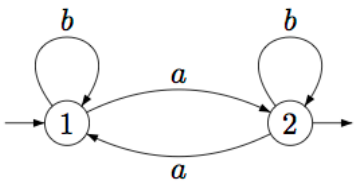
5 – Clôture des langages rationnels



Récapitulatif

- Par définition, l'ensemble des langages rationnels est clos par :
  - concaténation
  - union
  - étoile
- Cette classe est également close par :
  - complémentation
  - intersection
- Chaque langage rationnel est :
  - caractérisé (et reconnu) par son unique automate minimal
  - décrit par une infinité d'E.R.
  - reconnu par une infinité de A.F.D. ou de A.F.N.
  - engendré par une infinité de grammaires régulières ou non.

Clôture par complémentation

L langage rationnel sur Σ

$\Rightarrow$

$\overline{L} = \Sigma^*\backslash L$ langage rationnel sur Σ

déterministe + complet suffiraient ici

Construction :

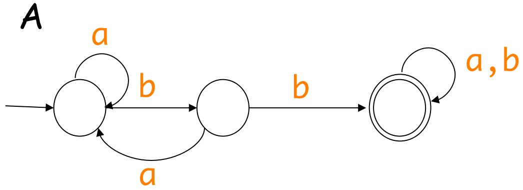
- L un L.R. sur $\Sigma \Rightarrow$ il existe un A.F.D. minimal $A = (Q, \Sigma, \delta, q_0, F)$ tel que $L(A) = L$
- on construit C à partir de A pour qu'il reconnaisse $\overline{L}$:

  $C = (Q, \Sigma, \delta, q_0, Q\backslash F)$

➔ *tout état non final est devenu final et vice-versa.*

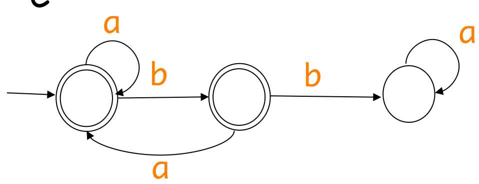
Exemple

A



$L(A) = (a+b)^*bb(a+b)^*$

C



$L(C) = \overline{L(A)}$

i.e. les mots exempts du facteur bb sur {a,b}

Cette construction n'est correcte que parce que A est complet et déterministe !

Clôture par intersection

L et K langages rationnels sur Σ

$\Rightarrow$

$L \cap K$ langage rationnel sur Σ

Démonstration :

- ensemble des L.R. clos par complémentation donc $\overline{L}$ et $\overline{K}$ rationnels
- ensemble des L.R. clos par union donc l'union $\overline{L} \cup \overline{K}$ est rationnelle
- de par la clôture par complémentation, rationalité du langage :

$$\overline{\overline{L} \cup \overline{K}}$$

- par identité ensembliste (De Morgan), on déduit que l'intersection $L \cap K$ est rationnelle.

Construction de l'intersection*

- Soient L et K deux langages rationnels et les A.F.D. :
  - $A = (Q, \Sigma, \delta, q_0, F)$ tel que $L(A) = L$
  - $B = (Q', \Sigma, \delta', q_0', F')$ tel que $L(B) = K$

alors :

$$C = (Q \times Q', \Sigma, \delta_C, [q_0, q_0'], F \times F')$$

avec pour tout couple p,q de $Q \times Q'$,
pour tout σ de Σ,

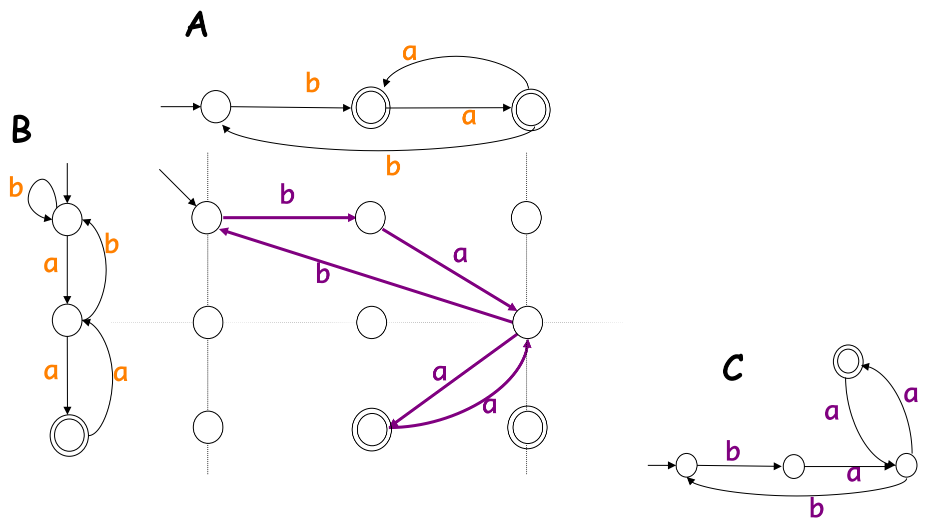
$$\delta_C([p,q], \sigma) = [\delta(p, \sigma), \delta'(q, \sigma)]$$

- *par construction*, l'automate fini C reconnaît le langage rationnel

$$L(C) = L \cap K$$

* On pourrait construire l'union de façon similaire : *cf. TD 5 !*

Exemple



Autres propriétés

L langage rationnel sur Σ

$\Rightarrow$

Pref(L) et Suff(L) sont des langages rationnels sur Σ

- soit un A.F.D. minimal $A = (Q, \Sigma, \delta, q_0, F)$ tel que $L(A) = L$ et p son éventuel état-puits. L'automate $P = (Q, \Sigma, \delta, q_0, Q \setminus \{p\})$ reconnaît le langage Pref(L)
- soit δ' la relation δ augmentée des transitions (q_0, ε, q) pour tout état q de Q, l'automate $S = (Q, \Sigma, \delta', q_0, F)$ est un A.F.N reconnaissant Suff(L)
- de façon similaire, on obtiendrait :

L langage rationnel sur Σ

$\Rightarrow$

pour tout $u \in \Sigma$, les quotients gauches $\{u\}^{-1}L$ ou les droits $L\{u\}^{-1}$ sont des langages rationnels sur Σ

Des langages non rationnels ?

Georg Ferdinand Ludwig Philip Cantor
(1845-1918)

On raisonne sur un alphabet donné

- par un argument de diagonalisation, on a obtenu (cours 2) que l'ensemble des langages *n'est pas* dénombrable

démontré en TD !
cf. TD 2 ex.4

- l'ensemble des expressions régulières est dénombrable
- donc l'ensemble des langages rationnels est dénombrable

➔ on en déduit qu'il existe des ***langages non rationnels***

D'un raisonnement similaire sur les automates, on aurait aussi pu déduire l'existence de langages non rationnels.

Problème

- Donnée un langage L sur l'alphabet Σ
- Problème le langage L est-il rationnel ?

Approche du problème :

- Le **lemme de l'étoile** va servir à montrer qu'un langage ***n'est pas*** rationnel
- dans le cas où le lemme s'avère inutilisable*, il faut une démonstration *ad hoc*

Ce lemme n'est pas une caractérisation des langages rationnels.
Il ne permet pas de montrer la non-rationalité de tous les langages non-rationnels.

Lemme de l'étoile : idée

- Soit *A* un automate fini comportant n états
- Lors de la lecture d'un mot de longueur m avec n ≤ m, l'automate *A* passe forcément deux fois par le même état
- Intuitivement, le lemme énonce :

  « quand un mot w a un facteur v de longueur plus grande que le nombre d'états de l'automate, la lecture de v boucle forcément quelque part ... »

Lemme de l'étoile

- Intuitivement, le lemme énonce :
  « quand un mot w a un facteur v de longueur plus grande que le nombre d'états de l'automate, la lecture de v boucle forcément quelque part ... »

Lemme de l'étoile
soit L un langage **rationnel**, il existe un entier n, n ≥ 1 tel que :
- pour tout mot w dans L de longueur ≥ n,
- pour toute factorisation de w en x v y avec v facteur de longueur ≥ n
- il existe une factorisation de v en r s t telle que :
  - i) $0 < |s| \leq n$
  - ii) $\forall\, i \geq 0,\ x\, r\, s^i\, t\, y \in L$

Démonstration

- $A = (Q, \Sigma, \delta, q_0, F)$ automate minimal à n états tel que $L = L(A)$
- prenons un mot $w = x\,v\,y$ de L avec $|v| \geq n$: $w = \sigma_1\sigma_2 \ldots \sigma_{|w|}$
  on sait que $\delta^*(q_0, \sigma_1\sigma_2\ldots\sigma_{|w|}) \in F$
- lors de la lecture du facteur v de w, A passe forcément 2 fois dans un même état q de Q
- il existe donc 2 lettres d'indices j et k avec $j<k$ et $|x|<j \leq |xv|$ et $|x|<k \leq |xv|$ telles que :

$$q = \delta^*(q_0, \sigma_1 \ldots \sigma_j) = \delta^*(q_0, \sigma_1 \ldots \sigma_k)$$

- si on pose : $x\,r = \sigma_1 \ldots \sigma_j$
  $s = \sigma_{j+1} \ldots \sigma_k$
  $t\,y = \sigma_{k+1} \ldots \sigma_{|w|}$

  pour tout $i \geq 0$, les mots $\boxed{x\,r\,s^i\,t\,y \in L}$
- $j<k$ donc $|s| > 0$ et $|s| \leq n$ en prenant j et k minimaux.

Un premier langage démontré non-rationnel

- soit le langage $L = \{ 0^m1^m, m \geq 0 \}$
- supposons *par l'absurde* qu'il est rationnel
- quel que soit l'entier $n \geq 1$ fourni par le lemme :
- soit le mot $w = 0^n1^n$ de L, de longueur $|w| = 2n \geq n$
- soit la factorisation de w en $x\,v\,y$ avec :

$$x = \varepsilon$$
$$v = 0^n$$
$$y = 1^n$$

- pour toute factorisation de v en r, s et t tels que :

$$0 < |s| \leq n$$

  on cherche un i tel que :

$$x\,r\,s^i\,t\,y \notin L$$

- si on prend $i = 0$: $z_0 = x\,r\,t\,y = 0^k1^n \notin L$ car $k < n$
  $\Rightarrow$ contradiction avec le lemme
- **Conclusion :** $L = \{0^m1^m, m \geq 0 \}$ n'est pas rationnel.

Autre idée ...

Utiliser les propriétés de clôture !

Elles peuvent suffire à démontrer qu'un langage n'est pas rationnel, dès lors que l'on connaît déjà des langages non rationnels ...

autant de 0 que de 1 !

Exemple :

- considérons le langage $L = \{ w \in \{0,1\}^*, |w|_0 = |w|_1 \}$
- est-il rationnel ? ... supposons que oui
- on sait que le langage K décrit par 0^*1^* est rationnel
- $K \cap L = \{ 0^n1^n, n \geq 0 \}$ donc $K \cap L$ n'est pas rationnel
- or on sait que l'intersection de 2 langages rationnels est un langage rationnel $\Rightarrow$ contradiction

Conclusion : $L = \{ w \in \{0,1\}^*, |w|_0 = |w|_1 \}$ n'est pas rationnel.