Chapitre 1.

Langages réguliers et Automates

I. Langages réguliers

L'ensemble des langages réguliers est stable par *union*, *concaténation*, *intersection*, *différence symétrique*, *complémentaire*, et par passage à l'*étoile*.

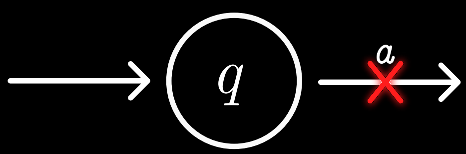
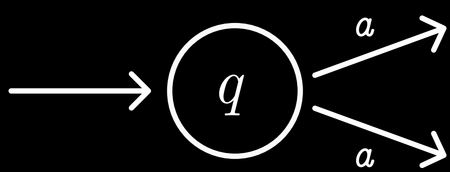
par propriétés des automates — par définition de LR

Tout langage *fini* est régulier.

Le langage d'une *expression régulière* est régulier.

II. Automates finis

À la lecture d'une lettre quelconque, et à un état quelconque :

il n'y a *pas* de chemin à suivre	il y a *plusieurs* chemins à suivre
	

Alors il n'est pas

COMPLET	DÉTERMINISTE

Mais,

on peut le compléter en ajoutant un état « poubelle ».	on peut le déterminiser en transformant considérant les ensembles des états possibles.

On passe de ... à ...

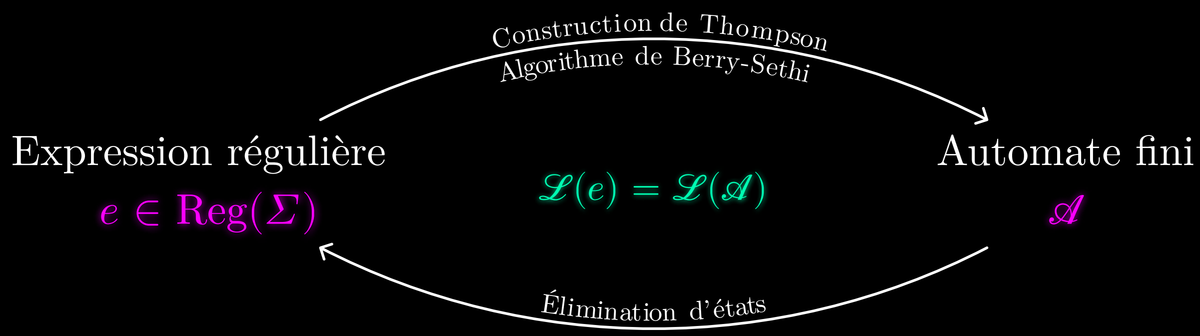
$n \to n+1$	$n \to 2^n$

Pour un automate *déterministe* et *complet*, on définit :

- la fonction de transition $\delta : Q \times \Sigma \to Q$, avec $\delta(q, a) = q'$ et $(q, a, q') \in \delta_{\mathscr{A}}$
- la fonction de transition étendue $\delta^\star : Q \times \Sigma^\star \to Q$, avec

$$\delta^\star(q, \varepsilon) = q \qquad \text{et} \qquad \delta^\star(q, w \cdot a) = \delta(\delta^\star(q, w), a)$$

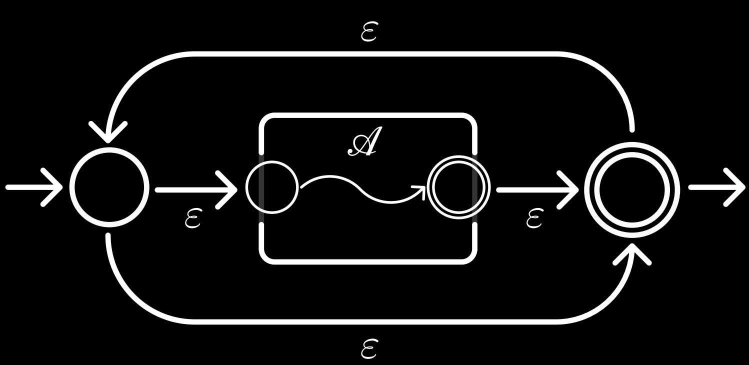
III. Théorème de Kleene



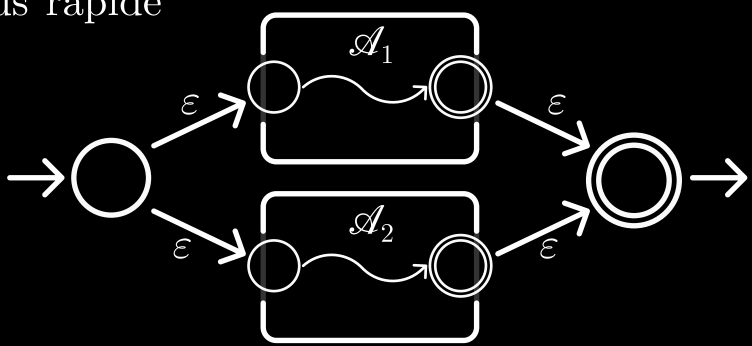
A. *Construction de Thompson*

On « déconstruit » l'expression régulière pour se trouver dans un des cas de base ci-dessous. Cette construction est généralement plus rapide que l'algorithme de Berry-Sethi.

Clôture par étoile :

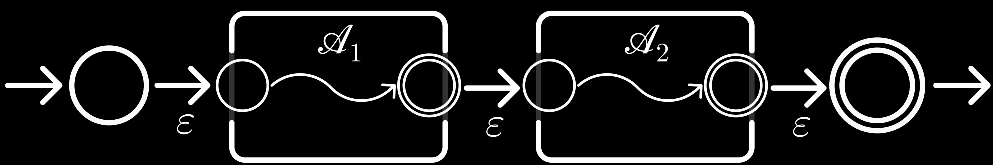


Clôture par union :

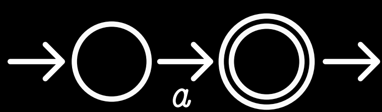


Langage vide $\varnothing$:

Clôture par concaténation :



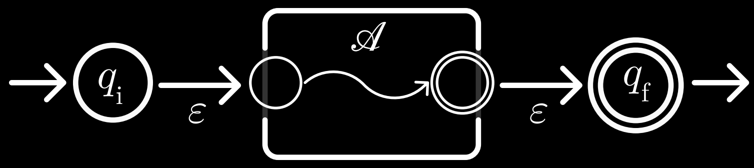
Langage $\{ a \}$:



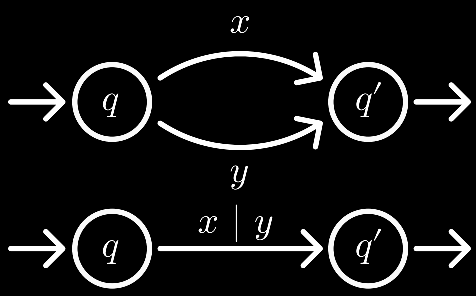
B. *Élimination d'états*

ALGORITHME D'ÉLIMINATION D'ÉTATS

- On « détoure » l'automate :



- Pour chaque état $q \in Q$, avec $q_\mathrm{i} \neq q \neq q_\mathrm{f}$:



- Combiner les transitions sortantes de q

  si $q \xrightarrow{x} q'$ et $q \xrightarrow{y} q'$

  alors on les remplace par $q \xrightarrow{x \mid y} q'$

- Éliminer l'état q

  si $p \xrightarrow{x} q \xrightarrow{y} r$

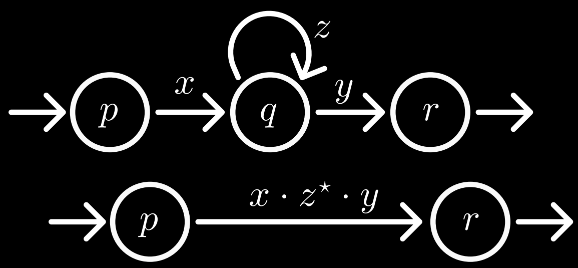
  alors

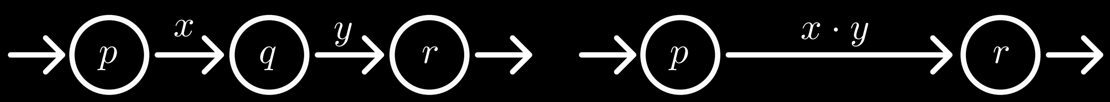
  si $q \xrightarrow{z} q$

  alors on les remplace par $p \xrightarrow{x \cdot z^\star \cdot y} r$

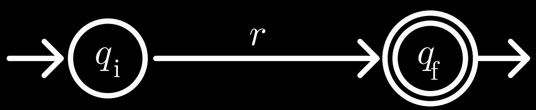
  sinon

  on les remplace par $p \xrightarrow{x \cdot y} r$





- On lit l'expression régulière : $\mathscr{L}(r) = \mathscr{L}(\mathscr{A})$



C. *Algorithme de Berry-Sethi, Construction de Glushkov*

ALGORITHME DE BERRY-SETHI

- On numérote les lettres présentes dans l'expression e pour obtenir f

$$e = aab(a|b)^\star \qquad \longrightarrow \qquad a_1a_2b_3(a_4|b_5)^\star = f$$

- On calcule <u>inductivement</u> $\Lambda(f)$, $P(f)$, $S(f)$, $F(f)$

$$\left(\!\!\left(\begin{array}{l} \Lambda \text{ : est-ce que } \varepsilon \text{ est accepté par } f \\ P \text{ : préfixes de } f \\ S \text{ : suffixes de } f \\ F \text{ : séquences de 2 lettres successives possibles} \end{array}\right.\right.$$

- On en déduit un automate $\mathscr{A}$ reconnaissant $\mathscr{L}(f)$
  - Les états sont des lettres numérotés, ou ε.
  - L'état ε est initial. Il peut-être final si $\Lambda(f) = \{\varepsilon\}$.
  - Les états sortants, ce sont les lettres suffixes.
  - Pour chaque facteur ab, on ajoute une transition de a à b étiquetée par a.

- On dénumérote l'automate $\mathscr{A}$

IV. Limites des langages réguliers

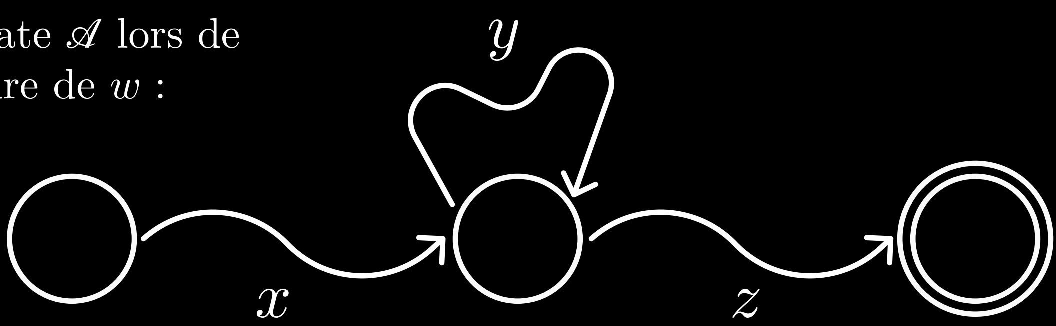
Le *lemme de l'étoile* :

Pour un automate $\mathscr{A}$ à n états, pour un mot w reconnu avec $|w| \geq n$, alors il existe trois mots x, y et z tels que :

$$
\begin{aligned}
x \cdot y \cdot z &= w && \text{« relation de \underline{structure} du mot »} \\
y &\neq \varepsilon && \text{« \underline{utilité} du lemme de l'étoile »} \\
|x \cdot y| &\leq n && \text{« \underline{contrainte} sur les trois mots »} \\
\forall p \in \mathbb{N},\ x \cdot y^p \cdot z &\in \mathscr{L}(\mathscr{A}) && \text{« \underline{récurrence} : ce qui montre l'absurdité »}
\end{aligned}
$$

On choisit le mot w, mais pas les sous-mots x, y et z.

Automate $\mathscr{A}$ lors de la lecture de w :



Par exemple, le langage $L = \{\ a^n\, b^n \mid n \in \mathbb{N}\ \}$ n'est pas régulier.