# Langages, Automates et Compilation

Dr Mouhamadou GAYE

UFR Sciences Et Technologies
Département d'Informatique
Licence 3 en Informatique Option Génie Logiciel

24 novembre 2020

# Grammaires formelles

### Définition 1.

Une grammaire est un quadriplet $G = (V_T, V_N, S, P)$, où

- $V_T$ est l'ensemble fini des symboles terminaux, appelé vocabulaire terminal ;
- $V_N$ est un ensemble fini (disjoint de $V_T$) de symboles dits non-terminaux, appelé vocabulaire non terminal ;
- S est un symbole initial non terminal appelé source ou axiome ;
- P est un ensemble de règles de productions, appelées aussi règles de réécriture.

# Règle de production

**Définition 2.**

Une règle de production $\alpha \rightarrow \beta$ précise que la séquence de symboles $\alpha$ ($\alpha \in (V_T \cup V_N)^+$) peut être remplacée par la séquence de symboles $\beta$ ($\beta \in (V_T \cup V_N)^*$).
$\alpha$ est appelé **partie gauche** de la production et $\beta$ **partie droite**.

**Convention** : On utilisera des mots commençant par une majuscule pour les non terminaux, et des lettres minuscules ou des symboles spéciaux (+, x, etc.) pour les terminaux.

**Exemple** : G = ($V_T$, $V_N$, S, P) avec

- $V_T = \{a, b, c, d, e\}$
- $V_N = \{S, A, B\}$
- $P = \{S \rightarrow aSb \mid cd \mid SAe$
  $A \rightarrow aAdB \mid \epsilon$
  $B \rightarrow bb\}$

**Définition 3.**

Une dérivation est l'application d'une ou de plusieurs règles de production à partir d'une mot de $(V_T \cup V_N)^+$.

- On note $\rightarrow$ une dérivation obtenue par l'application d'une seule règle de production ;
- $\overset{*}{\rightarrow}$ est utilisé pour noter une dérivation obtenue par l'application de n règles de production, $(n \geq 0)$ ;
- $\overset{+}{\rightarrow}$ est utilisé pour noter une dérivation obtenue par l'application de n règles de production, $(n > 0)$.

**Définition 4.**

Etant donné une grammaire G, on note L(G) le langage engendré par G défini par :

$$L(G) = \{\omega \in (V_T)^+ : S \xrightarrow{*} \omega\}$$

## Définition 5.

Un arbre de dérivation est un arbre tel que :

- la racine est l'axiome,
- les feuilles sont des symboles terminaux,
- les nœuds sont des symboles non terminaux,
- les fils d'un nœud X sont $\beta_1$, ..., $\beta_n$ si et seulement si $X \rightarrow \beta_1 \ldots \beta_n$ est une production ($\beta_i \in V_T \cup V_N$).

**Exemple :**

**Définition 6.**

Une grammaire $G = (V_T, V_N, S, P)$ est régulière à droite si toutes les productions sont de la forme :

$$A \rightarrow \alpha B, \ A \rightarrow \alpha \text{ ou } A \rightarrow \epsilon$$

avec $A, B \in V_N$ et $\alpha \in V_T$.

# Grammaire régulière à gauche

## Définition 7.

Une grammaire $G = (V_T, V_N, S, P)$ est régulière à gauche si toutes les productions sont de la forme :

$$A \longrightarrow B\alpha,\ A \longrightarrow \alpha \text{ ou } A \longrightarrow C$$

avec $A, B \in V_N$ et $\alpha \in V_T$.

**Définition 8.**

Une grammaire est régulière si elle est régulière à gauche ou régulière à droite.

**Définition 9.**

Une grammaire est ambiguë s'il existe un mot de L(G) ayant plusieurs arbres de dérivation distincts.

**Exemple :** La grammaire donnée par

$$\begin{aligned} P = \{ & \text{instr} \rightarrow \text{if (expr) instr else instr} \\ & \text{instr} \rightarrow \text{if (expr) instr} \\ & \text{instr} \rightarrow \ldots \\ & \text{expr} \rightarrow \ldots \} \end{aligned}$$

est ambiguë car le mot m = if $(x > 0)$ if $(y < 0)$ $a = 1$ else $a = 0$ a deux arbres de dérivation distincts.

# Hiérarchie de Chomsky

| **Grammaire** | **Règle de production** |
|---|---|
| générale ou de type 0 | $\alpha \rightarrow \beta$ |
| contextuelle ou de type 1 | $\alpha A \beta \rightarrow \alpha \gamma \beta$, avec $\gamma \neq \epsilon$ |
| algébrique ou de type 2 | $A \rightarrow \gamma$, avec $\gamma \neq \epsilon$ |
| rationnelle ou de type 3 | $A \rightarrow \alpha B$, $A \rightarrow \alpha$, avec $\alpha \neq \epsilon$ |
| à choix finis ou de type 4 | $A \rightarrow \alpha$, avec $\alpha \neq \epsilon$ |

# Construction d'un AFD à partir d'une grammaire régulière

Soit la grammaire régulière à droite $G = (V_T, V_N, S, P)$.
L'automate $A = (\Sigma, E, E_o, F, \delta)$ telle que :

- $E = V_N \cup \{X\}$
- $\Sigma = V_T$
- $E_o = S$
- dont les transitions sont :
  - $\delta(q_i, a) = q_j$ pour les productions $q_i \rightarrow aq_j$
  - $\delta(q_i, a) = q_i$ pour les productions $q_i \rightarrow aq_i$
- $F = \{X\} \cup \{A : A \rightarrow \epsilon\}$

reconnait le même langage que la grammaire G.