

# Langages et Compilation

Généralisation de la notion d'automate fini

D'expression régulière vers automate fini

## Le théorème de Kleene

Un langage est reconnaissable par automate fini déterministe si et seulement si une expression régulière le représente.

- ⇐ Tout langage décrit par une expression régulière est reconnaissable par automate fini :
- Les langages  $\emptyset$ ,  $\{\epsilon\}$  et  $\{a\}$  pour toute lettre  $a$  de  $\Sigma$  sont reconnaissables par AF.
  - La classe des langages reconnaissables par AF est close par union, concaténation et étoile.
- ⇒ On a un algorithme simple qui trouve une expression régulière à partir d'un automate fini

## Le théorème de Kleene

Un langage est reconnaissable par automate fini déterministe si et seulement si une expression régulière le représente.

Ce qui nous intéresse, c'est une construction effective, automatique et efficace qui traduise une expression régulière en automate fini.

Ces constructions utilisent des extensions des AF déterministes qui autorisent des transitions non-déterministes. Fait notable, ces extensions ne changent pas les capacités de reconnaissance et peuvent être rendues déterministes.

On va voir deux constructions : l'algorithme de Glushkov et celui de Thompson.

## Automate fini non déterministe

Déterminisation

Automate fini non-déterministes avec  $\varepsilon$ -transitions

D'expression régulière vers AF : l'algo de Glushkov

D'expression régulière vers AF : l'algo de Thompson

# Automate fini non déterministe

Un **automate fini non déterministe** (**AFN**) est un quintuplet  $(\Sigma, Q, \delta, I, F)$  où :

$\Sigma$  est un alphabet

$Q$  est un ensemble **fini** d'états

$\delta$  est la fonction de transition de  $Q \times \Sigma \rightarrow \mathcal{P}(Q)$   
( $\mathcal{P}(Q)$  l'ensemble de toutes les parties de  $Q$ )

$I \subset Q$  est l'ensemble des états initiaux

$F \subset Q$  est l'ensemble des états finaux

# Automate fini non déterministe

## Fonctionnement

Au temps initial, la machine est dans un des états initiaux  $q_0 \in I$  (**non déterminisme**)

Si à l'instant  $t$  la machine est dans l'état  $q$  et lit le symbole  $a$ , alors à l'instant  $t + 1$

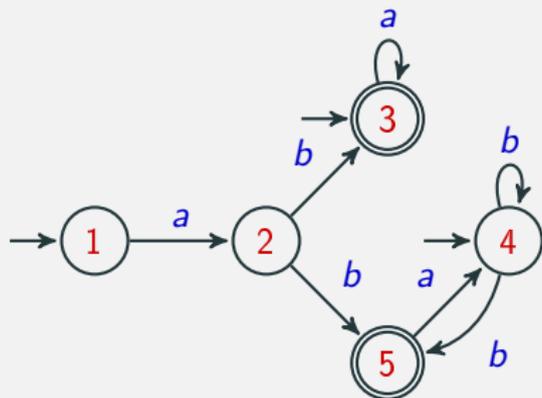
- si  $\delta(q, a) = \emptyset$  : la machine se bloque
- si  $\delta(q, a) \neq \emptyset$  : la machine entre dans l'un des états de  $\delta(q, a)$  (**non déterminisme**) et consomme  $a$ .

# Automate fini non déterministe

## Exemple

$(\{a, b\}, \{1, 2, 3, 4, 5\}, \delta, \{1, 3, 4\}, \{3, 5\})$  où  $\delta$  est définie par la table :

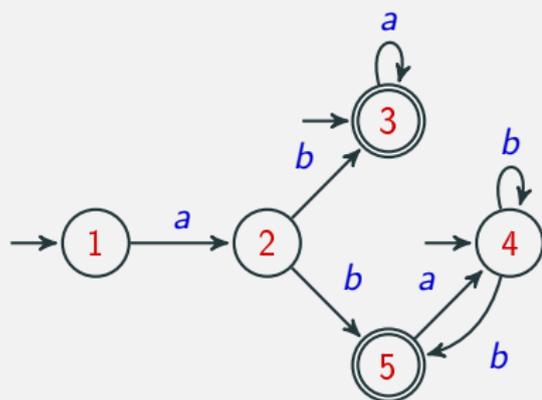
	$a$	$b$
$\rightarrow 1$	$\{2\}$	$\emptyset$
2	$\emptyset$	$\{3, 5\}$
$\rightarrow \textcircled{3}$	$\{3\}$	$\emptyset$
$\rightarrow 4$	$\emptyset$	$\{4, 5\}$
$\textcircled{5}$	$\{4\}$	$\emptyset$



Un mot  $u$  est **accepté** s'il existe, à partir d'un état initial  $q_0 \in I$ , un chemin étiqueté par  $u$  qui aboutit à un état d'acceptation  $f \in F$ .

# Automate fini non déterministe

## Exemple



*abab* est accepté : étiquète le chemin 1 2 5 4 5

autres calculs possibles 1 2 5 4 4 ( $4 \notin F$ ), 1 2 3 3 blocage, 3 3 blocage, 4 blocage

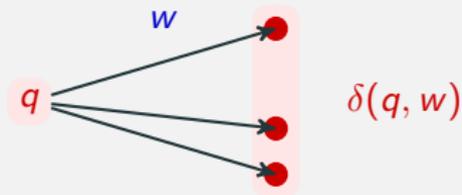
*ba* n'est pas accepté :

1 blocage, 3 blocage, 4 4 blocage, 4 5 4 ( $4 \notin F$ )

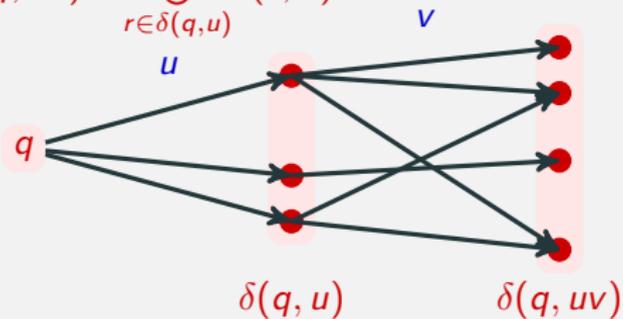
# Automate fini non déterministe

**Prolongement de la fonction de transition  $\delta$**  en une fonction de  $Q \times \Sigma^*$  dans  $\mathcal{P}(Q)$  :

$\delta(q, w)$  = l'ensemble des états accessibles en partant de  $q$  et après lecture du mot  $w$



On a pour tous mots  $u, v$  :  $\delta(q, uv) = \bigcup_{r \in \delta(q, u)} \delta(r, v)$



# Automate fini non déterministe

Le langage **reconnu** par un AFN  $\mathcal{A} = (\Sigma, Q, \delta, I, F)$  est l'ensemble des mots acceptés par  $\mathcal{A}$

$$L(\mathcal{A}) = \{w \in \Sigma^* : \bigcup_{i \in I} \delta(i, w) \cap F \neq \emptyset\}$$

Automate fini non déterministe

## Détermination

Automate fini non-déterministes avec  $\varepsilon$ -transitions

D'expression régulière vers AF : l'algo de Glushkov

D'expression régulière vers AF : l'algo de Thompson

# Automate fini non déterministe

Les automates finis déterministes et les automates finis non déterministes ont la même capacité de calcul :

## Le théorème de Rabin-Scott

Tout langage reconnu par un AFN l'est aussi par un AFD.

Preuve constructive : il existe un algorithme qui étant donné un AFN construit un AFD équivalent.

L'AFD simule en parallèle tous les calculs possibles de l'AFN.

Son état initial est l'ensemble des états initiaux de l'AFN.

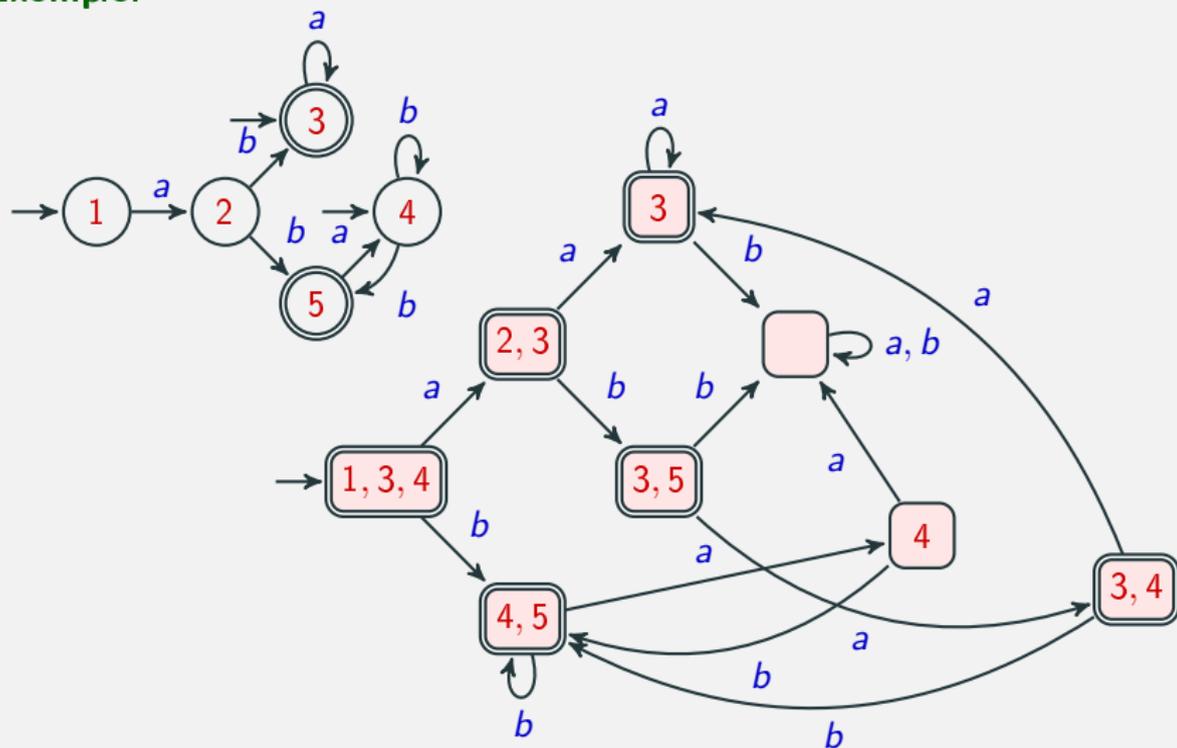
L'état atteint sur l'AFD après lecture du mot  $w$  correspond sur l'AFN à l'ensemble des états accessibles en partant d'un des états initiaux et après lecture du mot  $w$ .

Les états de l'AFD sont ainsi des ensembles d'états de l'AFN.

# Automate fini non déterministe

## Déterminisation

Exemple.



# Automate fini non déterministe

## Déterminisation

Si l'AFN  $\mathcal{A} = (\Sigma, Q, \delta, I, F)$  reconnaît  $L$  alors

l'AFD  $\mathcal{A}_{det} = (\Sigma, Q_{det}, \delta_{det}, i_{det}, F_{det})$  où :

$Q_{det} = \mathcal{P}(Q)$  l'ensemble des parties de  $Q$

$\delta_{det}$  la fonction de transition de  $\mathcal{P}(Q) \times \Sigma \rightarrow \mathcal{P}(Q)$   
définie par  $\delta_{det}(X, a) = \bigcup_{q \in X} \delta(q, a)$

$i_{det} = I$

$F_{det} = \{X \subset Q : X \cap F \neq \emptyset\}$  l'ensemble des parties contenant un  
état d'acceptation

reconnaît également  $L$ .

$$\begin{aligned} w \in L(\mathcal{A}) &\Leftrightarrow \bigcup_{i \in I} \delta(i, w) \cap F \neq \emptyset \Leftrightarrow \delta_{det}(I, w) \cap F \neq \emptyset \Leftrightarrow \delta_{det}(I, w) \in F_{det} \\ &\Leftrightarrow w \in L(\mathcal{A}_{det}) \end{aligned}$$

# Automate fini non déterministe

## Algorithme de détermination

Concrètement l'ensemble des états de l'AFD comprend uniquement les parties de  $Q$  accessibles de l'état initial.

On les caractérise via un parcours du graphe de l'AFD qui visite tous les sommets accessibles à partir de son état initial.

**Entrée** : l'AFN  $\mathcal{A} = (\Sigma, Q, \delta, I, F)$

**Sortie** : l'AFD  $\mathcal{A}_{\text{det}} = (\Sigma, Q_{\text{det}}, \delta, I, F_{\text{det}})$      $F_{\text{det}} = \{ X \subset Q_{\text{det}} : X \cap F \neq \emptyset \}$

**Initialisation**

$Q_{\text{det}} = \{I\}$

$\text{VoirSucc} = \{I\}$

**Traitement**

tant que  $\text{VoirSucc}$  n'est pas vide faire

extraire un élément  $X$  de  $\text{VoirSucc}$

pour toute lettre  $a$  de  $\Sigma$

$$T = \bigcup_{q \in X} \delta(q, a)$$

$$\delta_{\text{det}}(X, a) = T$$

si  $T$  n'est pas dans  $Q_{\text{det}}$  alors

ajouter  $T$  à  $Q_{\text{det}}$

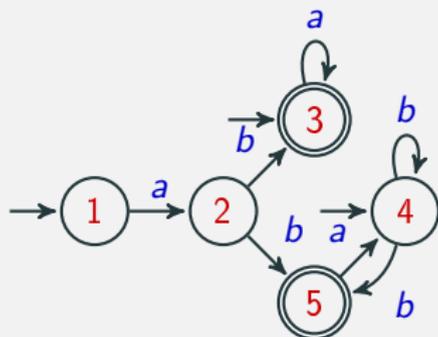
ajouter  $T$  à  $\text{VoirSucc}$

# Automate fini non déterministe

## Algorithme de détermination

### Exemple.

	<i>a</i>	<i>b</i>
→ 1	{2}	∅
2	∅	{3, 5}
→ 3	{3}	∅
→ 4	∅	{4, 5}
5	{4}	∅



	<i>a</i>	<i>b</i>
→ $I = \{1, 3, 4\}$	{2, 3}	{4, 5}
{2, 3}	{3}	{3, 5}
{4, 5}	{4}	{4, 5}
{3}	{3}	∅
{3, 5}	{3, 4}	∅
{4}	∅	{4, 5}
∅	∅	∅
{3, 4}	{3}	{4, 5}

# Automate fini non déterministe

## Algorithme de détermination

Quel est son coût ?

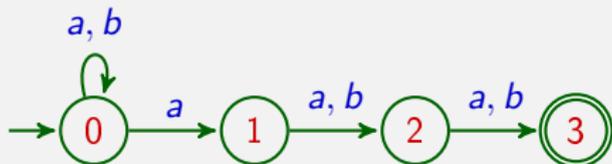
En  $O(|\Sigma| \times |Q_{det}|)$  : le coût du parcours du graphe de l'AFD qui a  $|Q_{det}|$  sommets et  $|Q_{det}| \times |\Sigma|$  arcs.

Hors le passage d'un AFN à un AFD peut coûter cher en nombre d'états.

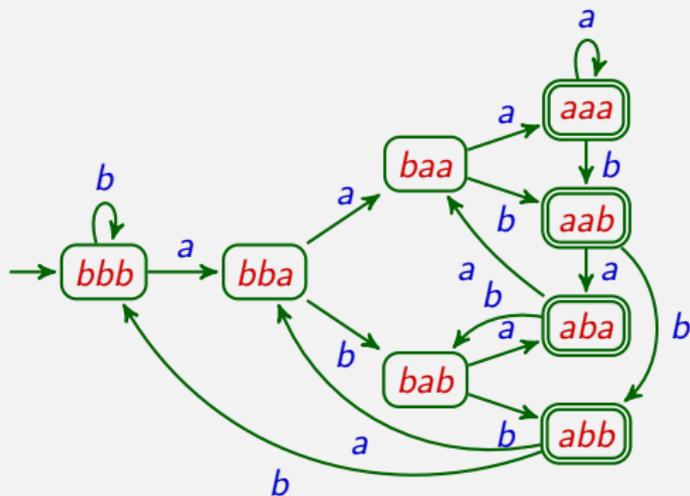
Dans le pire cas, c'est exponentiel  $|Q_{det}| = |\mathcal{P}(Q)| = 2^{|Q|}$ .

Exemple :  $(a + b)^* a (a + b)^{n-1}$

L'ensemble des mots dont la  $n$ -ème lettre en partant de la fin est un  $a$  est reconnu par un AFN avec  $n + 1$  mais tout AFD qui le reconnaît a au moins  $2^n$  états.



l'AFN pour  $(a + b)^* a(a + b)^2$



l'AFD  
coût de sa construction  $\sim 2^3$

Automate fini non déterministe

Déterminisation

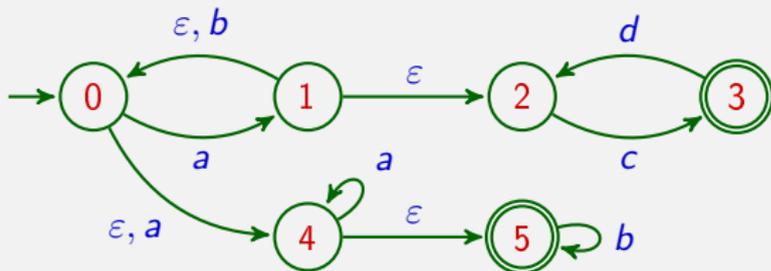
Automate fini non-déterministes avec  $\varepsilon$ -transitions

D'expression régulière vers AF : l'algo de Glushkov

D'expression régulière vers AF : l'algo de Thompson

# Automate fini non déterministe avec $\epsilon$ -transitions

Une généralisation des AFN avec des transitions étiquetées avec le mot vide  $\epsilon$ . Lorsque l'automate choisit une telle transition, il ne consomme pas de lettre.



Un mot  $w$  est **accepté**, s'il existe un chemin étiqueté par  $w$  d'un état initial à un état final.

(chemin de longueur  $|w| +$  le nombre de  $\epsilon$ -transitions empruntées)

$ac$  est accepté  $0 \xrightarrow{a} 1 \xrightarrow{\epsilon} 2 \xrightarrow{c} 3$

$aaba$  est accepté  $0 \xrightarrow{a} 1 \xrightarrow{\epsilon} 0 \xrightarrow{a} 1 \xrightarrow{b} 0 \xrightarrow{a} 4 \xrightarrow{\epsilon} 5$

$abba$  n'est pas accepté

# Automate fini non déterministe avec $\epsilon$ -transitions

Ajouter des  $\epsilon$ -transitions n'augmente pas les capacités de reconnaissance de l'automate :

## Proposition

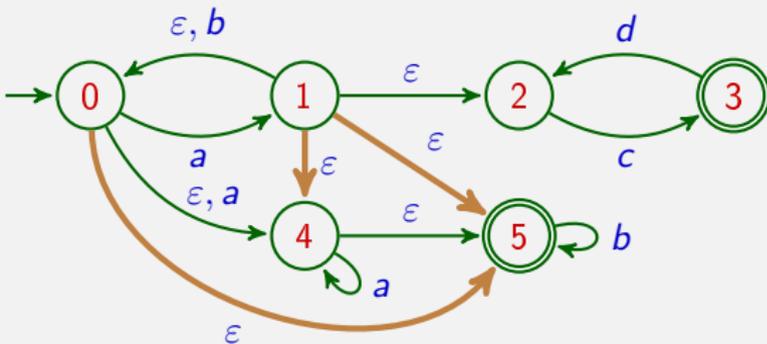
Tout langage reconnu par un AFN avec  $\epsilon$ -transitions l'est aussi par un AFN sans  $\epsilon$ -transitions.

Preuve constructive : il existe un algorithme qui étant donné un AFN avec  $\epsilon$ -transitions construit un AFN sans  $\epsilon$ -transitions équivalent et avec le même nombre d'états.

**Étape 1.** Fermeture transitive sur les  $\epsilon$ -transitions.

Pour chaque état  $q$ , on définit  $\text{Cloture}(q)$  l'ensemble des états accessibles à partir de  $q$  par des  $\epsilon$ -transitions.

On ajoute à l'automate initial une  $\epsilon$ -transition de  $q$  vers  $p$  pour tout  $p$  dans  $\text{Cloture}(q)$ .

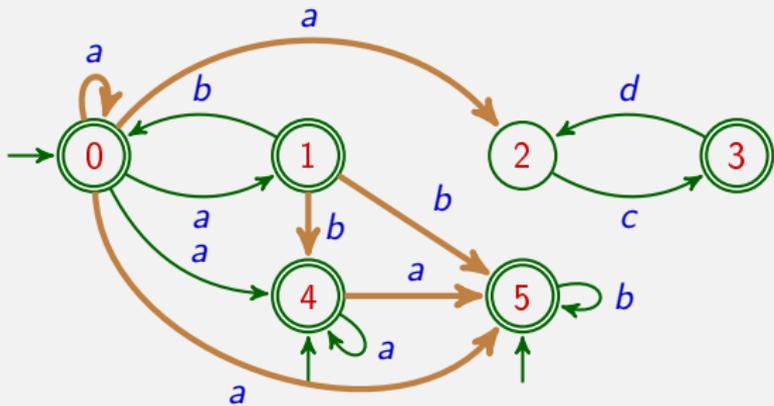
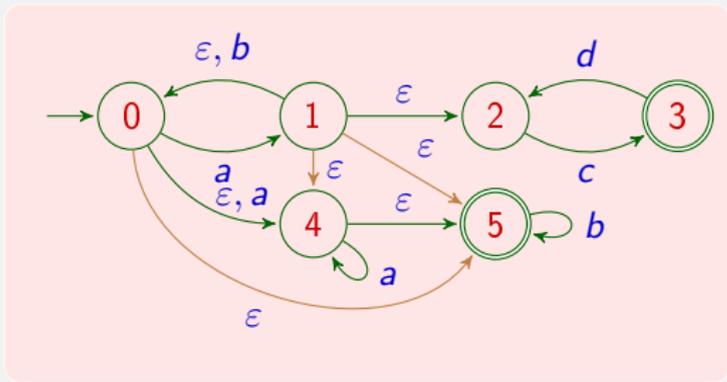


## Étape 2. Construction de l'AFN sans $\epsilon$ -transitions par fermeture avant.

On ajoute à l'automate de départ :

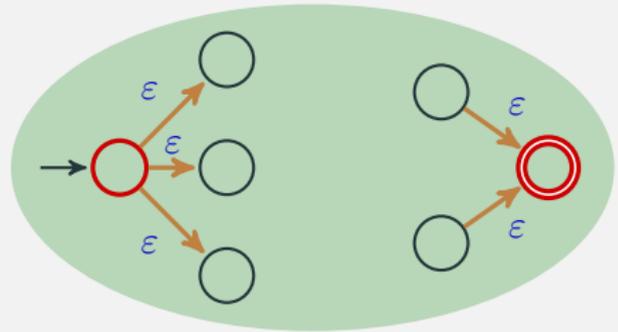
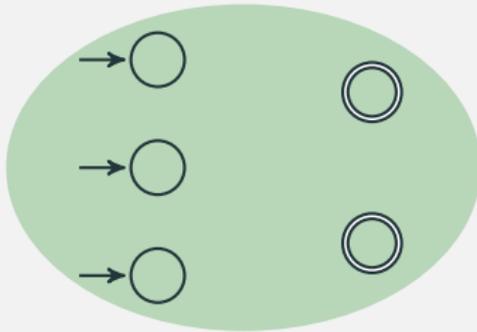
- dans l'ensemble des états initiaux : les états  $p$  extrémités d'une  $\epsilon$ -transition dont l'origine est un état initial ;
- dans l'ensemble des états finaux : les états  $p$  origines d'une  $\epsilon$ -transition dont l'extrémité est un état final ;
- dans l'ensemble des transitions : les transitions  $p \xrightarrow{a} r$  pour chaque séquence  $p \xrightarrow{a} q \xrightarrow{\epsilon} r$

On supprime ensuite les  $\epsilon$ -transitions.



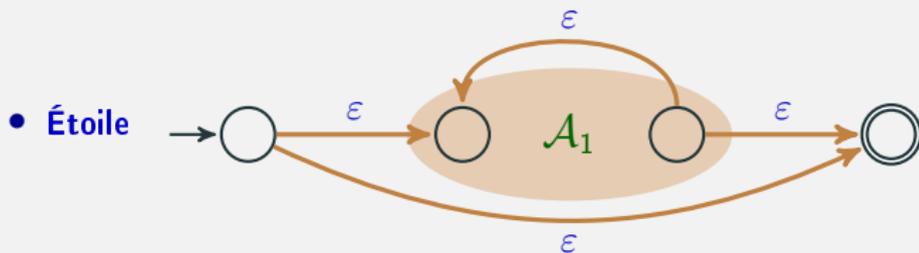
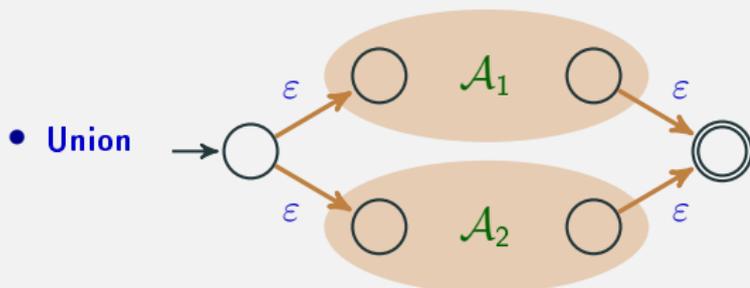
# Automate fini non déterministe avec $\epsilon$ -transitions

Grâce aux  $\epsilon$ -transitions, on peut toujours se ramener à des automates (avec  $\epsilon$ -transitions) qui ont un unique état initial et un unique état final.



# Automate fini non déterministe avec $\epsilon$ -transitions

## Propriétés de stabilité



Les trois objets introduits :

- AFD,
- AFN,
- AFN avec  $\epsilon$ -transitions

ont la même capacité de calcul.

Autrement dit, les classes de langages reconnaissables par AFD, AFN avec ou sans  $\epsilon$ -transitions sont identiques. Tous ces automates caractérisent l'ensemble des langages réguliers.

De plus, le théorème de Kleene assure l'équivalence entre expressions régulières et automates finis.

## Bilan

AFD, AFN avec ou sans  $\epsilon$ -transitions et expressions régulières définissent la même classe de langages.

Automate fini non déterministe

Déterminisation

Automate fini non-déterministes avec  $\varepsilon$ -transitions

D'expression régulière vers AF : l'algo de Glushkov

D'expression régulière vers AF : l'algo de Thompson

# D'expression régulière vers automate fini

## L'algorithme de Glushkov

**Entrée** : une expression régulière

$$(b + ab)^*(\varepsilon + ab)$$

**Étape 1.** Linéariser l'expression régulière.

On remplace toutes les lettres de l'expression par des symboles distincts ( $x_i$  pour la lettre en  $i$ -ème position)

$$(x_1 + x_2x_3)^*(\varepsilon + x_4x_5)$$

**Étape 2.** Déterminer

- **Premier** : l'ensemble des symboles pouvant commencer un mot
- **Dernier** : l'ensemble des symboles pouvant terminer un mot
- **Suivant** : pour tout symbole  $x_i$ , l'ensemble des symboles pouvant suivre  $x_i$

l'expression régulière linéarisée :  $(x_1 + x_2x_3)^*(\varepsilon + x_4x_5)$

**Premier** =  $\{x_1, x_2, x_4\}$

**Dernier** =  $\{x_1, x_3, x_5\}$

	<b>Suivant</b>
$x_1$	$x_1, x_2, x_4$
$x_2$	$x_3$
$x_3$	$x_1, x_2, x_4$
$x_4$	$x_5$
$x_5$	

Un mot  $w = a_1 \cdots a_n$  appartient au langage décrit par l'expression linéarisée si et seulement si

- $a_1 \in \text{Premier}$ ,
- $a_{i+1} \in \text{Suivant}(a_i)$  pour tout  $i = 1, \dots, n-1$  et
- $a_n \in \text{Dernier}$ .

$x_2x_3x_1x_1$  est un mot du langage :  $x_2 \in \text{Premier}$ ,  $x_3 \in \text{Suivant}(x_2)$ ,  
 $x_1 \in \text{Suivant}(x_3)$ ,  $x_1 \in \text{Suivant}(x_1)$ ,  $x_1 \in \text{Dernier}$ .

$x_4x_5x_1$  n'est pas un mot du langage :  $x_1 \notin \text{Suivant}(x_5)$ .

# D'expression régulière vers automate fini

## L'algorithme de Glushkov

### Étape 3. Construction de l'AFN

#### L'ensemble des états :

- un état initial  $0$
- un état  $i$  par symbole  $x_j$   
(l'automate sauvegarde le dernier symbole  $x_j$  lu via l'état  $i$ )

#### L'ensemble des états finaux :

- l'état initial  $0$  si  $\epsilon$  appartient au langage
- un état  $i$  pour tout  $x_j$  appartenant à **Dernier**

#### La fonction de transition :

- une transition de l'état initial  $0$  vers l'état  $i$  pour tout  $x_j$  appartenant à **Premier** et étiquetée par la lettre correspondant à  $x_j$
- une transition de l'état  $i$  vers l'état  $j$  pour tout  $x_j$  appartenant à **Suivant**( $x_j$ ) et étiquetée par la lettre correspondant à  $x_j$

$$(b + ab)^*(\epsilon + ab)$$

$$(x_1 + x_2x_3)^*(\epsilon + x_4x_5)$$

$$\text{Premier} = \{x_1, x_2, x_4\}$$

$$\text{Dernier} = \{x_1, x_3, x_5\}$$

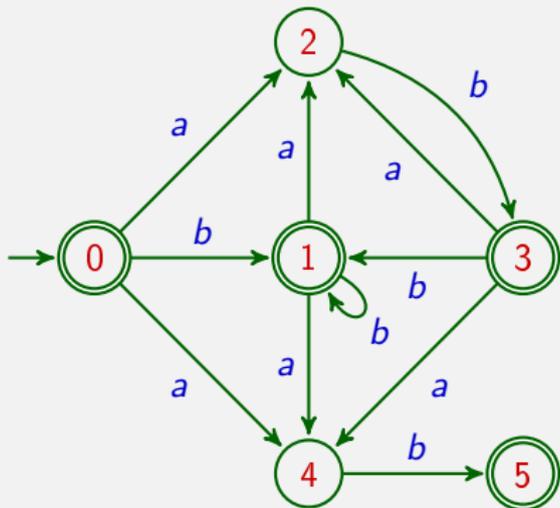
0 l'état initial

$$Q = \{0, 1, 2, 3, 4, 5\}$$

$\epsilon$  appartient au langage  
 $\Rightarrow$  l'état initial 0 est final

$$F = \{0, 1, 3, 5\}$$

	Suivant
$x_1$	$x_1, x_2, x_4$
$x_2$	$x_3$
$x_3$	$x_1, x_2, x_4$
$x_4$	$x_5$
$x_5$	



# D'expression régulière vers automate fini

## Complexité de l'algorithme de Glushkov

La **taille** d'une expression régulière, notée  $|r|$ , est son nombre de caractères  $\varepsilon$ ,  $a$  (lettre de l'alphabet),  $+$  et  $*$  (les parenthèses et la concaténation ne sont pas pris en compte).

$$|(b + ab)^*(\varepsilon + ab)| = 9$$

Le nombre d'états de l'automate est borné par  $|r| + 1$ .

Le nombre de transitions est quadratique dans le pire cas, mais linéaire dans le cas moyen (pour la distribution uniforme).

↪ Le temps de construction de l'AFN est

- en  $O(|r|^2)$  dans le pire cas
- en  $O(|r|)$  dans le cas moyen.

Automate fini non déterministe

Déterminisation

Automate fini non-déterministes avec  $\varepsilon$ -transitions

D'expression régulière vers AF : l'algo de Glushkov

D'expression régulière vers AF : l'algo de Thompson

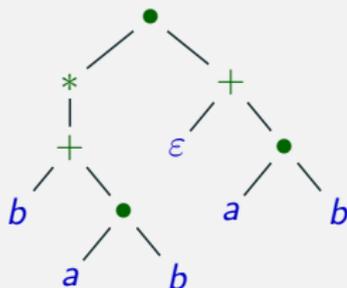
# D'expression régulière vers automate fini

Une autre variante : l'algorithme de Thompson

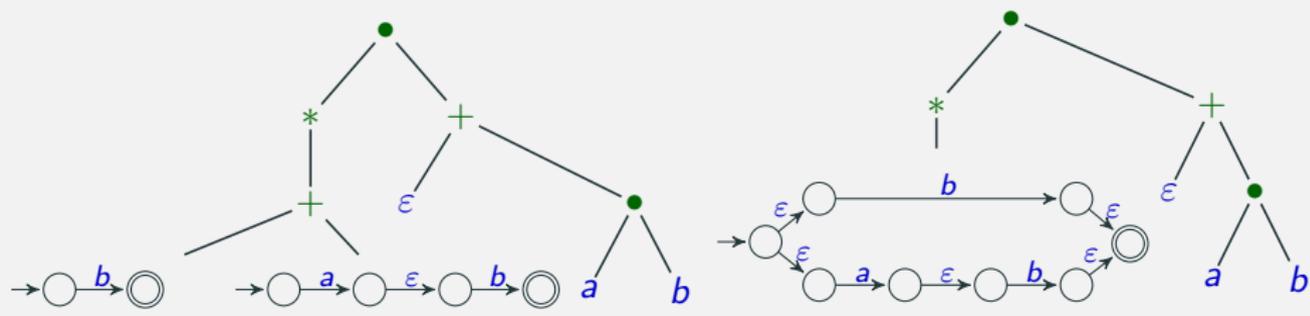
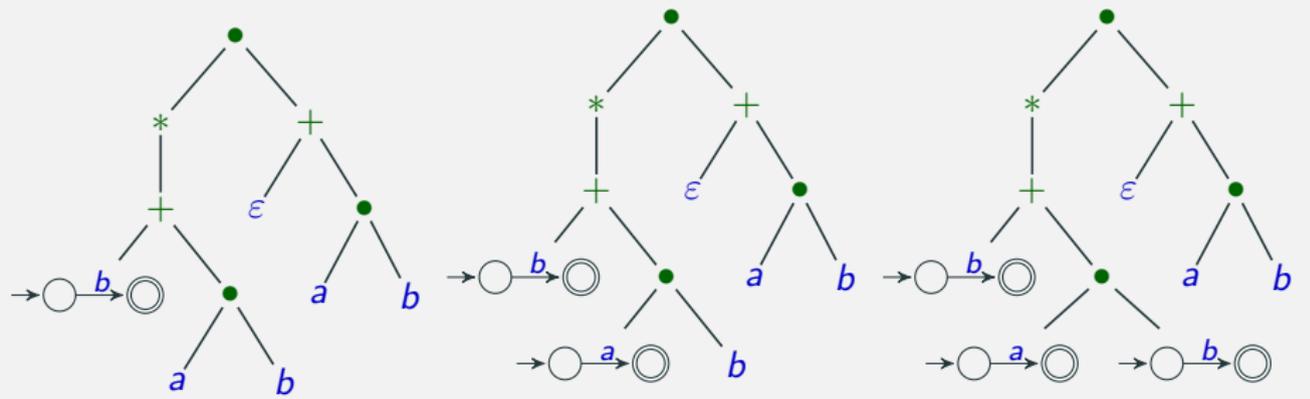
**Entrée** : une expression régulière

$$(b + ab)^*(\varepsilon + ab)$$

**Étape 1.** Construction de l'arbre syntaxique associé à l'expression.



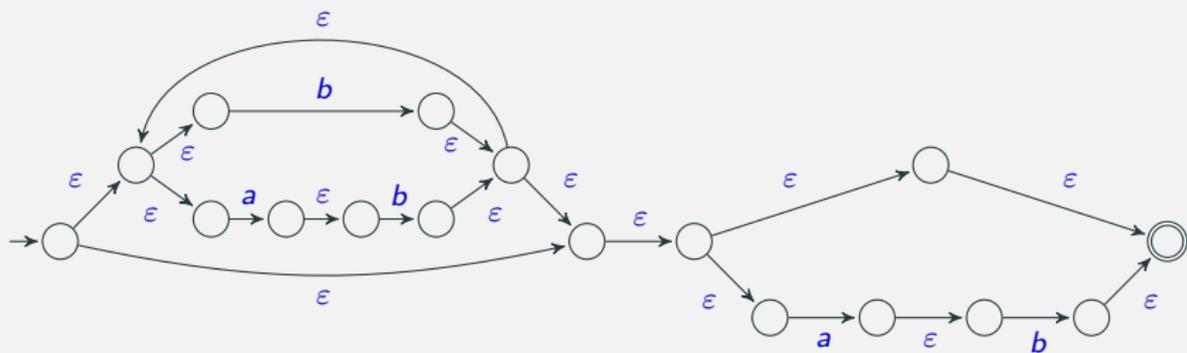
**Étape 2.** Construction d'un AFN avec  $\varepsilon$ -transitions via un parcours en profondeur postfixé de l'arbre. Chaque fragment a un seul état initial, un seul état final et aucun arc arrivant dans l'état initial ou partant de l'état final.



...

# D'expression régulière vers automate fini

## Complexité de l'algorithme de Thompson



Coût de l'algorithme de Thompson :

- Construction de l'arbre de syntaxe de taille  $O(|r|)$  en  $O(|r|)$  étapes
- Construction de l'AFN via un parcours postfixé de l'arbre de taille  $O(|r|)$  en  $O(|r|)$  étapes