Chapter 16: Non-Context-Free Languages

I. Theory of Automata

→ II. Theory of Formal Languages

III. Theory of Turing Machines …

- Theorem. Let G be a context-free grammar in Chomsky normal form. Let L be the subset of words generated by G which have derivations such that each production of the form:

  Nonterminal $\rightarrow$ Nonterminal Nonterminal

  is used at most once. L is a finite set of words.

Proof. At each step in a derivation, a nonterminal is replaced by either 2 nonterminals or one terminal. Thus, if there are p productions of the form:

Nonterminal $\rightarrow$ Nonterminal Nonterminal

a word in L contains at most $p+1$ letters.

The set of words that contain at most $p+1$ letters is finite

Example:

$S \rightarrow AB \,/\, a \quad A \rightarrow XY \,/\, a \quad Y \rightarrow SX \quad X \rightarrow a \,/\, b \quad B \rightarrow b$

- $S \Rightarrow AB \Rightarrow aB \Rightarrow ab$
- $S \Rightarrow AB \Rightarrow XYB \Rightarrow bYB \Rightarrow bSXB \Rightarrow baXB \Rightarrow baaB \Rightarrow baab$

  (1) (2) (3)
- $S \Rightarrow AB \Rightarrow XYB \Rightarrow XSXB \Rightarrow bSXB \Rightarrow baXB \Rightarrow baaB \Rightarrow baab$

  (1) (2) (3)

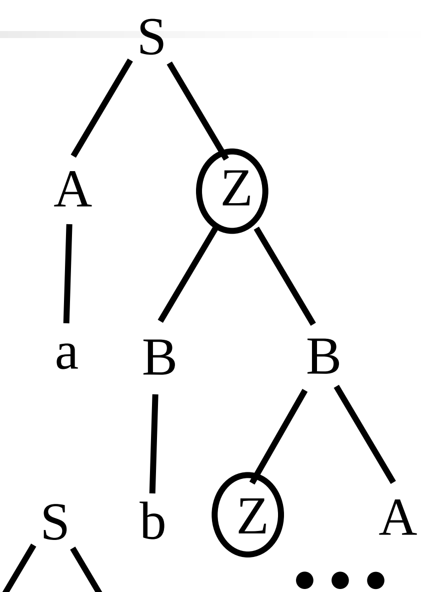
a branch: a path between the root and a leaf of a derivation tree

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

$A \to a \quad B \to b$

$S \Rightarrow AZ \Rightarrow aZ \Rightarrow aBB \Rightarrow abB \Rightarrow abZA\ldots$

the same nonterminal twice on the same branch (The 2nd is a tree descendant of the 1st. The nonterminal is self-embedded.)

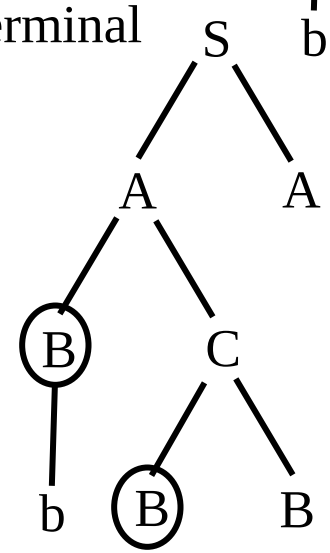


$S \to AA \quad A \to BC \quad C \to BB$

$A \to a \quad B \to b$

$S \Rightarrow AA \Rightarrow BCA \Rightarrow bCA \Rightarrow bBBA\ldots$

two different branches



<u>Remark:</u> drivation trees are binary

- Theorem. Let G be a context-free grammar in Chomsky normal form that has p productions of the form:

  Nonterminal $\rightarrow$ Nonterminal Nonterminal.

  Let w be a word such that length(w)>2^p. Then in every derivation tree for w there exists some nonterminal Z that appears twice on the same branch.

Example

1st rang 2^0 nodes

2nd rang 2^1 nodes

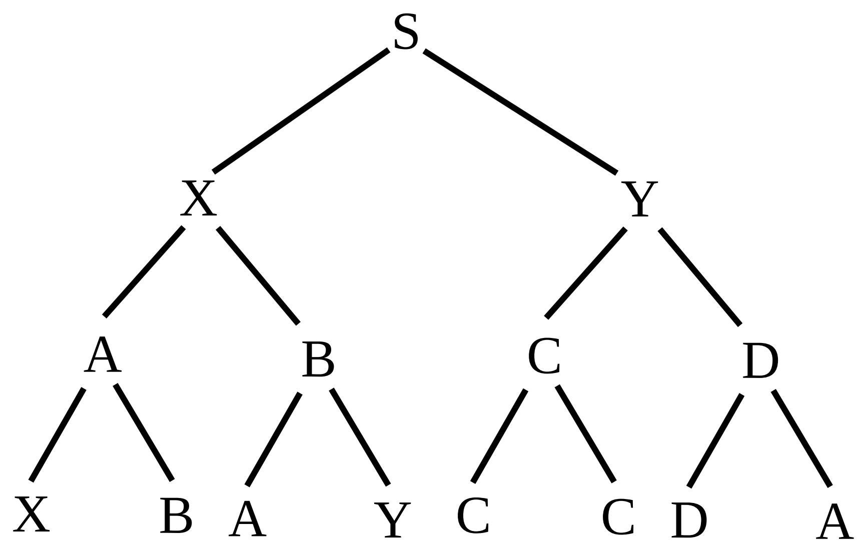
3rd rang 2^2 nodes

4th rang 2^3 nodes

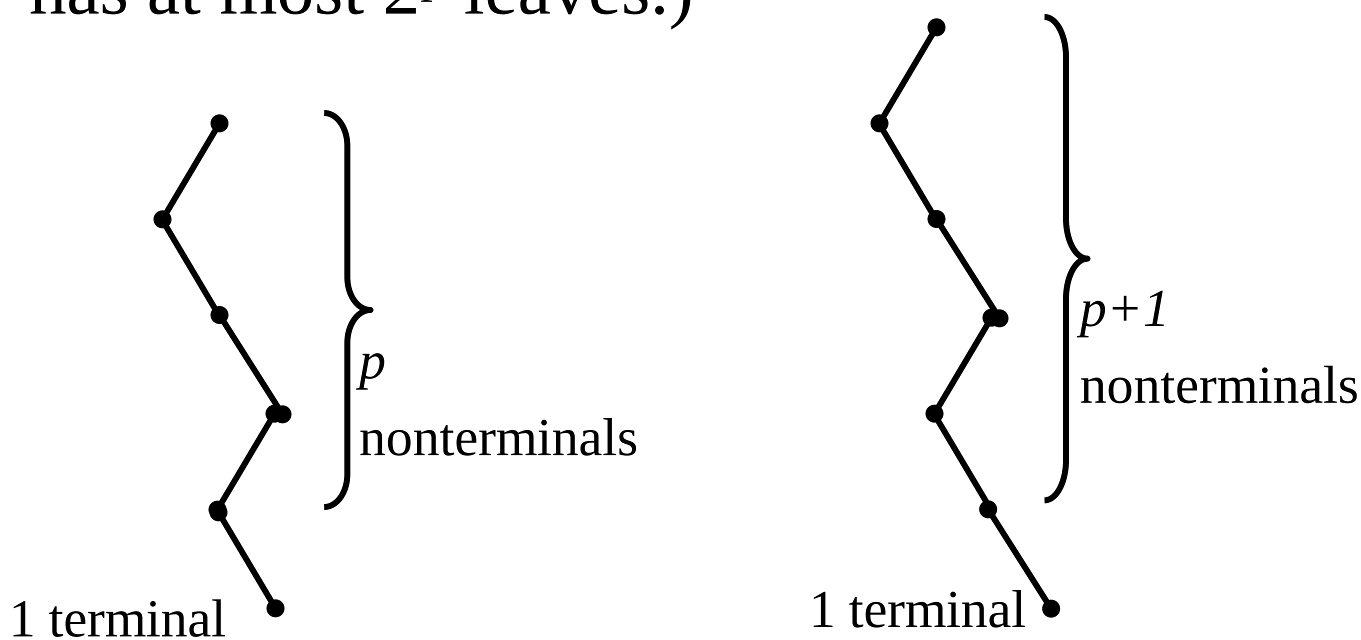
p=3

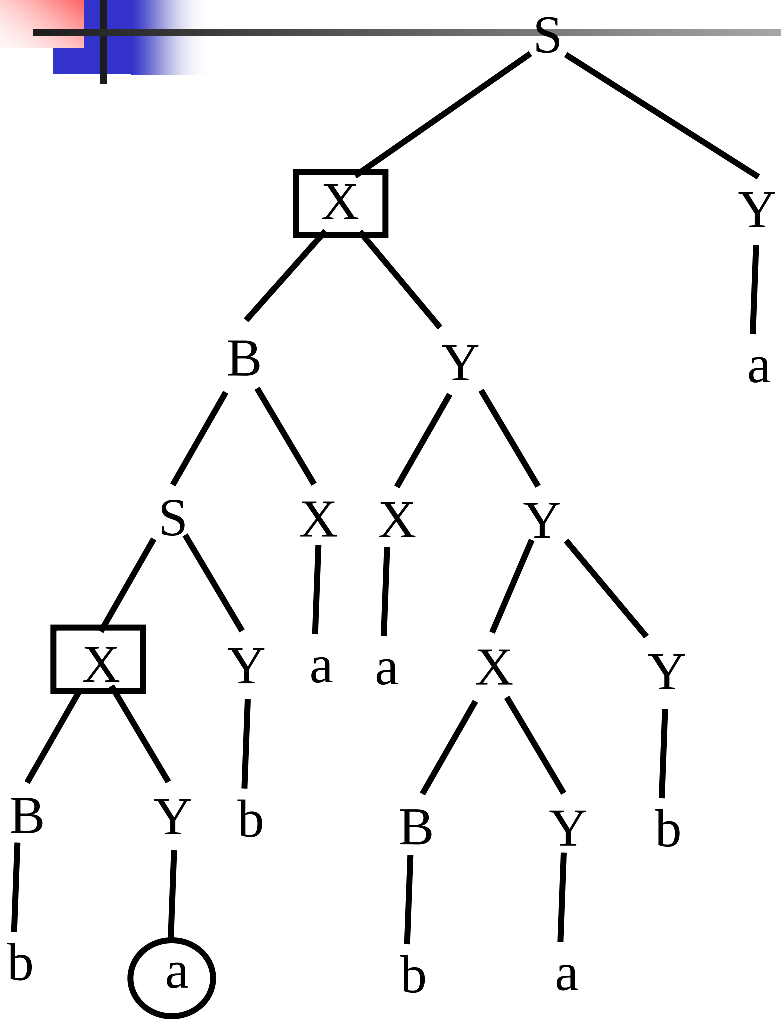
p+1 rows

2^p leaves, maximum



<u>Proof.</u> A tree that has more than 2^p leaves has more than $p+1$ rows. (A tree that has $p+1$ rows has at most 2^p leaves.)





$S \rightarrow XY$ $\quad B \rightarrow b$
$X \rightarrow BY$ $\quad X \rightarrow a$
$Y \rightarrow XY$ $\quad Y \rightarrow a$
$B \rightarrow SX$ $\quad Y \rightarrow b$

- p=4
- the branch that ends in a has 6 nonterminals (5 applications of productions containing only nonterminals)
- at least one production is used more than once

Example: PALINDROME – {Λ}

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

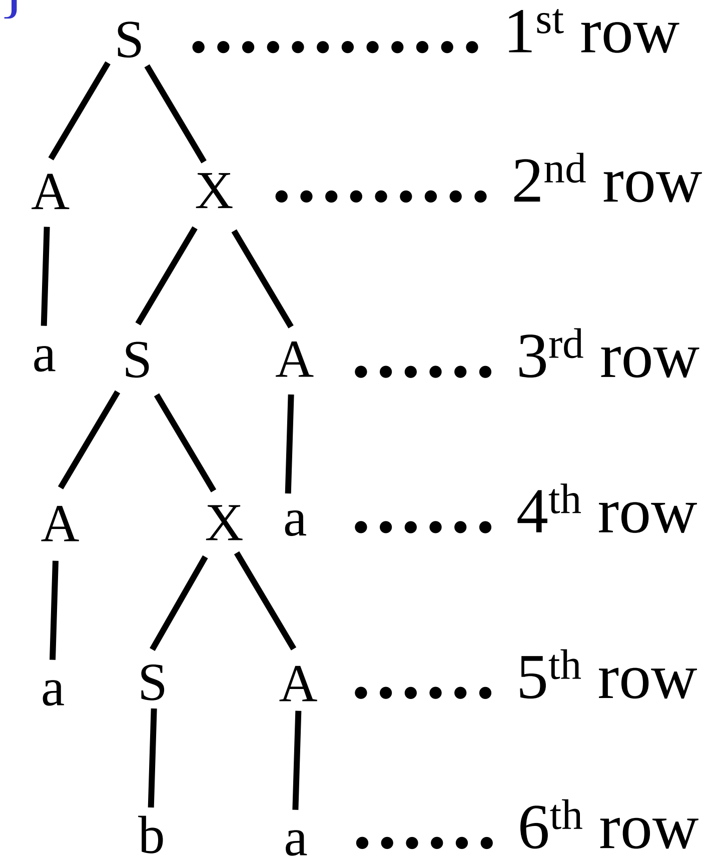
$X \rightarrow SA$

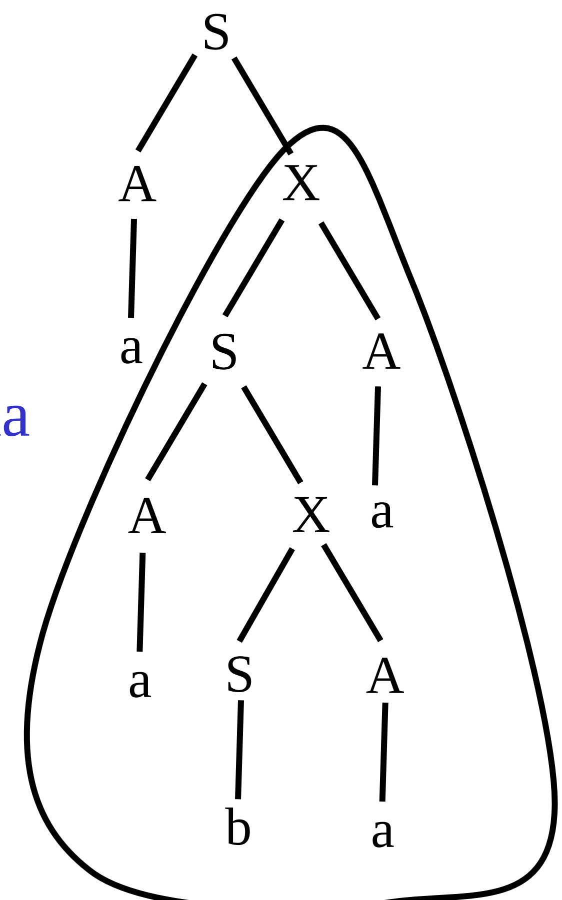
$Y \rightarrow SB$

$A \rightarrow a$

$B \rightarrow b$

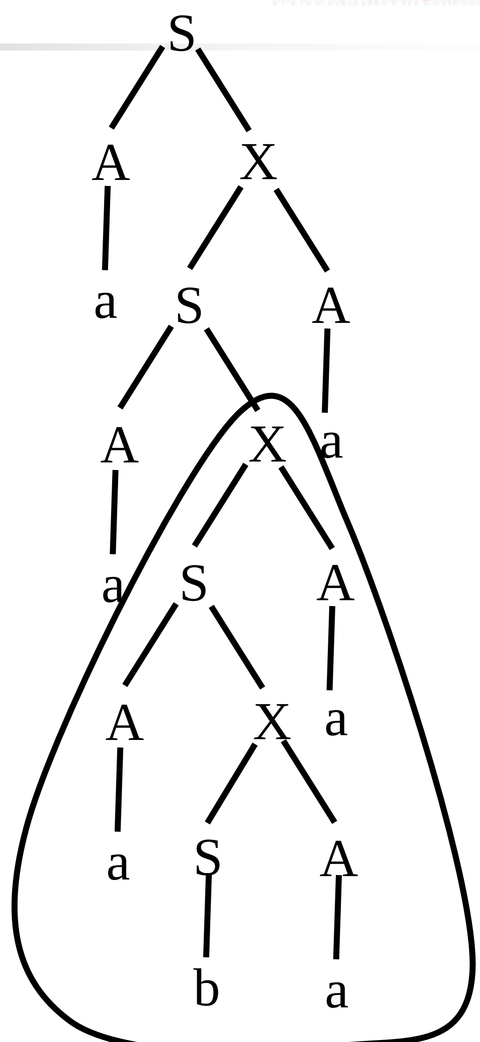
$2^6 = 64$





aabaa

aaabaaa



The pumping lemma (for context-free languages)

Let G be a context-free grammar in Chomsky normal form that has *p* productions of the form:

$$\text{Nonterminal} \rightarrow \text{Nonterminal Nonterminal.}$$

Let w be a word such that length(w)>2^p. Then w can be decomposed into 5 factors: w = uvxyz such that

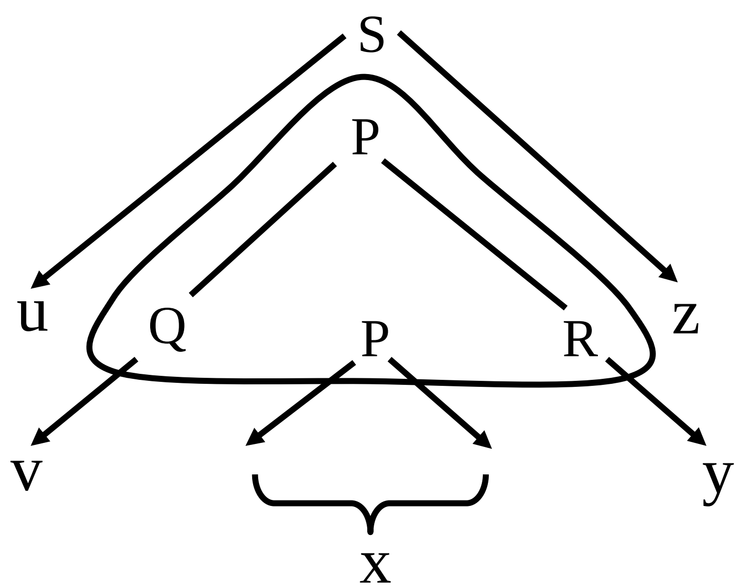
- x is not Λ
- at least one of v and y is not Λ
- for all $n \geq 1$, uv^nxy^nz is in the language generated by G.

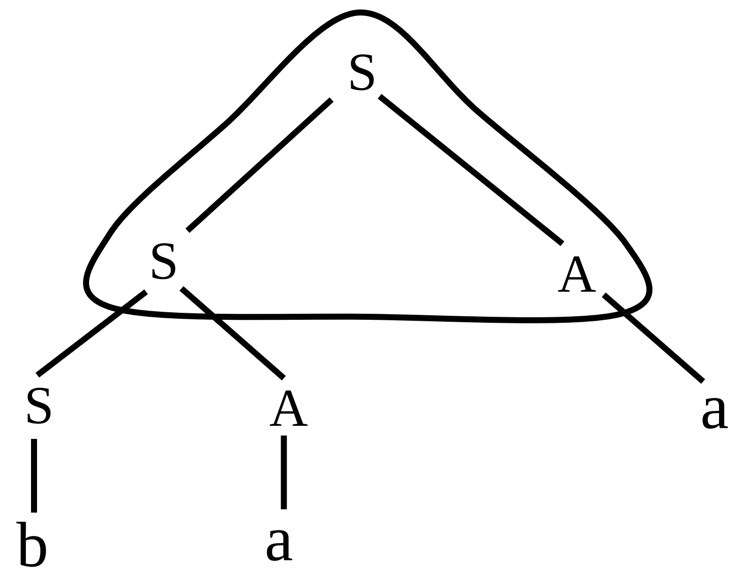
- Proof. From the previous theorem, there exists a non terminal P that occurs at least twice on the same branch.

$w = uvxyz$

$x \neq \Lambda$

either $v \neq \Lambda$, or $y \neq \Lambda$

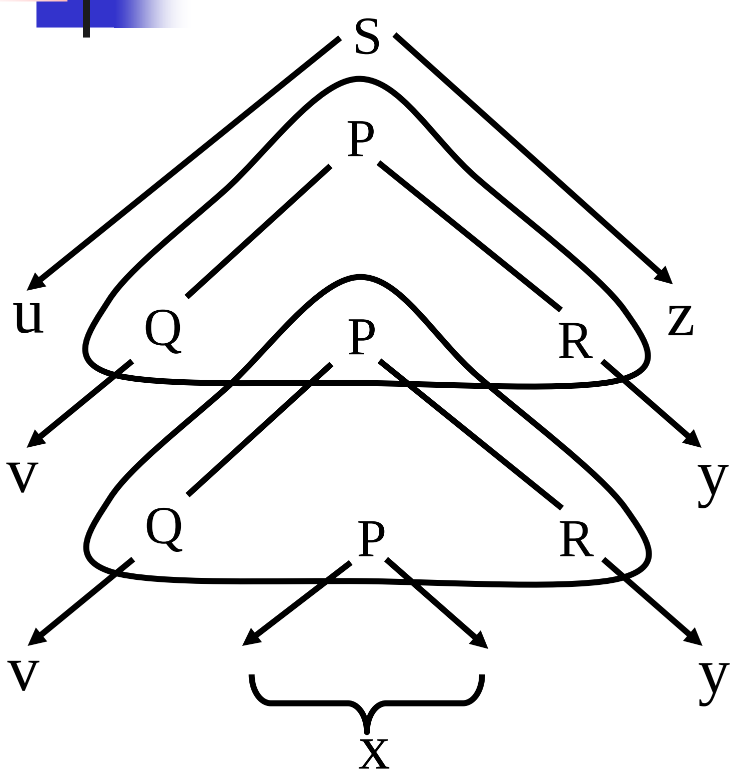




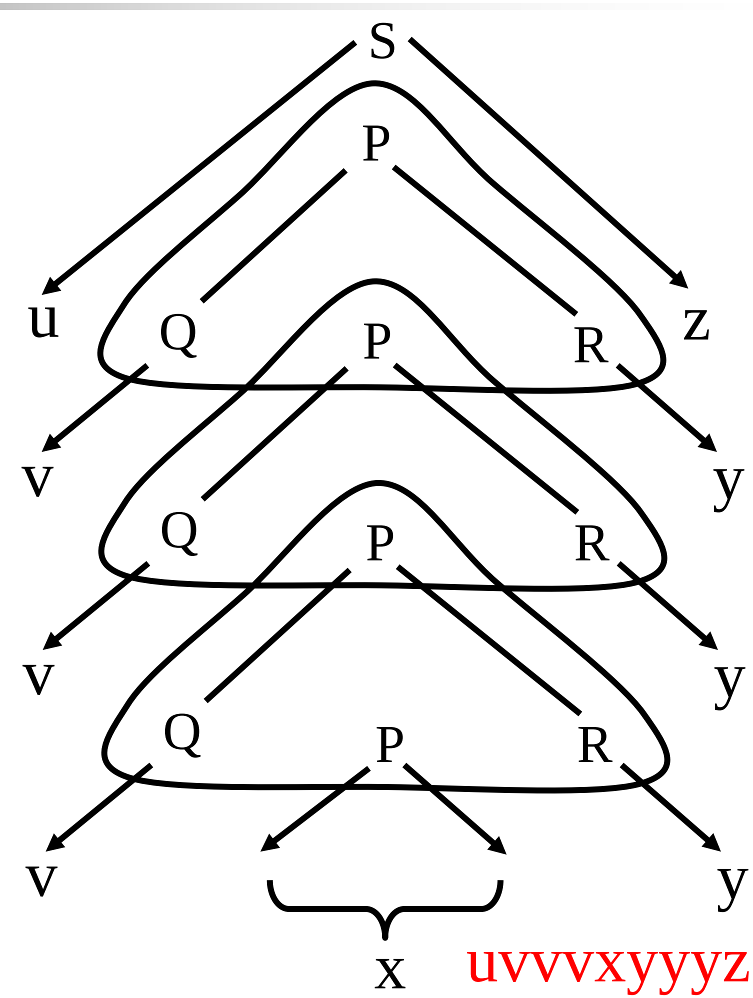
Note: u, z, and at least one of v and y could be Λ.

$u = \Lambda \quad v = \Lambda \quad x = ba \quad y = a \quad z = \Lambda$

Chapter 16: Non-Context-Free Languages



uvvxyyz



uvvvxyyyz

In general: uv^nxy^nz

Alternatively:

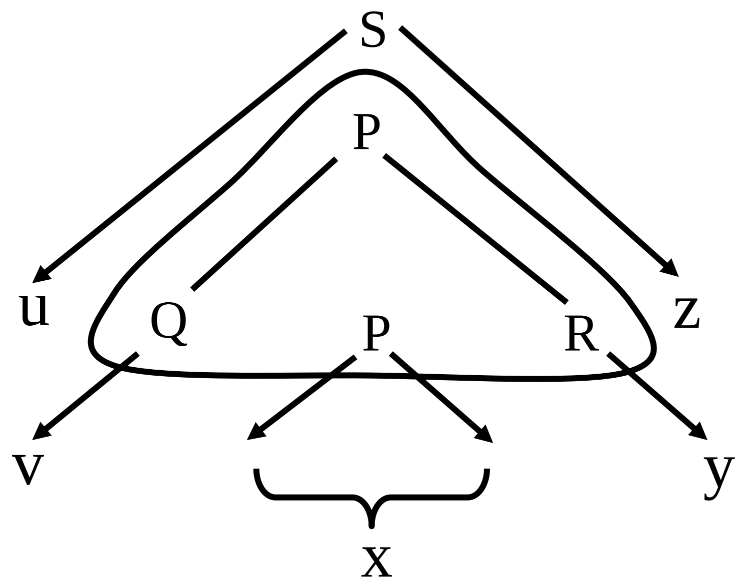
$S \Rightarrow^* uPz \qquad P \Rightarrow^* vPy \qquad P \Rightarrow^* x$

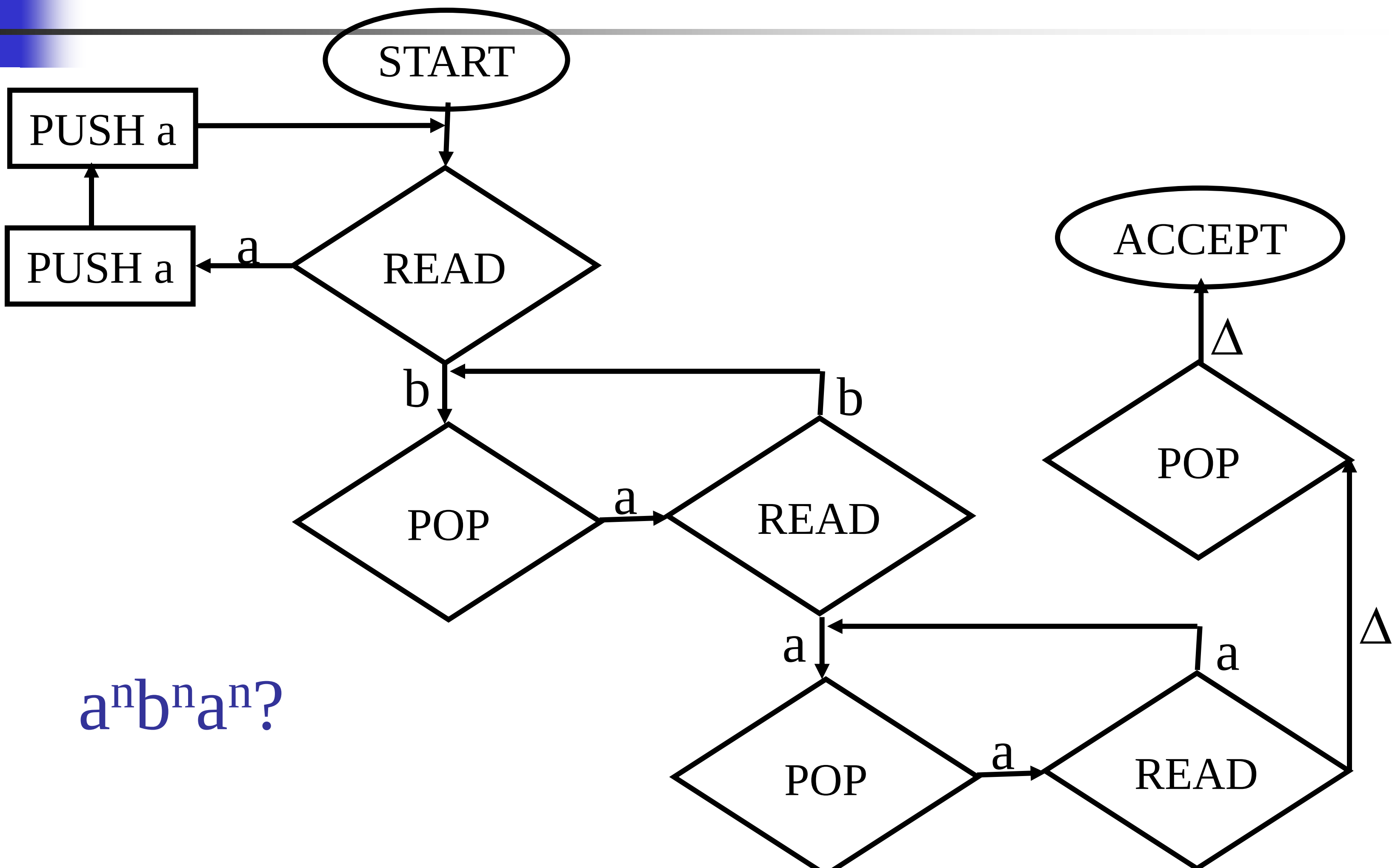
$S \Rightarrow^* uPz \Rightarrow^* uvPyz \Rightarrow^* uvxyz$

$w = uvxyz$

$x \neq \Lambda$

either $v \neq \Lambda$, or $y \neq \Lambda$




START
PUSH a
PUSH a
a
READ
ACCEPT
Δ
b
b
POP
POP
a
READ
Δ
a
a
POP
a
READ

$a^n b^n a^n$?