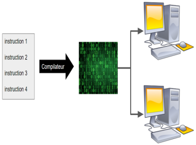


Grammaire & Langage

Université Assane Seck
UFR Sciences et technologie
Département Informatique

1. Introduction
2. La notion de grammaire
3. La notion de langage

Introduction

- Tout langage possède des règles qui indiquent la structure syntaxique
- La syntaxe d'un langage est décrite par une grammaire
- La grammaire décrit comment les unités lexicales doivent être agencées
- L'analyseur syntaxique reçoit une suite d'unités et vérifie si elle peut être engendrée par la grammaire
- Le problème est donc
  - Etant donné une grammaire G
  - Etant donné un mot m

  $\Longrightarrow$ Est-ce-que $m \in G$

Grammaire

Prenons par exemple le sous-ensemble suivant de la grammaire française :

- Le vocabulaire est défini par l'ensemble :
  T = { le, la, fille, jouet, regarde }
- Les catégories syntaxiques sont :
  - la phrase, notée PH
  - le groupe nominal, noté GN
  - le verbe, noté V
  - le déterminant, noté D
  - le nom, noté N

Grammaire

- Les règles permettant de combiner des éléments du vocabulaire et des catégories syntaxiques
  - PH $\longrightarrow$ GN V GN N $\longrightarrow$ fille
  - GN $\longrightarrow$ D N N $\longrightarrow$ jouet
  - D $\longrightarrow$ le V $\longrightarrow$ regarde
  - D $\longrightarrow$ la

  où le symbole $\longrightarrow$ est une abréviation de “peut être composé de”.

Grammaire

- La catégorie syntaxique de départ est la phrase PH
- La phrase “la fille regarde le jouet” est une phrase correcte pour la grammaire envisagée, comme le montre l’analyse suivante :

  PH $\Longrightarrow$ GN V GN $\Longrightarrow$ D N V GN $\Longrightarrow$ la N V GN $\Longrightarrow$ la fille V GN $\Longrightarrow$ la fille regarde GN $\Longrightarrow$ la fille regarde D N $\Longrightarrow$ la fille regarde le N $\Longrightarrow$ la fille regarde le jouet
- Le symbole $\Longrightarrow$ est une abréviation de “se dérive en”
- On distingue:
  - Les symboles terminaux: le vocabulaire du langage (la, fille, jouet, etc)
  - Les symboles non terminaux: les catégories syntaxiques

Grammaire

Définition 1: Une grammaire est un quadruplet G = (V_T, V_N, S, P) tel que

- V_T est un ensemble de symboles terminaux, appelé l'alphabet terminal
- V_N est un ensemble de symboles non terminaux, appelé l'alphabet ou le vocabulaire non terminal
- V_N et V_T sont disjoints
- P est un ensemble de règles dites de réécriture ou de production
- S $\in V_N$ est le symbole de départ ou axiome. C'est à partir de ce symbole non terminal que l'on commencera la génération de mots au moyen des règles de la grammaire.

Grammaire

Définition 2: Une règle de production $\alpha \longrightarrow \beta$ précise que la séquence de symboles $\alpha(\alpha \in V_T \bigcup V_N)^+$ peut être remplacée par la séquence de symboles $\beta(\beta \in V_T \bigcup V_N)^*$

- α est la partie gauche de la production
- β est la partie droite de la production

Grammaire

- Terminologie
  - une suite de symboles terminaux et non terminaux (un élément de $(V_N \bigcup V_T)^*$) est appelée une forme
  - une règle $u_1 \longrightarrow u$ telle que $u \in T^*$ est appelée une règle terminale
- Notation
  - $S \longrightarrow ab$, $S \longrightarrow aSb$, $S \longrightarrow c$ pourra s'écrire $S \longrightarrow ab|aSb|c$.

Quelques exemples

- Exemple 1
  - Symboles terminaux $V_T = \{a, b\}$
  - Symbole non terminaux $V_N = \{S\}$
  - Axiome: S
  - Règles de production
    $$\begin{cases} S & \longrightarrow & \varepsilon \\ S & \longrightarrow & aSb \end{cases}$$

    se résume en $S \longrightarrow \varepsilon | aSb$

Quelques exemples

Exemple 2

- G = (V_T, V_N, S, P)
- $V_T = \{il, elle, sympa, parle, est, devient, parle, court\}$
- $V_N = \{PHRASE, PRONOM, VERBE, COMPLEMENT, VERBETAT, VERBEACTION\}$
- $S = PHRASE$
- $P = \{PHRASE \longrightarrow PRONOM \quad VERBE\ COMPLEMENT$
- $PRONOM \longrightarrow il|elle$
- $VERBE \longrightarrow VERBETAT|VERBEACTION$
- $VERBETAT \longrightarrow est|devient|reste$
- $VERBEACTION \longrightarrow parle|court$
- $COMPLEMENT \longrightarrow sympa|vite\}$

Dérivation

Dérivation en une étape

Soient une grammaire G = (V_T, V_N, S, P), une forme non vide $u \in (V_T \bigcup V_N)^+$ et une forme éventuellement vide $v \in (V_T \bigcup V_N)^*$. La grammaire G permet de dériver v de u en une étape (noté $u \longrightarrow v$) si et seulement si :

- $u = xu^{'}y$ (u peut être décomposé en x, $u^{'}$ et y , x et y peuvent être vides),
- $v = xv^{'}y$ (v peut être décomposé en x,$u^{'}$ et y),
- $u^{'} \longrightarrow v^{'}$ est une règle de P.

Dérivation

Dérivation en plusieurs étapes

Une forme v peut être dérivée d'une forme u en plusieurs étapes :

- $v \xrightarrow{+} u$: si v peut être obtenue de u par une succession de 1 ou plusieurs dérivations en une étape,
- $v \xrightarrow{*} u$: si v peut être obtenue de u par une succession de 0, 1 ou plusieurs dérivations en une étape.

Langage

Langage

Le langage généré par une grammaire G = (V_T, V_N) est l'ensemble des mots sur V_T qui peuvent être dérivés à partir de S:
$L(G) = \{v \in V_T^* / v \xrightarrow{+} u\}$

Exercices

On considère la grammaire G = (V_T, V_N, Ph, P)

- $V_T = \{un, une, le, la, enfant, garcon, fille, cerise, haricot, cueille, mange\}$
- $V_N = \{Ph, Gn, Gv, Df, Dm, Nf, Nm, V\}$
- $S = Ph$
- $P = \{PHRASE \longrightarrow Gn \quad Gv$
- $Gv \longrightarrow Df\ Nf | Dm\ Nm$
- $Gv \longrightarrow V\ Gn$
- $Df \longrightarrow une | la$
- $Dm \longrightarrow un | le$
- $Nf \longrightarrow fille | cerise$
- $Nm \longrightarrow enfant | garcon | haricot$
- $V \longrightarrow cueille | mange\}$

La phrase “une cerise cueille un enfant” appartient-elle au langage L(G) ?

Exercices

On considère la grammaire G = (V_T, V_N, S, P)

- $V_T = \{b, c\}$
- $V_N = \{S\}$
- $S \longrightarrow bS|cc$

Déterminez $L(G)$.

Correction : $L(G) = \{b^n cc\}$

En effet, partant de l'axiome S, toute dérivation commencera nécessairement par appliquer 0, 1 ou plusieurs fois la première règle puis se terminera en appliquant la deuxième règle. On représentera cela en écrivant le schéma de dérivation suivant:

$S \xrightarrow{nfois(1)} b^n S \xrightarrow{1fois} b^n cc$

Exercices

On considère la grammaire G = (V_T, V_N, S, P)

- $V_T = \{b, c\}$
- $V_N = \{S\}$
- $S \longrightarrow bS|cc$

Déterminez $L(G)$.
Correction : $L(G) = \{b^n cc\}$
En effet, partant de l'axiome S, toute dérivation commencera nécessairement par appliquer 0, 1 ou plusieurs fois la première règle puis se terminera en appliquant la deuxième règle. On représentera cela en écrivant le schéma de dérivation suivant:
$S \xrightarrow{nfois(1)} b^n S \xrightarrow{1fois} b^n cc$

Exercices

On considère la grammaire G = (V_T, V_N, S, P)

- $V_T = \{0, 1\}$
- $V_N = \{S\}$
- $P = S \longrightarrow 0S|1S|0$

Déterminez $L(G)$.

Correction : $L(G) = \{u0/u \in \{0, 1\}^*$

En effet, partant de l'axiome S, toute dérivation commencera nécessairement par appliquer 0, 1 ou plusieurs fois la première ou la deuxième règle puis se terminera en appliquant la troisième règle. On représentera cela en écrivant le schéma de dérivation suivant :

$S \xrightarrow{nfois(1ou2)} u^nS \xrightarrow{1fois} u0$ avec $u \in \{0, 1\}^*$

Exercices

On considère la grammaire G = (V_T, V_N, S, P)

- $V_T = \{0, 1\}$
- $V_N = \{S\}$
- $P = S \longrightarrow 0S|1S|0$

Déterminez $L(G)$.
Correction : $L(G) = \{u0/u \in \{0, 1\}^*$
En effet, partant de l'axiome S, toute dérivation commencera nécessairement par appliquer 0, 1 ou plusieurs fois la première ou la deuxième règle puis se terminera en appliquant la troisième règle. On représentera cela en écrivant le schéma de dérivation suivant :
$S \xrightarrow{nfois(1ou2)} u^n S \xrightarrow{1fois} u0$ avec $u \in \{0, 1\}^*$

Exercices

Construire une grammaire pour le langage $L(G) = \{ab^na/n \in N\}$

Correction: on définit la grammaire suivante G = (V_T, V_N, S, P):

- $V_T = \{a, b\}$
- $V_N = \{S, U\}$
- $P = \{S \longrightarrow aUa$
  $U \longrightarrow bU|\varepsilon\}$

Exercices

Construire une grammaire pour le langage $L(G) = \{ab^n a / n \in N\}$
Correction: on définit la grammaire suivante G = (V_T, V_N, S, P):

- $V_T = \{a, b\}$
- $V_N = \{S, U\}$
- $P = \{S \longrightarrow aUa$
  $U \longrightarrow bU|\varepsilon\}$

Exercices

Construire une grammaire pour le langage $L(G) = \{0^{2n}1^n / n \geq 0\}$

Correction: on définit la grammaire suivante G = (V_T, V_N, S, P):

- $V_T = \{0, 1\}$
- $V_N = \{S\}$
- $P = \{S \longrightarrow 00S1|\varepsilon\}$

Exercices

Construire une grammaire pour le langage $L(G) = \{0^{2n}1^n / n \geq 0\}$
Correction: on définit la grammaire suivante G = (V_T, V_N, S, P):

- $V_T = \{0, 1\}$
- $V_N = \{S\}$
- $P = \{S \longrightarrow 00S1|\varepsilon\}$

Arbre de dérivation

On appelle arbre de dérivation (ou arbre syntaxique) tout arbre tel que:

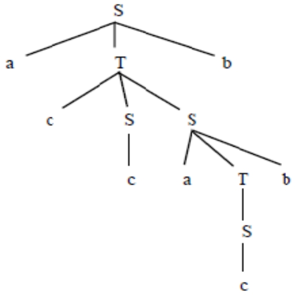
- La racine est l'axiome
- Les feuilles sont des symboles terminaux
- Les nœuds sont des symboles non terminaux
- Les fils d'un nœud X sont $\beta_0, .., \beta_n$ si et seulement si $X \longrightarrow \beta_0, .., \beta_n$ est une production (avec $\beta_i \in V_T \cup V_N$)

Arbre de dérivation

Soit la grammaire ayant pour axiome S et pour règle de production:

$$\begin{cases} S & \longrightarrow & aTb|c \\ T & \longrightarrow & cSS|S \end{cases}$$

Un arbre syntaxique pour le mot accacbb est :



$S \Longrightarrow aTb \Longrightarrow acSSb \Longrightarrow accSb \Longrightarrow accaTb \Longrightarrow accaSb \Longrightarrow accacbb$ (dérivations gauches)
$S \Longrightarrow aTb \Longrightarrow acSSb \Longrightarrow acsaTbb \Longrightarrow acSaSbb \Longrightarrow accaScbbb \Longrightarrow accacbb$(dérivations droites)

Exercices

Soit la grammaire suivante G = (V_T, V_N, S, P):

- $V_T = \{a, b\}$
- $V_N = \{S, U\}$
- $P = \{S \longrightarrow aS|aU$
  $U \longrightarrow bU|b\}$

Donnez l'arbre syntaxique pour chacun des mots suivant: aaabbb, aaabb

Exercices

Soit la grammaire suivante G = (V_T, V_N, S, P):

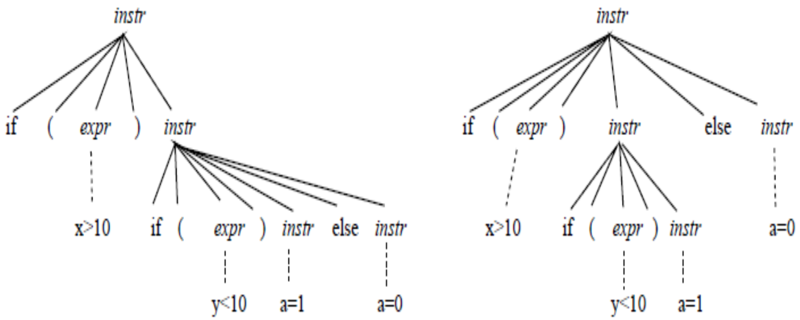
- $V_T = \{a, b, c\}$
- $V_N = \{S, U\}$
- $P = \{S \longrightarrow UbacU$
  $U \longrightarrow aU|bU|cU|\varepsilon\}$

Donnez l'arbre syntaxique pour le mot abccbaccabc

Grammaire ambiguë

- Une grammaire est **ambiguë** s'il existe un mot de $L(G)$ ayant plusieurs arbres syntaxiques.
- Exemple: Soit G la grammaire donnée par
  - instr $\longrightarrow$ if (expr) instr else instr
  - instr $\longrightarrow$ if (expr)
  - instr $\longrightarrow$ ...
  - expr $\longrightarrow$ ...
- Montrez que cette grammaire est ambiguë avec le mot ***m = if (x>10) if (y<0) a=1 else a=0***

Exercices


instr
if
(
expr
)
instr
x>10
if
(
expr
)
instr
else
instr
y<10
a=1
a=0
instr
if
(
expr
)
instr
else
instr
x>10
if
(
expr
)
instr
a=0
y<10
a=1

Exercices

Soit la grammaire suivante G = (V_T, V_N, S, P):

- $V_T = \{id, +, -, *, /, (,)\}$ où Id signifie identificateur
- $V_N = \{E, Op\}$
- $P = \{E \longrightarrow Id \;\mid\; (E) \;\mid\; E \;\; Op \;\; E$
  $Op \longrightarrow +| - | * |/\}$

Montrez que cette grammaire est ambiguë à travers le mot "Id + Id * Id".