Théorie des langages II

Wassim SAIDANE

01/03/2021

Note :

Ce cours est ma prise de note du cours de L3 infos de Théorie des langages II de Mamadou Kante

Chapitre 1 : Langages algébriques (ou hors contexte)

1.1 Grammaires algébriques

Une **grammaire algébrique** est un tuple (V, Σ, R, S) où :

- V : Ensemble fini, appelé ensemble de **varaiables** ou **non-terminaux**.
- Σ : Ensemble fini, appelé ensemble des **terminaux** ($\Sigma \bigcap V = \emptyset$).
- $S \in V$, appelé **non-terminal d'entrée (de départ)**.
- R est un ensemble de règles :

$$R \subseteq P(V \times (V \bigcup \Sigma)^*)$$

chaque règle dans R est écrite ainsi :

$$A- > w \text{ où } A \in V, w \in (V \bigcup \Sigma)^*$$

Si $u, v, w \in (V \bigcup \Sigma)^*$ et $A- >$(dérive)$w \in R$.
On dira que depuis uAv on peut dériver le mot uwv et on note : $uAb \Longrightarrow uwv$
On dira u dérive v, noté $u \Longrightarrow^* v$ si

- Soit $u = v$
- Soit il existe une séquance $u_1, u_2, ...u_k$ tel que :

$$u \Longrightarrow u_1, u_1 \Longrightarrow u_2, ..., u_i \Longrightarrow u_{i+1}, ..., u_{k-1} \Longrightarrow u_k, u_k \Longrightarrow v$$

Le langage d'une grammaire G, notée $L(G)$, c'est l'ensemble $\{w \in \Sigma^* : S \Longrightarrow^* w\}$
$L \subseteq \Sigma^*$ **est algébrique si il existe une grammaire algébrique** G **telle que** $L = L(G)$

<u>Exemples :</u>

$$G = (\{S\}, \{a, b\}, R, S)$$

$uAv \Longrightarrow uwv$ si $A- > w$

$$\left.\begin{array}{rl} R: & S \to \varepsilon \\ S & \to aSb \\ S & \to SS \end{array}\right\} S \Rightarrow \varepsilon | SS \mid aSb$$

Des mots générés par cette grammaire :

$$\varepsilon(\text{avec la règle S}) \to \varepsilon), ab(S \Longrightarrow aSb, aSb \Longrightarrow ab)$$

<u>Exemple 2 :</u>

$$EA = (\{E, T, F\}, \{+, *, a, (,)\}, R, \{E\})$$

$$E \to E + T \mid T$$

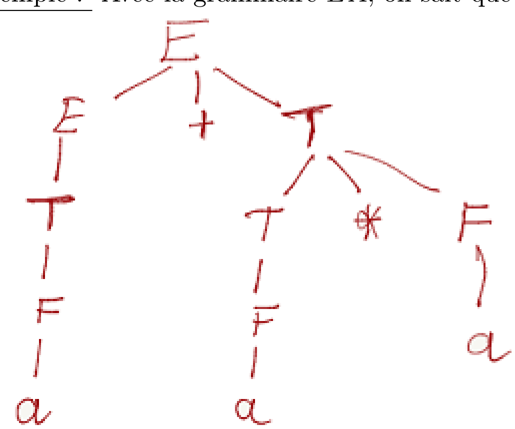
$$T \to T * F \mid F$$

$$F \to (E) \mid a$$

<u>Mots générés :</u> $a + a * a, a + a + a * a, ...(a + a) * a$

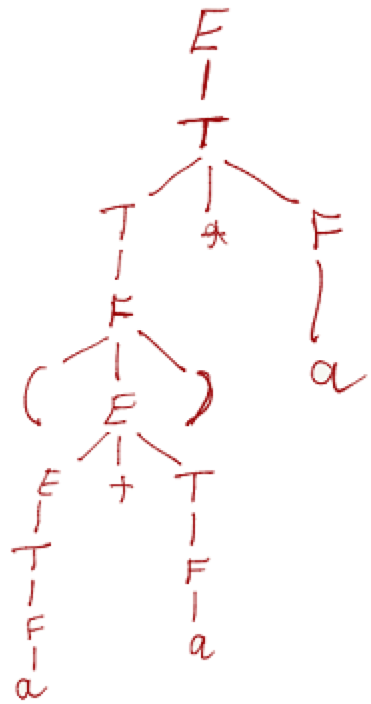
A chaque fois qu'un mot est généré (ou dérivé) avec une grammaire, on peut lui associé un <u>arbre de dérivation</u>, cet arbre décrit comment on construit le mot.
La racine : C'est le non-terminal d'entrée.
Les feuilles : Sont des terminaux.
Les noeuds internes : Sont des non-terminaux.

Si A est un noeud interne, et on applique la règle $A \to w, w \in (V \bigcup \Sigma)^*$
Les fils de A seront les lettres dans $V\Sigma$ qui composent le mot w.
<u>Exemple :</u> Avec la grammaire EA, on sait que E gènère $a + a * a$.



arbre de dérivation de a+a*a

1 Arbre de dérivation pour $(a+a)*a$



$$E \Rightarrow T \Rightarrow T*F \Rightarrow T*a \Rightarrow F*a \Rightarrow (E)*a \Rightarrow (E+T)*a \Rightarrow (T+T)*a \Rightarrow (F+T)*a \Rightarrow (F+F)*a \Rightarrow (a+F)*a \Rightarrow (a+a)*a$$

Un mot peut être dérivé de plusieurs façons différentes. Ceci peut induire un non-déterminisme sur la façon de savoir si un mot est généré par une grammaire. Une dérivation d'un mot (à partir d'une grammaire G) est une ddérivation gauche si à chaque étape de la dérivation, on remplace la variable la plus à gauche.

Définition 1.1 Un mot u est ambigüe pour une grammaire G si il existe deux dérivations gauches différentes du mot u à partir de G.
Une grammaire G est dite anbigûe si elle admet un mot anbigüe.

Proposition 1.1 : On ne peut pas rendre non ambigüe toutes les grammaires.
Preuve : Montrer que toute grammaire reconnaissant le langage $\{a^ib^jc^k \mid i = jouj = k\}$ est forcément ambigüe.

Un langage tel que toute grammaire la générant est ambigüe est dite : Langage anbigüe.

Les différentes dérivations gauches d'un mot (à partir d'une grammaire) sont représentées par les arbres de dérivation.
De façon équivalente, un mot est anbigüe si il admet deux arbres de dérivations différents.

1.2 Forme Normale de Chomsky

Définition 1.2 : Une grammaire est en forme normale de Chomsky si chaque règle est de la forme :

$$A \to BC$$

ou

$$A \to a$$

On accepte la règle $S \to \epsilon$

Théorème 1.1 : Tout langage algébrique est le langage d'une grammaire en forme normale de Chomsky.