1.2.4 Grammaires

Définition

Définition statique Une **grammaire formelle** est un quadruplet $\langle X, V, S, P\rangle$ où

- X est l'alphabet (du langage engendré)
- V est un alphabet disjoint de X dit « *non terminal* »
- $S \in V$ est un élément distingué de V, appelé *axiome*
- P est un ensemble de « *règles (de production)* », c'est-à-dire une partie finie du produit cartésien $(X \cup V)^*V(X \cup V)^* \times (X \cup V)^*$.

Exemple $\mathcal{G}_1 = \langle X, V, S, P\rangle$, avec : $X = \{l, c\}$
$V = \{S, A\}$
$P = \{(S, lA), (A, \varepsilon), (A, cA), (A, lA)\}$

Notation :
$$\begin{array}{lcllcllcl} S \longrightarrow lA & \text{ou} & S & \longrightarrow & lA & \text{ou} & S & \longrightarrow & lA \\ A \longrightarrow \varepsilon & & A & \longrightarrow & \varepsilon & & A & \longrightarrow & \varepsilon \,|\, cA \,|\, lA \\ A \longrightarrow cA & & & | & cA & & & & \\ A \longrightarrow lA & & & | & lA & & & & \end{array}$$

Exemple $\mathcal{G}_2 = \langle \{(,)\}, \{S\}, S, \{S \longrightarrow \varepsilon \,|\, (S)S\}\rangle$

Dérivation Soient $\mathcal{G} = \langle X, V, S, P\rangle$ une grammaire, $(f, g) \in (X \cup V)^*$, r une règle de production de P, de la forme $r : A \longrightarrow u$ ($A \in (X \cup V)^*V(X \cup V)^* \times (X \cup V)^*$).

- f se **réécrit** (ou **dérive immédiatement**) en g avec la règle r (notation $f \xrightarrow{r} g$) ssi $\exists v, w$ t.q. $f = vAw$ et $g = vuw$
- f se **réécrit** (ou **dérive**) en g dans la grammaire $\mathcal{G}$ (notation $f \xrightarrow{\mathcal{G}} g$) ssi $\exists r \in P$ t.q. $f \xrightarrow{r} g$.
- $f \xrightarrow{\mathcal{G}*} g$ si $f = g$
  $\exists f_1 = f, f_2, ..., f_n = g$ t.q. $f_{i-1} \longrightarrow f_i$

On note $L_{\mathcal{G}}(f)$ l'ensemble des terminaux engendrés par f dans la grammaire $\mathcal{G}$.
$L_{\mathcal{G}}(f) = \{g \in X^* / f \xrightarrow{\mathcal{G}*} g\}$
Par convention, on notera $L_{\mathcal{G}}$ le langage $L_{\mathcal{G}}(S)$.

Exemple de dérivation (avec $\mathcal{G}_2$ plus haut) : $S \rightarrow (S)S \rightarrow ()S \rightarrow () \in L_G(S)$, etc. (aussi ((())), ()()(), ((()()())), ...)

Hiérarchie de Chomsky

type 0 Aucune restriction sur $P \subset (X \cup V)^*V(X \cup V)^* \times (X \cup V)^*$.

type 1 (*grammaires contextuelles, context-sensitive*) Tout élément de P est de la forme (u_1Su_2, u_1mu_2), où u_1 et $u_2 \in (X \cup V)^*$, $S \in V$ et $m \in (X \cup V)^+$.

type 2 (*grammaires algébriques, context-free*) Tout élément de P est de la forme (S, m), où $S \in V$ et $m \in (X \cup V)^*$.

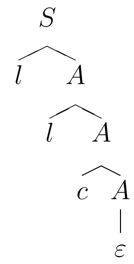
type 3 (*grammaires régulières, context-free*) Tout élément de P est de la forme (S, m), où $S \in V$ et $m \in X.V \cup X \cup \{\varepsilon\}$.

Arbre de dérivation

Exemple Reprenons la grammaire $\mathcal{G}_1$, soit le mot llc.

$$\begin{array}{lcl} S & \longrightarrow & lA \\ A & \longrightarrow & \varepsilon \,|\, cA \,|\, lA \end{array}$$

à plat : $S \rightarrow lA \rightarrow llA \rightarrow llcA \rightarrow llc\varepsilon = llc$

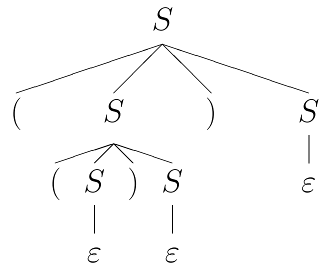


Exemple Grammaire $\mathcal{G}_2$, soit le mot (()).

$\langle \{(,)\}, \{S\}, S, \{S \longrightarrow \varepsilon \,|\, (S)S\}\rangle$

$S \rightarrow (S)S \rightarrow ((S)S)S \rightarrow ((S)S) \rightarrow ((S)) \rightarrow (())$

On ne voit pas clairement quel non terminal fait l'objet de la réécriture.



Exercices

1. Soit la grammaire $\mathcal{G}_2 = \langle \{(,)\}, \{S\}, S, \{S \longrightarrow \varepsilon \,|\, (S)S\}\rangle$. Quel est le langage engendré par $\mathcal{G}_2$? Quel est le type de la grammaire $\mathcal{G}_2$?
2. Soit la grammaire $\mathcal{G}_3 = \langle \{a, b\}, \{S, A\}, S, P\rangle$, avec $P = \left\{ \begin{array}{l} S \longrightarrow aSb \,|\, bSa \,|\, A \\ A \longrightarrow \varepsilon \,|\, a \,|\, b, \\ bbA \longrightarrow \varepsilon \\ Aaa \longrightarrow Saa \,|\, aa \end{array} \right\}$
   Quel est le type de cette grammaire ? Donnez trois dérivations possibles dans cette grammaire.
3. Soit la grammaire $S \rightarrow S + S \mid S \times S \mid x \mid y \mid z$. Donnez tous les arbres syntaxiques possibles pour l'expression $x + y \times z$.
4. Soit la grammaire $\begin{array}{ll} S & \rightarrow T2 \\ T & \rightarrow 0T1C \\ T & \rightarrow \varepsilon \\ C1 & \rightarrow 1C \\ C2 & \rightarrow 22 \\ 12 & \rightarrow 1 \end{array}$ $(X = \{0, 1\})$.
   Est-ce que le mot 001122 est engendré par cette grammaire ?