de Programmation de la Faculté des Sciences de Paris : enseignement de troisième niveau.

Rédigées à des fins essentiellement didactiques, ces Notions sur les Grammaires formelles ne prétendent pas à l'originalité scientifique; elles empruntent une grande part de leur substance à des ouvrages fondamentaux et « classiques » cités en bibliographie comme celui de M. Davis — Computability and Unsolvability [9] et ceux de N. Chomsky — entre autres les Formal Properties of Grammars [6]. Les emprunts que fait un Cours sont nécessairement nombreux, dans cet ordre d'idée les Auteurs tiennent à signaler qu'ils sont redevables à J. Piatrat pour ses conférences données dans le cadre du Centre de Linguistique Quantitative, déjà nommé, et à M. Nivat pour ses travaux relatifs aux codes ⁽²⁾ et à la transduction.

La théorie des grammaires formelles intéresse actuellement des personnes venues d'horizons variés et de formations très diverses. Cette hétérogénéité du public concerné appelait un choix, nous avons opté pour le ton le plus simple — ce qui, parfois, nous a conduits à sacrifier la parfaite rigueur mathématique à une présentation plus intuitive. Pour les mêmes raisons, nous avons laissé en exercices certaines démonstrations un peu techniques, tandis que d'autres font l'objet d'un simple renvoi à la bibliographie.

Nous tenons à remercier vivement MM. Salkoff et P. Schreiber dont les corrections effectuées au cours des traductions de cet ouvrage (en anglais et en allemand, respectivement) ont grandement contribué à améliorer cette seconde édition.

(1) Contrat N. S. F. : Transformations and Discourse Analysis Project.

(2) Contrat U. S. A. F. 61(052) — 945.

BIBLIOTHEQUE DU CERIST

PRÉFACE

L'étude traditionnelle du langage faisait une distinction importante entre « grammaire particulière » et « grammaire universelle ». Sous la première dénomination, on traitait des particularités propres à des langues prises individuellement, et sous la seconde, des caractères généraux, communs à toutes les langues naturelles. Le souci d'observer cette distinction est réapparu au premier plan des recherches, au cours des récentes années, et ce renouveau d'intérêt a coïncidé avec la naissance, en linguistique mathématique, d'une branche importante, qu'on appelle parfois « linguistique algébrique » afin de la distinguer tant des recherches statistiques sur le langage que des études de modèles probabilistes de performance. Il est à maints égards très naturel que ces faits se soient manifestés simultanément. La linguistique algébrique consiste en l'étude des propriétés formelles du langage obtenues par abstraction à partir de leur réalisation spécifique dans les langues particulières. Ainsi définie, elle ne saurait se distinguer en principe de la grammaire universelle, bien que la pratique ait vu se produire une division du travail entre les études, à caractère empirique, de grammaire universelle et celles, de nature purement mathématique, qui portent sur les structures formelles suggérées par la recherche linguistique. L'existence de cette division apparaît pour partie comme un signe du manque de maturité du domaine, pour partie comme un reflet de la diversité des motivations et intérêts qu'ont pu avoir tels chercheurs particuliers. Que l'on soit un jour en mesure de réduire substantiellement cette division et alors pourra naître une véritable linguistique mathématique théorique, qui étudiera de manière purement abstraite la classe des systèmes définis par les principes de la grammaire universelle, bref la classe des « langues possibles ».

Une seconde raison pour que coïncident la naissance de la linguistique algébrique et le renouveau d'intérêt envers la grammaire universelle peut être trouvée dans l'évolution de l'étude du langage à l'époque moderne. Il y a plus d'un siècle s'amorça le déclin des études sur la grammaire générale et raisonnée, la valeur et le sens en furent sérieusement mis en doute par les savants lorsqu'ils s'orientèrent d'abord vers les études historiques et comparatives, plus tard vers les études de langues conduites selon les traditions très fructueuses du « structuralisme » et du « descriptivisme ». Les études antérieures de grammaire générale et raisonnée furent tenues pour trop spéculatives, pour insuffisamment ancrées dans le fait linguistique. On ressentit qu'elles sous-estimaient sérieusement la variété possible des structures du langage et qu'elles imposaient arbitrairement un modèle de la structure du langage qui distordait les phénomènes réels des langues particulières. Certains spécialistes allèrent jusqu'à soutenir qu'on ne pouvait assigner aucune limite à la manière dont les langues diffèrent les unes des autres et qu'il n'existait pas de conditions générales significatives à quelque degré, qui limiteraient et contraindraient la forme de toute langue possible. En même temps, l'attention glissait vers la structure du son, ce que la tradition linguistique classique n'avait d'ailleurs pas négligé mais qu'elle avait considéré comme un simple auxiliaire à l'étude des propriétés syntaxiques profondes du langage. L'un des soucis majeurs dans la tradition classique concernait cette propriété du langage qu'on pourrait appeler « l'aspect créateur » et qui correspond au fait que le langage dispose de mécanismes récurrents, lui permettant l'expression d'un nombre infini de pensées, sensations, intentions, etc., indépendamment de tout contrôle par stimulation externe ou par états internes identifiables. « L'aspect créateur » du langage, de nombreux spécialistes modernes d'orientation béhaviouriste l'ont en fait nié, qui en sont venus à considérer le langage comme un système d'habitudes, un réseau de relations stimulus-réponse, ou quelque chose d'approchant. D'autres, un de Saussure par exemple, n'ont pas nié cet « aspect créateur », mais ils l'ont rattaché à la parole et l'ont considéré, soit comme secondaire, soit comme entièrement extérieur au domaine de l'étude de langage et, partant, comme non assujetti à des règles linguistiques strictes.

Il est certain que l'une au moins des raisons, la plus importante peut-être, de ce déplacement d'intérêt et de modification de doctrine tient au fait que les méthodes alors disponibles et la compréhension qu'on avait des processus formels ne permettaient pas d'aborder l'étude des mécanismes récurrents de la syntaxe d'une manière claire et féconde. Et c'est d'ailleurs depuis une génération seulement qu'ont été inventés et affinés les concepts adéquats, au cours de recherches portant sur le fondement des mathématiques. La façon dont on comprend maintenant les mécanismes récurrents et la nature des algorithmes — elle est le fruit du travail de ces trente dernières années — permet de reprendre l'étude de l'aspect créateur qu'on trouve dans l'emploi du langage et d'entreprendre la description précise des mécanismes dont chaque langage dispose pour l'utilisation libre et illimitée du langage.

Aujourd'hui, on désigne couramment l'étude de ces mécanismes par l'expression de « grammaire générative ». On considère la grammaire générative d'une langue naturelle comme le système des règles et processus qui caractérisent la classe potentiellement infinie des phrases de cette langue et qui attribuent à chacun de ces objets une description structurelle, laquelle en représente les propriétés essentielles : phonologiques, syntaxiques, sémantiques. Ainsi, l'étude de la grammaire générative devient réalisable du fait de progrès accomplis en mathématiques et dès lors, il n'est pas surprenant que l'intérêt porté aux propriétés formelles des grammaires et à la linguistique algébrique ait été une conséquence naturelle de cette nouvelle manière d'aborder l'étude du langage.

Je voudrais insister à nouveau sur le fait qu'un décalage sensible sépare encore les recherches, mathématiques d'une part, empiriques d'autre part, qui relèvent les unes et les autres du domaine appelé à devenir finalement une théorie mathématique de la grammaire universelle. Le schéma de description grammaticale qui semble être empiriquement motivé par les faits des langues particulières spécifie une classe de systèmes qui sont pour le moment bien trop complexes pour qu'une analyse mathématique fructueuse et significative puisse s'y appliquer. De plus, il est nécessaire de garder à l'esprit que toute

proposition qu'on pourrait avancer aujourd'hui concernant ce schéma universel serait à la fois hypothétique et quelque peu floue sur des points importants. En même temps, il y a eu d'intéressantes et suggestives études faites sur des schémas beaucoup plus restreints de description grammaticale — la théorie des langages « context-free » est l'exemple le plus important — mais il est absolument certain que ces systèmes ne sont pas empiriquement adéquats. Une théorie mathématique de la grammaire universelle est donc bien plus un espoir pour l'avenir qu'une réalité présente. Le mieux qu'on en puisse dire est que la recherche en cours apparaît comme une tentative faite dans la direction d'une telle théorie. Quant à moi, il me semble que cette tentative constitue aujourd'hui l'un des domaines d'étude les plus passionnants et que, si elle était couronnée de succès, elle placerait dans les prochaines années l'étude du langage sur un terrain ferme et tout nouveau.

NOAM CHOMSKY

TABLE DES MATIÈRES

	<i>Pages</i>
AVERTISSEMENT	1
PRÉFACE DE N. CHOMSKY	3

PREMIÈRE PARTIE

PRELIMINAIRES LOGICO-ALGEBRIQUES

Chapitre I. Mots. Monoïdes. Langages.

1.1. Le monoïde libre	11
1.2. Calcul sur les mots	13
1.3. Langages	16
1.4. Exercices	19

Chapitre II. Généralités sur les systèmes formels.

2.1. Présentation intuitive du calcul des propositions	21
2.2. La notion de système formel	25
2.3. Une variante formalisée du calcul des propositions	28
2.4. Définition des systèmes formels	30

Chapitre III. Systèmes combinatoires.

3.1. Définition des systèmes combinatoires	33
3.2. Systèmes normaux	38
3.3. Exercices	40
3.4. « Non influence » de l'alphabet	41

Chapitre IV. Algorithmes. Machines de Turing.

4.1. Algorithmes	43
4.2. Les machines de Turing	45
4.3. Machines de Turing « numériques »	50
4.4. Exercices	52

Chapitre V. Calculabilité. Décidabilité.

5.1. Calcul des fonctions	53
5.2. Opérations sur les fonctions	54
5.3. Techniques de Gödel	56

	<i>Pages</i>
5.4. Ensembles récursifs et ensembles récursivement énumérables	58
5.5. Remarques et compléments divers	60
<i>Chapitre VI. Systèmes combinatoires et machines de Turing. Problèmes indécidables</i>	
6.1. Systèmes et machines de Turing	65
6.2. Problèmes indécidables	69

DEUXIÈME PARTIE

QUELQUES CLASSES REMARQUABLES DE LANGAGES

<i>Chapitre VII. Les langages de Chomsky. Présentation et premières propriétés.</i>	
7.1. Grammaire et description	77
7.2. Définitions. Propriétés décidables	79
7.3. Propriétés de fermeture	84
7.4. Classes de C-langages	85
7.5. Exercices	87
<i>Chapitre VIII. Propriétés indécidables des C-grammaires.</i>	
8.1. Problèmes liés à l'intersection	89
8.2. Problèmes liés à l'ambiguïté	90
8.3. Ambiguïtés de C-langages	94
<i>Chapitre IX. Les automates à piles de mémoires.</i>	
9.1. Automate qui accepte les phrases d'un C-langage	99
9.2. Automate engendrant un langage	103
9.3. Classe des langages acceptés (engendrés) par la classe des automates à pile	104
9.4. Aperçu concernant les applications des C-langages	106
<i>Chapitre X. Langages de Kleene et automates finis.</i>	
10.1. Les K-grammaires	109
10.2. Les automates finis	110
10.3. Autres classes d'automates finis	112
10.4. Propriétés de fermeture. Caractérisation de Kleene	113
10.5. K-langages et C-langages	115
10.6. Transducteurs unilatères finis	115
<i>Chapitre XI. Langages définis par des systèmes d'équations.</i>	
11.1. Fonctions dont les arguments et les valeurs sont des langages	119
11.2. Langages et séries formelles	127
<i>Chapitre XII. Grammaires à règles contextuelles. Automates à mémoire linéairement bornée.</i>	
12.1. Grammaires à règles contextuelles	131

	<i>Pages</i>
12.2. Automate à mémoire linéairement bornée	132
12.3. Classification des automates	134
12.4. Exercices	135

TROISIÈME PARTIE

LE POINT DE VUE ALGÈBRE

Chapitre XIII. Morphismes de monoïdes.

13.1. Monoïdes quelconques	139
13.2. Congruence et équivalences associées à un langage	142
13.3. Notions sur les codes	145

Chapitre XIV. Compléments sur les langages de Kleene.

14.1. Les K-langages standard	147
14.2. Caractérisation et propriétés des K-langages standard	149
14.3. Caractérisation algébrique des K-langages	150
14.4. Transduction	155

Chapitre XV. Compléments sur les langages de Chomsky.

15.1. Langages de Dyck	159
15.2. C-langages standard	163
15.3. Équivalence des C-langages et des langages acceptés par des automates à pile	166
15.4. Exercices	167

Chapitre XVI. Langages algébriques.

16.1. Compléments sur les séries formelles	169
16.2. Séries algébriques	175
16.3. Applications aux langages	176
16.4. Exercices	179
16.5. Application des équations à la géométrie énumérative	179

Appendice. Grammaires transformationnelles.

1. Généralités sur les grammaires formelles et les langues naturelles	185
2. C-grammaires et transformations	186
3. Extension des grammaires	187
4. Problèmes liés aux transformations	189

<i>Bibliographie relative aux grammaires transformationnelles</i>	193
---	-----

<i>Bibliographie commentée</i>	195
--------------------------------------	-----