Automates & Lemme de l'étoile

> je suis pas sûre de la propriété à démontrer et pour le lemme de l'étoile j'ai compris le concept mais je comprends pas ce qui contredit exactement, il faut que la taille du langage soit supérieur au nombre d'états de l'automate acceptant ? Ou que le langage soit de longeur infini ?

Un langage fini est toujours régulier/rationnel/reconnaissable. Le lemme de l'étoile est un outil pour montrer qu'un langage (infini) n'est pas reconnaissable. Il fonctionne de la façon suivante: Ton but c'est de montrer que un langage L n'est pas reconnaissable. Tu lui donnes un langage L que tu suppose reconnaissable (et tu veux arriver à une contradiction). Il te donne un nombre N. Tu lui donnes un mot w dans L plus long que N. En échange, il te donne une décomposition, c'est à dire trois mots x, y, z tels que $w = x.y.z$ et l'information que : $y \neq \varepsilon$ et que $x.y^*.z \subseteq L$. Une fois que tu en es là, tu dois trouver le bon k pour montrer que $x.y^k.z \notin L$ pour arriver à une contradiction.

Donc, ce qui se contredit, c'est que le mot de départ w est bien dans L, mais le mot modifié $x.y^k.z$ lui n'est pas dans L mais devrait être dans L si L était bien reconnaissable!

Pour donner une intuition de ce que c'est qu'un langage non reconnaissable: en gros, comme il y a un ensemble fini d'état, la "mémoire est finie". Et en plus, on ne peut regarder que à un endroit dans le mot, donc on ne peut pas se souvenir de détails arbitrairement gros dans le début de la lecture d'un mot.

Voici un longue liste de langages non reconnaissables

- $\{a^n b^n, n \in \mathbb{N}\}$ voir plus loin pour la preuve
- $\{a^n b^m, n \in \mathbb{N}, m \in \mathbb{N}, n \neq m\}$ puisque les langages reconnaissables sont clot par complémentaire.
- $\{a^n b^m, n \in \mathbb{N}, m \in \mathbb{N}, n \leq m\}$.
- $\{u b^n, n \in \mathbb{N}, |u| = n\}$ par un raisonnement similaire au premier.
- $\{m \in A^* \mid |u|_a = |u|_b\}$ où le nombre de a et de b est le même. Le raisonnement est le même que le premier.
- $\{a^p, p \text{ premier}\}$ on l'a vu en TD
- $\left\{a^{n^2}, n \in \mathbb{N}\right\}$
- $\left\{a^{n^4}, n \in \mathbb{N}\right\}$
- $\left\{a^{n!}, n \in \mathbb{N}\right\}$

Et une liste de langages reconnaissables

- $\{abcdefg\}$ il y a un automate pour chaque mot
- $\{abc, def\}$ il y a un automate pour chaque langage fini
- A^* l'automate boucle
- $\{a^n b^m, n \equiv p \bmod 3\}$ si on ne peut pas se souvenir du nombre exact de a, on peut se souvenir du nombre de a modulo 3.
- Le langage associé à une expression régulière
- Les langages où il y a au moins/exactement/au plus k lettre b pour k fixé.
- $\mathrm{Pref}(L)$ où L est rationnel. Il suffit de determiniser l'automate, et de rendre acceptant tous les états qui peuvent mener vers un état final. C'est une construction extremement pratique pour faire de la prédiction du mot que tu es en train d'écrire!
- Le système de controle de vitesse de la ligne 14.

J'ai regardé ta preuve pour $L = \{a^n b^n, n \in \mathbb{N}\} = \{w \in A^* \mid \exists n, w = a^n.b^n\}$. Voici une version corrigée:

Preuve: On veut montrer que L n'est pas reconnaissable. Supposons qu'il est reconnaissable, et arrivons à une contradiction. Si L est reconnaissable, alors il existe un entier N donné par le lemme de l'étoile. On utilise le lemme de l'étoile sur $a^N.b^N$. On vérifie que $|a^N.b^N| = 2N \geq N$ et que $a^N.b^N$ est bien dans L. Le lemme est content, et te donne en change de ce mot trois mots x, y, z tels que $x.y.z = a^N.b^N$ avec $y \neq \varepsilon$ et $x.y^*.z \subseteq L$. À partir de là, on esssaye d'arriver à une contradiction. Dans ta preuve tu dis "bah du coup $y = a^k$ pour un certain k, donc on a fini". Mais non, tu ne sais pas que $y = a^k$. Le lemme de l'étoile ne te donne pas cette information. Il te dit juste que $x.y.z = a^N.b^N$ La question est, où se trouve la coupure du passage des a aux b dans $x.y.z$. Cela dépend de la taille de x et de y.

- Si $|x| + |y| \leq N$, alors y ne contient que des a. Donc $x.z = a^k.b^N$ avec $k < N$ puisque y est non vide. Ce mot n'appartient pas à L. Contradiction !
- Si $|x| < N$ et $|x + y| > N$, alors y est a cheval entre des a et des b alors $x.y.y.z = \cdots a^k b^{k'} a^k b^{k'} \cdots$ avec $k > 0$ et $k' > 0$. Ce mot n'est pas non plus dans L. Contradiction.
- Si $|x| \geq N$, alors y ne contient que des b, alors, $x.z = a^N.b^k$ avec $k < N$ puisque y est non vide. Ce mot n'appartient pas à L. Contradiction !

Dans tous les cas, on a une contradiction. Donc, notre supposition que le langage est bien reconnaissable était fausse, donc L n'est pas reconnaissable. ■

En fait on peut faire beaucoup plus rapide avec quelque chose qui rejoint plus l'intuition de ce qui se passe, en au lieu d'utiliser le lemme de l'étoile, on adapte sa preuve à la volée pour indiquer quelle partie du chemin on donne au lemme:

Preuve: Supposons qu'il existe un automate $\mathcal{A} = (S, I, F, T)$ tel que $L = \mathcal{L}(\mathcal{A})$. Alors, puisque $a^{|S|}.b^{|S|} \in L$, il existe un chemin acceptant $i \xrightarrow{a^{|S|}} s \xrightarrow{b^{|S|}} f$ avec $i \in I$ et $f \in F$. Puisque le chemin $i \xrightarrow{a^{|S|}} s$ est de taille $|S| + 1$, un état s' est visité deux fois. Il existe donc $x, y, z \in \mathbb{N}$ tels que $|S| = x + y + z$ et $y > 0$ et $i \xrightarrow{a^x} s' \xrightarrow{a^y} s' \xrightarrow{a^z} s$. On peut donc fabriquer le chemin suivant acceptant dans $\mathcal{A}$:

$$i \xrightarrow{a^x} s' \xrightarrow{a^z} s \xrightarrow{b^{|S|}} f$$

Donc $a^{|S| - y}.b^{|S|} \in \mathcal{L}(\mathcal{A})$, mais $y > 0$, donc $a^{|S| - y}.b^{|S|} \notin L$. Absurde! ■