Chapitre 3:

Les grammaires

Concepts de base

- Une grammaire est un formalisme permettant de spécifier des langages.
- Intuitivement, une grammaire est un système de règles permettant de réécrire un terme en un autre terme.
- Les règles de la grammaire sont composées d’une partie gauche et d’une partie droite reliées par une flèche où chacune est elle-même composée de symboles *terminaux* et de symboles *non terminaux*.
- Les symboles terminaux correspondent aux symboles qui forment les mots engendrés par la grammaires.

Concepts de base

- Les symboles non terminaux servent principalement à la réécriture.
- La réécriture d'un terme en un autre terme consiste à appliquer une règle au terme original, c'est-à-dire à substituer la partie gauche de la règle qui apparait dans le terme par la partie droite de la règle, et obtenir ainsi le nouveau terme.
- On commence la réécriture à partir d'une règle de départ particulière appelée *axiome* et on applique successivement les règles, on finit par obtenir un terme pour lequel aucune règle ne s'applique.

Concepts de base

- Un terme ainsi obtenu ne comportant aucun symbole non terminal est donc un mot défini sur l'alphabet des symboles terminaux.
- L'ensemble de ces mots forment alors un langage que l'on appelle *langage engendré* par la grammaire.

Concepts de base

- **Définition:**
- Une grammaire formelle est un quadruplet G = (X, N, S, P) où
- – X est un ensemble fini de symboles appelé alphabet terminal.
- – N, disjoint de X, est un ensemble de symboles non terminaux appelé alphabet non terminal. On note par V l'ensemble $X \cup N$
- – S est un élément de N: c'est l'axiome (ou le start)
- – P est un ensemble des règles de production (ou de réécriture)

- $P = \{ (u, v) \ / \ u \in V^+ \text{ et } v \in V^*\}$
- Les éléments de P seront notés $u \to v$ au lieu de (u, v)
- Un élément de P est une règle de production.

- **Exemple:**

- **Remarque:**

- Les symboles terminaux sont habituellement représentés par des chiffres ou des lettres minuscules et les symboles non terminaux par des lettres majuscules.

- **Définition (Dérivation en une étape)**

- Etant données deux mots f et g de V^*. Nous dirons que f se dérive directement en g pour la grammaire G et nous noterons $f \Rightarrow g$ s'il existe $z_1, z_2 \in V^*$, $u \in V^+$ et $v \in V^*$ tels que $f = z_1.u.z_2$ et $g = z_1.v.z_2$ et $(u \rightarrow v) \in P$.

- Exemple 1:

- **Exemple 2:**

- **Exercice 1:**

- Soit la grammaire suivante:
- G   1) $S \rightarrow \varepsilon$
-       2) $S \rightarrow aSb$

- **Exercice 2 (énoncé):**

- Soit G la grammaire suivante:
- G 1) S $\rightarrow$ aS
- 2) S $\rightarrow$ ε

1) Définir les termes de G (X, N, S, P)
2) Donner le langage de G
3) Donner une dérivation en une étape.

- **Exercice 2 (Solution):**

- **Définition (Dérivation en plusieurs étapes)**
- Etant donnés f et g deux mots de V*. On dit que g dérive de f à l'ordre k si et seulement si:
- – $k = 0$ alors $g = f$
- – $\exists f' \in V^* \,/\, f \rightarrow f'$ (f' dérive directement de f) et g dérive de f' à l'ordre k–1. On note $f \rightarrow_k g$
- – $\exists f_0, f_1, \ldots, f_k \in V^* \,/\, f_0 = f$ et $f_k = g$ et $\forall i \in \{1, \ldots, k-1\};\ f_{i-1} \rightarrow f_i$

- **Exemple:**

- **Définition (Squelette)**

- Si les règles de la grammaire sont numérotées, on peut associer à la dérivation la suite des numéros des règles utilisées. Cette suite s'appelle squelette de la dérivation.
- Exemple: $f_0 \rightarrow_1 f_1 \rightarrow_2 f_2 \rightarrow_3 f_3 \rightarrow_4 f_4$

- **Exemple:**

Soit G la grammaire suivante:

G 1) $S \rightarrow aSb$

2) $S \rightarrow \varepsilon$

Question: Donner une dérivation et sa squelette pour le mot aaaabbbb.

Solution

- **Définition (Langage engendré)**

- Le langage engendré par une grammaire G = (X, N, S, P), noté L(G) est l'ensemble des mots formés uniquement de terminaux que l'on peut dériver à partir de l'axiome S.

$$L(G) = \{w \in X^* \;/\; S \rightarrow^* w\}$$

- **Définition (Equivalence)**
- Deux grammaires G et G’ sont équivalentes si et seulement si L(G)= L(G’)

- **Exemple1:**

Exercice:

- Soient de grammaires G et G'.

G $S \rightarrow aS$ G' $S \rightarrow Sa$

$S \rightarrow \varepsilon$ $S \rightarrow a$

G et G'sont-elles équivalentes?

Solution:

- **Exercice:**
- G $\quad S \rightarrow AB$
- $\quad\quad A \rightarrow Aa \mid a$
- $\quad\quad B \rightarrow bB \mid b$

1) Définir G en terme de (X, N, S P)
2) Donner le langage L(G)
3) Donner une dérivation en une étape
4) Donner une dérivation en plusieurs étapes
5) Donner la définition de L(G) en fonction de L(G,A) et L(G, B).

- **Solution:**

Types de grammaire

- Soit G = (X, N, S, P)
- **Définition (Type 0)**
- Une grammaire formelle de type 0 est une grammaire où on n'introduit aucune contraintes sur la forme des règles de réécriture.
- **Exemple:**
- G est de type 0 si aucune contraintes n'ont été posées sur les règles. Toute grammaire n'est pas d type 3 ni de type 2 ni de type 1 alors elle est de type 0.

- **Définition (Type 1)**
- Soit G = (X, N, S, P) une grammaire formelle.
- Une grammaire formelle G de type 1 est dite contextuelle (context sensitif). G est contextuellement si chaque règle de production est de la forme : $u.A.v \rightarrow u.w.v$ avec $A \in N$, u,v et $w \in V^*$

- **Exemple:**

- **Définition (Type 2)**
- Soit G = (X, N, S, P) une grammaire formelle.
- Une grammaire formelle G de type 2 est dite non-contextuelle (context free). G est non-contextuelle si chaque règle de production est de la forme $A \rightarrow w$ avec $A \in N$ et $w \in V^*$.

- **Exemple:**

- **Définition (Type 3)**
- Soit $G = (X, N, S, P)$ une grammaire formelle.
- Une grammaire formelle G de type 3 est dite régulière si elle est régulière à droite ou à gauche.
- Une grammaire G est régulière à droite si toutes les règles de productions de la forme $A \rightarrow B$ possèdent les restrictions suivantes:
- 1) $|A| = 1$
- 2) $B=aC$, $B=a$ ou $B = \varepsilon$ avec $B \in N$ et $a \in X$

- Une grammaire G est régulière à gauche si toutes les règles de productions de la forme $A \rightarrow B$ possèdent les restrictions suivantes:
- 1) $|A| = 1$
- 2) $B=Ca$, $B=a$ ou $B = \varepsilon$ avec $B \in N$ et $a \in X$

- **Exemple:**

- **Exemple 2:**

- **Définition**
- Un langage est de type n si et seulement s'il peut être engendré par une grammaire de type n et ne peut pas être généré par une grammaire de type supérieur à n

- **Exercice:**
- Soit G la grammaire suivante:
- G 1) $S \rightarrow aSBc$
- 2) $S \rightarrow aBc$
- 3) $aB \rightarrow ab$
- 4) $bB \rightarrow bb$
- 5) $cC \rightarrow cc$
- 1) Donner la définition de G en terme de (X, N, S, P)
- 2) Quel est le type de G

- **Solution**

Dérivation, arbres et ambigüité

- Considérons la grammaire G:
- $S \rightarrow AB$
- G $A \rightarrow a$
- $B \rightarrow b$
- Soit le mot ab. Deux dérivations permettent de construire le mot ab.
- $S \rightarrow AB \rightarrow aB \rightarrow ab$ et $S \rightarrow AB \rightarrow Ab \rightarrow ab$
- Les deux dérivations différent uniquement par l'ordre dans lequel on dérive les non-terminaux A et B. Par contre dans les deux cas A et B sont remplacés par le même mot.

- Intuitivement ces deux dérivations sont identiques à l’ordre près des dérivations.
- Considérons par contre la grammaire G’
- G’ $S \rightarrow SbS$
- $S \rightarrow a$
- Soit le mot ababa. Ce mot peut être engendré de deux façons:
- $S \rightarrow \underline{S}bS \rightarrow \underline{S}bSbS \rightarrow ab\underline{S}bS \rightarrow abab\underline{S} \rightarrow ababa$
- $S \rightarrow Sb\underline{S} \rightarrow SbSb\underline{S} \rightarrow Sb\underline{S}ba \rightarrow \underline{S}baba \rightarrow ababa$

- Si les deux dérivations s'écrivent de façon identique dans le premier cas la première occurrence de S est dérivée en a, tandis que le deuxième cas la dernière occurrence de S est dérivée en a.
- Il s'agit de dérivations fondamentalement différentes.
- Un moyen pour lever cette ambigüité, due à une écriture linéaire des dérivations, est de choisir une représentation graphique par des arbres.

- **Définition (Arbre de dérivation)**
- Un arbre A est un arbre de dérivation dans G si et seulement si:
- – Tous les nœuds de A sont étiquetés par un symbole de $V \cup \varepsilon$.
- – La racine est étiquetée par S.
- – Si un nœud n n'est pas une feuille et porte l'étiquette X, alors $X \in V$.
- – Si n_1, n_2, …, n_k sont les fils de n dans A, d'étiquettes respectives X_1,.., X_n alors $X \rightarrow X_1 | \ldots | X_k$ est une production de G.

- Ainsi à toute production $A \rightarrow A_1 A_2 \ldots A_n$ avec $A_i \in V$, on associe l'arbre de la racine A et de feuilles A_i pour $i \in \{1,\ldots,n\}$

- **Exemple:**

- A toute étape de la dérivation on peut associer un arbre de dérivation. L'arbre obtenu est appelé arbre de dérivation.
- De même façon, on peut lever l'ambigüité en dérivant systématiquement le non terminal le plus à gauche ou le plus à droite du mot. On parle alors de dérivation la plus à gauche ou la plus à droite du mot.

- En effet, pour le mot ababa et la grammaire G' on a:
- Dérivation la plus à gauche:
S➔SbS➔SbSbS➔abSbS➔ababS➔ababa
- Dérvation la plus à droite:
S➔SbS➔SbSbS➔SbSba➔Sbaba➔ababa

- **Définition:**
- Une grammaire G est dite ambigüe s’il existe au moins un mot de L(G) auquel on peut associer au moins deux arbres de dérivations distincts.
- **Exemple:**

- **Exemple:**
- $$W = abc$$

Symboles non terminaux (Maj)   Symboles terminaux (Lettres min)