Institut International de Technologie

Cours Théorie des langages & automates

Les grammaires

IIT
INSTITUT INTERNATIONAL TECHNOLOGIE
Université Nord Américaine privée

Soufiene.Lajmi@gmail.com

Année universitaire 2020-2021

Plan

- **Introduction et définitions**
- **Dérivation**
- **Mot et langage générés par une grammaire**
- **Classification de Chomski**

Introduction

- Une grammaire est une notation pour la description de langages

- Idée de base:

  - utiliser des variables pour représenter des ensembles de mots (i.e., langages).
  - Ces variables sont définies récursivement les uns par les autres.
  - Les règles récursives (“productions”) ne permettent que la concaténation.
  - Les règles alternatives pour une variable permettent l’union.

Grammaire (définition)

- Une grammaire de réécriture est un 4-uplet

$$\mathbf{G = <V_N, V_T, S, P>}$$

- V_N : vocabulaire non terminal
- V_T : vocabulaire terminal
- $S \in V_N$: axiome ou symbole initial; racine de la grammaire; non terminal
- P : règles de production (ou de réécriture)
- $V_N \cap V_T = \emptyset$

Règle

- Une règle est un couple (φ, ψ) qu'on note en général :

$$\varphi \rightarrow \psi$$

où : $\varphi \in (V_N \cup V_T)^* .V_N.(V_N \cup V_T)^*$
et $\psi \in (V_N \cup V_T)^*$

- on dit que *φ se réécrit ψ*

- φ est appelé partie gauche de la règle
- ψ est appelé partie droite de la règle

Exemple de grammaire

- $V_N = \{S, A, B\}$
- $V_T = \{0, 1\}$
- $S \in V_N$: axiome
- P :

$$S \rightarrow 0A1B$$
$$1B \rightarrow 1ABB$$
$$1A \rightarrow A1$$
$$1B \rightarrow 11$$
$$0A \rightarrow 00$$

Écriture simplifiée

- Pour simplifier l'écriture, l'ensemble des règles suivantes:

$$\varphi \rightarrow \psi_1,\ \varphi \rightarrow \psi_2,\ \ldots,\ \varphi \rightarrow \psi_n$$

- Sera noté par :

$$\varphi \rightarrow \psi_1 | \psi_2 | \ \ldots | \psi_n$$

Dérivation immédiate

- $\varphi \Rightarrow \psi$ est dite *Dérivation immédiate* si et seulement si:
  - $\varphi = xuy$, $u \neq \varepsilon$
  - $\psi = xvy$
  - $u \rightarrow v \in P$

Dérivation à l'ordre k

- On dit que φ se dérive à l'ordre k en ψ et on note:

$$\varphi =>^{k} \psi$$

- Ssi il existe $\gamma \in (V_N \cup V_T)^*$ tel que:

$$\varphi =>^{k-1} \gamma \text{ et } \gamma => \psi$$

Dérivation

- On dit que φ se dérive en ψ et on note:

$$\varphi \Rightarrow^* \psi$$

- Ssi il existe $k \in \mathbb{N}$ tel que: $\varphi \Rightarrow^k \psi$

Mot généré par une grammaire

- On dit que $w \in V_T^*$ est engendré (généré) par une grammaire si on peut l'obtenir au moyen de réécritures à partir de l'axiome.
- Un mot appartenant à $(V_N \cup V_T)^*$ et dérivé de l'axiome est appelé une proto-phrase.
- ψ est une proto-phrase ssi $S \Rightarrow^* \psi$.
- Un mot généré par G est une proto-phrase de G ne contenant aucun symbole non terminal.

Langage généré par une grammaire

- Le langage généré par G, noté L(G) est l'ensemble des mots générés par G.

$$L(G) = \{w \in V_T^*;\ S \Rightarrow^* w\}$$

- Deux grammaires G1 et G2 sont équivalentes ssi

$$L(G1) = L(G2).$$

Exemples de langages générés par G

G = ({a, b}, {S, A} , S, P),

- P={S→bbA; A →aaA; A →aa}

G = ({0, 1}, {S} , S, P) :

- P={S→1S; S →0S; S →1}

G = ({a, b}, {S} , S, P) :

- P={S→aSb; S → ε}

Sens de la dérivation

- Une dérivation dans laquelle on réécrit toujours le non terminal le plus à gauche est appelée "dérivation la plus à gauche".

- De même, une dérivation dans laquelle on réécrit toujours le non terminal le plus à droite est dite "dérivation la plus à droite".

Sens de la dérivation

- Exemple :

  - Soit la grammaire G sur {a, b, c } dont les productions sont :
    - $S \rightarrow aBc \mid aCb \mid aAc$
    - $A \rightarrow BC$
    - $B \rightarrow aS \mid cS \mid \varepsilon$
    - $C \rightarrow aB \mid bB \mid \varepsilon$

Sens de la dérivation

- Dérivation la plus à gauche du mot w = acabbc :
  - $S \Rightarrow aBc$ $(S \rightarrow aBc)$
  - $aBc \Rightarrow acSc$ $(B \rightarrow cS)$
  - $acSc \Rightarrow acaCbc$ $(S \rightarrow aCb)$
  - $acaCbc \Rightarrow acabBbc$ $(C \rightarrow bB)$
  - $acabBbc \Rightarrow acabbc$ $(B \rightarrow \varepsilon)$

Sens de la dérivation

- Dérivation la plus à droite du mot w = acabbc :
  - $S \Rightarrow aAc$ $(S \rightarrow aBc)$
  - $aAc \Rightarrow aBCc$ $(A \rightarrow BC)$
  - $aBCc \Rightarrow aBbBc$ $(C \rightarrow bB)$
  - $aBbBc \Rightarrow aBbc$ $(B \rightarrow \varepsilon)$
  - $aBbc \Rightarrow acSbc$ $(B \rightarrow cS)$
  - $acSbc \Rightarrow acaCbbc$ $(S \rightarrow aCb)$
  - $acaCbbc \Rightarrow acabbc$ $(C \rightarrow \varepsilon)$

Classification de Chomsky

- Chomsky a identifié quatre classes de grammaires:
  - Grammaire de type 0
  - Grammaire de type 1
  - Grammaire de type 2
  - Grammaire de type 3

Grammaire de type 0 (sans restriction)

- Une grammaire est dite sans restriction ou de type 0 ssi:
  - Il n'y a aucune restriction sur les règles de production
- Rappel:
  - Une règle est un couple (φ, ψ) qu'on note en général :

$$\varphi \rightarrow \psi$$

où : $\varphi \in (V_N \cup V_T)^* . V_N . (V_N \cup V_T)^*$

et $\psi \in (V_N \cup V_T)^*$

Grammaire de type 1 (contextuelle)

- Une grammaire est dite contextuelle si et seulement si:
- Pour toute règle $\varphi \rightarrow \psi$ on a:

  - φ=gAd et
  - ψ=gBd
  - avec g, d, $B \in (V_N \cup V_T)^*$ et $A \in V_N$.

Grammaire de type 2 (hors-contexte)

Une grammaire hors-contexte (ou algébrique) $G = (V_T, V_N, P, S)$ est une grammaire dont les productions sont de la forme $\alpha \rightarrow \beta$, où :

1. $|\alpha| = 1$ et $\alpha \in V_N$
2. β est une séquence de terminaux ou de non terminaux

Le langage L(G) généré par une grammaire hors-contexte G est appelé *langage hors-contexte*.

G1 = ({a,b}, {S}, P1, S) avec P1 = {$S \rightarrow aSb$, $S \rightarrow ab$}

Grammaire de type 3 (régulière)

Une Grammaire peut être_régulière à droite ou à gauche

- Grammaire Régulière à gauche:

Toutes les productions $\alpha \to \beta$ sont telles que:

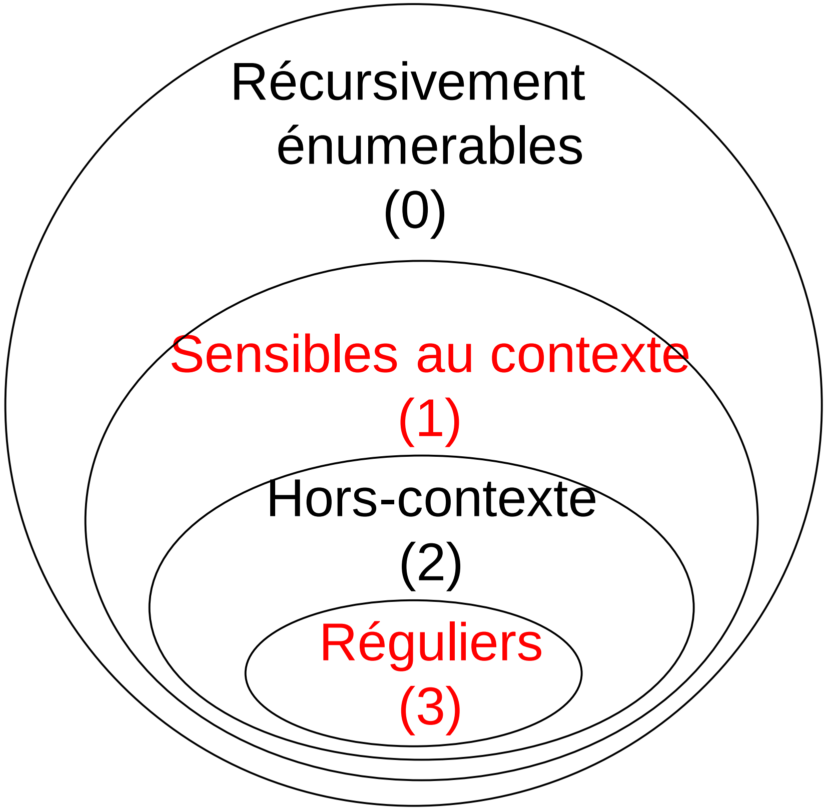
1. $|\alpha| = 1$,
2. $\beta = aB$, $\beta = a$ ou $\beta = \varepsilon$ avec $B \in V_N$ et $a \in V_T$

- Grammaire Régulière à Droite :

1. $|\alpha| = 1$,
2. $\beta = Ba$, $\beta = a$ ou $\beta = \varepsilon$ avec $B \in V_N$ et $a \in V_T$

Le langage L(G) généré par une grammaire régulière G est appelé *langage régulier*

Hiérarchie de Chomski



Conclusion

Questions ??