Maths 4.2

Au delà des langages réguliers

Florent Madelaine

BUT 2 Informatique

UNIVERSITÉ PARIS-EST CRÉTEIL

IUT SENART FONTAINEBLEAU

Plan

Introduction

Nous avons vu brièvement le modèle des automates finis et la correspondance avec les expressions régulières, avec des méthodes pour passer de l'un à l'autre, en particulier avec des ateliers disponibles dans JFLAP.

Nous voyons maintenant une méthode dite du lemme de la pompe qui permet parfois de montrer qu'un langage n'est pas régulier.

Objectif

Question

Étant donné un langage L, existe-t-il un automate fini qui reconnaît ce langage ?

Objectif

Question

Étant donné un langage L, existe-t-il un automate fini qui reconnaît ce langage ?

- Si le langage L est donné par une expression régulière, on sait que la réponse est OUI, et on sait même construire un automate fini qui reconnaît ce langage.
- Mais si la réponse est NON, autrement dit s'il n'existe aucun automate fini qui reconnaît le langage L, comment le prouver ?

Objectif

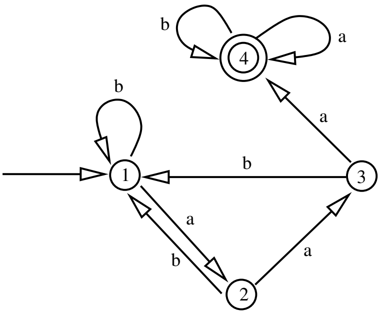
Question

Étant donné un langage L, existe-t-il un automate fini qui reconnaît ce langage ?

Outil pour le NON

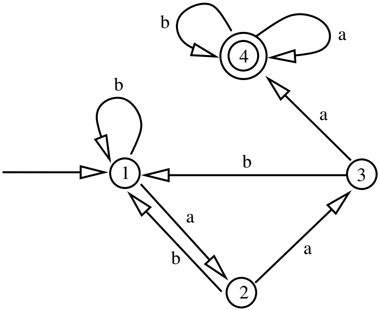
- On va voir une propriété (appelée le lemme de la pompe) qui est vraie pour tous les langages réguliers
- Par conséquent, si un langage ne vérifie pas cette propriété, il est impossible qu'il soit un langage régulier

Ce que pomper veut dire



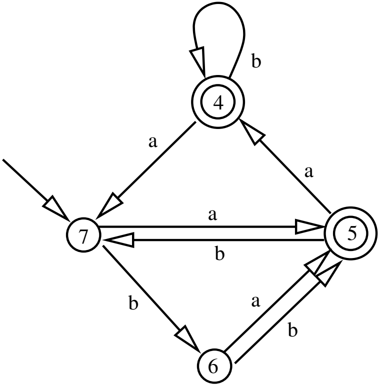
- *aabbaaaba* accepté
- *aabaabbaaaba* accepté
- *aabaabaabaabbaaaba* accepté
- N'importe quel mot de la forme $(aab \ldots aab)baaaba$ accepté
- Car il y a un cycle $1 \xrightarrow{a} 2 \xrightarrow{a} 3 \xrightarrow{b} 1$
- On dit qu'on peut pomper *aab* dans le mot *aabbaaaba*

Pomper ici ou pomper ailleurs



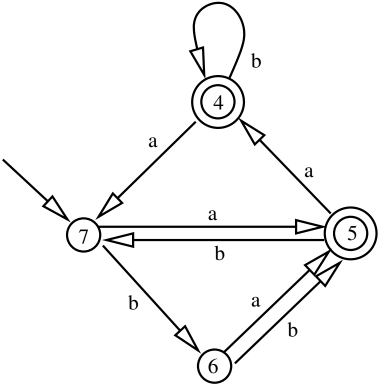
- *aabbaaaba* accepté
- *aabbbaaaba* accepté
- *aabbbbbaaaba* accepté
- N'importe quel mot de la forme $aab(b \ldots b)aaaba$ accepté
- Car il y a un cycle $1 \xrightarrow{b} 1$
- On peut aussi pomper *b* dans le mot *aabbaaaba*

Peut-on toujours pomper ?



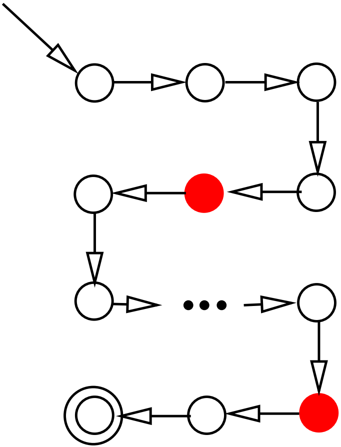
- *bbbaabaa* accepté
- Peut-on pomper quelque chose quelque part ?
- *bbbaabaa*
  (cycle $5 \xrightarrow{b} 7 \xrightarrow{a} 5$)
- ou *bbbaabaa*
  (cycle
  $5 \xrightarrow{a} 4 \xrightarrow{b} 4 \xrightarrow{a} 7 \xrightarrow{a} 5$)

Peut-on toujours pomper ?



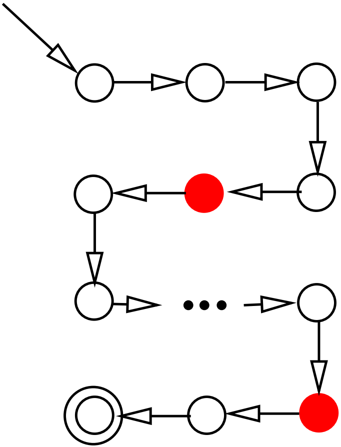
- *bb* accepté
- Peut-on pomper quelque chose quelque part ?
- Non : aucune partie de *bb* ne correspond à un cycle sur l'automate

Dans quels mots peut-on pomper ?



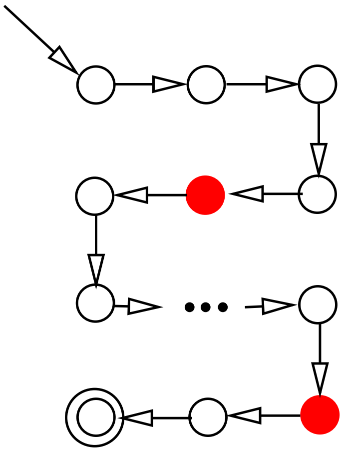
- On peut pomper chaque fois qu'une partie d'un mot accepté correspond à un cycle sur l'automate

Dans quels mots peut-on pomper ?



- Dès que la longueur d'un mot accepté est plus grande que le nombre d'états de l'automate, on doit forcément repasser par (au moins) un état en lisant ce mot $\Longrightarrow$ il y a un cycle

Dans quels mots peut-on pomper ?



- Finalement, dès que la longueur d’un mot accepté est plus grande que le nombre d’états de l’automate, on est sûr de pouvoir pomper une partie du mot

Lemme de la pompe

Théorème

Soit L un langage reconnu par un automate $\mathcal{A}$ avec n états. Alors on peut pomper quelque chose dans tout mot de L de longueur $\geqslant n$.

Autrement dit :
Tout mot w de L de longueur $\geqslant n$ se décompose sous la forme $w = uxv$ de telle sorte que :

1. $|ux| \leqslant n$
2. $x \neq \varepsilon$
3. Tout mot de la forme $u(x \dots x)v$ appartient à L

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \dots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \dots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

Utilisation du lemme de la pompe

$L = \{a^k b^k, k \geqslant 0\} = \{\varepsilon, ab, aabb, aaabbb, \ldots\}$

On va montrer que L n'est pas régulier par un raisonnement par l'absurde.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n b^n \in L$ avec $|w| = 2n \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} b^n$, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, c.-à-d. L n'est pas un langage régulier

D'autres langages non réguliers

- Le langage des palindromes
  - Mot qui se lit pareil de droite à gauche et de gauche à droite
  - En français : RADAR, KAYAK, RESSASSER, . . .
  - On note u^R le mot u renversé : bba devient abb
  - Tout mot w de la forme uu^R est un palindrome
- Le langage des parenthèses bien formées
  - $(())()(())$ ou $(((())))$
  - mais pas $())($ ni $(()(()$
- Le langage composé des mots qui contiennent le même nombre de a que de b
- etc.

Autre exemple

L = le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d}bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

$L =$ le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \dots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

L = le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \dots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d}bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

$L =$ le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

$L =$ le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n+2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \dots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d}bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

$L =$ le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \Longrightarrow$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \dots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d} bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

$L =$ le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \dots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d}bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

L = le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \ldots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d}bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Autre exemple

$L =$ le langage des palindromes

On va montrer par l'absurde que L n'est pas régulier.

- Supposons que L soit reconnu par un automate $\mathcal{A}$ et appelons n le nombre d'états de $\mathcal{A}$.
- Le mot $w = a^n bba^n \in L$ avec $|w| = 2n + 2 \geqslant n \implies$ pomper
- C.-à-d. $w = uxv$ avec $|ux| \leqslant n$, $x \neq \varepsilon$ et $u(x \dots x)v \in L$
- Donc $x = a^d$ avec $d > 0$
- Par exemple, $w' = uxxv$ appartient à L
- Mais $w' = a^{n+d}bba^n$ n'est pas un palindrome, donc w' n'appartient pas à L
- Contradiction
- Donc L n'est pas reconnu par $\mathcal{A}$

Conclusion : Il n'existe aucun automate fini qui reconnaisse le langage L, donc L n'est pas un langage régulier

Une arme absolue ?

- Si on montre qu'un langage ne satisfait pas le lemme de la pompe, c'est une preuve que ce langage n'est pas régulier
- MAIS il peut arriver qu'un langage qui satisfait le lemme de la pompe ne soit pas régulier **:-(**

L'arme absolue existe. Il s'agit du théorème de Myhill-Nerode, mais dépasse le cadre de ce cours.

Guide d’utilisation

Pour décider si L est régulier ou pas :

- On sait construire un automate fini qui reconnaît L
  - Conclusion : L est régulier
- On montre que L ne satisfait pas le lemme de la pompe
  - Conclusion : L n’est pas régulier
- On n’arrive pas à construire un automate fini qui accepte L, mais on montre que L satisfait le lemme de la pompe
  - On ne peut pas conclure

Un langage non régulier qui satisfait le lemme de la pompe pour tout $m \geqslant 3$

$L = \{uu^R v, u \neq \varepsilon, v \neq \varepsilon\}$

(un mot non vide, suivi de son renversé, suivi d'un autre mot non vide)

- Soit $w = uu^R v$ un mot de L (on note que $|w| \geqslant 3$)

Premier cas : u est une lettre

- On prend $x = uu^R$ et $y = v_1$ (première lettre de v)
- On a $|xy| = 3 \leqslant m$ et on peut pomper y en restant dans L

Un langage non régulier qui satisfait le lemme de la pompe pour tout $m \geqslant 3$

$L = \{uu^R v, u \neq \varepsilon, v \neq \varepsilon\}$

(un mot non vide, suivi de son renversé, suivi d'un autre mot non vide)

- Soit $w = uu^R v$ un mot de L (on note que $|w| \geqslant 3$)

Deuxième cas : $|u| \geqslant 2$

- On prend $x = \varepsilon$ et $y = u_1$ (première lettre de u)
- On a $|xy| = 1 \leqslant m$
- Pour $i > 0$, on a $xy^i z = u_1 \dots u_1 z \in L$
- Pour $i = 0$, on a $xz = u_{\geqslant 2} u^R v \in L$ car le début de u^R est le renversé de la fin de u

Un langage non régulier qui satisfait le lemme de la pompe pour tout $m \geqslant 3$

$L = \{uu^R v, u \neq \varepsilon, v \neq \varepsilon\}$

(un mot non vide, suivi de son renversé, suivi d'un autre mot non vide)

- Soit $w = uu^R v$ un mot de L (on note que $|w| \geqslant 3$)

Finalement :

- On peut pomper dans n'importe quel mot de L, c.-à-d. que L satisfait le lemme de la pompe
- Mais on n'arrive pas à construire un automate fini qui accepte L
- Donc on ne peut pas savoir si L est ou non régulier.
- En fait, c'est non, mais c'est une autre histoire...

Au delà des automates

Nous avons entrevu le modèles des automates et à l'instant ses limites.
Nous allons maintenant voir le modèle de Turing, plus puissant.