

INF 302 : LANGAGES & AUTOMATES

Chapitre 5 : Grammaires et grammaires régulières

Yliès Falcone

ylies.falcone@univ-grenoble-alpes.fr

www.ylies.fr

Univ. Grenoble-Alpes, Inria

Laboratoire d'Informatique de Grenoble - www.liglab.fr

Équipe de recherche LIG-Inria, CORSE - team.inria.fr/corse/

Année Académique 2017 - 2018

Plan Chap. 5 - Grammaires et grammaires régulières

- 1 Introduction et intuition
- 2 Grammaires : définition, dérivation et langage engendré
- 3 Grammaires régulières et automates à états finis
- 4 Grammaires et expressions régulières
- 5 Résumé

À propos des grammaires

Description informelle

Les grammaires permettent de décrire des langages de manière *générative*, en décrivant de *manière compacte* les différentes manières de générer *un mot*.

Les grammaires utilisent des *règles de production* permettant de décrire les transformations d'une partie d'un mot en cours de génération :

$$\text{Source} \rightarrow \text{Résultat}$$

À partir de sources, nous produisons des résultats (dont certains peuvent être ensuite utilisés comme sources).

Quand une source produit plusieurs résultats possibles, nous notons :

$$\text{Source} \rightarrow \text{Résultat-1} \mid \text{Résultat-2}$$

au lieu de

$$\text{Source} \rightarrow \text{Résultat-1}$$

$$\text{Source} \rightarrow \text{Résultat-2}$$

On parle de forme de Backus-Naur.

À propos des grammaires

Description informelle (suite)

Une grammaire définit donc des règles de production sous la forme :

$$\text{Source-1} \rightarrow \text{Résultat1-1} \mid \text{Résultat2-1} \mid \dots$$

$$\text{Source-2} \rightarrow \text{Résultat1-2} \mid \text{Résultat2-2} \mid \dots$$

$$\dots$$

$$\text{Source-n} \rightarrow \text{Résultat1-n} \mid \text{Résultat2-n} \mid \dots$$

où les résultats peuvent "contenir" des sources.

Une grammaire a une source unique au départ pour générer des mots (qui peuvent générer plusieurs sources), appelée *axiome*.

Le processus partant d'une source jusqu'à un résultat est appelé *dérivation*.

Le processus partant de l'axiome jusqu'à un résultat qui ne peut produire de nouvelles sources (cad un mot) est appelé *dérivation complète*.

Plan Chap. 5 - Grammaires et grammaires régulières

- 1 Introduction et intuition
- 2 Grammaires : définition, dérivation et langage engendré
- 3 Grammaires régulières et automates à états finis
- 4 Grammaires et expressions régulières
- 5 Résumé

Grammaire

Définition

Définition (Grammaire)

Une **grammaire** est un quadruplet (V_T, V_N, Z, P) tel que :

- V_T est un ensemble fini qui dénote l'ensemble des **symboles terminaux**,
- V_N est un ensemble fini qui dénote l'ensemble des **symboles non-terminaux**,
- $Z \in V_N$ l'**axiome** et
- $P \subseteq (V_T \cup V_N)^* \times (V_T \cup V_N)^*$ est l'ensemble des **règles de production**.

tels que V_T et V_N sont disjoints ($V_T \cap V_N = \emptyset$),

$V = V_T \cup V_N$ est le **vocabulaire** de la grammaire.

On note $x \rightarrow y$ au lieu de $(x, y) \in P$.

Exemple (Grammaire)

- $G_1 = (\{a, b\}, \{S\}, S, P)$ avec $P = \{S \rightarrow aSb, S \rightarrow \epsilon\}$.
- $G_2 = (\{a, b\}, \{S, A, B\}, S, P)$ avec $P = \{S \rightarrow ASB \mid \epsilon, A \rightarrow aA \mid a, B \rightarrow Bb \mid b\}$.

Dérivations en une et plusieurs étapes

Soient $G = (V_T, V_N, Z, P)$ une grammaire et $V = V_T \cup V_N$ son vocabulaire.

Définition (Relation de dérivation en une étape)

G permet de **dériver** un mot $u \in V^*$ en un mot $v \in V^*$ en une étape, noté $u \Rightarrow v$, s'il existe un règle $x \rightarrow y \in P$ telle que :

- $u = \alpha \cdot x \cdot \beta$,
- $v = \alpha \cdot y \cdot \beta$,

avec $\alpha, \beta \in V^*$. Les mots u et v sont appelés *formes sententielles*.

Exemple (Dérivations en une étape pour la grammaire G_2)

- $S \Rightarrow \epsilon$
- $S \Rightarrow ASB$
- $ASB \Rightarrow AB$
- $aASBBbbb \Rightarrow aaASBbbbb$

Définition (Relation de dérivation en plusieurs étapes)

La relation de dérivation en plusieurs étapes, notée \Rightarrow^* , est la *fermeture transitive* de \Rightarrow .

Exemple (Dérivation en plusieurs étapes)

- $S \Rightarrow^* \epsilon$
- $S \Rightarrow^* aaASBbbbb$
- $aaASBbbbb \Rightarrow^*$
- $S \Rightarrow^* ab$
- $S \Rightarrow^* aabb$
- $aaabbbbb$

Langage généré par une grammaire

Soient $G = (V_T, V_N, Z, P)$ une grammaire et $V = V_T \cup V_N$ son vocabulaire.

Définition (Langage engendré par une grammaire)

$$L(G) = \{w \in V_T^* \mid Z \Rightarrow^* w\}$$

Exemple (Grammaire)

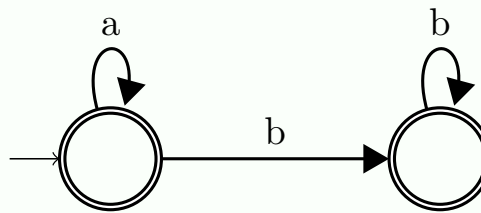
- Le langage généré par G_1 est $\{a^i b^i \mid i \in \mathbb{N}\}$.
- Le langage généré par G_2 est $\{a^i b^j \mid i, j \in \mathbb{N}\}$, aussi décrit par l'expression régulière $a^* b^*$.
- Le langage généré par $(\{a, b\}, \{S, A, B\}, S, P)$ avec $P = \{S \rightarrow AB, A \rightarrow aA \mid a, B \rightarrow Bb \mid b\}$ est $\{a^i b^j \mid i, j \in \mathbb{N} \setminus \{0\}\}$, aussi décrit par l'expression régulière $a^+ b^+$.

Grammaire

Lien avec les automates

L'ensemble des non-terminaux dans les grammaires « correspond » à l'alphabet des automates :

- Un automate reconnaît des mots sur son alphabet.
- Une grammaire génère des mots sur son ensemble de symboles terminaux.
- La grammaire $G_2 = (\{a, b\}, \{S, A, B\}, S, P)$ avec $P = \{S \rightarrow ASB \mid \epsilon, A \rightarrow aA \mid \epsilon, B \rightarrow Bb \mid \epsilon\}$ génère le langage $\{a^i b^j \mid i, j \in \mathbb{N}\}$, cad des mots sur l'ensemble de non-terminaux $\{a, b\}$.
- L'automate



reconnaît les mêmes mots sur $\{a, b\}$.

Classification de Chomsky des grammaires

Les contraintes sur la forme de la grammaire définissent son *type*.

Type 0, Grammaires *générales* :

$$x \rightarrow y, \text{ avec } x \in V^* \cdot V_N \cdot V^*, y \in V^*$$

Type 1, Grammaires *contextuelles* :

$$\alpha \cdot A \cdot \beta \rightarrow \alpha \cdot \gamma \cdot \beta, \text{ avec } \alpha, \beta, \gamma \in V^*, \gamma \neq \epsilon, A \in V_N$$

Type 2, Grammaires *non-contextuelles* : (langages hors-contextes, algébriques) $A \rightarrow \gamma$, avec $A \in V_N, \gamma \in V^*$

Type 3, Grammaires *régulières* : (langages rationnels)

- linéaires gauches, $A \rightarrow B \cdot a \mid A \rightarrow a \mid A \rightarrow \epsilon$
- linéaires droites, $A \rightarrow a \cdot B \mid A \rightarrow a \mid A \rightarrow \epsilon$

$$\text{avec } A, B \in V_N, a \in V_T$$

Cas des langages de programmation

- Les grammaires régulières sont utilisées pour le lexique (analyse lexicale).
- Les grammaires non-contextuelles pour décrire les (syntaxiquement) programmes corrects.

Exemples de grammaires non-contextuelles

Les grammaires suivantes sont décrites par leurs règles de production.

Exemple (Grammaire d'expressions arithmétiques)

$$\begin{array}{lll} E \rightarrow E + T & T \rightarrow T * F & F \rightarrow \text{IDF} \\ E \rightarrow T & T \rightarrow F & F \rightarrow \text{NUM} \\ & & F \rightarrow (E) \end{array}$$

aussi représenté

$$\begin{array}{l} E \rightarrow E + T \mid T \\ T \rightarrow T * F \mid F \\ F \rightarrow \text{IDF} \mid \text{NUM} \mid (E) \end{array}$$

Exemple (Le langage $\{a^n \cdot b^n \mid n \in \mathbb{N}\}$)

$$\begin{array}{ll} S \rightarrow \epsilon \\ S \rightarrow aSb \end{array}$$

Grammaires linéaires à droite

Les grammaires linéaires à *droite* sont t.q. les règles de production sont de la forme

- $A \rightarrow xB$, ou
- $A \rightarrow x$,

avec $x \in V_T^*$.

Exemple (Grammaire linéaire à droite)

- La grammaire

$$(\{a, b\}, \{S\}, S, P) \text{ avec } P = \{S \rightarrow baS, S \rightarrow a\}$$

génère le langage

$$\{(b \cdot a)^n \cdot a \in \{a, b\}^* \mid n \in \mathbb{N}\}.$$

- La grammaire

$$(\{a, b\}, \{S, A, B\}, S, P) \text{ avec } P = \{S \rightarrow A \mid B, A \rightarrow aA \mid \epsilon, B \rightarrow bB \mid \epsilon\}$$

génère le langage

$$\{a^n \in \{a, b\}^* \mid n \in \mathbb{N}\} \cup \{b^n \in \{a, b\}^* \mid n \in \mathbb{N}\}.$$

Grammaires linéaires à gauche

Les grammaires linéaires à *gauche* sont t.q. les règles de production sont de la forme

- $A \rightarrow Bx$, ou
- $A \rightarrow x$,

avec $x \in V_T^*$.

Exemple (Grammaire linéaire à gauche)

- La grammaire

$$(\{a, b\}, \{S\}, S, P) \text{ avec } P = \{S \rightarrow Sba, S \rightarrow a\}$$

génère le langage

$$\{a \cdot (b \cdot a)^n \in \{a, b\}^* \mid n \in \mathbb{N}\}.$$

- La grammaire

$$(\{a, b\}, \{S, A, B\}, S, P) \text{ avec } P = \{S \rightarrow Aab, A \rightarrow Aab \mid B, B \rightarrow a\}$$

génère le langage

$$\{a \cdot (a \cdot b)^n \in \{a, b\}^* \mid n \in \mathbb{N} \setminus \{0\}\}.$$

Grammaire linéaire

On peut ajouter à la hiérarchie de Chomsky le type des grammaires linéaires entre les grammaires de types 2 et 3.

Les grammaires linéaires sont telles que chaque règle de production a au plus un symbole non-terminal dans sa partie droite (il n'y a pas de contrainte sur la position de ce symbole non-terminal).

Exemple (Grammaire linéaire)

- La grammaire $(\{a, b\}, \{S\}, S, P)$ avec $P = \{S \rightarrow bSa, S \rightarrow \epsilon\}$ génère le langage

$$\{b^n a^n \in \{a, b\}^* \mid n \in \mathbb{N}\}.$$

- $(\{a, b\}, \{S\}, S, P)$ avec $P = \{S \rightarrow A, A \rightarrow aB \mid \epsilon, B \rightarrow Ab\}$ génère le langage

$$\{a^n b^n \in \{a, b\}^* \mid n \in \mathbb{N}\}.$$

Exemple (Grammaire non linéaire)

- La grammaire $(\{a, b\}, \{S, A, B\}, S, P)$ avec $P = \{S \rightarrow SS \mid \epsilon \mid aSb \mid bSa\}$ génère le langage

$$\{w \in \{a, b\}^* \mid |w|_a = |w|_b\}.$$

Plan Chap. 5 - Grammaires et grammaires régulières

- 1 Introduction et intuition
- 2 Grammaires : définition, dérivation et langage engendré
- 3 Grammaires régulières et automates à états finis
 - Des grammaires vers les automates
 - Des automates vers les grammaires
- 4 Grammaires et expressions régulières
- 5 Résumé

Grammaires régulières et automates à états finis

Équivalence entre grammaires régulières et AEFDs

- ① À chaque grammaire régulière, on peut associer un ϵ -AEFND qui reconnaît le même langage (correspondance entre vocabulaire terminal et état)
- ② Réciproquement, à chaque AEFD, on peut associer une grammaire régulière qui reconnaît le même langage.

Nous démontrons ces deux points en donnant des procédures de traduction entre grammaires et automates.

Corollaire

Les grammaires régulières génèrent les langages à états (et les langages réguliers par le théorème de Kleene).

Corollaire : retour sur la fermeture des langages réguliers

Étant données deux grammaires G et G' générant des langages L et L' , respectivement, nous pouvons trouver des grammaires générant les langages

- $L \cup L'$,
- $L \cap L'$,
- L^*
- \bar{L} ,
- $L \cdot L'$

à partir des grammaires G et G' . Nous trouverons ces grammaires en TD.

Plan Chap. 5 - Grammaires et grammaires régulières

- 1 Introduction et intuition
- 2 Grammaires : définition, dérivation et langage engendré
- 3 Grammaires régulières et automates à états finis
 - Des grammaires vers les automates
 - Des automates vers les grammaires
- 4 Grammaires et expressions régulières
- 5 Résumé

Grammaire linéaire à droite vers automate

Intuition

Intuitivement, nous construisons un ϵ -AEFND selon les étapes suivantes :

- 1 Création des états
 - un état par symbole non-terminal
 - l'axiome est l'état initial
 - un état final F qui n'est pas associé à un axiome.
- 2 Création des transitions associées aux règles de production :
 - Pour chaque règle $A \rightarrow xB$, nous créons une transition entre l'état A et l'état B étiquetée par x
 - Pour chaque règle $A \rightarrow x$, nous créons une transition entre l'état A et l'état F étiquetée par x .

L'automate obtenu est tel qu'il peut y avoir des mots (formés par la concaténation de plusieurs terminaux) sur les transitions.

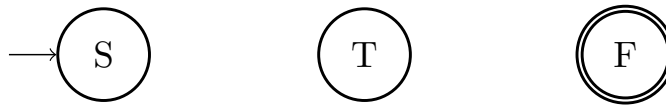
- 3 Nous créons ainsi des états et transitions permettant d'avoir uniquement un non-terminal étiquetant chaque transition.
 (Pour chaque transition $A \xrightarrow{x} X$ avec $x = x_0 \cdots x_n \in V_T^+$, $n \in \mathbb{N} \setminus \{0\}$ et X un état qui est soit associé à un non-terminal soit l'état F , nous créons les états intermédiaires $Ax_{\dots 0}X, Ax_{\dots 1}X, \dots, Ax_{\dots n-1}X$ (le nom assure l'unicité) et les transitions $A \xrightarrow{x_0} Ax_{\dots 0}X, Ax_{\dots 0}X \xrightarrow{x_1} Ax_{\dots 1}X, \dots, Ax_{\dots n-1}X \xrightarrow{x_n} X$.)

Grammaire linéaire à droite vers automate

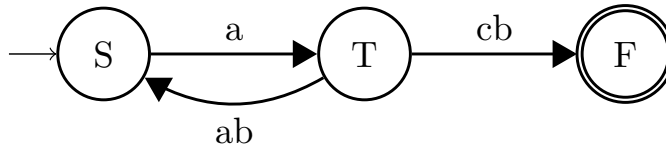
Intuition sur un exemple

Considérons la grammaire $(\{a, b\}, \{S, T\}, S, \{S \rightarrow aT, T \rightarrow abS \mid cb\})$.

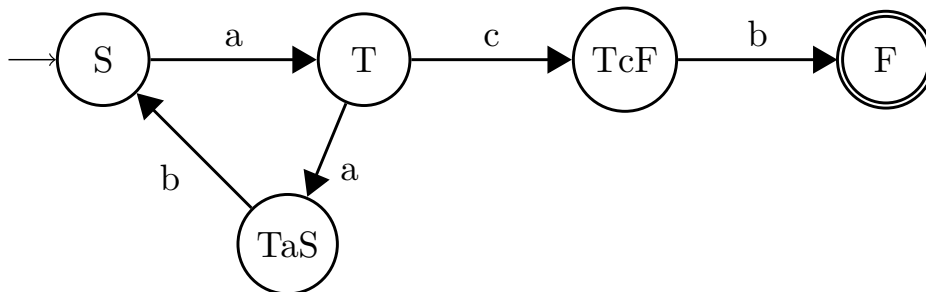
- ❶ Nous créons les états, déterminons l'état initial et ajoutons l'état final :



- ❷ Nous créons les transitions associées à chaque type de règle de production :



- ❸ Nous créons les états et transitions intermédiaires pour avoir uniquement un symbole étiquetant chaque transition :



Grammaire linéaire à droite vers automate

La transformation - première version

Soit $G = (V_T, V_N, Z, P)$ une grammaire régulière à droite.

Définition (Automate associé à une grammaire régulière droite)

L'" ϵ -AEFND" associé à G est $A_G = (V_N \cup \{F\}, Z, V_T, \Delta, \{F\})$ avec

$$\Delta = \begin{aligned} & \{(N, x, N') \mid (N, x \cdot N') \in P\} \\ & \cup \{(N, x, F) \mid (N, x) \in P\} \\ & \cup \{(N, \epsilon, F) \mid (N, \epsilon) \in P\} \end{aligned}$$

Ce qui est généré n'est pas stricto-sensu un ϵ -AEFND car Δ peut contenir des transitions étiquetées avec des mots de longueur strictement plus grande que 1.

Remarque Nous pouvons aussi transformer la grammaire (et obtenir une grammaire équivalente) afin d'assurer que toutes les règles sont de la forme $A \rightarrow x \cdot B$ ou $A \rightarrow x$ avec $|x| \leq 1$ en transformant chaque règle

$$A \rightarrow x_0 \cdots x_n B$$

en $A \rightarrow x_0 \cdot A_{x_0} B$, $A_{x_0} B \rightarrow x_1 \cdot A_{x_0 x_1} B$, \dots , $A_{x_0 \dots x_{n-1}} B \rightarrow x_n \cdot B$.

Ainsi, nous obtenons directement un ϵ -AEFND en appliquant la transformation ci-dessus à la grammaire transformée. □

Grammaire linéaire à droite vers automate

La transformation - seconde version

Soit $G = (V_T, V_N, Z, P)$ une grammaire régulière à droite.

Définition (Automate associé à une grammaire régulière droite)

L'ε-AEFND associé à G est $A_G = (V_N \cup \{F\} \cup \text{etats_inter}(\Delta'), Z, V_T, \Delta', \{F\})$ avec

$$\begin{aligned} \Delta = & \{(N, x, N') \mid (N, x \cdot N') \in P\} \\ & \cup \{(N, x, F) \mid (N, x) \in P\} \\ & \cup \{(N, \epsilon, F) \mid (N, \epsilon) \in P\} \end{aligned}$$

$$\begin{aligned} \Delta' = & \Delta \\ & \setminus \{(N, x, N') \in \Delta \mid |x| > 1\} \\ & \cup \{(N, x_0, Nx_0N'), (Nx_{n-1}N', x_n, N'), (Nx_{i-1}N', x_i, Nx_{i-1}N') \\ & \quad \mid (N, x, N') \in \Delta \text{ et } x = x_0 \cdots x_n \text{ et } |x| > 1 \text{ et } i \in [1, n-1]\} \\ \text{etats_inter}(\Delta') = & \{N, N' \mid \exists x \in V_T : (N, x, N') \in \Delta'\} \end{aligned}$$

Correction de la synthèse d'automate depuis des grammaires

Étant donnée une grammaire, le langage reconnu par l'automate obtenu à partir de cette grammaire est le langage généré par la grammaire.

Plan Chap. 5 - Grammaires et grammaires régulières

- 1 Introduction et intuition
- 2 Grammaires : définition, dérivation et langage engendré
- 3 Grammaires régulières et automates à états finis
 - Des grammaires vers les automates
 - Des automates vers les grammaires
- 4 Grammaires et expressions régulières
- 5 Résumé

Automate vers grammaire

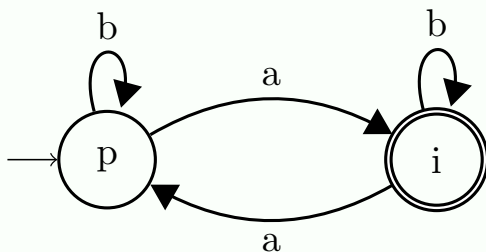
Intuition

Idée :

- L'ensemble des symboles non-terminaux est l'alphabet de l'automate.
- Associer un non-terminal par état.
- L'axiome est l'état initial.
- Les états accepteurs sont associés à des règles de production où la partie droite est ϵ .
- Pour chaque transition de q vers q' sur un symbole a , alors on a une règle de production $q \rightarrow a \cdot q'$.

Exemple (Automate vers grammaire)

Considérons l'automate



Nous lui associons la grammaire $(\{P, I\}, \{a, b\}, P, Prod)$, avec :

$$Prod = \{P \rightarrow b \cdot P \mid a \cdot I, I \rightarrow b \cdot I \mid a \cdot P \mid \epsilon\}.$$

Automate vers grammaire

Soit $A = (Q, q_0, \Sigma, \delta, F)$ un AEFD.

Définition (Grammaire régulière associée à un automate)

La grammaire associée à l'automate est (Q, Σ, q_0, P) avec

$$P = \{(q, a \cdot q') \mid (q, a, q') \in \delta\} \cup \{(f, \epsilon) \mid f \in F\}$$

Correction de la synthèse d'automate depuis des grammaires

Étant donné un automate, le langage généré par la grammaire obtenu à partir de l'automate est le langage reconnu par l'automate.

Remarque Cette transformation s'adapte aisément pour générer des grammaires à partir d'AEFNDs et ϵ -AEFNDs. □

Plan Chap. 5 - Grammaires et grammaires régulières

- 1 Introduction et intuition
- 2 Grammaires : définition, dérivation et langage engendré
- 3 Grammaires régulières et automates à états finis
- 4 Grammaires et expressions régulières**
 - Traduction d'expression régulière en grammaire
- 5 Résumé

Traduction d'expression régulière en grammaire

Rappel sur les expressions régulières

La syntaxe des expression régulières sur un alphabet Σ est définie inductivement par :

- ϵ et \emptyset sont des expressions régulières.
- Si $a \in \Sigma$ alors a est une expression régulière.
- Si e et e' sont des expressions régulières, alors $e + e'$ est une expression régulière.
- Si e et e' sont des expressions régulières, alors $e \cdot e'$ est une expression régulière.
- Si e est une expression régulière, alors e^* est une expression régulière.

Traduction d'expression régulière en grammaire

Rappel sur les expressions régulières

Définition (Traduction compositionnelle des expressions régulières en grammaire linéaire à droite)

Expressions régulières de base :

- Pour $\epsilon : (\{S\}, \emptyset, S, \{S \rightarrow \epsilon\})$
- Pour $\emptyset : (\{S\}, \emptyset, S, \emptyset)$
- Pour $a \in \Sigma : (\{S\}, \{a\}, S, \{S \rightarrow a\})$

Expressions régulières composées : soient $G_E = (N_E, T_E, S_E, P_E)$ et $G_{E'} = (N_{E'}, T_{E'}, S_{E'}, P_{E'})$ des grammaires linéaires à droite pour des expressions régulières e et e' .

Pour simplifier, nous supposons que $N_E \cap N_{E'} = \emptyset$.

- Pour $e + e' : (N_E \cup N_{E'}, T_E \cup T_{E'}, S, P_E \cup P_{E'} \cup \{S \rightarrow S_E \mid S_{E'}\})$
- Pour $e \cdot e' : (N_E \cup N_{E'}, T_E \cup T_{E'}, S, P'_E \cup P_{E'} \cup \{S \rightarrow S_E \mid S_{E'}\})$ avec $P'_E = \{A \rightarrow x \cdot S_{E'} \mid A \rightarrow x \in P_E\} \cup \{A \rightarrow x \cdot B \in P_E\}$.
- Pour $e^* : (N_E, T_E, S_E, P_E \cup \{A \rightarrow x \cdot S_E \mid A \rightarrow x \in P_E\} \cup \{S_E \rightarrow \epsilon\})$

Traduction d'expression régulière en grammaire

Rappel sur les expressions régulières

Exemple (Traduction compositionnelle des expressions régulières en grammaires linéaires à droite)

Pour $(a + b)^*$.

- Grammaire pour $a : (\{S_a\}, \{a\}, S_a, \{S_a \rightarrow a\})$
- Grammaire pour $b : (\{S_b\}, \{b\}, S_b, \{S_b \rightarrow b\})$
- Grammaire pour $a + b : (\{S, S_a, S_b\}, \{a, b\}, S, \{S \rightarrow S_a \mid S_b, S_a \rightarrow a, S_b \rightarrow b\})$
- Grammaire pour $(a + b)^* :$
 $(\{S, S_a, S_b\}, \{a, b\}, S, \{S \rightarrow S_a \mid S_b \mid \epsilon, S_a \rightarrow a \cdot S, S_b \rightarrow b \cdot S\})$

Plan Chap. 5 - Grammaires et grammaires régulières

- 1 Introduction et intuition
- 2 Grammaires : définition, dérivation et langage engendré
- 3 Grammaires régulières et automates à états finis
- 4 Grammaires et expressions régulières
- 5 **Résumé**

Résumé

Les grammaires permettent de décrire les langages de manière **générative**.
(Les automates permettent de décrire les langages (à états) de manière *opérationnelle*.
Les expressions régulières permettent de décrire les langages (réguliers) de manière *déclarative*.)

Grammaire et grammaires régulières

- Grammaires : définition, dérivation et langage généré.
- Types de grammaires et classification de Chomsky.
- Équivalence entre grammaires régulières et AEFDs :
 - une grammaire régulière est linéaire à droite ou à gauche, on parle de grammaire régulière à droite ou à gauche,
 - traduction d'une grammaire régulière à droite vers un ϵ -AEFND,
 - traduction d'une grammaire régulière à gauche vers un ϵ -AEFND,
 - traduction d'un ϵ -AEFND/AEFND/AEFD vers une grammaire linéaire à droite.