Théorème 1 (Intersection avec un langage algébrique, [Car, p. 98])**.** Soient L un langage algébrique et K un rationnel. Alors $L \cap K$ est algébrique.

Démonstration. Soit $G = (\Sigma, V, S, P)$ une grammaire algébrique engendrant L. Pour simplifier les notations, on suppose que G est en forme normale quadratique. Soit $\mathcal{A} \coloneqq (Q, \Sigma, \delta, i, f)$ un automate normalisé acceptant le langage K. Nous montrons que $L \cap K$ est engendré, à ε près, par la grammaire $G' \coloneqq (\Sigma, V', S', P')$, où

$$\begin{aligned} V' &\coloneqq \{A_{p,q} \mid (A,p,q) \in V \times Q^2\}, \\ P' &\coloneqq \{A_{p,q} \to a \mid A \to a \in P \text{ et } p \xrightarrow{a} q \in \delta\} \\ &\qquad \cup \{A_{p,q} \to B_{p,r}C_{r,q} \mid A \to BC \in P \text{ et } r \in Q\}, \end{aligned}$$

et

$$S' \coloneqq S_{i,f}.$$

Soit $w = w_1 \cdots w_n \in \Sigma^\star$. Supposons que $w \in \mathrm{L}_{G'}(S')$. On a alors un arbre de dérivation T' de w' dans G' où les pères des feuilles sont ${w_1}_{p_0,p_1}{w_2}_{p_1,p_2} \cdots {w_n}_{p_{n-1},p_n}$ avec $p_0 = i$, $p_n = f$. Par construction de G' le chemin

$$i = p_0 \xrightarrow{w_1} p_1 \xrightarrow{w_2} \cdots \xrightarrow{w_n} p_n = f$$

est dans $\mathcal{A}$, donc $w \in K$. Aussi, en supprimant les branches unaires et les indices figurant dans les étiquettes des nœuds de T' on obtient un arbre de dérivation T de w dans la grammaire G, donc $w \in L$. Ainsi $w \in L \cap K$. Réciproquement, on montre que $\mathrm{L}_G(A) \cap \mathrm{L}_{\mathcal{A}}(p,q) \subseteq \mathrm{L}_{G'}(A_{p,q})$ par récurrence sur la longueur de la dérivation. Si $A \to_G a$ et $p \xrightarrow{a}_{\mathcal{A}} q$ alors $A_{p,q} \to_{G'} a$ par construction. Si la dérivation est de longueur au moins 2 alors $A \to_G BC \to_G^+ w =: w_1w_2$, où $B \to_G^+ w_1$ et $C \to_G^+ w_2$ par le lemme fondamental. En identifiant $r \in Q$ tel que $p \xrightarrow{w_1}_{\mathcal{A}} r \xrightarrow{w_2}_{\mathcal{A}} q$ on a alors, en utilisant l'hypothèse de récurrence (deux fois)

$$A_{p,q} \to_{G'} B_{p,r}C_{r,q} \to_{G'}^+ w_1C_{r,q} \to_{G'}^+ w_1w_2,$$

et l'étape d'induction est donc démontrée. En particulier $L \cap K = \mathrm{L}_G(S) \cap \mathrm{L}_{\mathcal{A}}(i,f) \subseteq \mathrm{L}_{G'}(S_{i,f}) = \mathrm{L}_{G'}(S')$. Finalement $L \cap K = \mathrm{L}_{G'}(S')$. ∎

Remarque 2. Si $\mathcal{A}$ est déterministe et G est non ambiguë, la grammaire G' est non ambiguë.

Contre-exemple 3. Les langages $L_1 \coloneqq \{a^nb^nc^m \mid n,m \geqslant 0\}$ et $L_2 \coloneqq \{a^nb^mc^m \mid n,m \geqslant 0\}$ sont engendrés par les grammaires

$$G_1\colon \begin{cases} S_1 \to S_1c \mid T_1 \\ T_1 \to aT_1b \mid \varepsilon \end{cases}, \qquad G_2\colon \begin{cases} S_2 \to aS_2 \mid T_2 \\ T_2 \to bT_2c \mid \varepsilon \end{cases}.$$

Pourtant $L_1 \cap L_2 = \{a^nb^nc^n \mid n \geqslant 0\} =: L$ n'est pas algébrique d'après le lemme d'Ogden : si L est engendré par une grammaire $G \coloneqq (\Sigma, V, S, P)$ alors il existe $k \geqslant 1$ tel que $w = \underline{a^k}b^kc^k \in L$ (où les a sont distingués) se factorise en $w = \alpha u \beta v \gamma$ avec

1. $S \to_G^* \alpha T \gamma$ et $T \to_G^* uTv$, $T \to_G^* \beta$,
2. soit α, u, β, soit β, v, γ contiennent des lettres distinguées,
3. $u\beta v$ contient moins de k lettres distinguées.

Nécessairement $u, v \in a^\star \cup b^\star \cup c^\star$, et $u \in a^+$ ou $v \in a^+$. Par conséquent $\alpha u^2 \beta v^2 \gamma$ ne peut être dans L ; contradiction. ∎

Références. [Car, p. 98]

910 Langages algébriques. Exemples et applications.

[Car] Olivier Carton : *Langages formels, calculabilité et complexité.*