NGUYEN Thi Minh Tuyen

Théorie des automates et langages formels

Langages qui ne sont pas non-contextuels

Théorème 1

Soit G une grammaire non contextuelle sous la forme normale de Chomsky. Soit L l'ensemble de tous les mots qui ont une dérivation où chaque production est sous la forme: Variable →Variable Variable

est utilisé au plus une fois. Alors L est un langage fini.

Démonstration. A chaque étape dans une dérivation, une variable est remplacée par soit 2 variables soit un terminal. Donc, s'il y a p productions de la forme:

Variable →Variable Variable

un mot de L contient au plus p+1 lettres. L'ensemble des mots qui contiennent au plus p+1 lettres est fini.

Exemple

$S \rightarrow AB \mid a \qquad (1)$

$A \rightarrow XY \mid a \qquad (2)$

$Y \rightarrow SX \qquad (3)$

$X \rightarrow a \mid b$

$B \rightarrow b$

- $S \Rightarrow AB \Rightarrow aB \Rightarrow ab$
- $\underline{S} \Rightarrow \underline{A}B \Rightarrow XYB \Rightarrow b\underline{Y}B \Rightarrow bSXB \Rightarrow baXB \Rightarrow baaB \Rightarrow baab$

  (1) (2) (3)
- $S \Rightarrow \underline{A}B \Rightarrow X\underline{Y}B \Rightarrow XSXB \Rightarrow bSXB \Rightarrow baXB \Rightarrow baaB \Rightarrow baab$

  (1) (2) (3)

Branche

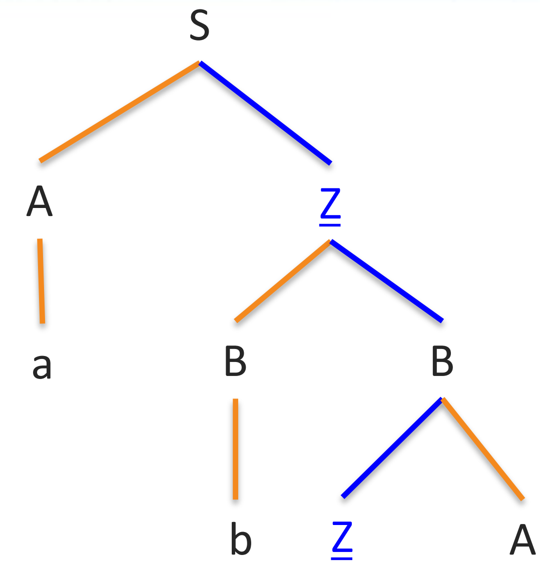
une branche: un chemin entre la racine et une feuille d'un arbre

$S \rightarrow AZ$ $\quad$ $Z \rightarrow BB$ $\quad$ $B \rightarrow ZA$

$A \rightarrow a$ $\quad$ $B \rightarrow b$

$S \Rightarrow A\underline{Z} \Rightarrow aZ \Rightarrow aBB \Rightarrow abB \Rightarrow ab\underline{Z}A\ldots$

Z est une variable qui apparaît 2 fois sur la même branche.

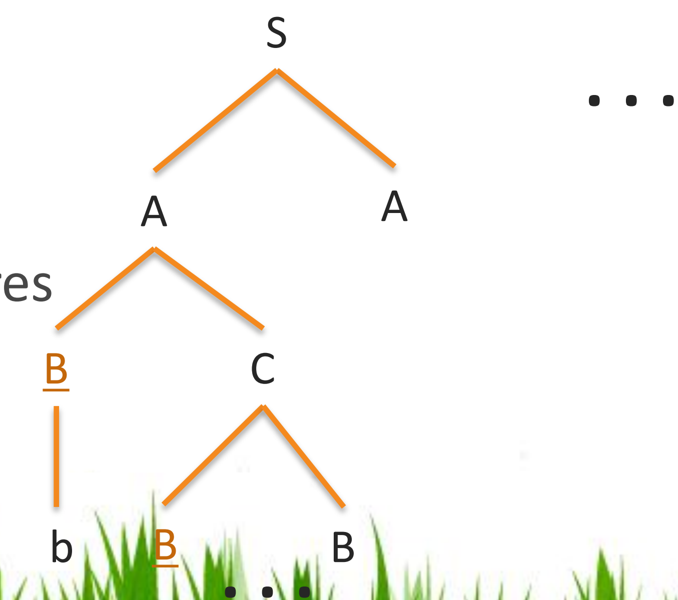


$S \rightarrow AA$ $\quad$ $A \rightarrow BC$ $\quad$ $C \rightarrow BB$

$A \rightarrow a$ $\quad$ $B \rightarrow b$

$S \Rightarrow AA \Rightarrow \underline{B}CA \Rightarrow bCA \Rightarrow b\underline{B}BA\ldots$

B apparaît sur 2 branches différentes



<u>Remarque</u>: les arbres de dérivations sont binaires

Théorème 2

Soit G une grammaire non contextuelle sous la forme normale de Chomsky et que contient p productions de la forme:

Variable →Variable Variable.

Soit w un mot généré par G tel que longueur(w)>2^p.

Tout arbre de dérivation pour w contient une branche tel qu'il existe une variable Z qui apparaît au moins 2 fois sur la même branche.

Exemple

1^{ere} rang — 2^0 noeuds

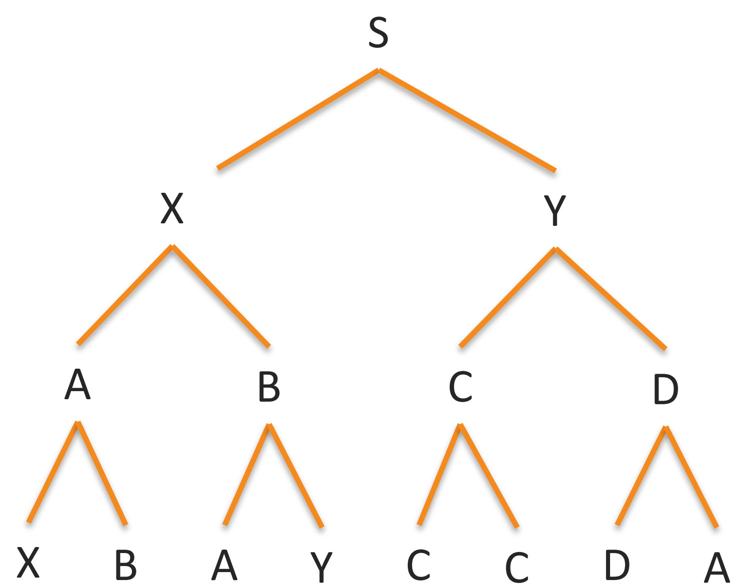
2^{eme} rang — 2^1 noeuds

3^{eme} rang — 2^2 noeuds

4^{eme} rang — 2^3 noeuds

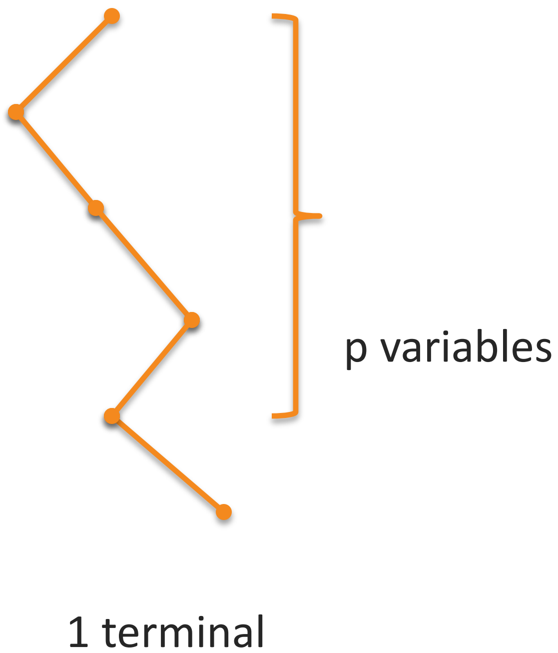
(p=3)

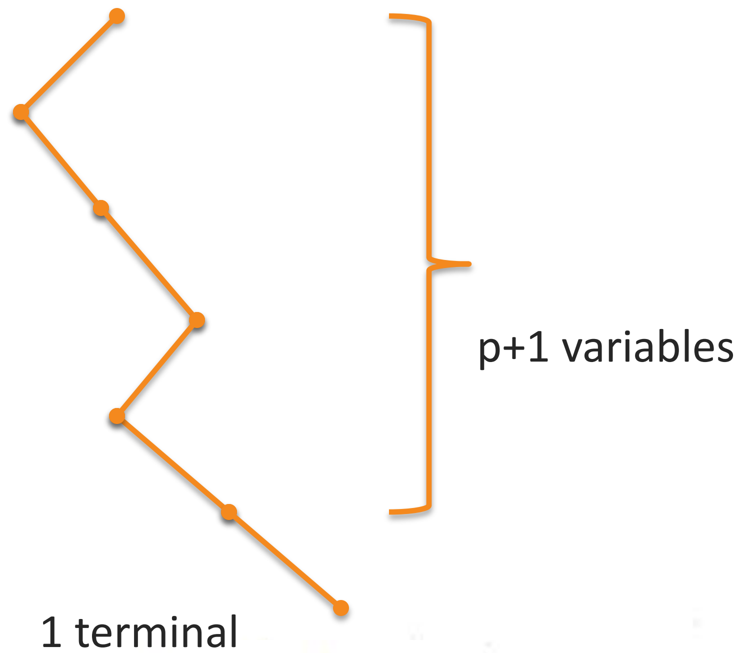
p+1 rangs — 2^p feuilles aux maximum



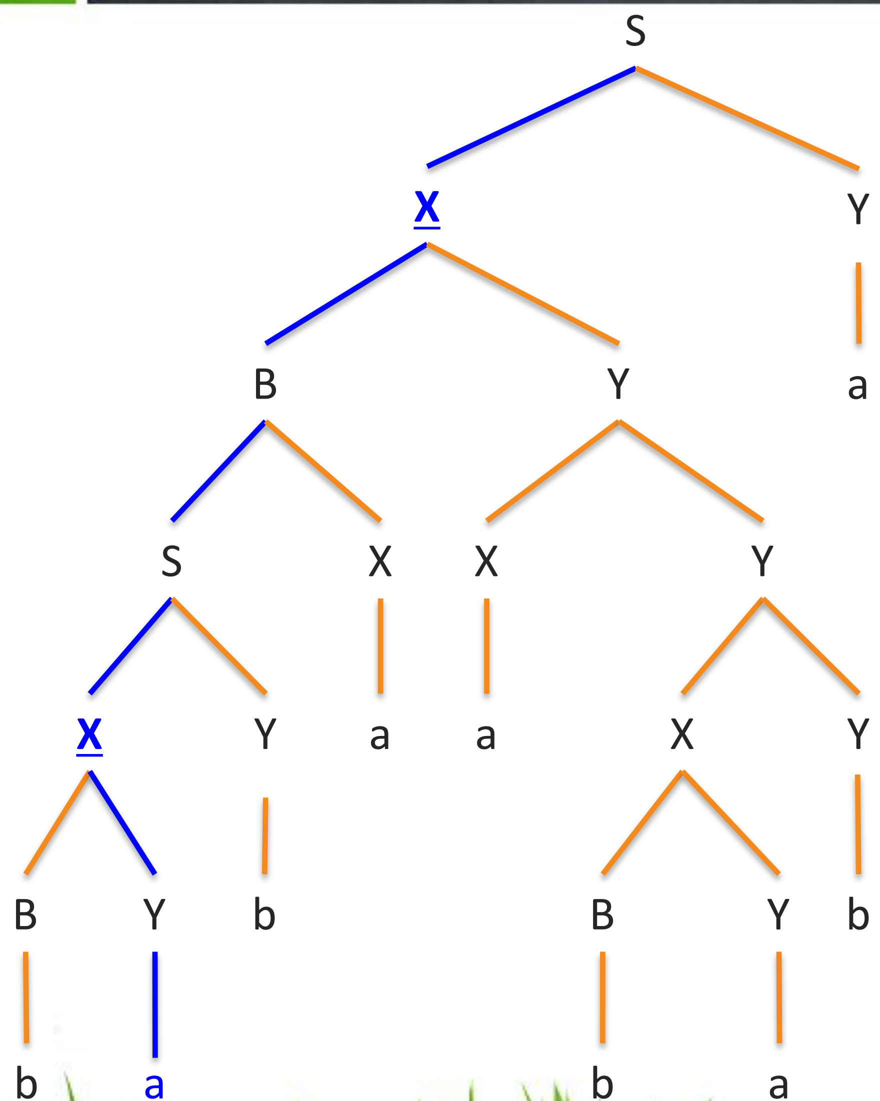
Démonstration [1]

Un arbre contenant plus de 2^p feuilles contient plus de $p+1$ rangs. (Un arbre contenant $p+1$ rangs contient 2^p feuilles au maximum.)





Démonstration [2]



S → XY   B → b

X → BY   X → a

Y → XY   Y → a

B → SX   Y → b

- p=4
- La branche qui se termine par a contient 6 variables (5 applications des productions qui ne contiennent que des variables)

→Au moins une production est utilisée plus qu'une fois

Exemple: PALINDROME – {Λ}

$S \rightarrow AX|BY|AA|BB|a|b$

$X \rightarrow SA$

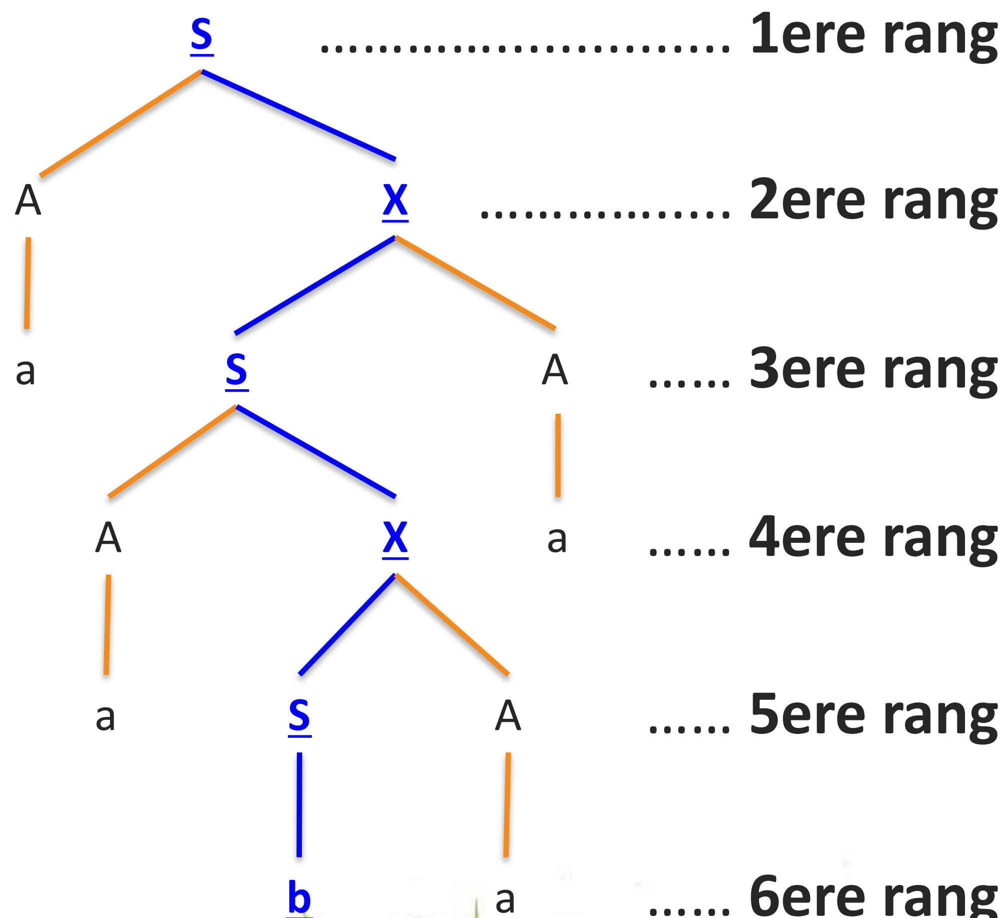
$Y \rightarrow SB$

$A \rightarrow a$

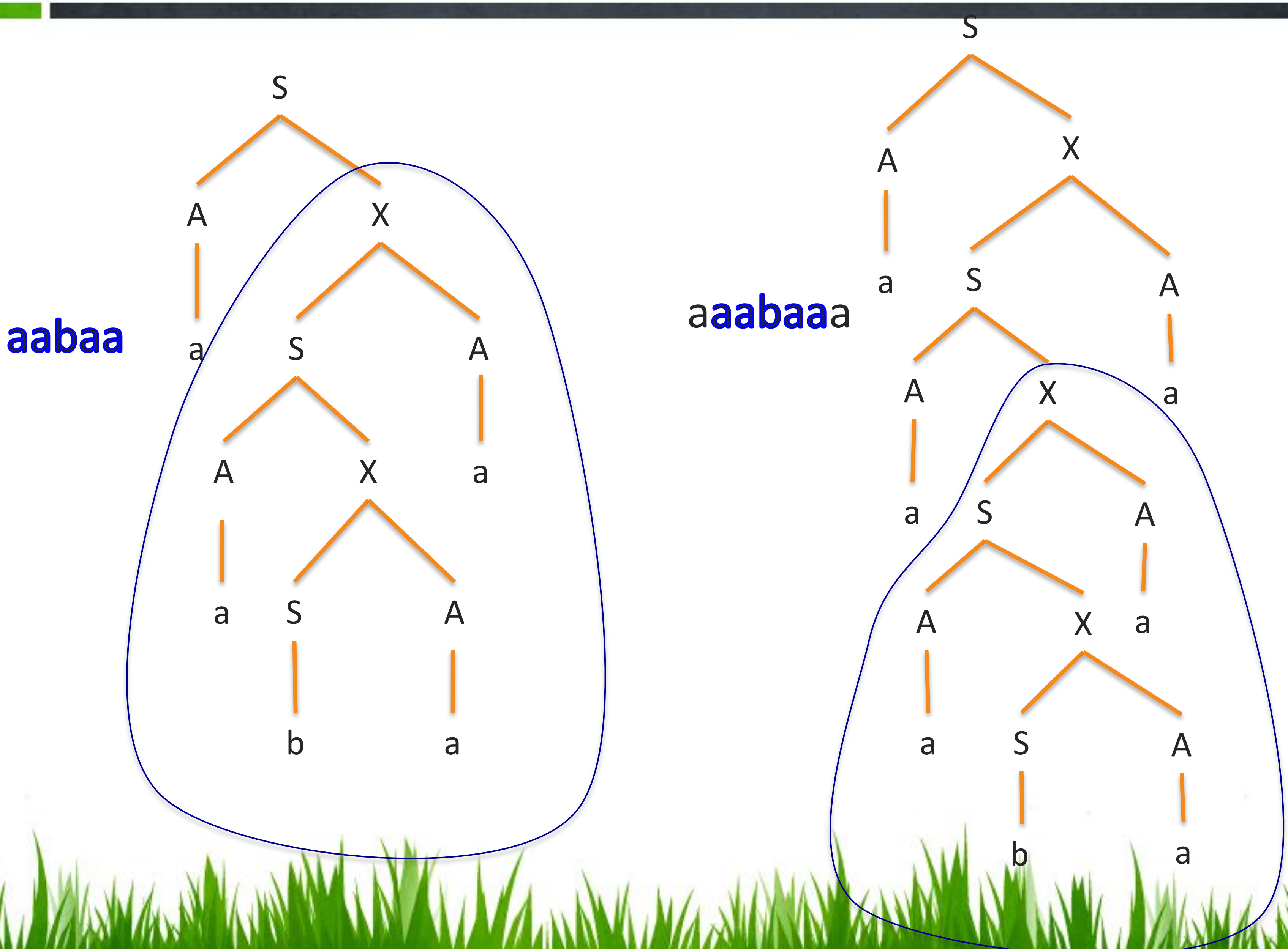
$B \rightarrow b$

- $p = 6$

$2^6=64$



Exemple: PALINDROME – {Λ}



Le lemme de l'étoile (pour les langages non contextuels)

Soit G une grammaire non contextuelle sous la forme normale de Chomsky et qui contient p productions de la forme:

Variable →Variable Variable.

Soit w un mot tel que longueur(w)>2^p. Le mot w peut être décomposer en 5 facteurs: w = uvxyz tel que

- x n'est pas Λ
- Au moins un des deux mots v et y n'est pas Λ
- Pour tout $n \geq 1$, uv^nxy^nz est dans le langage engendré par G.

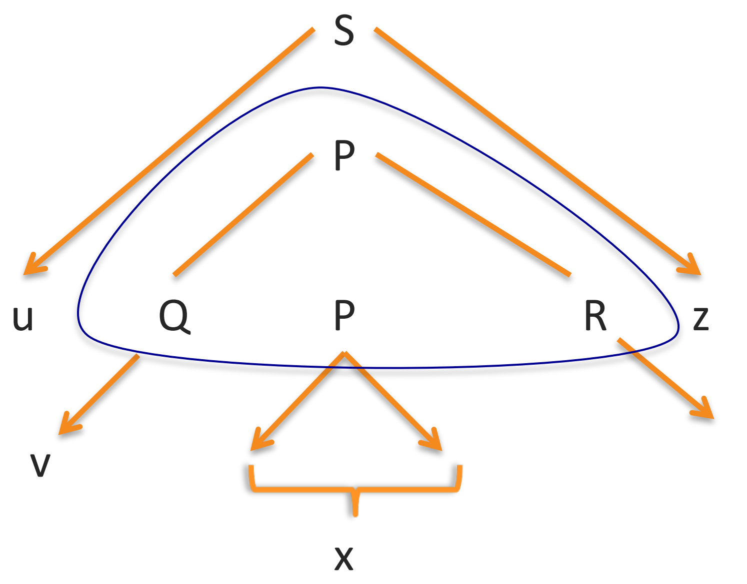
Démonstration

Par le théorème précédent, il existe une variable P qui apparaît au moins 2 fois sur la même branche.

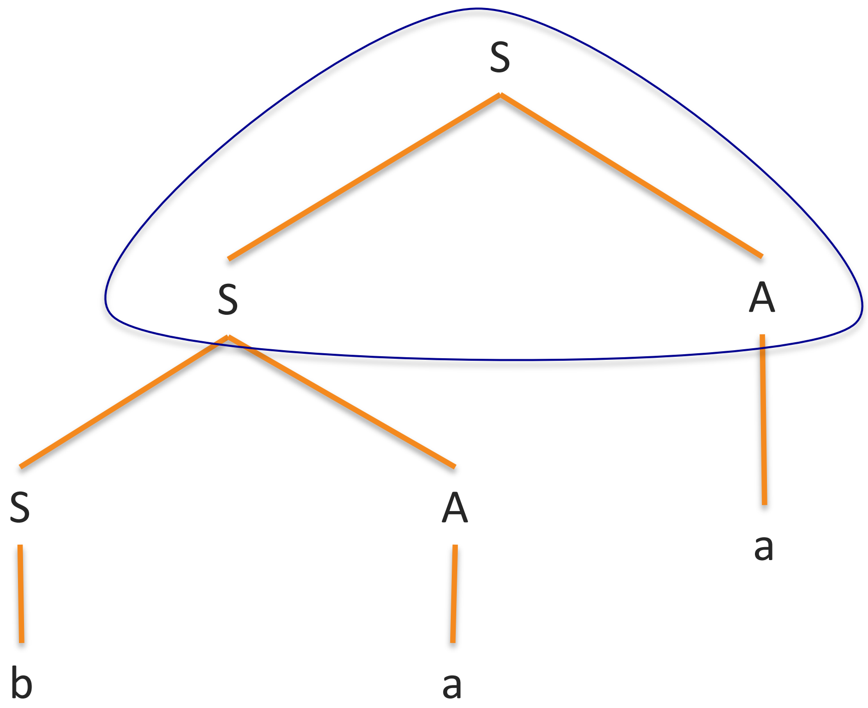
$w = uvxyz$

$x \neq \Lambda$

soit $v \neq \Lambda$, soit $y \neq \Lambda$



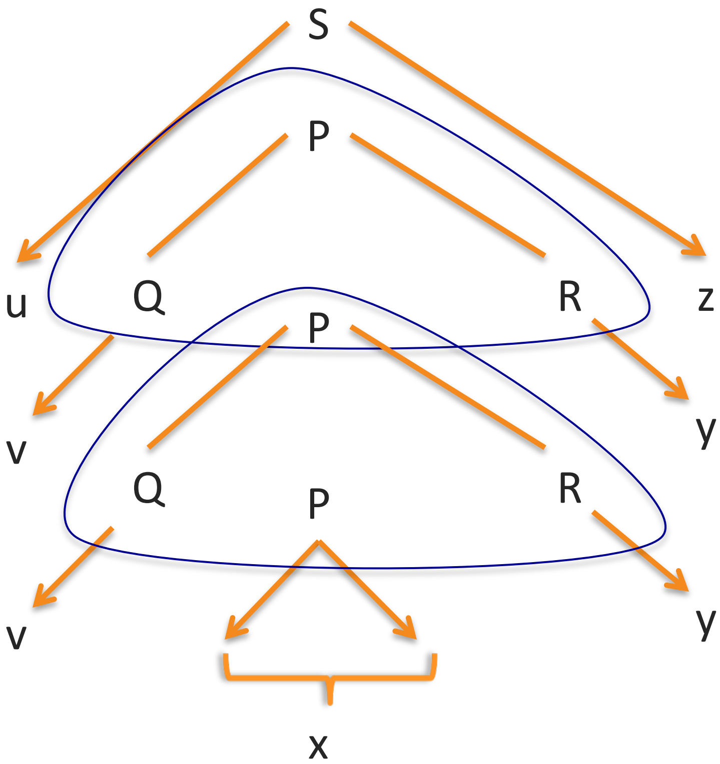
Démonstration



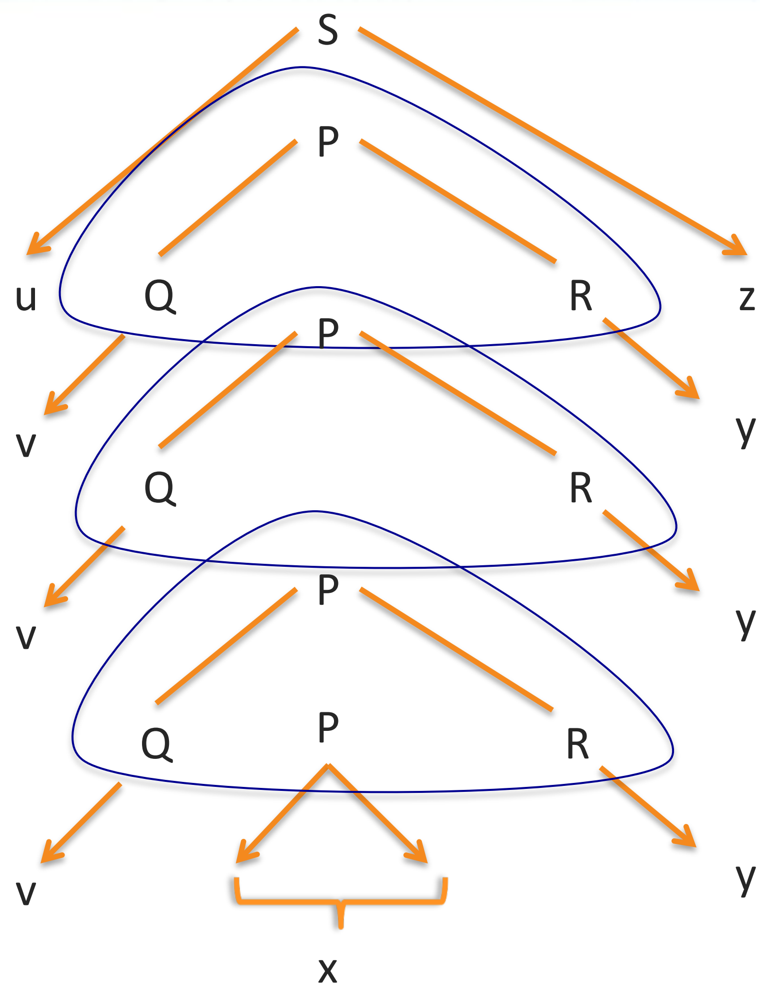
<u>Remarque</u>: u, z, et un des deux mots v ou y peut être Λ.

u = Λ v = Λ x = ba y = a z = Λ

Démonstration



uvvxyyz



En général: uv^nxy^nz

uvvvxyyyz

Une autre alternative

$S \Rightarrow {}^* uPz$

$P \Rightarrow^* vPy$

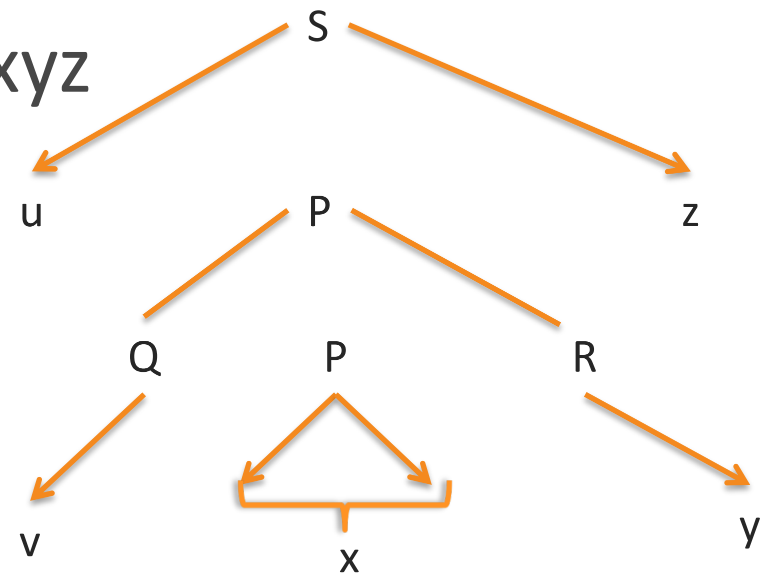
$P \Rightarrow^* x$

$S \Rightarrow^* uPz \Rightarrow^* uvPyz \Rightarrow^* uvxyz$

$w = uvxyz$

$x \neq \Lambda$

soit $v \neq \Lambda$, soit $y \neq \Lambda$



Question?