Université Joseph Fourier **Polytech'Grenoble**
Département RICM 4

Langages et Traducteurs

Grammaires Hors-Contexte et Automates à Pile

Grammaires Hors-Contexte

Une *grammaire hors-contexte* est un quadruplet $G = (V_t, V_n, Z, P)$ tel que :
- V_t, *vocabulaire terminal*, est un ensemble fini de symboles (qui seront notés en lettres minuscules) ;
- V_n, *vocabulaire non-terminal*, est un ensemble fini de symboles (qui seront notés en lettres majuscules) ;
- $Z \in V_n$ est l'*axiome* de la grammaire ;
- P est l'ensemble des *règles* (ou *productions*) de la grammaire : $P \subseteq V_n \times (V_t \cup V_n)^*$
  Les éléments (X, α) de P seront notés $X \to \alpha$.

Dérivations

Soit $G = (V_t, V_n, Z, P)$ une grammaire hors-contexte, w et w' des mots sur $V_t \cup V_n$ $(w, w' \in (V_t \cup V_n)^*)$.

On dit alors que :
- w' est un *dérivé direct* par G de w, noté $w \Rightarrow_G w'$, si et seulement si, il existe $X \in V_n$, il existe $u, t, v \in (V_t \cup V_n)^*$ tels que $w = uXv$ et $w' = utv$ et $X \to t \in P$.
- w' est un *dérivé* par G de w, noté $w \Rightarrow_G^* w'$, si et seulement si il existe $w_0, w_1, \ldots, w_n \in (V_t \cup V_n)^*$ tels que $w_0 = w$, $w_n = w'$ et pour tout i w_{i+1} est un dérivé direct de w_i.

Remarque : $\Rightarrow_G$ et $\Rightarrow_G^*$ sont deux relations définies sur $(Vt \cup Vn)^*$, $\Rightarrow_G^*$ étant la fermeture transitive de $\Rightarrow_G$.

Langages hors-contexte

Soit $G = (V_t, V_n, Z, P)$ une grammaire hors-contexte. Le *langage défini* par G (on dit aussi langage *reconnu* par G) est l'ensemble $L(G)$ tel que :

$$L(G) = \{w \in V_t^* \mid Z \Rightarrow_G^* w\}$$

Un langage L est dit *hors-contexte* s'il existe une grammaire hors-contexte qui le reconnaît.

Remarque : Tout langage régulier est un langage hors-contexte, il existe des langages hors-contexte qui ne sont pas réguliers.

Arbres de dérivation

Soit $G = (V_t, V_n, Z, P)$ une grammaire et w un mot de $L(G)$. On appelle *arbre de dérivation* (ou *arbre syntaxique*) de w dans G tout arbre n-aire A dont les noeuds sont des éléments de $(V_t \cup V_n)$ et tel que :
- la racine de A est Z (l'axiome de G) ;

- les feuilles de A sont des éléments de $V_t \cup \{\varepsilon\}$ et la séquence gauche-droite des feuilles de A est égale à w ;
- les noeuds non feuilles de A sont des éléments de V_n et si un noeud non feuille X a pour fils u_1, u_2, $\dots u_n$ dans A alors la règle $X \to u_1.u_2 \dots u_n$ doit appartenir à P.

A tout mot de $L(G)$ on peut associer un arbre de dérivation. Si de plus cet arbre est *unique* alors la grammaire est dite *non ambigüe.*

Automates à Piles

Un *automate à pile* est un 7-uplet $M = (Q, V, \Gamma, \Delta, Z_0, q_0, F)$ tel que :
- Q est un ensemble fini d'états ;
- V est l'alphabet d'entrée ;
- Γ est l'alphabet de pile ;
- $Z_0 \in \Gamma$ est le symbole initial de pile ;
- $q_0 \in Q$ est l'état initial ;
- $F \subseteq Q$ est l'ensemble des états accepteurs
- $\Delta \subseteq ((Q \times V^* \times \Gamma) \times (Q \times \Gamma^*))$ est la relation de transition.

Un automate à pile permet de reconnaître un ensemble de mots de V^* (le langage accepté, ou reconnu par l'automate) : à chaque transition $((q, u, \beta), (q', \gamma)) \in \Delta$ l'automate passe de l'état q à l'état q' en lisant le préfixe u sur la chaîne d'entrée et en remplaçant le sommet de pile β par γ.

Une *configuration* d'un automate à pile est un élément $(q, w, \alpha) \in Q \times V^* \times \Gamma^*$ dans laquelle q désigne l'état courant de l'automate, w la chaîne d'entrée et α le contenu de la pile.

Une configuration $c' = (q', w', \alpha')$ est dite *dérivable* d'une configuration $c = (q, w, \alpha)$ si et seulement si il existe une transition $((q, u, \beta), (q', \gamma)) \in \Delta$ telle que
- $w = uw'$ (le préfixe u a été lu sur la chaîne d'entrée) ;
- $\alpha = \beta.\delta$ et $\alpha' = \gamma.\delta$ (le sommet de pile β a été remplacé par γ).

Un automate à pile est dit *déterministe* si et seulement si pour chaque configuration c il existe *au plus* une configuration c' dérivable à partir de c.

Un mot w de V^* est *reconnu* (ou *accepté*) par M si et seulement si il existe une suite de configurations c_0, c_1, $\dots c_n$ telles que :
- $c_0 = (q_0, w, Z_0)$ (l'automate est dans son état initial, la pile ne contient que le symbol initial, la chaîne d'entrée contient w) ;
- $c_n = (q, \varepsilon, \gamma)$ avec $p \in F$ (l'automate a atteint un état final, la chaîne d'entrée est vide)
- c_{i+1} est une configuration dérivable à partir de c_i.

Le *langage accepté ou reconnu par* M, $L(M)$, est alors défini par : $L(M) = \{w \in V^* \mid w \text{ est reconnu par } M\}$

Automates à pile et langages hors-contexte

On a les propriétés suivantes :
- le langage reconnu par un automate à pile est un langage hors-contexte ;
- pour tout langage hors-contexte il existe un automate à pile qui le reconnaît ;
- il n'y a pas équivalence entre automates à pile déterministes et non-déterministes : il existe des langages hors-contexe non reconnaissables par un automate à pile déterministe.