Équivalence AP & algébriques

Problème

- On souhaite un analogue algébrique au théorème de Kleene $(Rat(\Sigma)=Rec(\Sigma))$

- Montrer que tout langage algébrique est reconnaissable par AP (non-déterministe) et réciproquement, tout langage reconnaissable par un AP est algébrique !

Algébrique$\Rightarrow$ reconnaissable

Si L est algébrique, alors il existe un automate à pile qui reconnaît L par pile vide.

Principe de construction de l'automate :

- L algébrique $\Rightarrow$ il existe une grammaire algébrique G qui l 'engendre.
- L'automate à pile M acceptera le mot w si G engendre w.
- M va simuler une suite de dérivations gauches de G qui engendrent w.

Hypothèses de départ :

- le mot vide n'est pas dans L(G)
- G sous forme normale de Greibach (ou presque)

Pour simplifier

- Règles de G sous la forme $X \rightarrow a\gamma$ avec $a \in T$ et $\gamma \in N^*$
  (il suffit d 'ajouter dans γ des règles $C_a \rightarrow a$ (comme en FNC))

Construction

Donnée $G=(N,T,S,R)$, on construit

$$M=(Q=\{q\},\Sigma = T,\Gamma =N, \delta,q,S,\emptyset)$$

- Si $X\rightarrow a\gamma \in R$ on ajoute la transition

$$\{(q,\gamma)\} \in \delta(q,a,X)$$

- M simule les dérivations gauches de G
- Les mots engendrés par dérivations gauches sont de la forme $m\alpha$ pour $m\in T^*$ et $\alpha \in N^*$
- M mémorise α dans la pile après avoir traité m.

Construction

- Pour finir la preuve, il faudrait montrer
  - par récurrence sur le nombre de transitions que
  - par récurrence sur la longueur de la dérivation que

$$S \to^* x\alpha \text{ par dérivation gauche SSI } (q,x,S) \to^* (q,\varepsilon,\alpha)$$

- Pour terminer, notons que

$$x \in L(G) \Leftrightarrow (q,x,S) \to (q,\varepsilon,\varepsilon)$$

Ce qui est le cas pour $\alpha=\varepsilon$ i.e. que

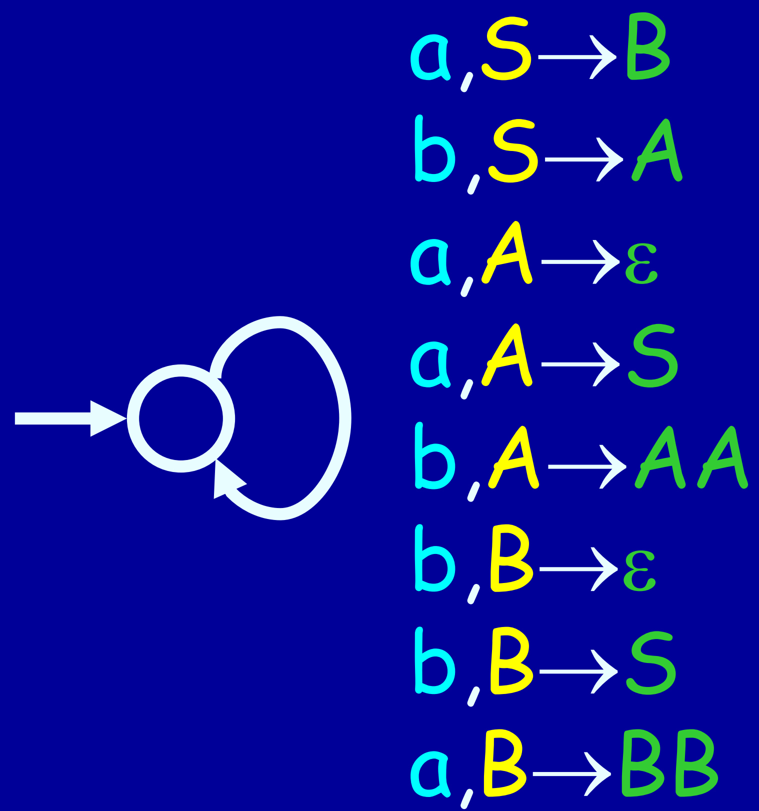
  - x peut être engendré par dérivation gauche par G
  - x est reconnu par pile vide avec M

Exemple

- Grammaire $|w|_a=|w|_b$
- $S \to aB \mid bA$
- $A \to a \mid aS \mid bAA$
- $B \to b \mid bS \mid aBB$

- $X \to a\gamma$

$\{(q,\gamma)\} \in \delta(q,a,X)$



Caractériser les algébriques

Le but :

L algébrique SSI il existe un AP qui le reconnaît.

On a vu

L algébrique $\Rightarrow \exists$ un AP qui reconnaît L.

reste à montrer :

L reconnu par pile vide par AP $\Rightarrow$ L algébrique

Idée

- **Donnée** : un AP M
- **Résultat** : G grammaire algébrique tq L(G)=L(M)
- G engendre w ssi w reconnu par M.
- A chaque paire d'états (p,q) de M, et à chaque symbôle de pile (c.a.d. $A \in \Gamma$) on associe une variable de la forme $A_{p,q}$
- G engendre w depuis $A_{p,q}$ SSI $(p,w,A) \rightarrow^* (q,\varepsilon,\varepsilon)$: une dérivation gauche du mot w est la simulation du fonctionnement de l'AP sur l'entrée w.

Construction

$$M=(Q,\Sigma,\Gamma,\delta,i,Z,\varnothing) \rightarrow G=(N,T,S,R)$$

- $T=\Sigma$,
- $N=\{A_{p,q} : p,q \in Q,\ A \in \Gamma\} \cup \{S\}$

  Une variable $A_{q,p}$ correspond aux mots qu'on peu obtenir par pile vide si on commence en état q et on termine en état p

- Au démarrage :
  - $S \rightarrow Z_{i,q},\ \forall q \in Q$

Construction (2)

- Au démarrage :
  - $S \rightarrow Z_{i,q}, \forall q \in Q$

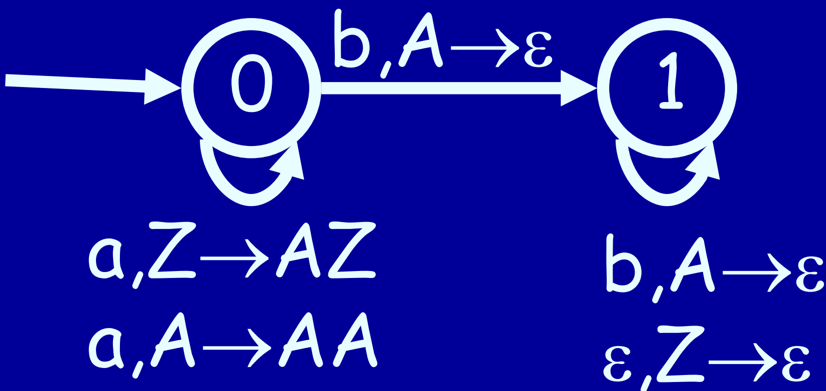
- Régime de croisière :
  - $(q,\varepsilon) \in \delta(p,a,A)$ □ $A_{p,q} \rightarrow a$

  - $(q,B_1B_2\ldots B_m) \in \delta(p,a,A)$ $(m>0)$

$$A_{p,r} \rightarrow a\ B_{1,q,q_2}\ B_{2,q_2,q_3}\ \ldots\ B_{m,q_m,r}$$

$$\forall q_2,\ldots,q_m,r \in Q,\ a \in \Sigma \cup \{\varepsilon\},\ A,B_1,\ldots,B_m \in \Gamma$$

Exemple



Démarrage :

$S \to Z_{0,*}$

$S \to Z_{0,0}$

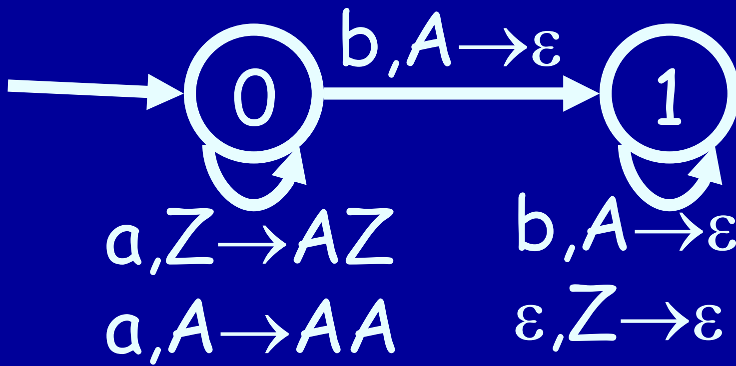
$S \to Z_{0,1}$

- Transition de 0 à 1

$\delta(0,b,A)=(1,\varepsilon)$

$A_{0,1} \to b$

Exemple (la boucle en 0)



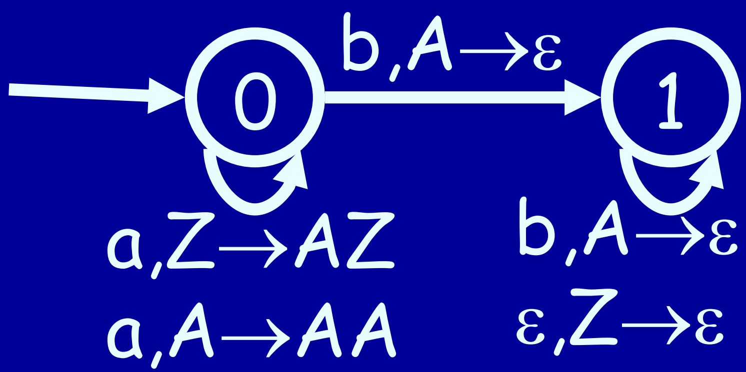
$\delta(0,a,Z)=(0,AZ)$

- $r=0$
  - $Z_{0,0} \to a\ A_{0,*}\ Z_{*,0}$
  - $Z_{0,0} \to a\ A_{0,0}\ Z_{0,0}$
  - $Z_{0,0} \to a\ A_{0,1}\ Z_{1,0}$
- $r=1$
- $Z_{0,1} \to a\ A_{0,*}\ Z_{*,1}$
  - $Z_{0,1} \to a\ A_{0,0}\ Z_{0,1}$
  - $Z_{0,1} \to a\ A_{0,1}\ Z_{1,1}$

$\delta(0,a,A)=(0,AA)$

- $r=0$
  - $A_{0,0} \to a\ A_{0,*}\ A_{*,0}$
  - $A_{0,0} \to a\ A_{0,0}\ A_{0,0}$
  - $A_{0,0} \to a\ A_{0,1}\ A_{1,0}$
- $r=1$
  - $A_{0,1} \to a\ A_{0,*}\ A_{*,1}$
  - $A_{0,1} \to a\ A_{0,0}\ A_{0,1}$
  - $A_{0,1} \to a\ A_{0,1}\ A_{1,1}$

Exemple (la boucle en 1)



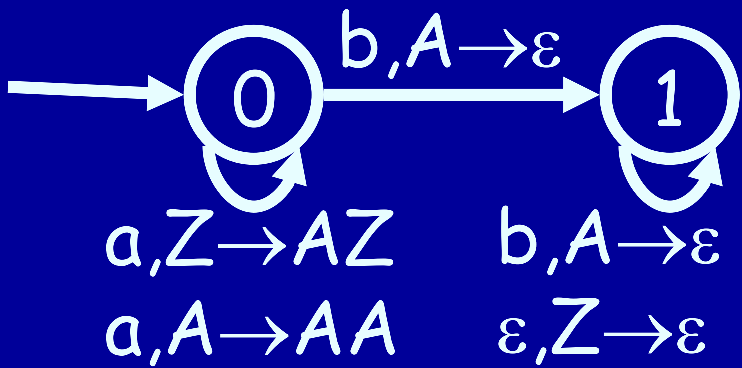
- $\delta(1,b,A)=(1,\varepsilon)$

$A_{1,1} \rightarrow b$

- $\delta(1,\varepsilon,Z)=(1,\varepsilon)$

$Z_{1,1} \rightarrow \varepsilon$

Exemple (résumé)



Notation

$Z_{0,0}$ .. A
$Z_{0,1}$ .. B
$A_{0,0}$ .. C
$A_{0,1}$ .. D
$Z_{1,0}$ .. E
$Z_{1,1}$ .. F
$A_{1,0}$ .. G
$A_{1,1}$ .. H

$S \to Z_{0,0}$
$S \to Z_{0,1}$
$Z_{0,0} \to a\ A_{0,0}\ Z_{0,0}$
$Z_{0,0} \to a\ A_{0,1}\ Z_{1,0}$
$Z_{0,1} \to a\ A_{0,0}\ Z_{0,1}$
$Z_{0,1} \to a\ A_{0,1}\ Z_{1,1}$
$A_{0,0} \to a\ A_{0,0}\ A_{0,0}$
$A_{0,0} \to a\ A_{0,1}\ A_{1,0}$
$A_{0,1} \to a\ A_{0,0}\ A_{0,1}$
$A_{0,1} \to a\ A_{0,1}\ A_{1,1}$
$A_{0,1} \to b$
$A_{1,1} \to b$
$Z_{1,1} \to \varepsilon$

$S \to A$
$S \to B$
$A \to aCA$
$A \to aDE$
$B \to aCB$
$B \to aDF$
$C \to aCC$
$C \to aDG$
$D \to aCD$
$D \to aDH$
$D \to b$
$H \to b$
$F \to \varepsilon$

Exemple

$S \to A$
$S \to B$
$A \to aCA$
$A \to aDE$
$B \to aCB$
$B \to aDF$
$C \to aCC$
$C \to aDG$
$D \to aCD$
$D \to aDH$
$D \to b$
$H \to b$
$F \to \varepsilon$

$S \to A|B$
$A \to aCA|aDE$
$B \to aCB|aDF$
$C \to aCC|aDG$
$D \to aCD|aDH|b$
$H \to b$
$F \to \varepsilon$

$S \to A|B$
$A \to aCA|aDE$
$B \to aCB|aD$
$C \to aCC|aDG$
$D \to aCD|aDH|b$
$H \to b$

$S \to A|B$
$A \to aCA|aDE$
$B \to aCB|aD$
$C \to aCC|aDG$
$D \to aCD|aDb|b$

Productifs : D, B, S

$S \to B$
$B \to aD$
$D \to aDb|b$

$$\boxed{\begin{array}{l} S \to aD \\ D \to aDb|b \end{array}}$$

Engendre $\{a^n b^n : n > 0\}$

$(p,x,A)\rightarrow^{i}(q,\varepsilon,\ \varepsilon) \Rightarrow A_{p,q}\rightarrow^{*} x$

Par récurrence sur i

- **Base** : i=1 $(p,x,A)\rightarrow^{1}(q,\varepsilon,\ \varepsilon) \Rightarrow A_{p,q}\rightarrow x$
  $(q,\ \varepsilon)\in\delta(p,x,A)$ pour $x\in T\cup\{\varepsilon\}$ et $A_{p,q}\rightarrow x$ règle de G
- **Induction** : i>1 (HR) $(p,x,A)\rightarrow^{i}(q,\varepsilon,\ \varepsilon) \Rightarrow A_{p,q}\rightarrow^{*} x$
  vrai pour toute lecture de longueur au plus i
- Soit x=ay et

$$(p,ay,A)\rightarrow(q_1,y,B_1\ldots B_m)\rightarrow^{i-1}(q,\varepsilon,\varepsilon)$$

- **Ruse** : découpage de y : $y=y_1\ldots y_m$: y_j a pour effet de dépiler B_j éventuellement après plusieurs transitions

$(p,x,A)\rightarrow^{i}(q,\varepsilon,\ \varepsilon) \Rightarrow A_{p,q} \rightarrow^{*} x$

$$(p,ay_1\ldots y_m,A)\rightarrow(q_1,\ y_1\ldots y_m,B_1\ldots B_m)\rightarrow^{i-1}(q,\varepsilon,\varepsilon)$$

- Il existe des états $q_2,\ldots,q_{m+1}$ tels que $q_{m+1}=q$ et pour lesquels

$$(q_j,y_j,B_j)\rightarrow^{*}(q_{j+1},\varepsilon,\varepsilon)$$

 en moins de i transitions. Par (HR), dérivation de G: pour $0<j<m+1$

$$B_{j,q_j,q_{j+1}} \rightarrow^{*} y_j$$

$$(p,ay_1\ldots y_m,A)\rightarrow(q_1,\ y_1\ldots y_m,B_1\ldots B_m)\rightarrow^{i-1}(q,\varepsilon,\varepsilon)$$

$$B_{j,q_j,q_{j+1}} \rightarrow^{*} y_j \text{ pour } 0<j<m+1$$

- Donc, avec la construction pour la 1ère transition,

$$A_{p,q} \rightarrow a\ B_{1,q_1,q_2}\ B_{2,q_2,q_3} \ldots B_{m,q_m,q}$$

$$A_{p,q} \rightarrow^{*} a\ y_1 y_2 \ldots y_m$$

Réciproquement

- On montre $A_{p,q} \rightarrow^{i} x \Rightarrow (p,x,A) \rightarrow^{*} (q,\varepsilon, \varepsilon)$ par récurrence sur i.
  - Base i=1 : $A_{p,q} \rightarrow x \Rightarrow (p,x,A) \rightarrow^{*} (q,\varepsilon, \varepsilon)$
    $A_{p,q} \rightarrow x$ règle de G, $x \in T \cup \{\varepsilon\}$ doit être reconnu par M, donc

$$(q,\varepsilon) \in \delta(p,x,A)$$

  - Induction : (HR) vraie pour toute dérivation de longueur au plus i.

$$A_{p,q} \rightarrow a\ B_{1,q_1,q_2}\ B_{2,q_2,q_3} \ldots B_{m,q_m,q} \rightarrow^{i-1} x$$

Réciproquement

- Et $x=ax_1 \ldots x_m$ avec $B_{j,q_j,q_{j+1}} \rightarrow^* x_j$ pour $0<j<m+1$ en moins de i étapes
- Par (HR), $(q_j, x_j, B_j) \rightarrow^* (q_{j+1}, \varepsilon, \varepsilon)$
- On ajoute $B_{j+1} \ldots B_m$ au fond de chaque pile,

$$(q_j,x_j,B_j) \rightarrow^* (q_{j+1},\varepsilon,\varepsilon) \rightarrow (q_j,x_j,B_jB_{j+1}\ldots B_m) \rightarrow^* (q_{j+1},\varepsilon,B_{j+1}\ldots B_m) \quad (1)$$

- Et par la première étape de la génération de x par $A_{p,q}$ on a

$$(p, x=ax_1\ldots x_m, A) \rightarrow (q_1, x_1\ldots x_m, B_1\ldots B_m) \quad (2)$$

- En combinant (1) et (2) on obtient

$$(p, x=ax_1\ldots x_m, A) \rightarrow^* (q, \varepsilon, \varepsilon)$$

Conclusion

$$A_{p,q} \rightarrow^* x \text{ SSI } (p,x,A) \rightarrow^* (q,\varepsilon,\varepsilon)$$

Pour p=i, A=Z, on a que

$$Z_{i,q} \rightarrow^* x \text{ SSI } (i,x,Z) \rightarrow^* (q,\varepsilon,\varepsilon)$$

Or par construction de G, $S \rightarrow Z_{i,p}$

$$S \rightarrow^* x \text{ SSI } (i,x,Z) \rightarrow^* (p,\varepsilon,\varepsilon) \Leftrightarrow L(G)=L(M)$$

Donc, L algébrique SSI L reconnaissable par AP

Conclusion

- Deux manières de caractériser les langages algébriques :
  - Par un mécanisme de génération : les grammaires algébriques
  - Par un mécanisme de reconnaissance : les automates à pile