Department of Computer & Information Sciences
Pakistan Institute of Engineering and Applied Sciences

The Pumping Lemma for Context-Free Languages

Umar Faiz

http://www.pieas.edu.pk/umarfaiz/cis317

Pumping Lemma for CFLs

Take an infinite context-free language. It generates an infinite number of different strings

Example:

$$S \to AB$$
$$A \to aBb$$
$$B \to Sb$$
$$B \to b$$

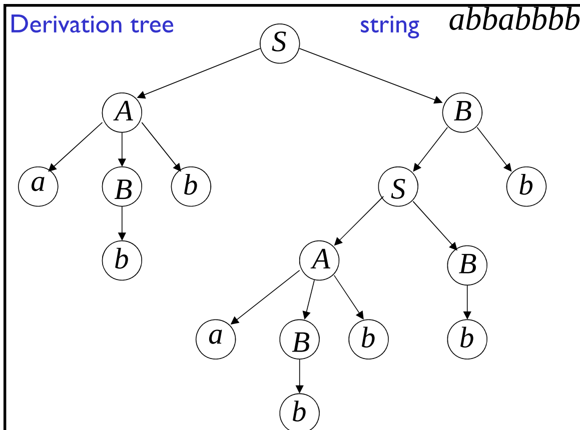
In a derivation of a long string, variables are repeated

$$S \to AB$$
$$A \to aBb$$
$$B \to Sb$$
$$B \to b$$

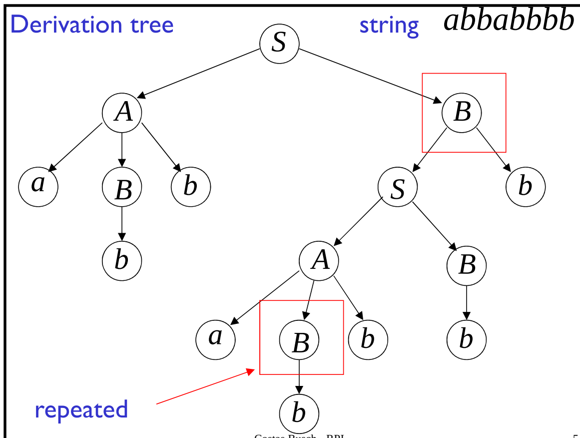
A derivation:

$$S \Rightarrow AB \Rightarrow aBbB \Rightarrow abb\underline{B}$$
$$\Rightarrow abbSb \Rightarrow abbABb \Rightarrow abbaBbBb \Rightarrow$$
$$\Rightarrow abbabb\underline{B}b \Rightarrow abbabbbb$$

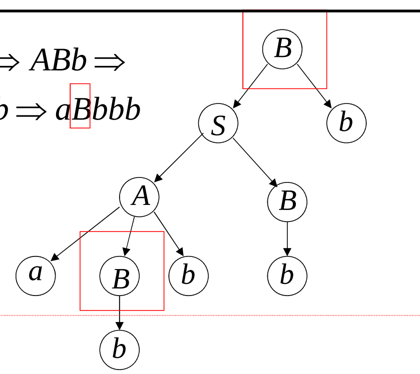
Derivation tree string $abbabbbb$



Derivation tree string $abbabbbb$

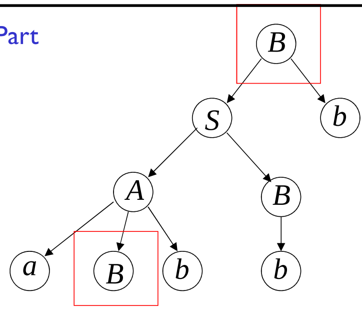


$$B \Rightarrow Sb \Rightarrow ABb \Rightarrow$$
$$\Rightarrow aBbBb \Rightarrow aBbbb$$



$$B \stackrel{*}{\Rightarrow} aBbbb \qquad B \Rightarrow b$$

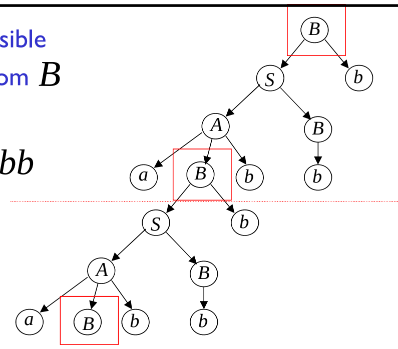
Repeated Part



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

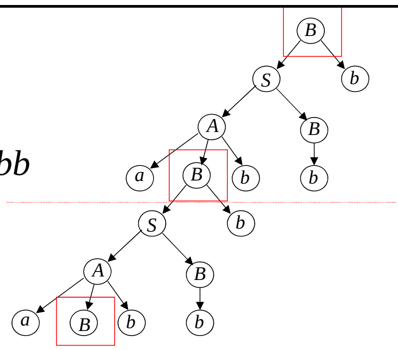
Another possible derivation from B

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



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

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



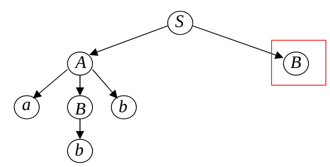
$$B \overset{*}{\Rightarrow} (a)B(bbb) \overset{*}{\Rightarrow} (a)^2 B(bbb)^2$$

A Derivation from S

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

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

$$B \Rightarrow b$$

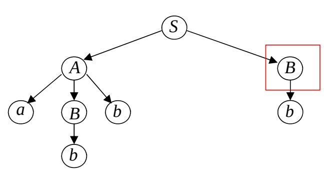


$$S \overset{*}{\Rightarrow} abbBb$$

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

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

$$B \Rightarrow b$$



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

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

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

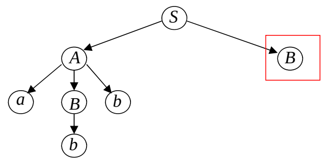
$$abb(a)^0 b(bbb)^0 \in L(G)$$

A Derivation from S

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

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

$B \Rightarrow b$

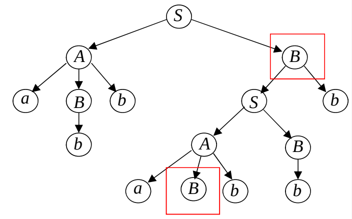


$$S \overset{*}{\Rightarrow} abbBb$$

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

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

$B \Rightarrow b$

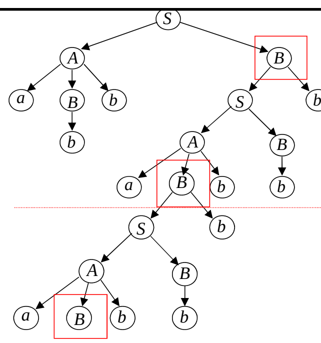


$$S \overset{*}{\Rightarrow} abbBb \overset{*}{\Rightarrow} abbaBbbb$$

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

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

$B \Rightarrow b$

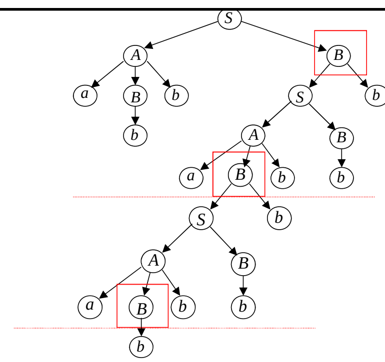


$$S \overset{*}{\Rightarrow} abb(a)B(bbb) \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2$$

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

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

$B \Rightarrow b$



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

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

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

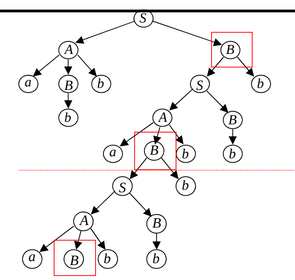
$$abb(a)^2 b(bbb)^2 \in L(G)$$

A Derivation from S

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

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

$B \Rightarrow b$

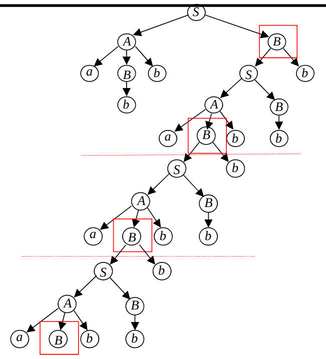


$$S \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2$$

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

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

$B \Rightarrow b$

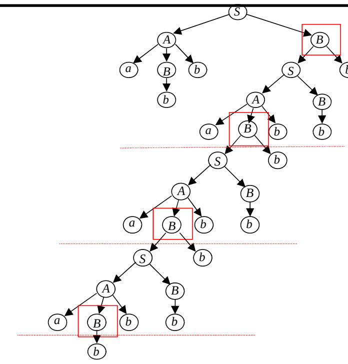


$$S \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2 \overset{*}{\Rightarrow} abb(a)^3 B(bbb)^3$$

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

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

$B \Rightarrow b$



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

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

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

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

In General:

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

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

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

Consider now an infinite context-free language L

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

Take G so that it has no unit-productions
no λ -productions

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

Let $m = p + 1$

Example G

$S \rightarrow AB$
$A \rightarrow aBb$
$B \rightarrow Sb$
$B \rightarrow b$

$p = 4 \times 3 = 12$

$m = p + 1 = 13$

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

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

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

$$S = v_1$$

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

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

⇓

$|w| < k \cdot f$

⇓

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

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

$p < k \cdot f$

⇓

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

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

$k >$ Number of productions in grammar

⇓

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

$S \rightarrow r_1$

$A \rightarrow r_2$

$B \rightarrow r_2$

$\ldots$

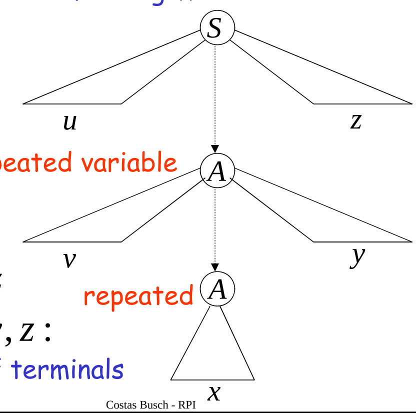
$w \in L(G)$ $|w| \geq m$

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

Derivation tree of string w



Last repeated variable

repeated

$w = uvxyz$

u, v, x, y, z :

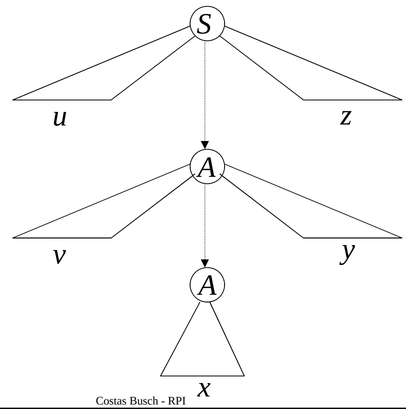
Strings of terminals

Possible derivations:

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

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

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



We know:

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

This string is also generated:

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

$$uv^0xy^0z$$

We know:

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

This string is also generated:

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

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

We know:

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

This string is also generated:

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

$$uv^2xy^2z$$

We know:

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

This string is also generated:

$$S \overset{*}{\Rightarrow} uAz \Rightarrow uvAyz \overset{*}{\Rightarrow} uvvAyyz \overset{*}{\Rightarrow}$$
$$\overset{*}{\Rightarrow} uvvvAyyyz \overset{*}{\Rightarrow} uvvvxyyyz$$

$$uv^3xy^3z$$

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

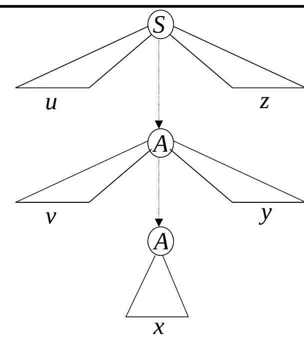
Therefore,

knowing that $uvxyz \in L(G)$

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

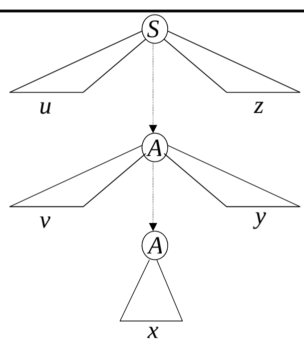
$L(G) = L - \{\lambda\}$

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



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 $|vy| \geq 1$

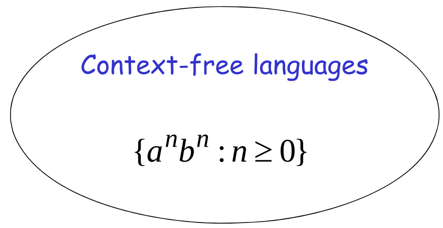
and it must be:

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

Applications of The Pumping Lemma

Non-context free languages

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



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

$m \quad m \quad m$

$aaa...aaa\ bbb...bbb\ ccc...ccc$

$u \quad vxy \quad z$

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

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

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

Case 1: v and y consist from only a

$m \quad m \quad m$

$aaa...aaa\ bbb...bbb\ ccc...ccc$

$u \quad vxy \quad z$

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

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

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

Case 1: Repeating v and y

$k \ge 1$

$m + k \quad m \quad m$

$aaaaaa...aaaaaa\ bbb...bbb\ ccc...ccc$

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

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

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

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

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

$k \ge 1$

$m + k \quad m \quad m$

$aaaaaa...aaaaaa\ bbb...bbb\ ccc...ccc$

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

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

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

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

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!!!

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

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

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

Case 2: vxy is within b^m

$m \quad m \quad m$

$aaa...aaa\, bbb...bbb\, ccc...ccc$

$u \quad vxy \quad z$

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

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

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

Case 2: Similar analysis with case 1

$m \quad m \quad m$

$aaa...aaa\, bbb...bbb\, ccc...ccc$

$u \quad vxy \quad z$

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

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

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

Case 3: vxy is within c^m

$m \quad m \quad m$

$aaa...aaa\, bbb...bbb\, ccc...ccc$

$u \quad vxy \quad z$

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

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

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

Case 3: Similar analysis with case 1

$m \quad m \quad m$

$aaa...aaa\, bbb...bbb\, ccc...ccc$

$u \quad vxy \quad z$

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

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

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

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

$m \quad m \quad m$

$aaa...aaa\, bbb...bbb\, ccc...ccc$

$u \quad vxy \quad z$

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

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

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

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

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

$u \qquad vxy \qquad z$

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

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

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

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

$k_1 + k_2 \ge 1$

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

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

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

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

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

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

$k_1 + k_2 \ge 1$

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

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

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

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

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

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!!!

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

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

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

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

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

$u \qquad vxy \qquad z$

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

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

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

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

$k_1 + k_2 + k \ge 1$

$$\overbrace{aaa...aaaa}^{m}\overbrace{abb}^{k_1}\overbrace{aabb}^{k_2}\,\overbrace{bbbbbbb...bbb}^{m+k}\,\overbrace{ccc...ccc}^{m}$$

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

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

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

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

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

$k_1 + k_2 + k \geq 1$

$m \quad k_1 \quad k_2 \quad m+k \quad m$

aaa...aaaaabbaabb bbbbbbb...bbb ccc...ccc

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

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

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

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

Case 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!!!

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

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

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

Case 4: Possibility 3: v contains only a
y contains a and b

$m \quad m \quad m$

aaa...aaa bbb...bbb ccc...ccc

$u \qquad vxy \qquad z$

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

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

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

Case 4: Possibility 3: v contains only a
y contains a and b

Similar analysis with Possibility 2

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

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

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

Case 5: vxy overlaps b^m and c^m

$m \quad m \quad m$

aaa...aaa bbb...bbb ccc...ccc

$u \qquad vxy \qquad z$

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

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

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

Case 5: Similar analysis with case 4

$m \quad m \quad m$

aaa...aaa bbb...bbb ccc...ccc

$u \qquad vxy \qquad z$

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)

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