SÉMINAIRE DUBREIL.
ALGÈBRE ET THÉORIE
DES NOMBRES

LUC BOASSON

Familles de langages

Séminaire Dubreil. Algèbre et théorie des nombres, tome 25, nº 1 (1971-1972), exp. nº 15, p. 1-5
<http://www.numdam.org/item?id=SD_1971-1972__25_1_A14_0>

Séminaire DUBREIL
(Algèbre)
25e année, 1971/72, n° 15, 5 p.

15 mai 1972

FAMILLES DE LANGAGES

par Luc BOASSON

Le but de ce bref exposé n'est pas de présenter un résultat nouveau précis, mais plutôt, de donner un aperçu des différents problèmes que nous traitons à l'heure actuelle en théorie des langages formels. Nous nous contentons de donner quelques résultats (dont certains sont classiques) qui permettent de se faire une idée du type de questions auxquelles nous essayons de répondre.

L'exposé est divisé en trois parties principales : la première concerne les transductions rationnelles, la deuxième les familles de langages, et la dernière les langages algébriques (= "context-free").

1. Transductions rationnelles.

Définition 1 [6]. - Etant donné un monoïde M, on note $\mathrm{Rat}(M)$ la famille des parties rationnelles de M, i. e. la plus petite famille de parties de M contenant les parties finies et qui soit fermée par union, produit et étoile (passage au sous-monoïde engendré).

Dans le cas particulier de M monoïde libre finement engendré, soit $M = X^*$, on retrouve la famille de langages rationnels [11] qui fait l'objet de nombreuses études.

Définition 2. - Etant donnés deux monoïdes libres X^* et Y^*, on appelle transduction toute partie du monoïde $X^* \times Y^*$.

A une telle transduction $\hat{\tau}$, on associe deux applications, l'une τ de X^* dans les parties de Y^*, définie par

$$f \in X^* , \quad \tau f = \{g \in Y^* \mid (f , g) \in \hat{\tau}\} ,$$

l'autre τ^{-1} de Y^* dans les parties de X^*, définie par

$$g \in Y^* , \quad \tau^{-1} g = \{f \in X^* \mid (f , g) \in \hat{\tau}\} .$$

(N.-B. : En général, τ^{-1} n'est pas l'inverse de τ au sens des applications.)

Une transduction $\tau \subset X^* \times Y^*$ est dite [12] :

- rationnelle si, et seulement si, $\hat{\tau} \in \mathrm{Rat}(X^* \times Y^*)$,
- d'image finie si, et seulement si, $\forall f \in X^*$, τf est une partie finie de Y^*,
- fidèle si, et seulement si, τ^{-1} est d'image finie,
- continue si, et seulement si, $(1 \in \tau f) \Rightarrow (f = 1)$ (1 étant élément neutre de X^*).

Les transductions rationnelles sont rendues manipulables par le résultat suivant.

THÉORÈME 1 (NIVAT [12]). - La transduction τ de X^* dans Y^* est rationnelle si, et seulement si, il existe un alphabet fini Z, un langage rationnel K de Z^*, et deux homomorphismes φ et ψ de Z^* dans X^* et dans Y^* respectivement, tels que

$$\hat{\tau} = \{(\varphi h \, , \, \psi h) \mid h \in K\} \, .$$

On trouvera dans [1] une caractérisation analogue des transductions rationnelles fidèles et continues. Quelques exemples de transductions rationnelles peuvent être trouvés dans [7]. Les transductions rationnelles constituent l'outil essentiel que nous utilisons dans l'étude des familles de langages.

2. Familles de langages.

Définition 3. - On appelle famille de langages $\mathcal{L}$ une famille de parties du monoïde libre Σ^* (où Σ est cette fois infini) tel que, pour chaque élément L de $\mathcal{L}$, on puisse trouver une partie finie X_L de Σ telle que $L \subset X_L^*$.

Définition 4 [6]. - Une famille de langage $\mathcal{L}$ constitue un cône rationnel (resp. un cône rationnel fidèle) si, et seulement si, $\mathcal{L}$ est fermé par transduction rationnelle (resp. par transduction rationnelle fidèle et continue).

Définition 5 [9]. - Une famille de langage $\mathcal{L}$ constitue une famille agréable de langages, en abrégé une FAL (resp. une FAL fidèle) si, et seulement si, $\mathcal{L}$ est fermée par transduction rationnelle (resp. transduction rationnelle fidèle et continue), par union, produit et étoile (resp. étoile propre).

Rappelons que si $A \subset X^*$, on note

$$A^* = \cup_{n \geq 0} A^n \quad \text{et} \quad A^+ = \cup_{n \geq 1} A^n \, .$$

On désigne par "étoile propre" l'opération $+$.

Etant donnée une famille de langages $\mathcal{L}$ quelconque, on note respectivement par $\mathfrak{C}(\mathcal{L})$, $\mathfrak{C}_f(\mathcal{L})$, $\mathfrak{F}(\mathcal{L})$ et $\mathfrak{F}_f(\mathcal{L})$, le plus petit cône rationnel, le plus petit cône rationnel fidèle, la plus petite FAL et la plus petite FAL fidèle contenant $\mathcal{L}$.

Définition 6 [9]. - Un cône rationnel $\mathcal{L}$ (resp. une FAL $\mathcal{L}$) est dit principal s'il existe une famille de langages réduite à un élément L telle que $\mathcal{L} = \mathfrak{C}(\{L\})$ (resp. $\mathcal{L} = \mathfrak{F}(\{L\})$).

Nous écrirons $\mathcal{L} = \mathfrak{C}(L)$ et $\mathcal{L} = \mathfrak{F}(L)$. Nous ne donnons pas la définition de la principalité pour les cônes et les FAL fidèles : elle est identique à celle-ci.

Les deux notions de cône rationnel et de FAL, introduite de façon indépendante, sont cependant très liées ainsi qu'il résulte du théorème 2 ci-dessous.

THÉORÈME 2 [10]. – *Etant donnée un famille de langages* $\mathcal{L}$, *on a* :

$$\mathfrak{F}(\mathcal{L}) = \mathrm{Rat} \,\square \mathcal{C}(\mathcal{L})$$

$$\mathfrak{F}_f(\mathcal{L}) = \mathrm{Rat}^+ \,\square \mathcal{C}_f(\mathcal{L}) ,$$

où $\mathrm{Rat} \,\square\mathcal{C}(\mathcal{L})$ (resp. $\mathrm{Rat}^+ \,\square\mathcal{C}_f(\mathcal{L})$) désigne la fermeture par union, produit et étoile (resp. étoile propre) de $\mathcal{C}(\mathcal{L})$ (resp. de $\mathcal{C}_f(\mathcal{L})$).

3. Les langages algébriques.

En général, nous nous intéressons aux familles de langages inclus dans celle des langages algébriques. Nous nous bornerons ici à donner quelques résultats concernant cette dernière famille et à indiquer quelques autres familles que nous avons étudiées.

En ce qui concerne la définition des langages algébriques (= "context-free"), nous renvoyons à [8] par exemple. Nous nous bornons à rappeler ici la définition des langages de Dyck : Soit $Z_n = \{z_1 ,\dots, z_n , \bar{z}_1 ,\dots, \bar{z}_n\}$ un alphabet à $2n$ lettres. On appelle langage de Dyck D_n^* (resp. langage de Dyck restreint $D_n'^*$) la classe de 1 dans la congruence de Thue engendrée par les $2n$ relations

$$\sigma_i \,\bar{\sigma}_i \equiv \bar{\sigma}_i \,\sigma_i \equiv 1$$

(resp. par les n relations $\sigma_i \,\bar{\sigma}_i \equiv 1$). Ces langages jouent un rôle particulièrement important en théorie des langages en raison du résultat suivant :

THÉORÈME 3 (CHOMSKY-SCHÜTZENBERGER [5]). – *Le langage* $L \subset X^*$ *est algébrique si, et seulement si, il existe* $n \in \mathbb{N}$, *un alphabet* Z_n , *un homomorphisme* φ *de* Z_n^* *dans* X^* , *et un langage rationnel* $K \subset Z_n^*$ *tels que* $L = \varphi(D_n^* \cap K)$.

Sachant par ailleurs que la famille des langages algébriques constitue un cône rationnel [8], on déduit de ce théorème le résultat suivant.

THÉORÈME 4. – *Quel que soit* n *entier,* $n \geqslant 2$, $D_n'^*$ *(resp.* D_n^**) constitue un générateur principal du cône des langages algébriques.*

Le cas particulier de $n = 1$ a conduit à l'étude des cônes $\mathcal{C}(D_1^*)$ et $\mathcal{C}(D_1'^*)$ que l'on trouvera dans [10].

La famille des langages algébriques constituant également une FAL, il est remarquable que $\mathcal{C}(D_n'^*) = \mathfrak{F}(D_n'^*)$, et l'on peut se demander s'il existe un langage L qui soit générateur principal de la FAL des langages algébriques sans l'être du cône rationnel qu'ils constituent. La réponse nous est fournie par le résultat suivant :

THÉORÈME 5 [4]. – *Quel que soit* L , *générateur principal de la FAL des algébriques, il est générateur principal du cône rationnel qu'ils constituent.*

Ce théorème résulte d'une propriété plus générale :

Définition 7 [4]. - Une famille de langage $\mathfrak{L}$ est dite translatable si, et seulement si, quel que soit $L \in \mathfrak{L}$, $L \subset X_L^*$, on a

$$[a , L , b] = \{a^n f b^n \mid n \in \underset{\sim}{N} , f \in L\} \in \mathfrak{L} ,$$

où a et b sont deux nouveaux symboles.

Définition 8 [4]. - La famille $\mathfrak{F}$ constitue une FAL conique si, et seulement si, pour toute famille $\mathfrak{L}$, telle que $\mathfrak{F} = \mathfrak{F}(\mathfrak{L})$, on a $\mathfrak{F} = \mathfrak{C}(\mathfrak{L})$.

On peut alors énoncer un théorème.

THÉORÈME 6 [4]. - Si $\mathfrak{L}$ constitue une FAL translatable, $\mathfrak{L}$ est une FAL conique.

Pour terminer, il importe de souligner que la famille des langages algébriques donne lieu également à l'étude, en général très difficile, de propriétés de structures des langages qui la composent. On trouvera un tel exemple en particulier dans [3].

BIBLIOGRAPHIE

[1] BOASSON (Luc). - Cônes rationnels et familles agréables de langages. Application aux langages à compteur, Thèse 3e cycle, Math., Univ. Paris-7, 1971.

[2] BOASSON (Luc). - Deux théorèmes d'itération pour certaines familles de langages (à paraître). [Voir aussi : An iteration theorem for one counter languages, "Proceedings of the 3rd Annual ACM symposium on theory of computing" [3-5 mai 1971. Shaker Heights], p. 116-120. - New York, Association for Computing Machinery, 1971]

[3] BOASSON (Luc). - Un critère de rationnalité pour les langages algébriques, "Automata, languages and programming" [1972. Rocquencourt], p. 359-365. - Amsterdam, North-Holland publishing Company ; New York, American Elsevier publishing Company, 1973.

[4] BOASSON (Luc) et NIVAT (M.). - Sur diverses familles de langages fermées par transduction rationnelle (à paraître).

[5] CHOMSKY (N.) et SCHÜTZENBERGER (M. P.). - The algebraic theory of context free languages, "Computer programming and formal systems", p. 118-161. - Amsterdam, North-Holland publishing Company, 1963.

[6] EILENBERG (S.). - Algebraic aspects of automata theory, "Actes du Congrès international des Mathématiciens" [1970. Nice], Tome 3, p. 265-267. - Paris, Gauthier-Villars, 1971.

[7] FLIESS (M.). - Exemples de transductions, "Automaten Theorie und Formale Sprachen" [1969. Oberwolfach], p. 155-168 (Tagungsbericht. Math. Forschungsinst. Oberwolfach).

[8] GINSBURG (S.). - The mathematical theory of context-free languages. - New York, McGraw-Hill Book Company, 1966.

[9] GINSBURG (S.), GREIBACH (S.) and HOPCROFT (J.). - Studies in abstract families of languages. - Providence, American mathematical Society, 1969 (Memoirs of the A. M. S., 87).

[10] GINSBURG (S.) and GREIBACH (S.). - Principal AFL, J. Comput. System. Science, t. 4, 1970, p. 308-338.

[11] GINZBURG (A.). - Algebraic theory of automata. - New York, London, Academic Press, 1968.

[12] NIVAT (M.). - Transductions des langages de Chomsky, Annales Inst. Fourier, Grenoble, t. 18, 1968, fasc. 1, p. 339-455 (Thèse Sc. math. Paris, 1967).

Luc BOASSON
Université de Paris-7
Mathématiques, Tour 55
2 place Jussieu
75221 PARIS CEDEX 05