Dans ce qui suit, nous allons introduire les grammaires ainsi que leur relation aux langages qu'elles décrivent. Nous allons également présenter la classification de Chomsky qui classe les grammaires selon leur type.

1 Grammaire

Définition : Une grammaire est un quadruplet G <X,V, P, S> où :

- X est l'alphabet,
- V est l'ensemble des variables, $X \cap V = \varnothing$
- $S \in V$ est l'axiome,
- P est l'ensemble des règles de production.

P est un sous-ensemble de : $(X \cup V)^*.V.(X \cup V)^* \times (X \cup V)^*$. Les éléments de P sont des couples (α, β) où $\alpha \in (X \cup V)^*.V.(X \cup V)^*$ et $\beta\ (X \cup V)^*$.

α est appelé membre gauche de la règle de production et β son membre droit. On dit que α génère β.

<u>Représentation des éléments de P</u> :

Il existe plusieurs représentations des éléments de P :

1. $\alpha \rightarrow \beta$. (C'est la notation que nous utiliserons dans ce cours)

2. Forme normale de Backus : Dans la FNB, une règle de production est représentée de la manière suivante : **< α > ::= β**. C'est la notation utilisée pour définir les langages de programmation.

.

<u>Exemple 1:</u> Soit G la grammaire, où X= {a, b}, V={S, A} et P est défini comme suit :

P = { $S \rightarrow ab\,S\,A \,/\, \varepsilon$,
$A \rightarrow Aa \,/\, \varepsilon$ }

2 Mot généré par une grammaire

Définition : Soit G<X,V, P, S> une grammaire, on dit qu'un mot $u \in (X \cup V)^*.V.(X \cup V)^*$ **se dérive directement** en un mot $v \in (X \cup V)^*$, que l'on note $u \vdash_G v$ si et seulement si :

$u = u_1\,A\,u_2$ et $v = u_1\,\beta\,u_2$ tq $\alpha \rightarrow \beta \in P$

avec $u_1, u_2 \in (X \cup V)^*$, $A \in V$ et $\beta \in (X \cup V)^*$.

Exemple 2: Soit G la grammaire définie dans l'exemple 1, nous avons

$$\underbrace{\text{ababab}S}_{u_1}\ \underbrace{A}_{\alpha}\ \underbrace{AAa}_{u_2} \;\vdash_G\; \underbrace{\text{ababab}S}_{u_1}\ \underbrace{Aa}_{\beta}\ \underbrace{AAa}_{u_2} \quad \text{et} \quad A \to Aa \in P$$

Définition : Soit G<X,V, P, S> une grammaire, on dit qu'un mot $u \in (X \cup V)^*.V.(X \cup V)^*$ se dérive en k étapes en un mot $v \in (X \cup V)^*$, que l'on note $u \vdash_G^k v$, s'il exite $u_1, u_2, \ldots u_k$ tq :

$$u_0 = u_1 \vdash_G u_2 \vdash_G u_3 \ldots \vdash_G u_k = v \text{ tq } u_i \vdash_G u_{i+1} \; 1 \le i \le k\text{-}1 \text{ est une dérivation directe.}$$

La relation *se dérive en* est la fermeture réflexive et transitive de la relation se dérive directement en.

Exemple 3: Soit G la grammaire définie dans l'exemple 1, nous avons

$$\text{abab } S \text{ } AAa \vdash_G^{4} \text{ababab}Saa$$

3 Langage engendré par une grammaire

Définition : Soit G<X,V, P, S> une grammaire. On dit qu'un mot $w \in X^*$ est généré par G s'il existe une dérivation pour ce mot à partir de l'axiome que l'on note : $S \vdash_G^* w$.

Définition : Le langage généré par une grammaire G<X,V, P, S> est défini comme suit :

$L(G) = \{w \in X^* \text{ tq } S \vdash_G^* w\}$.

Exemple 4: Le langage engendré par la grammaire du PASCAL est l'ensemble de tous les programmes qui sont correctement écrits en code source PASCAL.

Exemple 5: Le langage engendré par la grammaire G< X={a,b},V={S},{S → abS / ε}, S> est le suivant :

$L(G) = \{(ab)^i, i \ge 0\}$.

Définition : Soient G_1<X,V, P, S> et G_2<X,V, P, S> deux grammaires. G_1 et G_2 sont équivalentes si et seulement si $L(G_1) = L(G_2)$.

4 Classification de Chomsky

La classification des grammaires se fait selon la nature des règles de productions. La classification des grammaires se transmet aux langages. Un langage de type *i* est généré par une grammaire de type *i*. Chomsky a classé les grammaires en 4 classes que nous présentons ci-dessous.

Les grammaires régulières (Type3)

Un langage régulier est généré par une grammaire régulière droite ou une grammaire régulière gauche. Pour ces grammaires, il y des restrictions aussi bien sur le membre gauche d'une règle de production (α) que sur le membre droit (β).

Définition: Une grammaire G <X, V, P, S> est régulière droite si et seulement si toutes les productions de P sont de la forme :

$A \rightarrow wB / w$ où $A, B \in V$ et $w \in X^*$

Exemple 6: Soit G <X,V,P,S> la grammaire suivante où X={a, b}, V = {S, A}, et

$P = \{ S \rightarrow a S / A$
$A \rightarrow b A / B$
$B \rightarrow c B / \varepsilon \}$

G est une grammaire régulière droite engendrant le langage régulier $L(G)=\{a^n b^m c^p, n, m,p \geq 0\}$.

Définition: Une grammaire G <X, V, P, S> est régulière gauche si et seulement si toutes les productions de P sont de la forme :

$A \rightarrow B w / w$ où $A, B \in V$ et $w \in X^*$

Exemple 7: Soit G <X,V,P,S> la grammaire suivante où X={0, 1}, V = {S, A}, et

$P = \{ S \rightarrow S 1 / A 1$
$A \rightarrow A 0 / \varepsilon \}$

G est une grammaire régulière gauche engendrant le langage $L(G) = \{0^i 1^j, i \geq 0$ et $j > 0\}$.

Les grammaires langages algébriques (A contexte libre-Type 2)

Cette classe est également appelée classe des langages à contexte libre. Un langage algébrique, ou à contexte libre, est généré par une grammaire algébrique. Pour ces grammaires, il y a une restriction sur le membre gauche des règles de production (α), $\alpha \in V$, mais aucune sur le membre droit (β).

Définition: Une grammaire G <X, V, P, S> est algébrique si et seulement si toutes les productions de P sont de la forme :

$A \rightarrow \beta$ avec $A \in V$ et $\beta \in (X \cup V)^*$

Exemple 8: Soit G <X,V,P,S> la grammaire suivante où X={a, b}, V = {S}, et
$P = \{ S \rightarrow AB$
$A \rightarrow aAb / \varepsilon$
$B \rightarrow cB / \varepsilon \}$

G est une grammaire algébrique engendrant le langage algébrique $L(G) = \{a^n b^n c^m, n, m \geq 0\}$.

La classe des langages à contexte lié (Type 1)

Un langage à contexte lié est généré par une grammaire monotone ou par une grammaire à contexte lié

Définition: Une grammaire G <X, V, P, S> est monotone si et seulement si toutes les productions de P sont de la forme :

$\alpha \to \beta$ avec $\alpha \in (X \cup V)^*.V.(X \cup V)^*$ et $\beta \in (X \cup V)^+$ et tq $|\alpha| \le |\beta|$.

Remarque : On permet d'avoir $S \to \varepsilon$ dans le cas ou le mot vide appartient au langage. Dans ce cas, S ne devra apparaître dans aucun membre droit de production.

<u>Exemple 8:</u> Soit G <X,V,P,S> la grammaire suivante où X={a, b}, V = {S}, et

P = { $S \to aAbc$ / abc
$A \to aATb$ / aTb
$Tb \to bT$
$Tc \to cc$ }

G est une grammaire monotone engendrant le langage à contexte lié $L(G) = \{a^n b^n c^n, n > 0\}$.

Définition: Une grammaire G <X, V, P, S> est à contexte lié si et seulement si toutes les productions de P sont de la forme :

$y_1 A y_2 \to y_1 \beta y_2$ $y_1, y_2 \in (X \cup V)^*$, $A \in V$ et $\beta \in (X \cup V)^+$

Remarque : Par définition, toute grammaire à contexte lié est monotone.

Théorème : A toute grammaire monotone G<X, V, P, S>, il existe une grammaire G'<X, V',P',S'> à contexte lié équivalente (La démonstration sera faite en cours).

La classe des langages sans restrictions

La forme des productions dans P ne fait l'objet d'aucune restriction.

Définition: Une grammaire G <X,V, P, S> est sans restriction si et seulement si toutes les productions sont de la forme :

$\alpha \to \beta$ où $\alpha \in (X \cup V)^*.V.(X \cup V)^*$ et $\beta \in (X \cup V)^*$

Série d'exercices

1.1. Donner les grammaires engendrant les langages suivants:

- G_1 tel que $L(G_1) = \{a^i b, i \geq 0\}$
- G_2 tel que $L(G_2) = \{a^n b^p / n > p\}$
- G_3 tel que $L(G_3) = \{ a^n b^p / n \neq p\}$
- G_4 tel que $L(G_4) = \{ a^i (ab)^j c^k, \quad i, k \geq 0 \text{ et } j > 0 \}$
- G_5 tel que $L(G_5) = \{ a^i (ab)^i c^k, \ i, k > 0\}$
- G_6 tel que $L(G_6) = \{ a^n b^n c^n / n > 0 \}$
- G_7 tel que $L(G_7) = \{ 0^i 1^j 2^k, i, j, k \geq 0 \text{ et } k = \max(i,j) \}$
- G_8 tel que $L(G_8) = \{w \in \{a, b, c\}^* / |w|_a \equiv 1[2] \text{ et } |w|_b \equiv 1[2]\}$
- Le complément de $L = \{a^n b^n / n \geq 0 \}$

1.2. Soit X={0,1}, donner les grammaires engendrant respectivement:

- les mots binaires divisibles par 2
- les mots divisibles par 3
- les mots binaires divisibles par 6
- les mots binaires non divisibles par 20

1.3. Soit $X = \{a, b\}$, donner la grammaire du langage suivant :

$L = \{|w|_a - 2|w|_b \equiv 1[4]\}$

1.4. Trouver la grammaire G engendrant $L(G) = \{w / |w|_a = |w|_b \text{ et } w \in \{a, b\}^* \}$

1.5. Donner la grammaire engendrant ce langage Dych2 suivant :

$L = \{w \in \{a, b\}^* / |w|_a = |w|_b \text{ et } \forall w_1 \text{ facteur gauche de } w \text{ alors } |w_1|_a \geq |w_1|_b\}$

1.6. Trouver les grammaires qui engendrent les langages suivants:

- $L_1 = \{ a^n b^i c^n d^j / n, i, j \geq 0\} \cup \{ a^i b^n c^j d^n / n, i, j \geq 0\}$
- $L_2 = \{ a^n b^p c^q / n, q \geq 0, p \geq n+q \}$
- $L_3 = \{ a^n b^p / n \neq p+2 \}$
- $L_4 = \{ wcw^R / w \in \{a, b\}^* \}$
- $L_5 = \{ w w^R / w \in \{a, b\}^* \}$
- $L_6 = \{ w w / w \in \{a, b\}^* \}$
- $L_7 = \{0^i 1^j 2^k / i \leq k \text{ ou } j \leq k \}$
- $L_8 = L^*$ tq $L = \{w \in \{a, b\}^* / |w|_a \neq |w|_b \}$

- $L_9 = \{a^i b^j a^i b^j, i, j \geq 1\}$
- $L_{10} = \{a^i b^j (ab)^{|i-j|}, i, j \geq 0\}$

1.7. Déterminer le langage L(G) où G est une grammaire dont les productions sont les suivantes :

$S \rightarrow Sa / bS / a / b$

1.8. Donner la grammaire du le langage suivant $L=\{a^n b^p a^q$ tq $p = 2n+q \}$

1.9. Soit G <{a, b}, {S, S1, S2, S3, A, B}, P, S> la grammaire telle que :

$$
\begin{aligned}
P = \{S &\rightarrow S1 / S2 / S3 \\
S1 &\rightarrow A\ A \\
S2 &\rightarrow A\ B\ A \\
S3 &\rightarrow a\ S3\ b / AA \\
A &\rightarrow a\ A\ b / \varepsilon \\
B &\rightarrow b\ B\ a / \varepsilon\}
\end{aligned}
$$

1. De quel type est cette grammaire ?
2. Donner le langage généré par cette grammaire (Justifier).

1.10. Soit G <X, {S, A, B, C, D}, P, S> la grammaire algébrique suivante :

$$
\begin{aligned}
P = \{ \quad S &\rightarrow A\ B / C\ D \\
A &\rightarrow 0\ A\ 1 / \varepsilon \\
B &\rightarrow 2\ B / \varepsilon \\
C &\rightarrow 0\ C / \varepsilon \\
D &\rightarrow 1\ D\ 2 / \varepsilon
\end{aligned}
$$

Trouver le langage engendré par G. (Justifier)

1.11. Soit le langage suivant $L = \{w / w \in \{0, 1\}^*$ et $|w|_0$ est pair$\}$

- Donner une grammaire G_1 de type 2 qui engendre L.
- Donner une grammaire G_2 régulière droite telle que $L(G_2)=L$

1.12. Soit G< X, V, S, P> où X = {a, b} , V = {S, A, B} et

$P= \{S \rightarrow aSb / aa\ S/S\ bb/ aa\ A/ \varepsilon\}$

- Les mots suivants appartiennent-ils au langage L(G) ?
  abab, aabb, aaaaab, aabbb
- Donner L le langage généré par la grammaire.
- Démontrer que L(G) = L.

1.13. Soit $L = \{0^n 1^n 0^m \; n, m \geq 1\}$

- Donner la grammaire engendrant L
- Donner les grammaires engendrant les langages suivants :
  - $L_1 = Init(L) = \{w \,/\, wx \in L)$
  - $L_2 = Fin(L) = \{w \,/\, xw \in L)$

EXERCICES EMD

1.14. Comparer les trois langages suivants :

- $L_1 = \{ww^R / \, w \in \{0, 1\}^*\}$
- $L_2 = \{(01)^i \, (10)^j \, (01)^k \, (10)^m \,/\, i, j, k, m \geq 0\}$
- $L_3 = \{(01)^i \, (10)^j \, (01)^j \, (10)^i \,/\, i, j \geq 0\}$

1.15. Soit G la grammaire suivante :

$S \rightarrow S_1 \,/\, S_2$	$S_2 \rightarrow B \; b \; B_2$
$S_1 \rightarrow A_1 \; A_3$	$B \rightarrow B_1 B_3 / \varepsilon$
$A_1 \rightarrow aA_1a / \, A_2 / \varepsilon$	$B_1 \rightarrow bB_1b / \, B_2 / \varepsilon$
$A_2 \rightarrow bA_2 \,/ \varepsilon$	$B_2 \rightarrow aB_2 \,/ \varepsilon$
$A_3 \rightarrow bA_3 \,/\, S_1 / \varepsilon$	$B_3 \rightarrow aB_3 \,/\, B \,/ \varepsilon$

1. A quelle classe appartient cette grammaire?
2. Donner le langage engendré par cette grammaire.
3. Montrer que L(G) = L
4. A quelle classe appartient ce langage ? Donner sa grammaire G'.

1.16. Soient les langages suivants :

$L_1 = \{d^n w \,/\, w \in \{a,b\}^*, n \geq 1 \text{ et } n+|w| \equiv 0[3]\}$
$L_2 = \{d^{2n} \, a^i \, b^{n+2} \,/\, n \geq 1 \text{ et } i \geq 0\}$

1. Donner une grammaire régulière qui génère L_1.
2. Donner une grammaire qui génère L_2.

Soit le langage $L_3 = L_1 \cap L_2$

3. Trouver le langage L_3.
4. Donner une grammaire qui le génère.

1.17. Donner la grammaire du langage L = $\{a^{i^2} > 0\}$.

Soient L_1, L_2 et L_3 les langages suivants

$L_1 = X^*aa$

$L_2 = b^*a(b^*a \cup a^+b^+a)^*a^+$

$L_3 = (b \cup ab \cup aa^+b)^*a \cup (b \cup ab \cup aa^+b)^*$

1. Donner une grammaire qui engendre L_1.
2. Comparer L_1 et L_2
3. Comparer L_1 et L_3
4. Donnez une grammaire qui engendre L_3

1.18. Soit la grammaire G<{a, b}, {S, A, B_1, B_2, C, D }, S, P> suivante

P = {

$S \rightarrow aA \,/\, bB_1$

$A \rightarrow aS \,/\, bB_2$

$B_1 \rightarrow bB_1 \,/\, aC$

$C \rightarrow aD$

$D \rightarrow aC \,/\, \varepsilon$

$B_2 \rightarrow bB_2 \,/\, aD$ }

1. Donner le langage engendré par cette grammaire (Justifier).
2. Donner la grammaire du langage $L = L(G) \cap \{w \in \{a, b\}^* \,/\, |w| \equiv 0[2]\}$

1.19. Donner les grammaires des langages suivants

$L_1 = \{0^n 0^{2n} 0^{3n} \mid n >= 0\}$.

$L_2 = \{0^n 1 0^{2n} 1 0^{3n} \mid n >= 0\}$

$L_3 = \{0^n 1^m 2^k \,/\, k > n + m$ et k n'est pas un multiple de 3$\}$