

# Grammaires non-contextuelles

Quentin Fortier

October 1, 2024

# Définitions : Grammaire non-contextuelle

## Définition : Grammaire non-contextuelle

Une grammaire non-contextuelle (ou : hors-contexte) est un quadruplet  $G = (\Sigma, V, R, S)$  où :

- $V$  est un ensemble fini de variables
- $\Sigma$  est un alphabet fini de terminaux
- $R \subset V \times (V \cup \Sigma)^*$  est un ensemble fini de règles de production, chaque règle  $(X, \alpha) \in R$  étant notée  $X \rightarrow \alpha$
- $S \in V$  est le symbole initial

## Remarques :

- Par convention, on note les variables en majuscules et les terminaux en minuscules.
- On peut noter  $X \rightarrow \alpha \mid \beta$  au lieu de  $X \rightarrow \alpha, X \rightarrow \beta$ .

# Définitions : Dérivation

Soit  $G = (\Sigma, V, R, S)$  une grammaire non-contextuelle.

## Définition : Dérivation

Soient  $\alpha, \beta \in (V \cup \Sigma)^*$ .

- On note  $\alpha \Rightarrow \beta$  s'il existe une règle  $X \rightarrow \gamma$  telle que  $\alpha = \alpha_1 X \alpha_2$  et  $\beta = \alpha_1 \gamma \alpha_2$  avec  $\alpha_1, \alpha_2 \in (V \cup \Sigma)^*$ .

On dit alors qu'on a une dérivation immédiate de  $\alpha$  en  $\beta$ .

- On note  $\alpha \Rightarrow^n \beta$  s'il existe des mots  $\gamma_0 = \alpha, \gamma_1, \dots, \gamma_n = \beta$  tels que  $\gamma_0 \Rightarrow \gamma_1 \Rightarrow \dots \Rightarrow \gamma_n$ .

On dit alors qu'on a une dérivation de longueur  $n$  de  $\alpha$  en  $\beta$ .

- On note  $\alpha \Rightarrow^* \beta$  si  $\alpha \Rightarrow^n \beta$  pour un certain  $n \in \mathbb{N}$ . On parle alors de dérivation de  $\alpha$  en  $\beta$ .

# Définitions : Langage engendré

Soit  $G = (\Sigma, V, R, S)$  une grammaire non-contextuelle.

## Définition : Mot généré

On dit que  $G$  génère un mot  $w \in \Sigma^*$  si  $S \Rightarrow^* w$ .

Attention : les mots générés sont des mots de  $\Sigma^*$ , qui ne doivent donc contenir que des symboles terminaux.

## Définition : Langage engendré

L'ensemble  $L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$  est le langage engendré par  $G$ .

## Définition : Langage non-contextuel

Un langage  $L$  est dit non-contextuel (ou : algébrique) s'il existe une grammaire non-contextuelle  $G$  telle que  $L = L(G)$ .

Remarque : deux grammaires différentes peuvent engendrer le même langage.

# Définitions : Langage engendré

## Exemples :

- Soit  $G = (\Sigma = \{a, b\}, V = \{S\}, R = \{S \rightarrow aaS \mid b\}, S)$ .  
 $G$  génère  $aab$  car  $S \Rightarrow aaS \Rightarrow aab$ .  
 $L(G) =$

# Définitions : Langage engendré

## Exemples :

- Soit  $G = (\Sigma = \{a, b\}, V = \{S\}, R = \{S \rightarrow aaS \mid b\}, S)$ .  
 $G$  génère  $aab$  car  $S \Rightarrow aaS \Rightarrow aab$ .  
 $L(G) = (aa)^*b$ .
- Soit  $G = (\{a, b\}, \{S\}, \{S \rightarrow aSb \mid \varepsilon\}, S)$ .  
 $L(G) =$

# Définitions : Langage engendré

## Exemples :

- Soit  $G = (\Sigma = \{a, b\}, V = \{S\}, R = \{S \rightarrow aaS \mid b\}, S)$ .  
 $G$  génère  $aab$  car  $S \Rightarrow aaS \Rightarrow aab$ .  
 $L(G) = (aa)^*b$ .
- Soit  $G = (\{a, b\}, \{S\}, \{S \rightarrow aSb \mid \varepsilon\}, S)$ .  
 $L(G) = \{a^n b^n \mid n \in \mathbb{N}\}$ .
- Soit  $G = (\{x, y, \top, \perp, \vee, \wedge, \neg\}, \{S\}, R, S)$  avec  $P$  contenant les règles suivantes :  $S \rightarrow \top \mid \perp \mid x \mid y \mid \neg S \mid S \vee S \mid S \wedge S$ .  
 $L(G)$  est

# Définitions : Langage engendré

## Exemples :

- Soit  $G = (\Sigma = \{a, b\}, V = \{S\}, R = \{S \rightarrow aaS \mid b\}, S)$ .  
 $G$  génère  $aab$  car  $S \Rightarrow aaS \Rightarrow aab$ .  
 $L(G) = (aa)^*b$ .
- Soit  $G = (\{a, b\}, \{S\}, \{S \rightarrow aSb \mid \varepsilon\}, S)$ .  
 $L(G) = \{a^n b^n \mid n \in \mathbb{N}\}$ .
- Soit  $G = (\{x, y, \top, \perp, \vee, \wedge, \neg\}, \{S\}, R, S)$  avec  $P$  contenant les règles suivantes :  $S \rightarrow \top \mid \perp \mid x \mid y \mid \neg S \mid S \vee S \mid S \wedge S$ .  
 $L(G)$  est l'ensemble des formules logiques bien formées, avec  $x$  et  $y$  comme variables propositionnelles.



# Définitions : Langage engendré

On peut décrire un langage de programmation à l'aide d'une grammaire (BNF) : [en C](#) et [en OCaml](#).

# Définitions : Langage engendré

## Exercice

Donner des grammaires engendrant les langages suivants sur  $\{a, b\}$  :

- ①  $L_1 = ab^*a$ .
- ②  $L_2 =$  ensemble des mots dont la taille est un multiple de 3.
- ③  $L_3 =$  ensemble des mots ayant  $bbb$  comme facteur.
- ④  $L_4 =$  ensemble des expressions arithmétiques bien formées, comme  $4 + 3 \times 2$ .
- ⑤  $L_5 =$  ensembles des palindromes, c'est-à-dire des mots qui se lisent de la même façon de gauche à droite et de droite à gauche.
- ⑥  $L_6 =$  ensembles des mots qui ne sont pas des palindromes.

# Définitions : Langage engendré

On peut montrer rigoureusement  $L(G) = L$  par double inclusion.

- $L(G) \subset L$  : montrer que si  $S \Rightarrow^n u$  alors  $u \in L$ , par récurrence sur  $n$ .
- $L \subset L(G)$  : montrer que si  $u \in L$  alors  $u \in L(G)$ , par récurrence sur  $|u|$ .

# Définitions : Langage engendré

On utilise alors souvent le théorème « évident » suivant :

## Théorème

Soit  $G = (\Sigma, V, R, S)$  une grammaire,  $\alpha_1, \alpha_2, \beta \in (V \cup \Sigma)^*$  et  $n \in \mathbb{N}$ .

Si  $\alpha_1 \alpha_2 \Rightarrow^n \beta$  alors il existe  $\beta_1, \beta_2 \in (V \cup \Sigma)^*$ ,  $k, p \in \mathbb{N}$  tels que :

- $\alpha_1 \Rightarrow^k \beta_1$
- $\alpha_2 \Rightarrow^p \beta_2$
- $n = k + p$

## Définitions : Langage engendré

### Exercice

Soit  $G$  la grammaire définie par les règles  $S \rightarrow aSbS \mid bSaS \mid \varepsilon$ .  
Déterminer  $L(G)$ , en le démontrant.

## Théorème

L'ensemble des langages non-contextuels est stable par union, concaténation et étoile.

C'est-à-dire : si  $L_1$  et  $L_2$  sont des langages non-contextuels alors  $L_1 \cup L_2$ ,  $L_1 L_2$  et  $L_1^*$  sont des langages non-contextuels.

## Théorème

Tout langage régulier est non-contextuel.

## Théorème

Tout langage régulier est non-contextuel.

Remarque : la réciproque est fausse car  $\{a^n b^n \mid n \in \mathbb{N}\}$  est non-contextuel mais n'est pas régulier.



## Définition : Grammaire régulière (HP)

Une grammaire est dite régulière (à droite) si chaque règle est de la forme  $X \rightarrow aY$ ,  $X \rightarrow a$  ou  $X \rightarrow \varepsilon$ .

## Exercice

Donner des grammaires régulières engendrant les langages suivants sur  $\{a, b\}$  :

- 1 Ensemble des mots finissant par  $a$ .
- 2 Ensemble des mots de taille un multiple de 3.
- 3  $a^*ba^*$ .

## Théorème

Un langage est régulier si et seulement s'il est engendré par une grammaire régulière.