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 → 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 → 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:

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

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

  (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 → AZ Z → BB B → ZA
A → a B → b

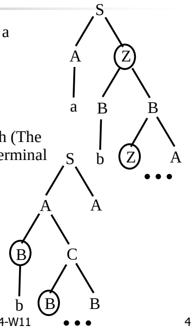
S ⇒ A(Z) ⇒ aZ ⇒ aBB ⇒ abB ⇒ ab(Z)A…

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

S → AA A → BC C → BB
A → a B → b

S ⇒ AA ⇒ (B)CA ⇒ bCA ⇒ b(B)BA…

two different branches



Remark: drivation trees are binary

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

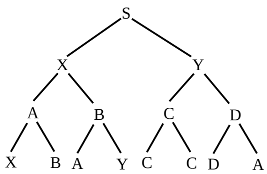
  Nonterminal → Nonterminal Nonterminal.

  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.

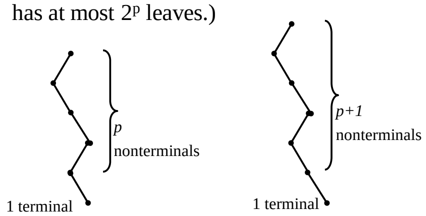
Example

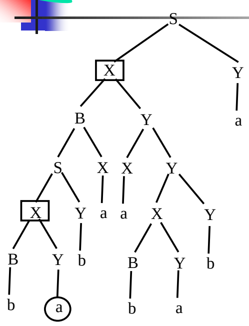
1st rang 2^0 nodes
2nd rang 2^1 nodes
3rd rang 2^2 nodes
4th rang 2^3 nodes

p=3
p+1 rows
2^p leaves, maximum



Proof. 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.)





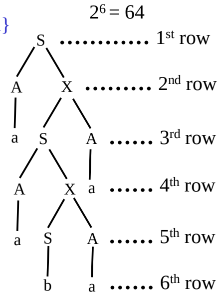
$S \to XY$ $B \to b$
$X \to BY$ $X \to a$
$Y \to XY$ $Y \to a$
$B \to SX$ $Y \to b$

- 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

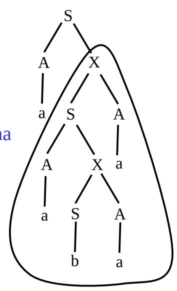
Example: PALINDROME – {Λ}

$S \to AX/BY/a/b/AA/BB$
$X \to SA$
$Y \to SB$
$A \to a$
$B \to b$

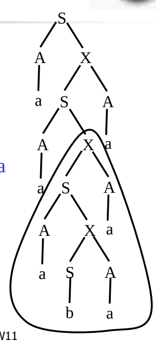
$2^6 = 64$



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:

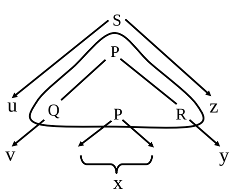
Nonterminal → Nonterminal Nonterminal.

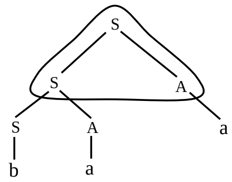
Let w be a word such that length(w)>2^p. Then w can be decomposed into 5 factors: w = uvxyz such that

- x is not Λ
- at least one of v and y is not Λ
- for all n ≥ 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.

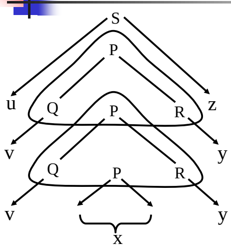
w = uvxyz
x ≠ Λ
either v ≠ Λ, or y ≠ Λ





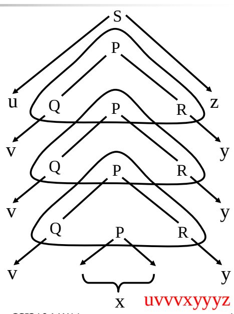
Note: u, z, and at least one of v and y could be Λ.

u = Λ   v = Λ   x = ba   y = a   z = Λ



uvvxyyz

In general: uv^nxy^nz



uvvvxyyyz

Alternatively:

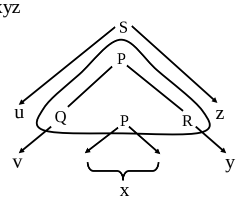
$S \Rightarrow^* uPz$ $P \Rightarrow^* vPy$ $P \Rightarrow^* x$

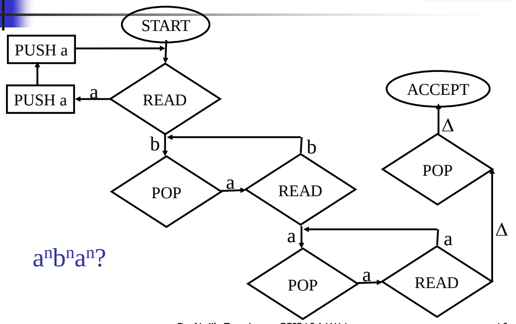
$S \Rightarrow^* uPz \Rightarrow^* uvPyz \Rightarrow^* uvxyz$

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
a
READ
POP
a
a
Δ
POP
a
READ

$a^nb^na^n$?