Problème

- On souhaite un analogue algébrique au théorème de Kleene **(**Rat(Σ)=Rec(Σ)**)**
- Montrer que tout langage algébrique est reconnaissable par AP (non-déterministe) et réciproquement, tout langage reconnaissable par un AP est algébrique !

2

Algébrique⇒ reconnaissable

Théorème : 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 ⇒ 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))

3

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.

4

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

$S\rightarrow^* x\alpha$ par dérivation gauche SSI $(q,x,S)\rightarrow^*(q,\varepsilon,\alpha)$

- Pour terminer, notons que

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

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

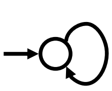
- x peut être engendré par dérivation gauche par *G*
- x est reconnu par pile vide avec M

5

Exemple

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

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



a,S→B
b,S→A
a,A→ε
a,A→S
b,A→AA
b,B→ε
b,B→S
a,B→BB

6

Intérêt de la preuve

- Etant donnée un grammaire algébrique sous forme normale de Greibach, on a construit un automate à pile qui reconnaît les mots du langage engendré par la grammaire
- C'est PRECISEMENT ce que fait un outil d'analyse syntaxique utilisé dans un compilateur
- Pratiquement, Un outil d 'analyse syntaxique ne transforme pas la grammaire pour la mettre sous FNG
- But de l'analyse syntaxique :
  - Faire la même chose en conservant la structure de la grammaire.

7

Caractériser les algébriques

Le but :

Théorème :
L algébrique SSI il existe un AP qui le reconnaît.

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

reste à montrer :

Théorème :
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, on associe une variable de la forme [q,A,p] pour $A\in\Gamma$

les variables de G correspondent à l'état de la pile

- G engendre w depuis [q,A,p] SSI $(q,w,A)\to^*(p,\varepsilon,\varepsilon)$: une dérivation gauche du mot w est la simulation du fonctionnement de l'AP sur l'entrée w.

Construction

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

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

  Une variable [q,A,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 :
  - $S\to[i,Z,q],\ \forall q\in Q$

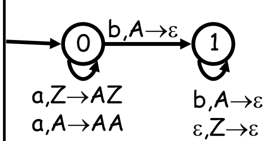
Construction (2)

- Au démarrage :
  - $S\to[i,Z,q],\ \forall q\in Q$
- Régime de croisière :
  - $(p,\varepsilon)\in\delta(q,a,A) \Rightarrow [q,A,p]\to a$
  - $(p,B_1B_2\ldots B_m)\in\delta(q,a,A)$ (m>0)

    $[q,A,r]\to a[p,B_1,q_2][q_2,B_2,q_3]\ldots[q_m,B_m,r]$

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

Exemple



Démarrage :

$S\to[0,Z,*]$

$S\to[0,Z,0]$

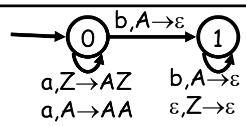
$S\to[0,Z,1]$

- Transition de 0 à 1

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

  $[0,A,1]\to b$

Exemple (la boucle en 0)



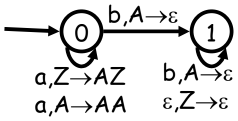
$\delta(0,a,Z)=(0,AZ)$

- r=0
- $[0,Z,0]\to a[0,A,*][*,Z,0]$
  $[0,Z,0]\to a[0,A,0][0,Z,0]$
  $[0,Z,0]\to a[0,A,1][1,Z,0]$
- r=1
- $[0,Z,1]\to a[0,A,*][*,Z,1]$
  $[0,Z,1]\to a[0,A,0][0,Z,1]$
  $[0,Z,1]\to a[0,A,1][1,Z,1]$

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

- r=0
- $[0,A,0]\to a[0,A,*][*,A,0]$
  $[0,A,0]\to a[0,A,0][0,A,0]$
  $[0,A,0]\to a[0,A,1][1,A,0]$
- r=1
- $[0,A,1]\to a[0,A,*][*,A,1]$
  $[0,A,1]\to a[0,A,0][0,A,1]$
  $[0,A,1]\to a[0,A,1][1,A,1]$

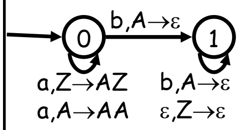
Exemple (la boucle en 1)



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

- $\delta(1,\varepsilon,Z)=(1,\varepsilon)$
  [1,Z,1]→ε

Exemple (résumé)



Notation
[0,Z,0] .. A
[0,Z,1] .. B
[0,A,0] .. C
[0,A,1] .. D
[1,Z,0] .. E
[1,Z,1] .. F
[1,A,0] .. G
[1,A,1] .. H

S→[0,Z,0]
S→[0,Z,1]
[0,Z,0]→a[0,A,0][0,Z,0]
[0,Z,0]→a[0,A,1][1,Z,0]
[0,Z,1]→a[0,A,0][0,Z,1]
[0,Z,1]→a[0,A,1][1,Z,1]
[0,A,0]→a[0,A,0][0,A,0]
[0,A,0]→a[0,A,1][1,A,0]
[0,A,1]→a[0,A,0][0,A,1]
[0,A,1]→a[0,A,1][1,A,1]
[0,A,1]→b
[1,A,1]→b
[1,Z,1]→ε

S→A
S→B
A→aCA
A→aDE
B→aCB
B→aDF
C→aCC
C→aDG
D→aCD
D→aDH
D→b
H→b
F→ε

Exemple

S→A
S→B
A→aCA
A→aDE
B→aCB
B→aDF
C→aCC
C→aDG
D→aCD
D→aDH
D→b
H→b
F→ε

S→A|B
A→aCA|aDE
B→aCB|aDF
C→aCC|aDG
D→aCD|aDH|b
H→b
F→ε

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

S→A|B
A→aCA|aDE
B→aCB|aD
C→aCC|aDG
D→aCD|aDb|b

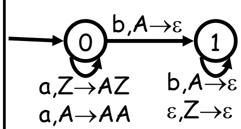
Productifs : D, B, S

S→B
B→aD
D→aDb|b

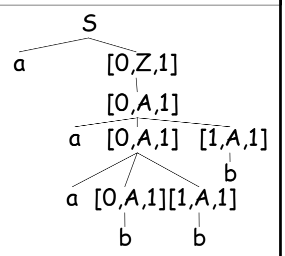
S→aD
D→aDb|b

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

Exemple

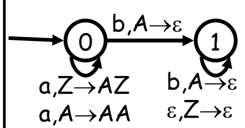


S→a[0,Z,1]
[0,Z,1]→[0,A,1]
[0,A,1]→a[0,A,1][1,A,1] | b
[1,A,1]→b



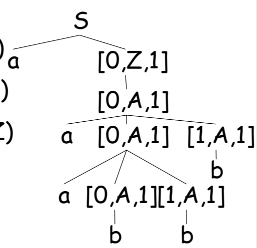
Dérivation de a^3b^3

Exemple

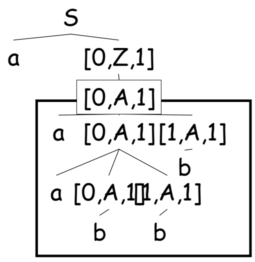


[q,A,p]→* w
SSI
(q,w,A)→*(p,ε, ε)

(0,aaabbb,Z)
(0,aabbb,AZ)
(0,abbb,AAZ)
(0,bbb,AAAZ)
(1,bb,AAZ)
(1,b,AZ)
(1,ε,Z)
(1,ε,ε)



Exemple



[0,A,1]→* a^2b^3 SSI $(0,a^2b^3,A)$→*(1,ε, ε)

(0,aaabbb,Z)
(0,aabbb,AZ)
(0,abbb,AAZ)
(0,bbb,AAAZ)
(1,bb,AAZ)
(1,b,AZ)
(1,ε,Z)
(1,ε,ε)

$(q,x,A) \to^i (p,\varepsilon,\ \varepsilon) \Rightarrow [q,A,p] \to^* x$

Par récurrence sur i

- Base : i=1 $(q,x,A)\to^1(p,\varepsilon,\varepsilon) \Rightarrow [q,A,p]\to^* x$
  $(p,\varepsilon)\in\delta(q,x,A)$ pour $x\in T\cup\{\varepsilon\}$ et $[q,A,p]\to x$ règle de G
- Induction : i>1 (HR) $(q,x,A)\to^i(p,\varepsilon,\varepsilon) \Rightarrow [q,A,p]\to^* x$ vrai pour toute lecture de longueur au plus i
- Soit x=ay et
  $$(q,ay,A)\to(q_1,y,B_1\ldots B_m)\to^{i-1}(p,\varepsilon,\varepsilon)$$
- Ruse : découpage de y : $y=y_1\ldots y_m$: y_j a pour effet de dépiler B_j éventuellement après plusieurs transitions

20

$(q,x,A) \to^i (p,\varepsilon,\ \varepsilon) \Rightarrow [q,A,p] \to^* x$

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

- Il existe des états $q_2,\ldots,q_{m+1}$ tels que q_{m+1}=p et pour lesquels
  $$(q_j,y_j,B_j)\to^*(q_{j+1},\varepsilon,\varepsilon)$$
  en moins de i transitions. Par (HR), dérivation de G: pour 0<j<m+1
  $$[q_j,\ B_j,\ q_{j+1}] \to^* y_j$$
  $$(q,ay_1\ldots y_m,A)\to(q_1,\ y_1\ldots y_m,B_1\ldots B_m)\to^{i-1}(p,\varepsilon,\varepsilon)$$
  $$[q_j,\ B_j,\ q_{j+1}] \to^* y_j \text{ pour } 0<j<m+1$$
- Donc, avec la construction pour la 1ère transition,
  $$[q,A,p]\to a[q_1,\ B_1,\ q_2]\ [q_2,\ B_2,\ q_3]\ldots\ [q_m,\ B_m,\ p]$$
  $$[q,A,p]\to^*\ a\ y_1 y_2\ldots\ y_m$$

21

Réciproquement

- On montre $[q,A,p]\to^i x \Rightarrow (q,x,A)\to^*(p,\varepsilon,\ \varepsilon)$ par récurrence sur i.
  - Base i=1 : $[q,A,p]\to x \Rightarrow (q,x,A)\to^*(p,\varepsilon,\ \varepsilon)$
    $[q,A,p]\to x$ règle de G, $x\in T\cup\{\varepsilon\}$ doit être reconnu par M, donc
    $$(p,\varepsilon) \in \delta(q,x,A)$$
  - Induction : (HR) vraie pour toute dérivation de longueur au plus i.
  $$[q,A,p]\to a[q_1,\ B_1,\ q_2]\ [q_2,\ B_2,\ q_3]\ldots\ [q_m,\ B_m,\ p=q_{m+1}] \to^{i-1}x$$

22

Réciproquement

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

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

- Et par la première étape de la génération de x par [q,A,p] on a
  $$(q,\ x=ax_1\ldots x_m,\ A\ ) \to (q_1,\ x_1\ldots x_m,\ B_1\ldots\ B_m\ ) \qquad (2)$$
- En combinant (1) et (2) on obtient
  $$(q,\ x=ax_1\ldots x_m,\ A\ ) \to^*(p,\ \varepsilon,\ \varepsilon)$$

23

Conclusion

$$[q,A,p] \to^* x \text{ SSI } (q,x,A)\to^* (p,\varepsilon,\ \varepsilon)$$

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

$$[i,Z,p] \to^* x \text{ SSI } (i,x,Z)\to^* (p,\varepsilon,\ \varepsilon)$$

Or par construction de G, $S \to [i,Z,p]$

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

Donc, L algébrique SSI L reconnaissable par AP

24

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

25