THÉORIE DES LANGAGES : ANALYSE LEXICALE – PARTIE 2

L3 INFORMATIQUE – 2020/21

Eric Andres
eric.andres@univ-poitiers.fr

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Def : S un alphabet, on appelle **classe des langages rationnels** Rat(Σ) l'ensemble des langages vérifiant :
  - Tout sous-ensemble fini de S* est dans Rat(Σ)
  - Pour L1 $\in$ Rat(Σ) et L2 $\in$ Rat(Σ) alors
    - L1 $\cup$ L2 $\in$ Rat(Σ)
    - L1 $\bullet$ L2 $\in$ Rat(Σ)
    - L1* $\in$ Rat(Σ) (et bien sûr L2* aussi)
  - Rat(Σ) est le plus petit ensemble de langages possibles vérifiant cela.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Def : Expression régulière

  La définition d'un langage rationnel en

  - Un ensemble fini de mots (y compris)
  - Concaténation
  - Union
  - Etoile

  Et donc nécessairement fini. Une telle expression s'appelle **« expression régulière »**

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

Propriété : Un langage est rationnel ssi il existe une expression rationnel le définissant

exemple :

- **ab*** : attention, c'est un abus : on devrait écrire **{a}.{b}*** = {a, ab, abb, abbb, …}
- **(ab)*** = {ε, ab, abab, ababab, … }
- **a+ba*** = {a} ∪ {b}{a}* = {a, b, ba, baa, baaa, …}

aa* = a+ : Deux expressions rationnelles peuvent être différentes et représenter le même langage !

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

Propriété :

- Si L $\in$ Rat(Σ) et si R: $\Sigma^* \rightarrow \Pi^*$ est un mophisme de monoïde

  alors R(L) $\in$ Rat(Π)

- L'image miroir d'un langage rationnel est aussi rationnel (on verra la preuve en TD)

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

Propriété : Lemme de l'étoile

Soit L un langage rationnel, alors il existe un entier k tel que :

- $\forall$u e L, tq |u| > k, il existe trois mots v,w et z tel que :
  - u = v • w • z
  - w ≠ ε
  - v • w* • z ∈ L

Le lemme de l'étoile est couramment utilisé pour montrer qu'un langage donné n'est pas rationnel (en raisonnant par l'absurde). En revanche, il ne peut être employé pour démontrer qu'un langage est rationnel. En effet, il énonce une condition, nécessaire certes, mais non suffisante, de rationalité

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Exemple : $L = \{a^n\ b^n : n \in N\}$ est-il rationnel ?

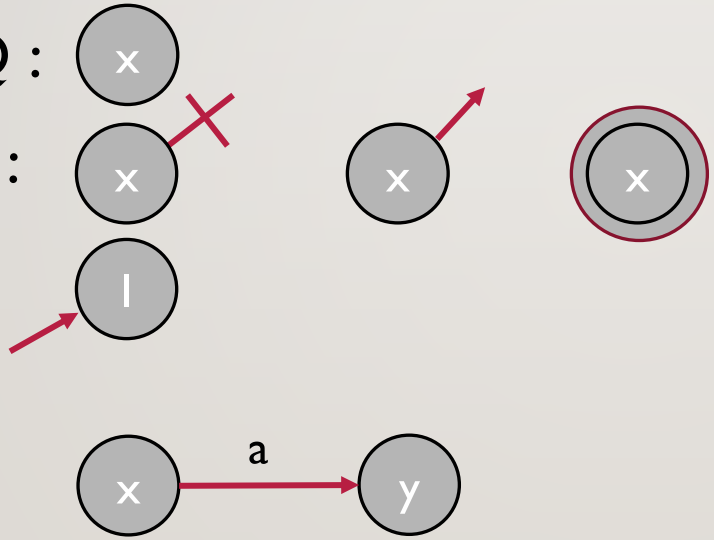
Parfois c'est plus compliqué : le langage de tous les palindromes n'est pas rationnel mais on peut toujours trouver un découpage $v \bullet w \bullet z$ qui respecte toutes les conditions.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

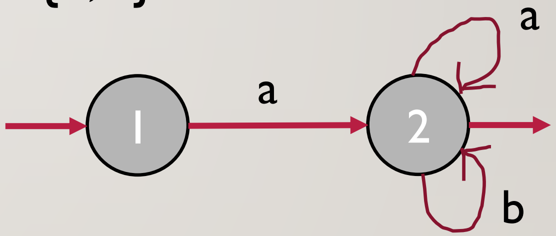
- Def : On appelle **automate fini** un quintuplet (Σ, Q, I, T, d) où
  - Σ est un alphabet fini
  - $Q \subset N$ est un ensemble fini d'états
  - $I \in Q$ est l'état initial (en général un seul)
  - $T \subset Q$ est l'ensemble des états terminaux
  - $d \subset Q \times S \times Q$ est l'ensemble des transitions

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Image pour les automates :
  - $x \in Q$:
  - $x \in T$:
  - I
  - (x,a,y)



Exemple d'automate fini : les mots commençant par a sur $\Sigma=\{a,b\}$

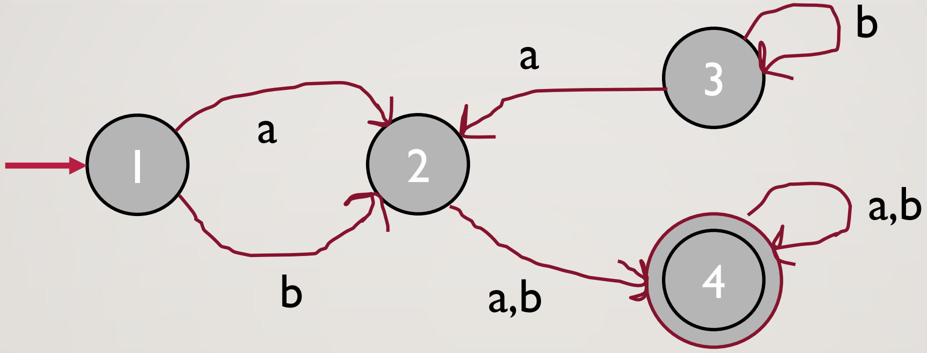


LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Def :
  - Automate Complet : de chaque état part une transition pour chaque lettre de l'alphabet
  - Automate Déterministe : si pour chaque état il n'existe qu'une seule transition possible (et pas de transition en ε)
  - Automate Monogène : Chaque état est atteignable (depuis l'état initial)

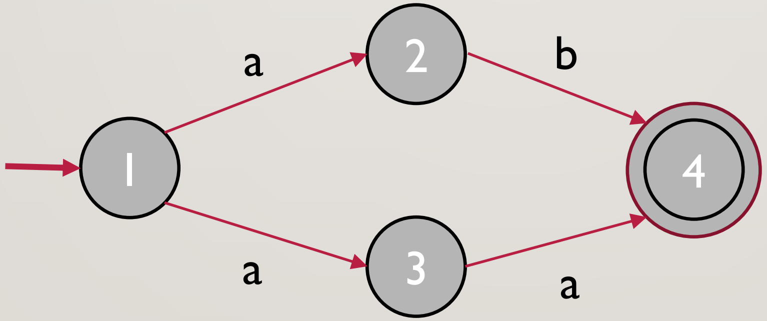
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Exemple :



Complet, déterministe mais pas monogène.

L = (a+b)(a+b)(a+b)*



Pas déterministe

L = {aa, ab} = a(a+b)

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Def : On appelle chemin dans un automate toute suite d'états partant d'un état i et atteignant un état j.

- Def : Un chemin valide va de l'état initial à un état final.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Def : A chaque chemin correspond un mot (par les lettres des transitions). Un mot reconnu est un mot d'un chemin valide.
- Def : On appelle langage accepté/reconnu par un automate A l'ensemble des mots reconnus de A : noté L(A)

Notez que pour les automates non déterministes, il peut y avoir plusieurs chemins correspondant au même mot. Le mot sera reconnu si au moins un des chemins correspondant à ce mot est un chemin valide.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Def : Soit Σ un alphabet et $L \subset \Sigma^*$, alors L est un langage reconnaissable ssi il existe un automate fini sur Σ tel que $L=L(A)$

L'ensemble des langages reconnaissables est noté $Rec(\Sigma)$

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Proposition :
  - $L1 \bullet L2 \in Rec(\Sigma)$
  - $L1 \cup L2 \in Rec(\Sigma)$
  - $L1 \cap L2 \in Rec(\Sigma)$
  - $\Sigma^* - L1 \in Rec(\Sigma)$ Le complémentaire
  - $L1^* \in Rec(\Sigma)$

Preuve : Il suffit pour chaque cas d'appliquer la définition, i.e. trouver un automate qui reconnait le langage considéré.

Corollaire : Les reconnaissables sont fermés par $\cup$, $\bullet$ et $*$ donc $Rat(\Sigma) \subset Rec(\Sigma)$

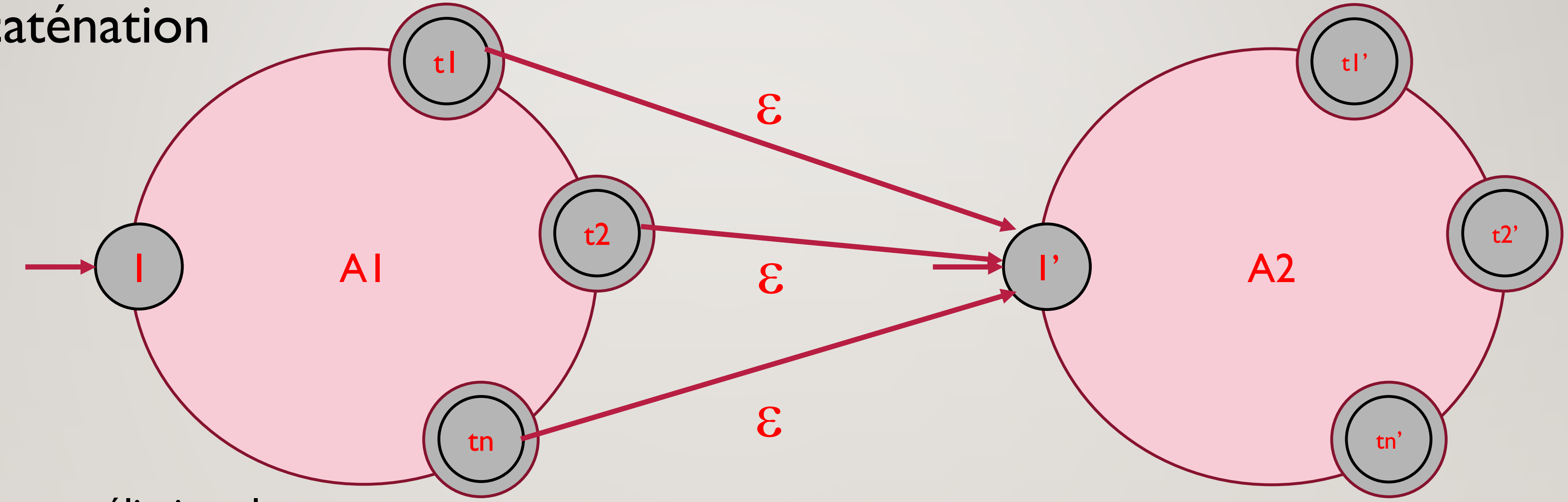
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Proposition :
  - $L1 \bullet L2 \in Rec(\Sigma)$
  - $L1 \cup L2 \in Rec(\Sigma)$
  - $L1 \cap L2 \in Rec(\Sigma)$
  - $\Sigma^* - L1 \in Rec(\Sigma)$
  - $L1^* \in Rec(\Sigma)$

Preuve : Il suffit pour chaque cas d'appliquer la définition, i.e. trouver un automate qui reconnait le langage considéré.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Concaténation



Nous verrons comment éliminer les transitions en ε. En particulier il faut faire un petit peu attention aux états terminaux

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

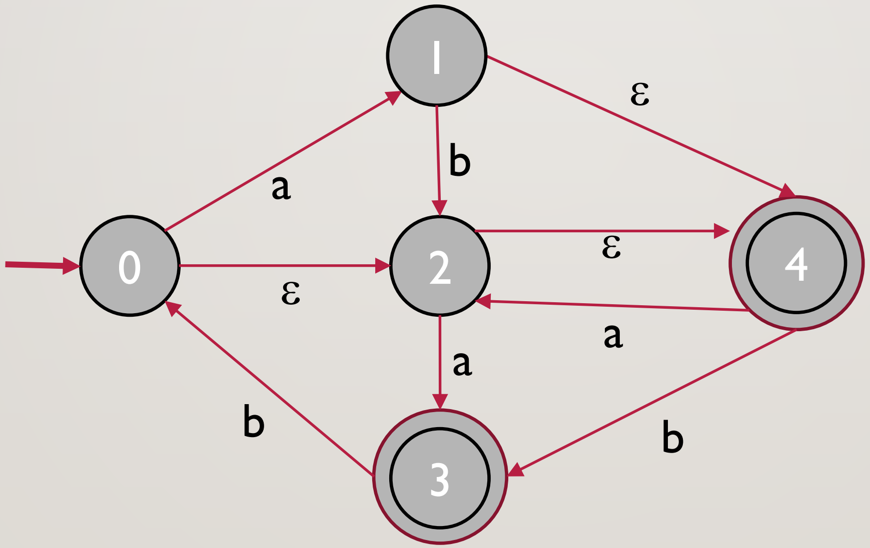
- Elimination des ε-transitions
  - Pour éliminer les ε-transitions, il faut comprendre qu'une e-transition c'est comme se téléporter d'un état à un autre sans consommer de lettre. Pour les éliminer, on va calculer la ε-fermeture càd qu'on va regarder à partir d'un état vers quels autres états on peut aller avec ε et appliquer l'algorithme de fermeture transitive.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Elimination des ε-transitions : Exemple : A={{a,b},{0,1,2,3,4},0,{3,4},d}

d : table de transitions

	a	b	ε
0	1		2
1		2	4
2	3		4
3		0	
4	2	3	



ε-cloture pour chaque état

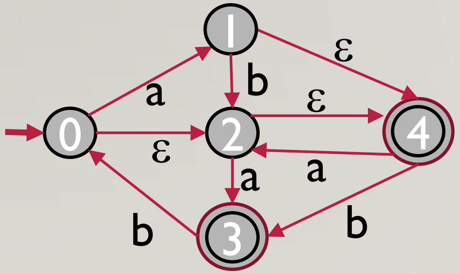
Etats	Atteints par ε
0	0, 2, 4
1	1, 4
2	2, 4
3	3
4	4

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Elimination des ε-transitions :

	a	b	e
0	1		2
1		2	4
2	3		4
3		0	
4	2	3	

Etats	Atteints par ε
0	0, 2, 4
1	1, 4
2	2, 4
3	3
4	4



Exemple au tableau

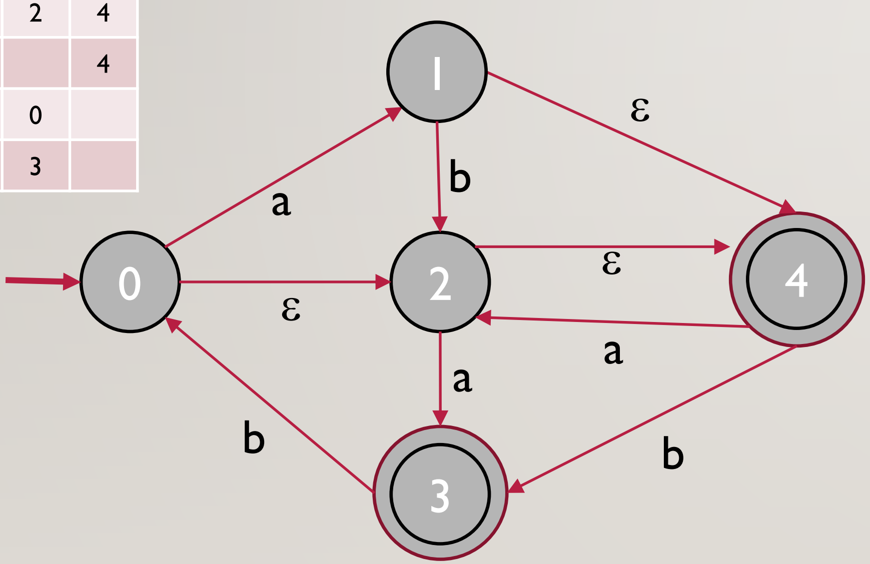
Dernière partie : Si on a $\mathbf{s} \xrightarrow{\varepsilon} \mathbf{t}$ avec **t** état terminal alors **s** devient aussi terminal.

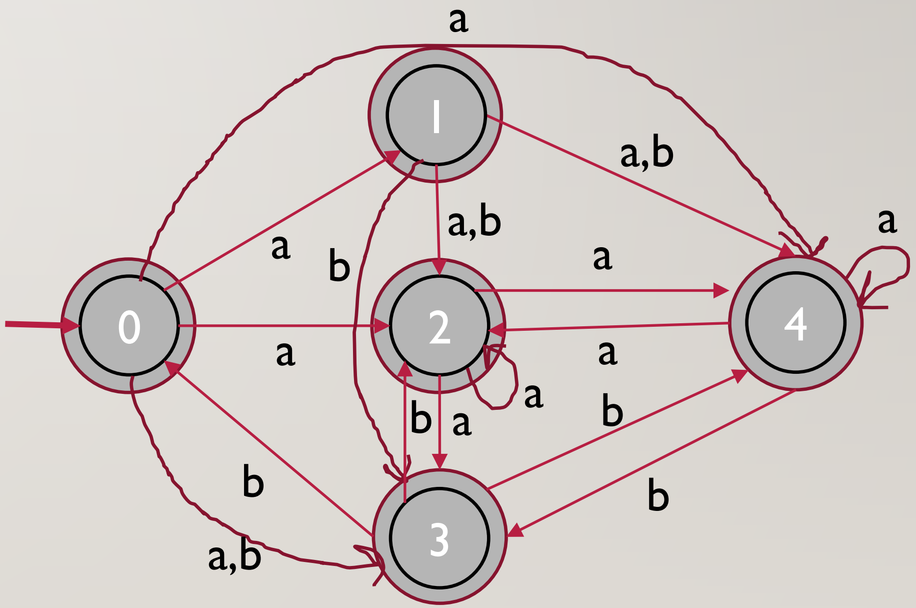
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Elimination des ε-transitions :

	a	b	e
0	1		2
1		2	4
2	3		4
3		0	
4	2	3	

	a	b
0	1,2,3,4	3
1	2,4	2,3,4
2	2,3,4	3
3		0,2,4
4	2,4	3





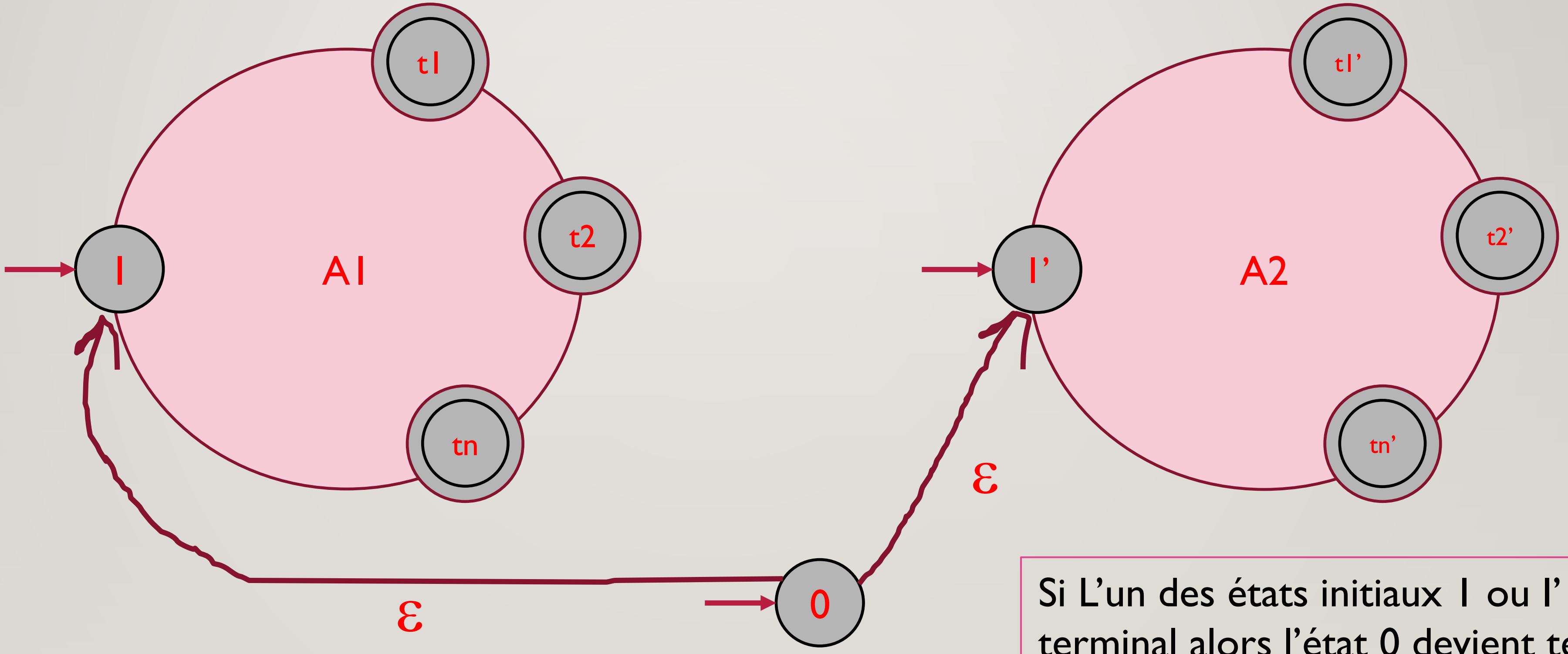
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Proposition :
  - $L1 \bullet L2 \in Rec(\Sigma)$
  - $L1 \cup L2 \in Rec(\Sigma)$
  - $L1 \cap L2 \in Rec(\Sigma)$
  - $\Sigma^* - L1 \in Rec(\Sigma)$
  - $L1^* \in Rec(\Sigma)$

Preuve : Il suffit pour chaque cas d'appliquer la définition, i.e. trouver un automate qui reconnait le langage considéré.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Union



Si L'un des états initiaux 1 ou I' est terminal alors l'état 0 devient terminal

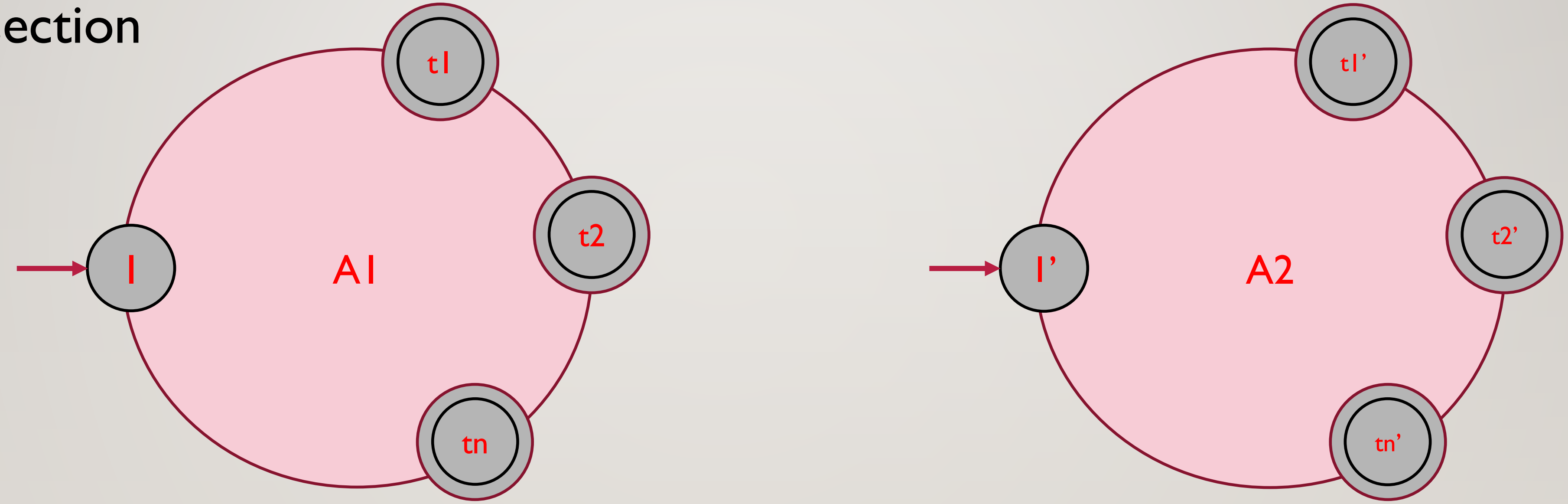
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Proposition :
  - $L1 \bullet L2 \in Rec(\Sigma)$
  - $L1 \cup L2 \in Rec(\Sigma)$
  - $L1 \cap L2 \in Rec(\Sigma)$
  - $\Sigma^* - L1 \in Rec(\Sigma)$
  - $L1^* \in Rec(\Sigma)$

Preuve : Il suffit pour chaque cas d'appliquer la définition, i.e. trouver un automate qui reconnait le langage considéré.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Intersection



L'idée générale est la suivante : On avance en // sur les **deux automates complets et déterministes**. Si on stoppe en même temps sur un état terminal de chaque coté alors le mot est reconnu par les deux. **On verra comment ça marche pour la déterminisation.**

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Comment rendre un automate complet ? Ici $\Sigma = \{a,b\}$

Pour compléter un automate on crée un PUIT qui correspond à tous les chemins manquants

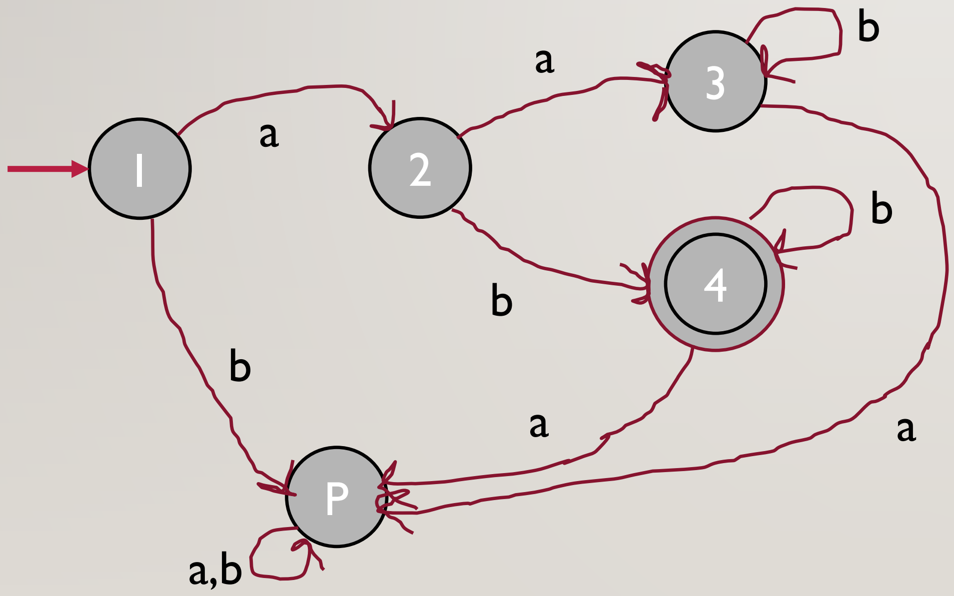


Table de Transition

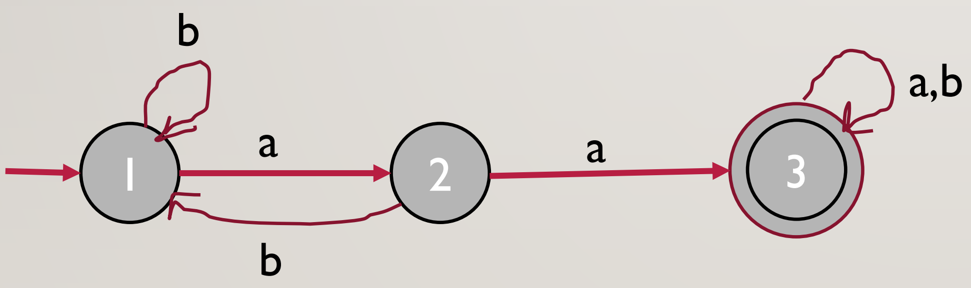
	a	b
1	2	P
2	3	4
3	P	3
4	P	4
P	P	P

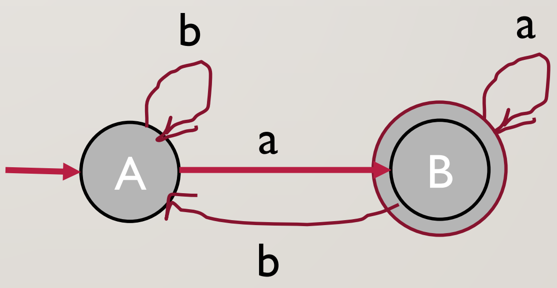
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Exemple d'Intersection : $\Sigma=\{a,b\}$
- Langage L1 = mots contenant aa = $(a+b)^*\ aa\ (a+b)^*$
- Langage L2 = mots se terminant par a = $(a+b)^*\ a$

A1

	a	b
1	2	1
2	3	1
3	3	3





A2

	a	b
A	B	A
B	B	A

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Exemple d'Intersection : Σ={a,b}

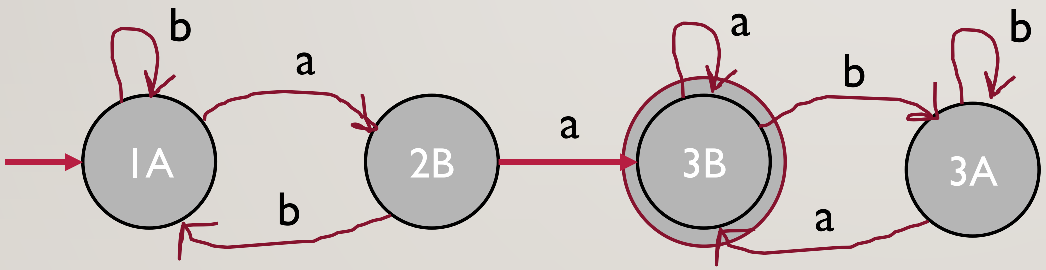
A1

	a	b
1	2	1
2	3	1
3	3	3

	a	b
A	B	A
2	B	A

A2

	a	b
1A	2B	1A
2B	3B	1A
3B	3B	3A
3A	3B	3A



L'état terminal est un état terminal pour les 2 automates (pour l'intersection, un mot doit être reconnu par les 2) : ici 3 et B : 3B

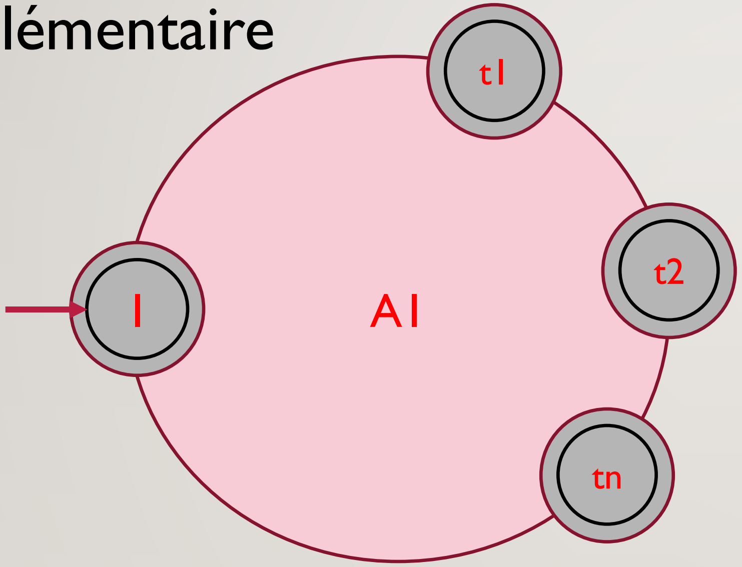
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Proposition :
  - $L1 \bullet L2 \in Rec(\Sigma)$
  - $L1 \cup L2 \in Rec(\Sigma)$
  - $L1 \cap L2 \in Rec(\Sigma)$
  - $\Sigma^* - L1 \in Rec(\Sigma)$ Le complémentaire
  - $L1^* \in Rec(\Sigma)$

Preuve : Il suffit pour chaque cas d'appliquer la définition, i.e. trouver un automate qui reconnait le langage considéré.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Complémentaire

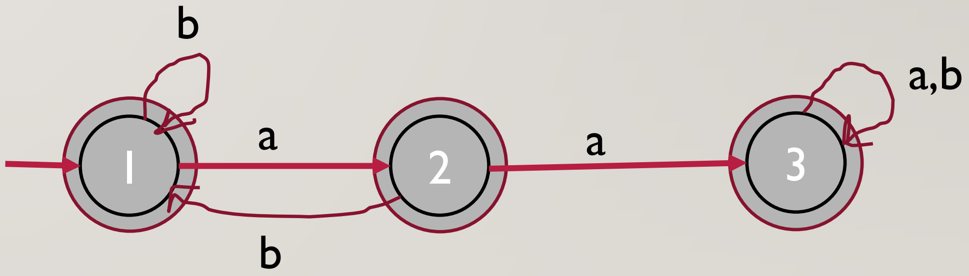


Le langage complémentaire de L se définit par Σ^* - L

Pour les expressions rationnelles, décrire le langage complémentaire peut être compliqué :

Ex: L = mots contenant aa sur $\Sigma=\{a,b\}$: $(a+b)^*aa(a+b)^*$

Ex: $\overline{L}$ = mots sans facteur aa : $b^*\ a(b^+\ a)^*$: pas immédiat



Il suffit de changer les états terminaux en non-terminaux et vice-versa

si l'automate est complet et déterministe

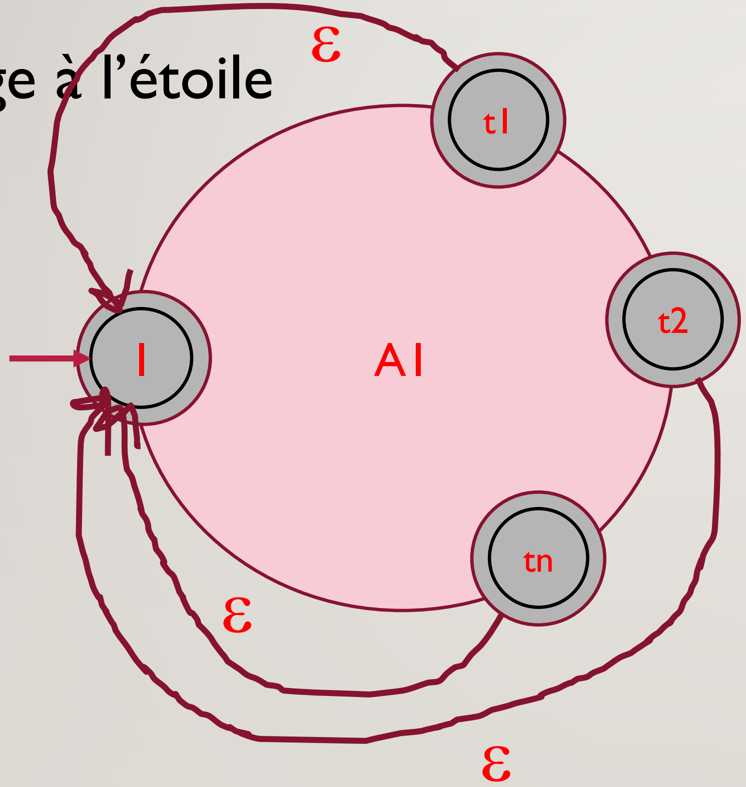
LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Proposition :
  - $L1 \bullet L2 \in Rec(\Sigma)$
  - $L1 \cup L2 \in Rec(\Sigma)$
  - $L1 \cap L2 \in Rec(\Sigma)$
  - $\Sigma^* - L1 \in Rec(\Sigma)$
  - $L1^* \in Rec(\Sigma)$

Preuve : Il suffit pour chaque cas d'appliquer la définition, i.e. trouver un automate qui reconnait le langage considéré.

LANGAGES RATIONNELS ET LANGAGES RECONNAISSABLES

- Passage à l'étoile



L'idée c'est de fusionner les états terminaux avec l'état initial pour pouvoir boucler sur l'automate et reconnaître un nouveau mot sur l'automate.

Evidemment, l'état initial devient terminal.