The Pumping Lemma
for
Context-Free Languages

class 13

1

Take an **infinite** context-free language

Generates an infinite number
of different strings

Example:

$S \to AB$

$A \to aBb$

$B \to Sb$

$B \to b$

2

$S \to AB$

$A \to aBb$

$B \to Sb$

$B \to b$

In a derivation of a long string,
variables are repeated

A derivation:

$S \Rightarrow AB \Rightarrow aBbB \Rightarrow abb\underline{B}$

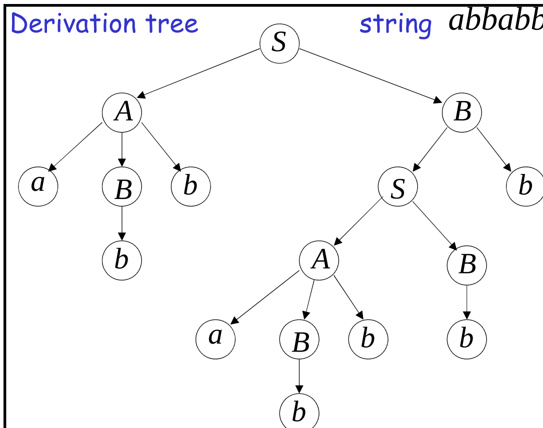
$\Rightarrow abbSb \Rightarrow abbABb \Rightarrow abbaBbBb \Rightarrow$

$\Rightarrow abbabb\underline{B}b \Rightarrow abbabbbb$

3

Derivation tree

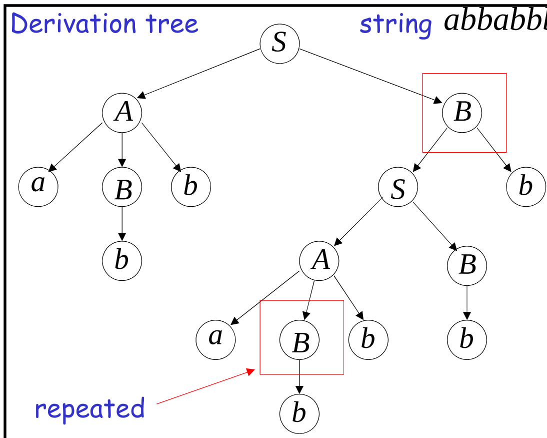
string $abbabbbb$



4

Derivation tree

string $abbabbbb$

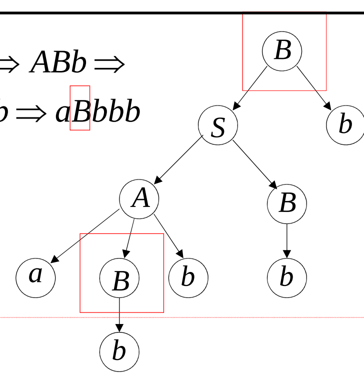


repeated

5

$\boxed{B} \Rightarrow Sb \Rightarrow ABb \Rightarrow$

$\Rightarrow aBbBb \Rightarrow a\boxed{B}bbb$

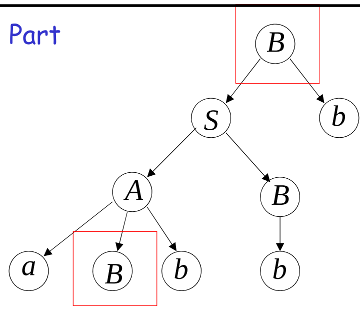


$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$

6

Repeated Part

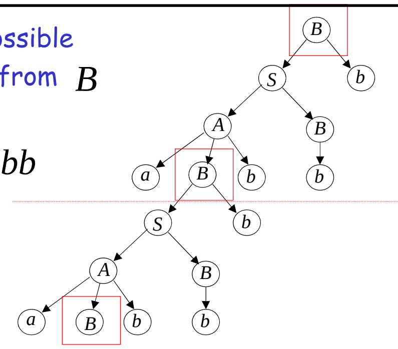


$$B \overset{*}{\Rightarrow} aBbbb$$

7

Another possible derivation from B

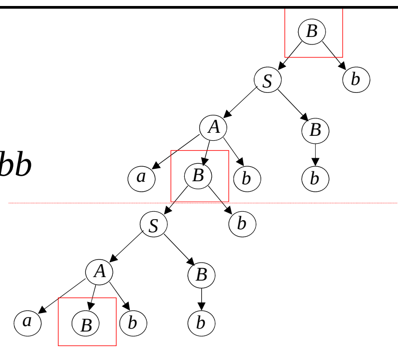
$$B \overset{*}{\Rightarrow} aBbbb$$



$$B \overset{*}{\Rightarrow} aBbbb \overset{*}{\Rightarrow} aaBbbbbbb$$

8

$$B \overset{*}{\Rightarrow} aBbbb$$



$$B \overset{*}{\Rightarrow} (a)B(bbb) \overset{*}{\Rightarrow} (a)^2 B(bbb)^2$$

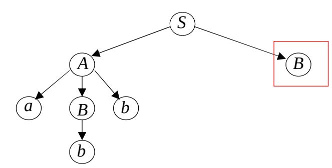
9

A Derivation from S

$$S \overset{*}{\Rightarrow} abbBb$$

$$B \overset{*}{\Rightarrow} aBbbb$$

$$B \Rightarrow b$$



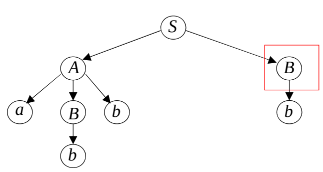
$$S \overset{*}{\Rightarrow} abbBb$$

10

$$S \overset{*}{\Rightarrow} abbBb$$

$$B \overset{*}{\Rightarrow} aBbbb$$

$$B \Rightarrow b$$



$$S \overset{*}{\Rightarrow} abbBb \Rightarrow abbbb \quad = abb(a)^0 b(bbb)^0$$

11

$S \overset{*}{\Rightarrow} abbBb$ $\quad B \overset{*}{\Rightarrow} aBbbb$ $\quad B \Rightarrow b$

$$\Downarrow$$

$$S \overset{*}{\Rightarrow} abb(a)^0 b(bbb)^0$$

$$\Downarrow$$

$$abb(a)^0 b(bbb)^0 \in L(G)$$

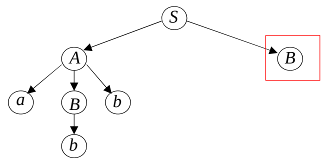
12

A Derivation from S

$S \overset{*}{\Rightarrow} abbBb$

$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$

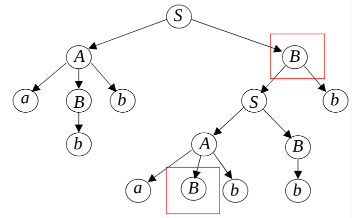


$$S \overset{*}{\Rightarrow} abbBb$$

$S \overset{*}{\Rightarrow} abbBb$

$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$

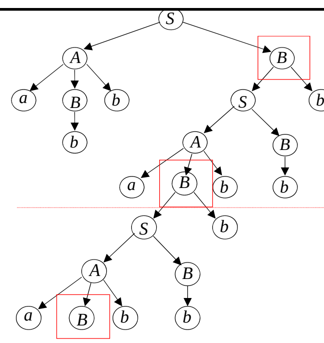


$$S \overset{*}{\Rightarrow} abbBb \overset{*}{\Rightarrow} abbaBbbb$$

$S \overset{*}{\Rightarrow} abbBb$

$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$

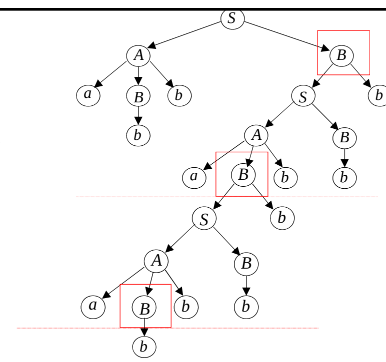


$$S \overset{*}{\Rightarrow} abb(a)B(bbb) \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2$$

$S \overset{*}{\Rightarrow} abbBb$

$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$



$$S \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2 \overset{*}{\Rightarrow} abb(a)^2 b(bbb)^2$$

$S \overset{*}{\Rightarrow} abbBb$ $\quad$ $B \overset{*}{\Rightarrow} aBbbb$ $\quad$ $B \Rightarrow b$

$$S \overset{*}{\Rightarrow} abb(a)^2 b(bbb)^2$$

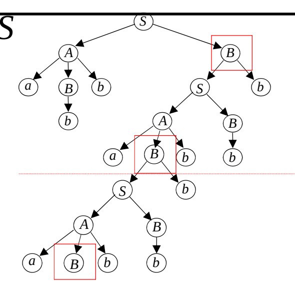
$$abb(a)^2 b(bbb)^2 \in L(G)$$

A Derivation from S

$S \overset{*}{\Rightarrow} abbBb$

$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$

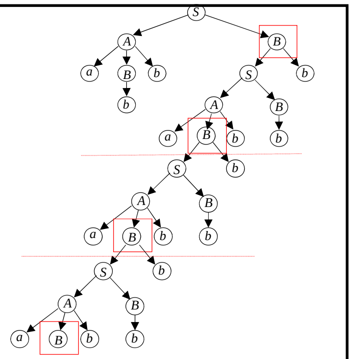


$$S \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2$$

$S \overset{*}{\Rightarrow} abbBb$

$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$



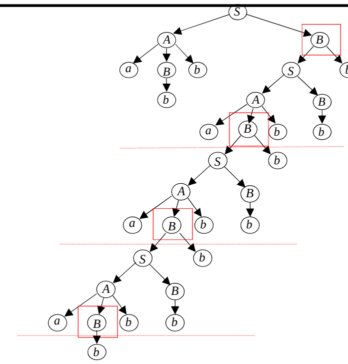
$$S \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2 \overset{*}{\Rightarrow} abb(a)^3 B(bbb)^3$$

19

$S \overset{*}{\Rightarrow} abbBb$

$B \overset{*}{\Rightarrow} aBbbb$

$B \Rightarrow b$



$$S \overset{*}{\Rightarrow} abb(a)^3 B(bbb)^3 \Rightarrow abb(a)^3 b(bbb)^3$$

20

$S \overset{*}{\Rightarrow} abbBb \qquad B \overset{*}{\Rightarrow} aBbbb \qquad B \Rightarrow b$

$$S \overset{*}{\Rightarrow} abb(a)^3 b(bbb)^3$$

$$abb(a)^3 b(bbb)^3 \in L(G)$$

21

In General:

$S \overset{*}{\Rightarrow} abbBb \qquad B \overset{*}{\Rightarrow} aBbbb \qquad B \Rightarrow b$

$$S \overset{*}{\Rightarrow} abb(a)^i b(bbb)^i$$

$$abb(a)^i b(bbb)^i \in L(G) \qquad i \geq 0$$

22

Consider now an infinite context-free language L

Let G be the grammar of $L - \{\lambda\}$

Take G so that I has no unit-productions

no λ-productions

23

Let p = (Number of productions) × (Largest right side of a production)

Let $m = p + 1$

Example G: $S \to AB$

$A \to aBb$

$B \to Sb$

$B \to b$

$p = 4 \times 3 = 12$

$m = p + 1 = 13$

24

Take a string $w \in L(G)$
with length $|w| \geq m$

We will show:
in the derivation of w
a variable of G is repeated

25

$$S \overset{*}{\Rightarrow} w$$

$$v_1 \Rightarrow v_2 \Rightarrow \cdots \Rightarrow v_k \Rightarrow w$$

$$S = v_1$$

26

$$v_1 \Rightarrow v_2 \Rightarrow \cdots \Rightarrow v_k \Rightarrow w$$

$|v_i| < |v_{i+1}| + f$ ← maximum right hand side of any production

$$\Downarrow$$

$$|w| < k \cdot f$$

$$\Downarrow$$

$m \leq |w| \leq k \cdot f$ ⇒ $p < k \cdot f$

27

$$v_1 \Rightarrow v_2 \Rightarrow \cdots \Rightarrow v_k \Rightarrow w$$

$$p < k \cdot f$$

$$\Downarrow$$

$k > \frac{p}{f}$ ← Number of productions in grammar

28

$$v_1 \Rightarrow v_2 \Rightarrow \cdots \Rightarrow v_k \Rightarrow w$$

$k >$ Number of productions in grammar

$$\Downarrow$$

Some production must be repeated

$$v_1 \Rightarrow \cdots \Rightarrow a_1 A a_2 \Rightarrow \cdots \Rightarrow a_3 A a_4 \Rightarrow \cdots \Rightarrow w$$

Repeated variable

$S \rightarrow r_1$
$A \rightarrow r_2$
$B \rightarrow r_2$
$\ldots$

29

$w \in L(G)$ $\quad |w| \geq m$

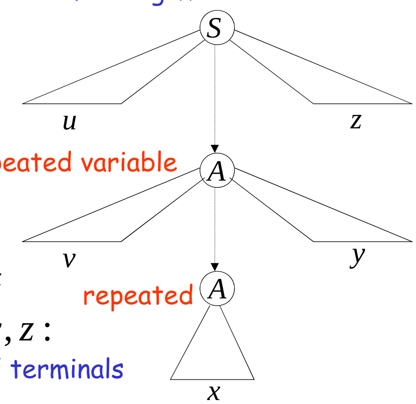
Derivation of string w

$$S \Rightarrow \cdots \Rightarrow a_1 A a_2 \Rightarrow \cdots \Rightarrow a_3 A a_4 \Rightarrow \cdots \Rightarrow w$$

Some variable is repeated

30

Derivation tree of string w



Last repeated variable

repeated

$w = uvxyz$

u, v, x, y, z :

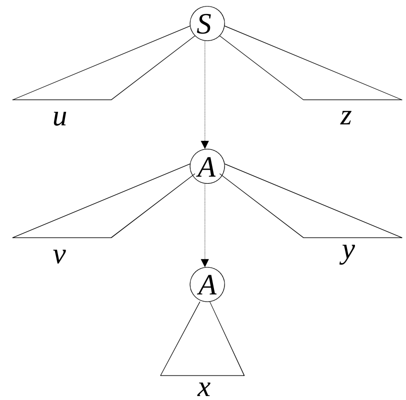
Strings of terminals

Possible derivations:

$S \stackrel{*}{\Rightarrow} uAz$

$A \stackrel{*}{\Rightarrow} vAy$

$A \stackrel{*}{\Rightarrow} x$



We know:

$S \stackrel{*}{\Rightarrow} uAz \qquad A \stackrel{*}{\Rightarrow} vAy \qquad A \stackrel{*}{\Rightarrow} x$

This string is also generated:

$$S \stackrel{*}{\Rightarrow} uAz \stackrel{*}{\Rightarrow} uxz$$

$$uv^0xy^0z$$

We know:

$S \stackrel{*}{\Rightarrow} uAz \qquad A \stackrel{*}{\Rightarrow} vAy \qquad A \stackrel{*}{\Rightarrow} x$

This string is also generated:

$$S \stackrel{*}{\Rightarrow} uAz \stackrel{*}{\Rightarrow} uvAyz \stackrel{*}{\Rightarrow} uvxyz$$

The original $w = uv^1xy^1z$

We know:

$S \stackrel{*}{\Rightarrow} uAz \qquad A \stackrel{*}{\Rightarrow} vAy \qquad A \stackrel{*}{\Rightarrow} x$

This string is also generated:

$$S \stackrel{*}{\Rightarrow} uAz \stackrel{*}{\Rightarrow} uvAyz \stackrel{*}{\Rightarrow} uvvAyyz \stackrel{*}{\Rightarrow} uvvxyyz$$

$$uv^2xy^2z$$

We know:

$S \stackrel{*}{\Rightarrow} uAz \qquad A \stackrel{*}{\Rightarrow} vAy \qquad A \stackrel{*}{\Rightarrow} x$

This string is also generated:

$$S \stackrel{*}{\Rightarrow} uAz \Rightarrow uvAyz \stackrel{*}{\Rightarrow} uvvAyyz \stackrel{*}{\Rightarrow}$$

$$\stackrel{*}{\Rightarrow} uvvvAyyyz \stackrel{*}{\Rightarrow} uvvvxyyyz$$

$$uv^3xy^3z$$

We know:

$$S \overset{*}{\Rightarrow} uAz \qquad A \overset{*}{\Rightarrow} vAy \qquad A \overset{*}{\Rightarrow} x$$

This string is also generated:

$$S \overset{*}{\Rightarrow} uAz \overset{*}{\Rightarrow} uvAyz \overset{*}{\Rightarrow} uvvAyyz \overset{*}{\Rightarrow}$$
$$\overset{*}{\Rightarrow} uvvvAyyyz \overset{*}{\Rightarrow} \ldots$$
$$\overset{*}{\Rightarrow} uvvv\cdots vAy\cdots yyyz \overset{*}{\Rightarrow}$$
$$\overset{*}{\Rightarrow} uvvv\cdots vxy\cdots yyyz$$

$$uv^i xy^i z$$

Therefore, any string of the form

$$uv^i xy^i z \qquad i \geq 0$$

is generated by the grammar G

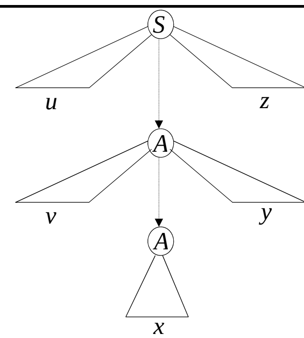
Therefore,

knowing that $uvxyz \in L(G)$

we also know that $uv^i xy^i z \in L(G)$

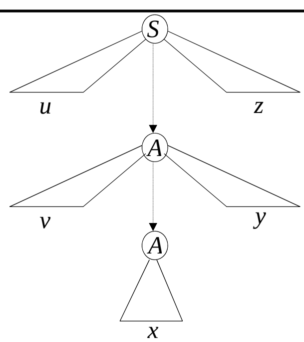
$L(G) = L - \{\lambda\}$

$$uv^i xy^i z \in L$$



Observation: $|vxy| \leq m$

Since A is the last repeated variable



Observation: $|vy| \geq 1$

Since there are no unit or λ-productions

The Pumping Lemma:

For infinite context-free language L

there exists an integer m such that

for any string $w \in L, \quad |w| \geq m$

we can write $w = uvxyz$

with lengths $|vxy| \leq m$ and $|vy| \geq 1$

and it must be:

$$uv^i xy^i z \in L, \quad \text{for all } i \geq 0$$

Applications of The Pumping Lemma

Non-context free languages

$\{a^n b^n c^n : n \geq 0\}$

Context-free languages

$\{a^n b^n : n \geq 0\}$

Theorem: The language

$$L = \{a^n b^n c^n : n \geq 0\}$$

is **not** context free

Proof: Use the Pumping Lemma for context-free languages

$$L = \{a^n b^n c^n : n \geq 0\}$$

Assume for contradiction that L is context-free

Since L is context-free and infinite we can apply the pumping lemma

$$L = \{a^n b^n c^n : n \geq 0\}$$

Pumping Lemma gives a magic number m such that:

Pick any string $w \in L$ with length $|w| \geq m$

We pick: $w = a^m b^m c^m$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$$w = a^m b^m c^m$$

We can write: $w = uvxyz$

with lengths $|vxy| \leq m$ and $|vy| \geq 1$

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Pumping Lemma says:

$uv^i xy^i z \in L$ for all $i \ge 0$

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

We examine **all** the possible locations of string vxy in w

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 1: vxy is within a^m

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$u \quad vxy \qquad\qquad z$$

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 1: v and y consist from only a

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$u \quad vxy \qquad\qquad z$$

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 1: Repeating v and y

$k \ge 1$

$$\overbrace{aaaaaa...aaaaaa}^{m+k}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$u \qquad v^2xy^2 \qquad\qquad z$$

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 1: From Pumping Lemma: $uv^2xy^2z \in L$

$k \ge 1$

$$\overbrace{aaaaaa...aaaaaa}^{m+k}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$u \qquad v^2xy^2 \qquad\qquad z$$

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 1: From Pumping Lemma: $uv^2xy^2z \in L$

$k \geq 1$

However: $uv^2xy^2z = a^{m+k}b^m c^m \notin L$

Contradiction!!!

55

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 2: vxy is within b^m

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$u \qquad vxy \qquad z$

56

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 2: Similar analysis with case 1

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$u \qquad vxy \qquad z$

57

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 3: vxy is within c^m

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$u \qquad vxy \quad z$

58

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 3: Similar analysis with case 1

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$u \qquad vxy \quad z$

59

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 4: vxy overlaps a^m and b^m

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$u \qquad vxy \qquad z$

60

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w=a^m b^m c^m$

$w=uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 4: Possibility 1: v contains only a
y contains only b

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$u \qquad vxy \qquad z$$

61

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w=a^m b^m c^m$

$w=uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 4: Possibility 1: v contains only a
y contains only b

$k_1+k_2 \ge 1$

$$\overbrace{aaa...aaaaaaa}^{m+k_1}\ \overbrace{bbbbbbb...bbb}^{m+k_2}\ \overbrace{ccc...ccc}^{m}$$

$$u \qquad v^2xy^2 \qquad z$$

62

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w=a^m b^m c^m$

$w=uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 4: From Pumping Lemma: $uv^2xy^2z \in L$

$k_1+k_2 \ge 1$

$$\overbrace{aaa...aaaaaaa}^{m+k_1}\ \overbrace{bbbbbbb...bbb}^{m+k_2}\ \overbrace{ccc...ccc}^{m}$$

$$u \qquad v^2xy^2 \qquad z$$

63

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w=a^m b^m c^m$

$w=uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 4: From Pumping Lemma: $uv^2xy^2z \in L$

$k_1+k_2 \ge 1$

However: $uv^2xy^2z = a^{m+k_1} b^{m+k_2} c^m \notin L$

Contradiction!!!

64

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w=a^m b^m c^m$

$w=uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 4: Possibility 2: v contains a and b
y contains only b

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$u \qquad vxy \qquad z$$

65

$$L=\{a^n b^n c^n : n \ge 0\}$$

$w=a^m b^m c^m$

$w=uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$

Case 4: Possibility 2: v contains a and b
y contains only b

$k_1+k_2+k \ge 1$

$$\overbrace{aaa...aaaaa}^{m}\overbrace{bb}^{k_1}\overbrace{aa}^{k_2}\overbrace{bb\ bbbbbbb...bbb}^{m+k}\ \overbrace{ccc...ccc}^{m}$$

$$u \qquad v^2xy^2 \qquad z$$

66

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 4: From Pumping Lemma: $uv^2xy^2z \in L$

$k_1 + k_2 + k \geq 1$

$$\underbrace{aa}_{u}\overbrace{a...aaaaa}^{m}\underbrace{\overbrace{bb}^{k_1}\overbrace{aa}^{k_2}\overbrace{bb\ bbbbbbb...bbb}^{m+k}}_{v^2xy^2}\underbrace{\overbrace{ccc...ccc}^{m}}_{z}$$

67

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 4: From Pumping Lemma: $uv^2xy^2z \in L$

However: $k_1 + k_2 + k \geq 1$

$$uv^2xy^2z = a^m b^{k_1} a^{k_2} b^{m+k} c^m \notin L$$

Contradiction!!!

68

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 4: Possibility 3: v contains only a

y contains a and b

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

u vxy z

69

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 4: Possibility 3: v contains only a

y contains a and b

Similar analysis with Possibility 2

70

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

$\boxed{\textbf{Case 5:}\ vxy \text{ overlaps } b^m \text{ and } c^m}$

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

u vxy z

71

$$L = \{a^n b^n c^n : n \geq 0\}$$

$w = a^m b^m c^m$

$w = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$

Case 5: Similar analysis with case 4

$$\overbrace{aaa...aaa}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

u vxy z

72

There are no other cases to consider

(since $|vxy| \leq m$, string vxy cannot

overlap a^m, b^m and c^m at the same time)

73

In all cases we obtained a contradiction

Therefore: The original assumption that

$L = \{a^n b^n c^n : n \geq 0\}$

is context-free must be wrong

Conclusion: L is not context-free

74