

Grammaires

Sommaire

1. Définitions
2. Dérivations
3. Classification des grammaires
4. Représentations des grammaires hors contexte
 1. Notation BNF
 2. Diagrammes de conway
 3. Arbres de dérivation

1-définitions

Une **grammaire formelle**, appelée aussi **grammaire de Chomsky**, G est une description de la forme des symboles et des phrases d'un langage noté $L(G)$. Elle est définie par un quadruplet $G = (V_T, V_N, S, R)$ où :

- V_T est un ensemble fini non vide : le **vocabulaire terminal** ou **alphabet** ;
 $L(G) \subseteq V_T^*$
- V_N est un ensemble fini non vide : le **vocabulaire non terminal** (on parle aussi de **variable** ou de **catégorie syntaxique**) ; On a $V_T \cap V_N = \emptyset$ et on note $V = V_T \cup V_N$ le **vocabulaire** de la grammaire ;
- $S \in V_N$ est un symbole non terminal particulier appelé **l'axiome** ;
- R est un ensemble de **règles**, appelées **règles de production** (ou **règles de réécriture**), de la forme $\alpha \rightarrow \beta$ avec $\alpha \in V^+$ et $\beta \in V^*$

Exemples

- $G_0 = (\{0, 1\}, \{S, X\}, S, R)$ avec R :
 - $R1 : S \rightarrow 0X1$
 - $R2 : 0X \rightarrow 00X1$
 - $R3 : X \rightarrow \varepsilon$
- $G_1 = (\{0, 1\}, \{S, X\}, S, R)$ avec :
 - $R = \{S \rightarrow 0X1 ; 0X \rightarrow 00X1 ; 0X \rightarrow 001\}$
 - ou $R = \{S \rightarrow 0X1 ; 0X \rightarrow 00X1 \mid 001\}$ (équivalent)
- $G_2 = (\{0, 1\}, \{S\}, S, R)$ avec :
 - $R = \{S \rightarrow 0S1 \mid 01\}$
- $G_3 = (\{0, 1\}, \{S, X\}, S, R)$ avec :
 - $R = \{S \rightarrow 0S \mid 0X ; X \rightarrow 1X \mid 1\}$

Dérivations

- Soit $G = (V_T, V_N, S, R)$ une grammaire.
- On appelle **dérivation de longueur n** ($n \in \mathbb{N}$) toute suite $s = u_0, \dots, u_n$ de chaînes de V^* telle que $u_0 =_G \Rightarrow u_1$ et ... et $u_{n-1} =_G \Rightarrow u_n$ (ce que l'on abrège par $u_0 =_G \Rightarrow \dots =_G \Rightarrow u_n$)
- $a =_n \Rightarrow b$ désigne une dérivation de longueur n .
- $a =_* \Rightarrow b$ désigne une dérivation de longueur quelconque. Autrement dit, $=_* \Rightarrow$ est la fermeture réflexive et transitive de $=_G \Rightarrow$.
- $a =_{+} \Rightarrow b$ désigne une dérivation de longueur supérieure à 1. Autrement dit, $=_{+} \Rightarrow$ est la fermeture transitive de $=_G \Rightarrow$.
- Si $a =_* \Rightarrow b$, on dit alors que a **engendre** b ou, symétriquement, que b **dérive de** a . Si de plus $a =_{+} \Rightarrow b$, on dit "engendre" (resp. "dérive") "de façon non triviale".

- Soit $G = (V_T, V_N, S, R)$ une grammaire :
- Une chaîne a est une forme sententielle (proto-phrase) pour la grammaire G si et seulement si $S \xRightarrow{*} a$.
- Le langage **engendré** par G est l'ensemble des chaînes de symboles terminaux que l'on peut dériver à partir de l'axiome S :
 $L(G) = \{x \in V_T^* \mid S \xRightarrow{+} x\}$.
- Deux grammaires G_1 et G_2 sont dites équivalentes si $L(G_1) = L(G_2)$.

4. Classification des grammaires

- **Type 0 : sans contraintes**

Les règles sont de la forme:

– - $a \rightarrow b$ avec $a \in V^+$ et $b \in V^*$.

- Type 1 (**contextuelle**, context sensitive, sensible au contexte): - $a \rightarrow b \mid |a| \leq |b|$.
- - $aAb \rightarrow aBb$ avec $a, b \in V^*$, $A \in V_N$ et $B \in V^+$ tel que $|B| \geq 1$ (c'est-à-dire que A peut être remplacé par B dans le contexte aAb).

l'axiome, s'il n'est pas en partie droite d'une autre règle que celle qui le définit, peut engendrer ε . Cette exception est aussi valable pour les types suivants

- **Type 2 : hors-contexte**

(context free,
algébrique)

– $A \rightarrow a$ avec $A \in V_N$ et $a \in V^+$.

- **Type 3 : linéaires**

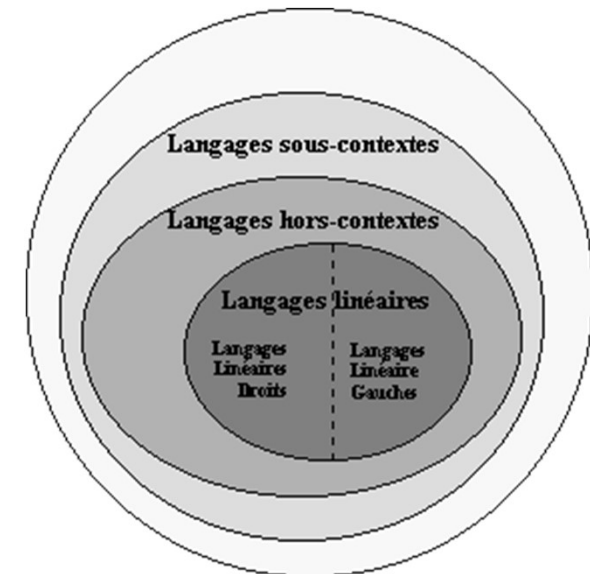
(régulières)

[à] droite (resp. [à] gauche):

– - uniquement : $A \rightarrow x_1 \mid x_2 B$ (resp. $A \rightarrow x_1 \mid B x_2$)
avec $A, B \in V_N$ et $x_1 \in V_T^+$, $x_2 \in V_T^*$.

Exemples

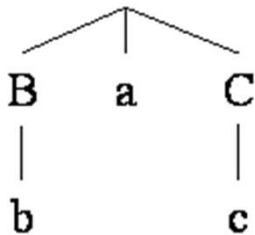
- règles $S \rightarrow OS \mid 1S \mid \varepsilon$ est linéaire droite.
- G_0 et G_4 sont de type 0,
- G_1 est de type 1, G_2 est de type 2 et G_3 est de type 3.



4-3 Arbres de dérivation

Tout mot du langage $L(G)$ peut être obtenu après un ensemble de dérivation à partir de l'axiome: $S \Rightarrow^+ a$ ($a \in L(G) \subseteq V_T^*$).

Une telle dérivation peut être aussi représentée par un arbre de dérivation (appelé aussi « arbre syntaxique » ou « arbre d'analyse »), exemple : Par exemple, en considérant les règles $R1 : A \rightarrow BaC$, $R2 : B \rightarrow b$ et $R3 : C \rightarrow c$, à la dérivation $A \Rightarrow^{R1} BaC \Rightarrow^{R2} baC \Rightarrow^{R3} bac$ correspond l'arbre suivant :



la représentation des dérivations sous forme d'arbres ne mémorise pas l'ordre dans lequel les règles sont appliquées. Par exemple, l'arbre ci-dessus correspond également à la dérivation $A \Rightarrow^{R1} BaC \Rightarrow^{R3} Bac \Rightarrow^{R2} bac$.

Remarques: La racine de l'arbre est l'axiome. Les feuilles sont les symboles terminaux de la phrase.

Une grammaire est dite ambiguë si au moins une phrase de son langage a plus d'un arbre de dérivation.

Exemple

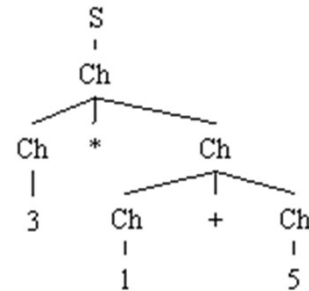
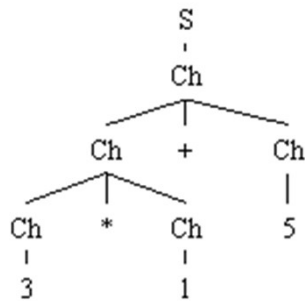
soit la grammaire $G_6 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, *\}, \{S, Ch\}, R$,
avec R :

$S \rightarrow Ch$

$Ch \rightarrow Ch + Ch \mid Ch * Ch$

$Ch \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

Par cette grammaire, nous pouvons obtenir pour le mot "3*1+5"
les arbres de dérivation suivants :

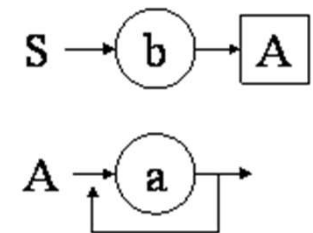


5- représentation des grammaires

- **5.1 norme BNF**
- Elle est particulièrement utile pour analyser un langage défini par une grammaire hors-contexte
- Les symboles terminaux sont encadrés par des guillemets ("..." ou <<...>>), les symboles non terminaux ne sont pas distingués
- L'axiome est la partie gauche de la première règle énoncée. La disjonction est notée par une barre verticale (|). La flèche des règles est remplacée par le signe "::=". Par exemple, le tableau suivant illustre la forme BNF pour les règles $\{S \rightarrow bA ; A \rightarrow aA ; A \rightarrow a\}$.
 - $S ::= B A$
 - $A ::= " a " A \mid A$

5.2 les diagrammes de conway

- Elle décrit graphiquement les grammaires hors-contexte
- La partie gauche des règles est notée à gauche du schéma. Les symboles terminaux sont écrits dans des cercles et les symboles non terminaux sont écrits dans des rectangles.



Exercices et tests :

Exercice 5.1. Pour chacune des grammaires suivantes :

1. Les écrire sous forme BNF

2. Donner les arbres de dérivation associés aux mots donnés à l'exercice 3.1. dans le cas où la grammaire est de type 2 ou 3. Les représenter à l'aide des diagrammes de Conway

Gex1 = ($\{a,b,c\}$, $\{S\}$, S , $\{S \rightarrow aSbSa \mid c\}$)

Gex2 = ($\{a,b,ch,d\}$, $\{S,A,B,C\}$, S , $\{S \rightarrow BCaCbbA ; A \rightarrow CaCbb \mid \varepsilon ; Ca \rightarrow ba ; Cbb \rightarrow da ; B \rightarrow cha\}$)

Gex3 = ($\{a,b\}$, $\{S,A\}$, S , $\{S \rightarrow Aa \mid bA ; A \rightarrow Sa \mid bS\}$)

Exercice 5.2. Soit la grammaire ($\{0,1,2,3,4,5,6,7,8,9,;, -, +\}$, $\{S, Exp, Ch\}$, S , R) définie par les règles suivantes :

• $S \rightarrow Exp ; S \mid Exp$

• $Exp \rightarrow Exp + Exp \mid Exp - Exp \mid Ch$

• $Ch \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

1. Écrire ces règles sous forme BNF

2. Donner les diagrammes de Conway correspondants

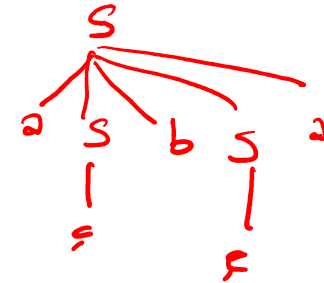
Solutions

Solution de l'exercice 5.1

$G_{ex1} = (\{a,b,c\}, \{S\}, S, \{S \rightarrow aSbSa \mid c\})$ (type 2)

$S ::= "a" S b S "a" \mid "c"$

le mot "acbca" donne l'arbre de dérivation suivant :



$G_{ex2} = (\{a,b,ch,d\}, \{S,A,B,C\}, S, \{S \rightarrow BCaCbbA ; A \rightarrow CaCbb \mid \varepsilon ;$
 $Ca \rightarrow ba ; Cbb \rightarrow da ; B \rightarrow cha\})$ (type 0)

$S ::= B C "a" C "b" "b" A$

$A ::= C "a" C "b" "b" \mid \varepsilon$

$C "a" ::= "b" "a"$

$C "b" "b" ::= "d" "a"$

$B ::= "ch" "a"$

$G_{ex3} = (\{a,b\}, \{S,A\}, S, \{S \rightarrow Aa \mid bA ; A \rightarrow Sa \mid bS\})$ (type 2)

$S ::= A "a" \mid "b" A$

$A ::= S "a" \mid "b" S$

Le langage étant vide, il n'y a pas de mots à dériver !

- $S \rightarrow \text{Exp} ; S \mid \text{Exp}$
- $\text{Exp} \rightarrow \text{Exp} + \text{Exp} \mid \text{Exp} - \text{Exp} \mid \text{Ch}$
- $\text{Ch} \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

1. Ecrire ces règles sous forme BNF

2. Donner les diagrammes de Conway correspondants

$S ::= \text{Exp} ";" S \mid \text{Exp}$

$\text{Exp} ::= \text{Exp} "+" \text{Exp} \mid \text{Exp} "-" \text{Exp} \mid \text{Ch}$

$\text{Ch} ::= "0" \mid "1" \mid "2" \mid "3" \mid "4" \mid "5" \mid "6" \mid "7" \mid "8" \mid "9"$

