Automate à pile = grammaire hors-contexte

Benjamin Hellouin

Carton, 2.6

Théorème 1. *Les langages reconnus par un automate à pile sont exactement les langages algébriques.*

Preuve. ($\Rightarrow$)
Soit $G = (\Sigma, V, S_0, P)$ une grammaire hors contexte (algébrique). On suppose, sans perte de généralité, que toute règle de G est de la forme $S \to w$ avec $w \in V^*$ ou $S \to a$ avec $a \in \Sigma$. Cette transformation se fait très simplement en ajoutant une variable V_a pour chaque lettre $a \in \Sigma$, de remplacer à droite a par V_a et d'ajouter les règles $V_a \to a$.

On construit un automate d'alphabet d'entrée Σ, d'alphabet de pile V. L'automate ne comporte qu'un état q_0, son symbole de pile initial est S_0 et son ensemble de transitions est donné par

$$\{q_0, S \xrightarrow{a} q_0, \varepsilon \mid a \in \Sigma, S \to a \in P\} \cup \{q_0, S \xrightarrow{\varepsilon} q_0, w \mid w \in V^*, S \to w \in P\}.$$

On peut facilement prouver (par récurrence sur la longueur du calcul ou la longueur de la dérivation) qu'un mot est accepté en n étapes si et seulement si il est engendré en utilisant n règles. Le résultat s'ensuit.

($\Leftarrow$)
La réciproque est plus difficile et utilise la méthode des triples de Ginsburg. Soit un langage accepté par un automate à pile $A = (\Sigma, V, Q, z_0, q_0, \to)$, qui accepte par pile vide. On peut supposer que chaque transition de l'automate empile au plus deux symboles, au besoin en la décomposant en plusieurs transitions et en ajoutant un état intermédiaire. Soit q_0 et z_0 l'état initial, respectivement le symbole de pile initial de l'automate A.

Si q et q' sont deux états et $z \in V$, on note $L_{q,q',z} = \{w \in \Sigma^* \mid (q,z) \xrightarrow{w} (q', \varepsilon)\}$, autrement dit l'ensemble des mots tels qu'il existe un calcul allant de (q,z) à (q',ε) étiqueté par ce mot. On a alors $L = \bigcup_{q \in Q} L_{q_0,q,z_0}$.

Examinons maintenant les calculs possibles de (q,z) à (q',ε). La première transition peut soit dépiler z, soit remplacer z par z_1, soit remplacer z par $z_1 z_2$. On a donc l'égalité suivante :

$$L_{q,q',z} = \{x \in V \mid q, z \xrightarrow{x} q', \varepsilon\}$$

$$\cup \bigcup_{x,q_1,z_1} x L_{q_1,q',z_1} \mid q, z \xrightarrow{x} q_1, z_1$$

$$\cup \bigcup_{x,q_1,q_2,z_1,z_2} xL_{q_1,q_2,z_1}L_{q_2,q',z_2} \mid q,z \xrightarrow{x} q_1, z_1z_2.$$

Cette égalité se traduit immédiatement en une grammaire ayant comme variables $\{V_{q,q',z} \mid q,q' \in Q, z \in V\} \cup S$, de symbole initial S, avec les règles :

- $S \to V_{q_0,q,z_0}, q \in Q$;
- $V_{q,q',z} \to x$ si $q,z \xrightarrow{x} q', \varepsilon$;
- $V_{q,q',z} \to xV_{q_1,q',z_1}$ si $q,z \xrightarrow{x} q_1, z_1$;
- $V_{q,q',z} \to xV_{q_1,q_2,z_1}V_{q_2,q',z_2}$ si $q,z \xrightarrow{x} q_1, z_1z_2$;

Les égalités précédentes expriment le fait que cette grammaire engendre exactement $L(A)$.