Pumping Lemma for Context-free Languages

Take an **infinite** context-free language

Generates an infinite number of different strings

Example:

$$S \to ABE \mid bBd$$

$$A \to Aa \mid a$$

$$B \to bSD \mid cc$$

$$D \to Dd \mid d$$

$$E \to eE \mid e$$

Smallest length string accepted?

aabbccdd ee

In a derivation of a "long" enough string, variables are repeated

A possible derivation:

$$S \Rightarrow A\boxed{B}E \Rightarrow AaBE \Rightarrow aaBE$$

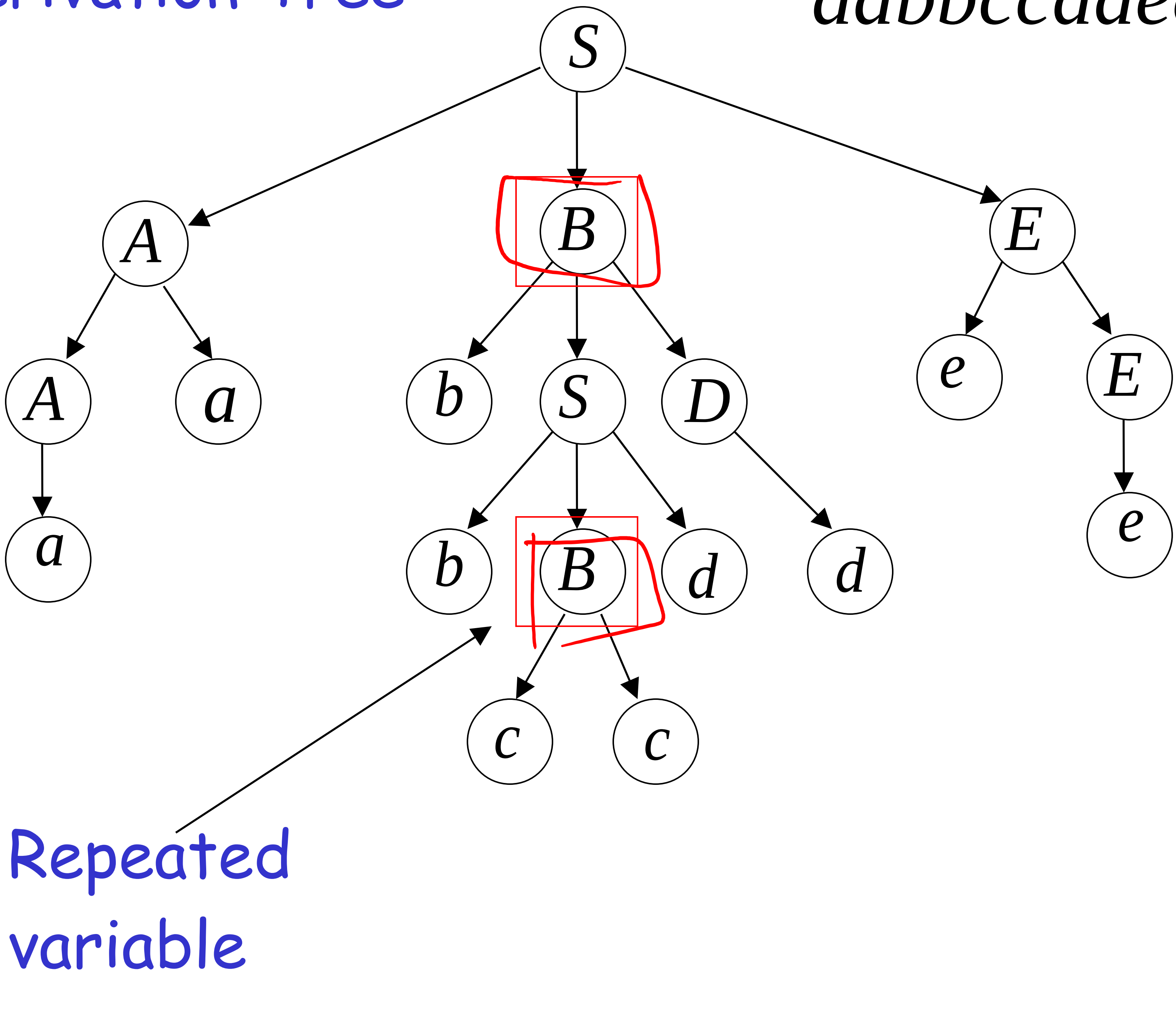
$$\Rightarrow aabSDE \Rightarrow aabb\boxed{B}dDE \Rightarrow$$

$$\Rightarrow aaabbccdDE \Rightarrow aabbccddE$$

$$\Rightarrow aabbccddeE \Rightarrow \underline{aabbccddee}$$

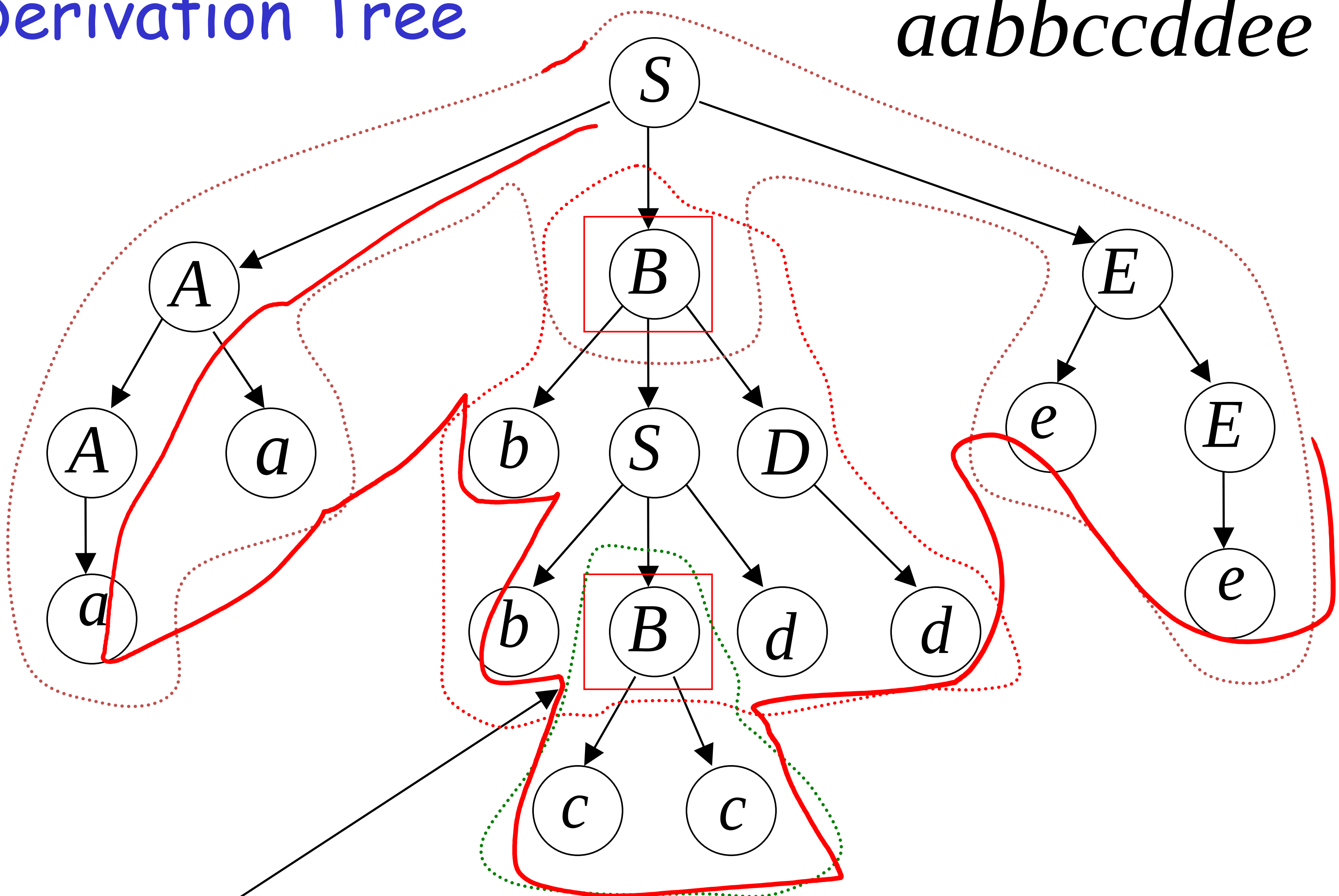
Derivation Tree

$aabbccddee$



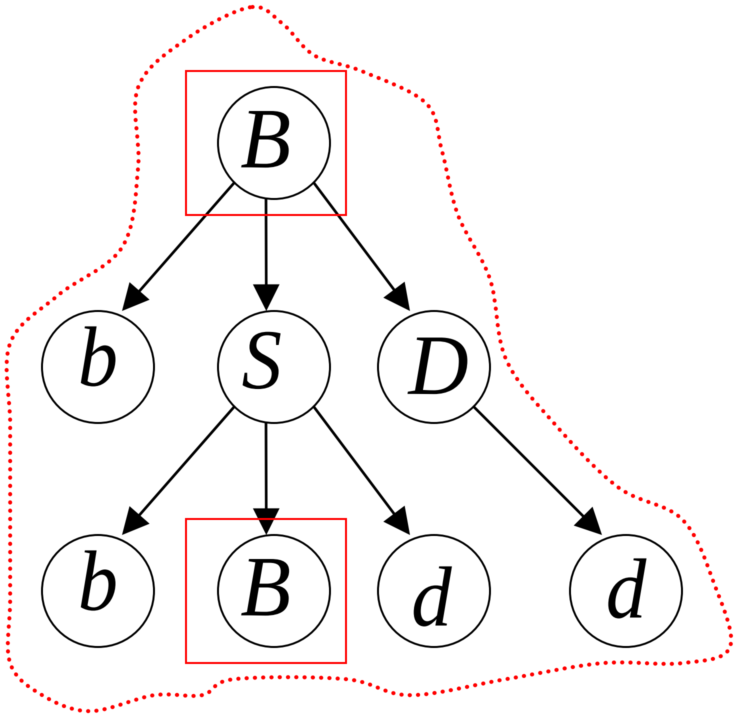
Derivation Tree

$aabbccddee$



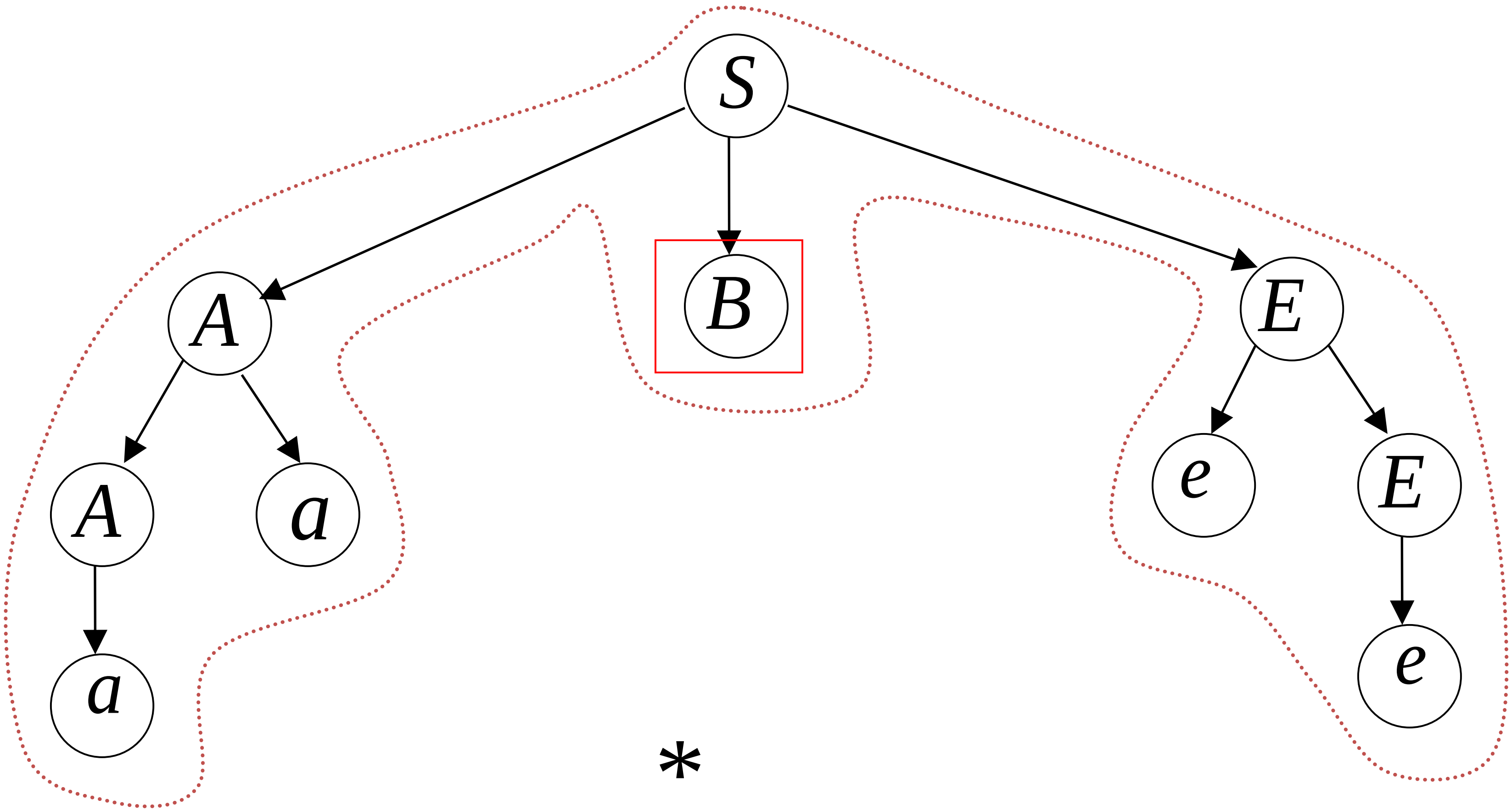
Repeated variable

$$B \Rightarrow bSD \Rightarrow bbBdD \Rightarrow bbBdd$$

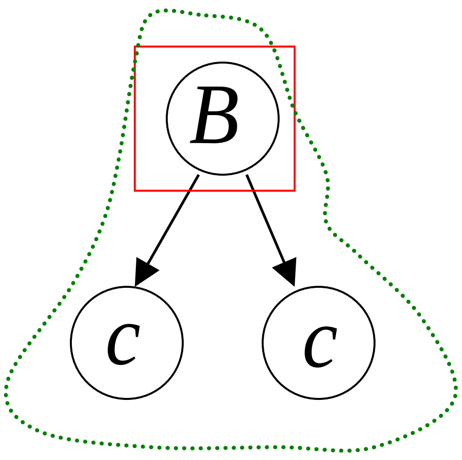


$$B \overset{*}{\Rightarrow} bbBdd$$

$$S \Rightarrow ABE \Rightarrow AaBE \Rightarrow aaBE \Rightarrow aaBeE \Rightarrow aaBee$$

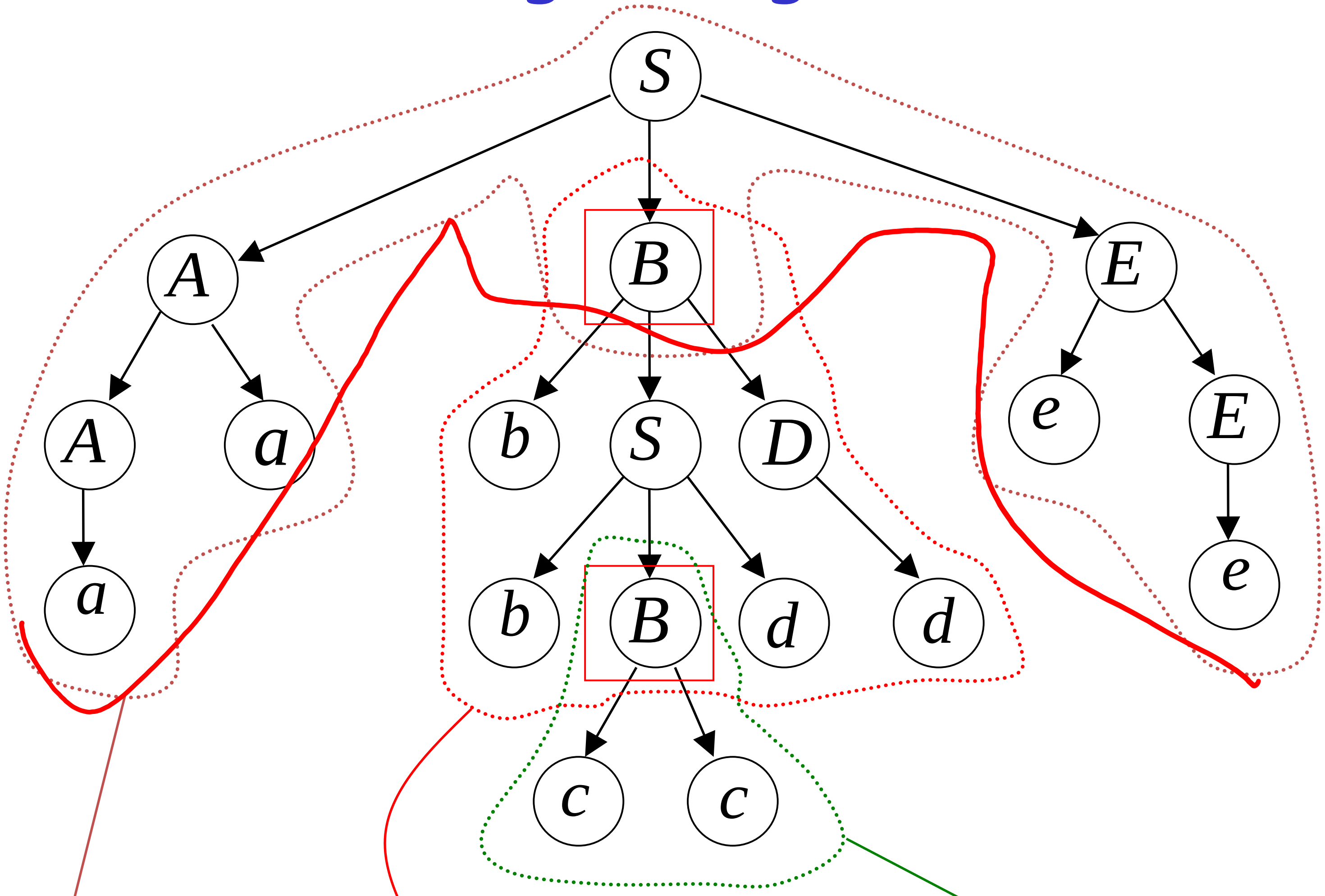


$$S \overset{*}{\Rightarrow} aaBee$$



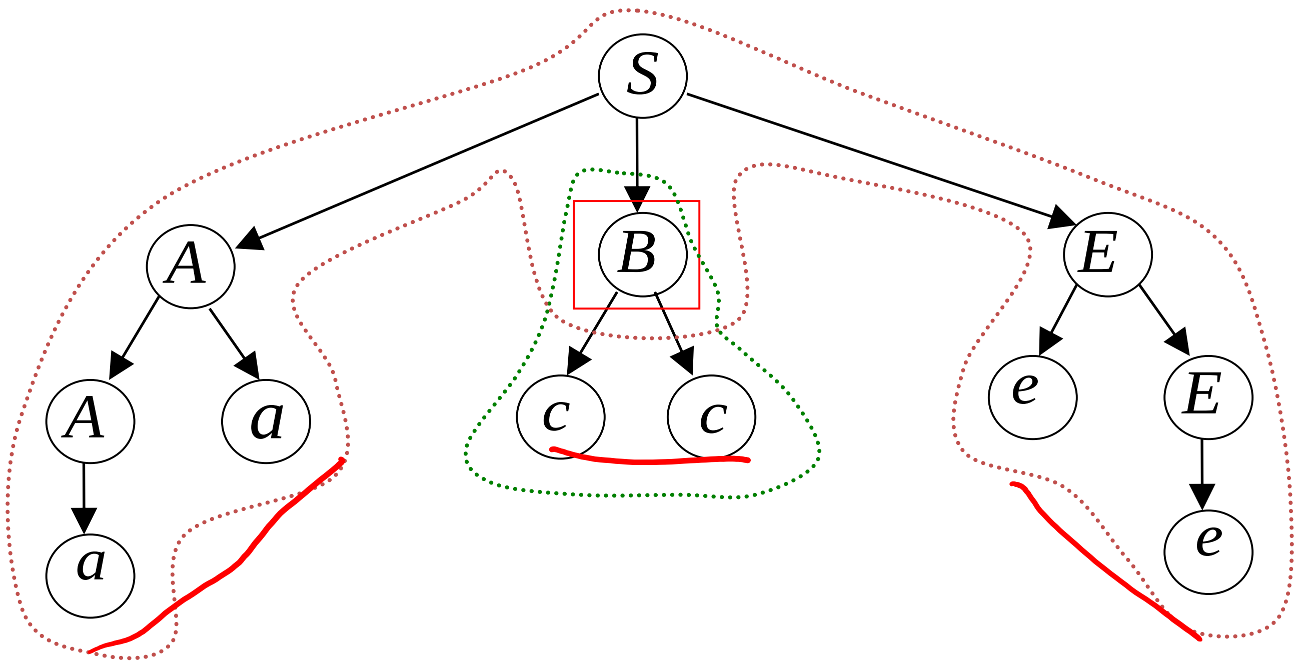
$$B \Rightarrow cc$$

Putting all together



$$S \overset{*}{\Rightarrow} aaBee \qquad B \overset{*}{\Rightarrow} bbBdd \qquad B \Rightarrow cc$$

We can remove the middle part

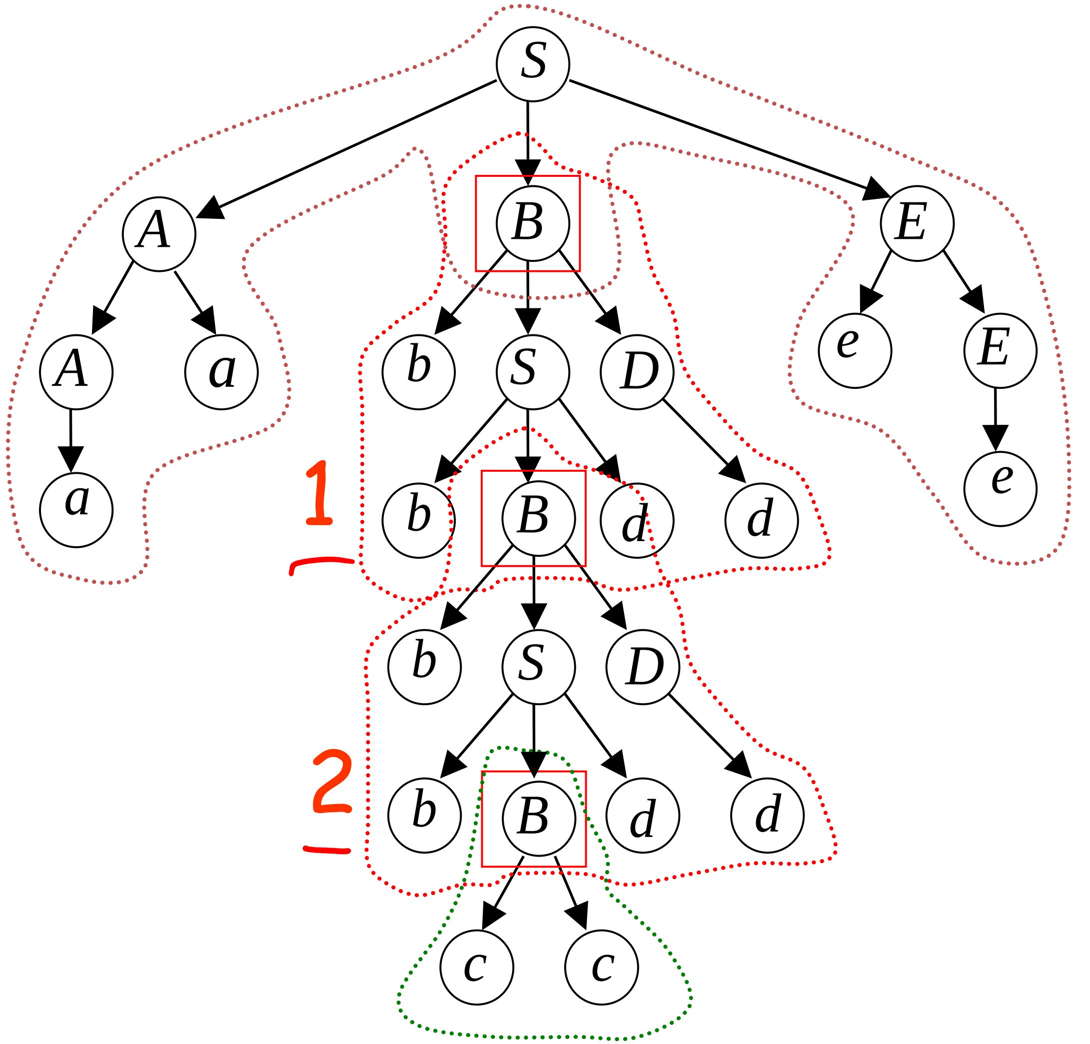


$$S \overset{*}{\Rightarrow} aaBee \qquad B \overset{*}{\Rightarrow} bbBdd \qquad B \Rightarrow cc$$

$$S \overset{*}{\Rightarrow} aaBee \overset{*}{\Rightarrow} aaccee \quad = aa(bb)^0 cc(dd)^0 ee$$

$$aa(bb)^0 cc(dd)^0 ee \in L(G)$$

We can repeated middle part two times



$$S \overset{*}{\Rightarrow} aa(\underbrace{bb}_{})^2 cc(\underbrace{dd}_{})^2 ee$$

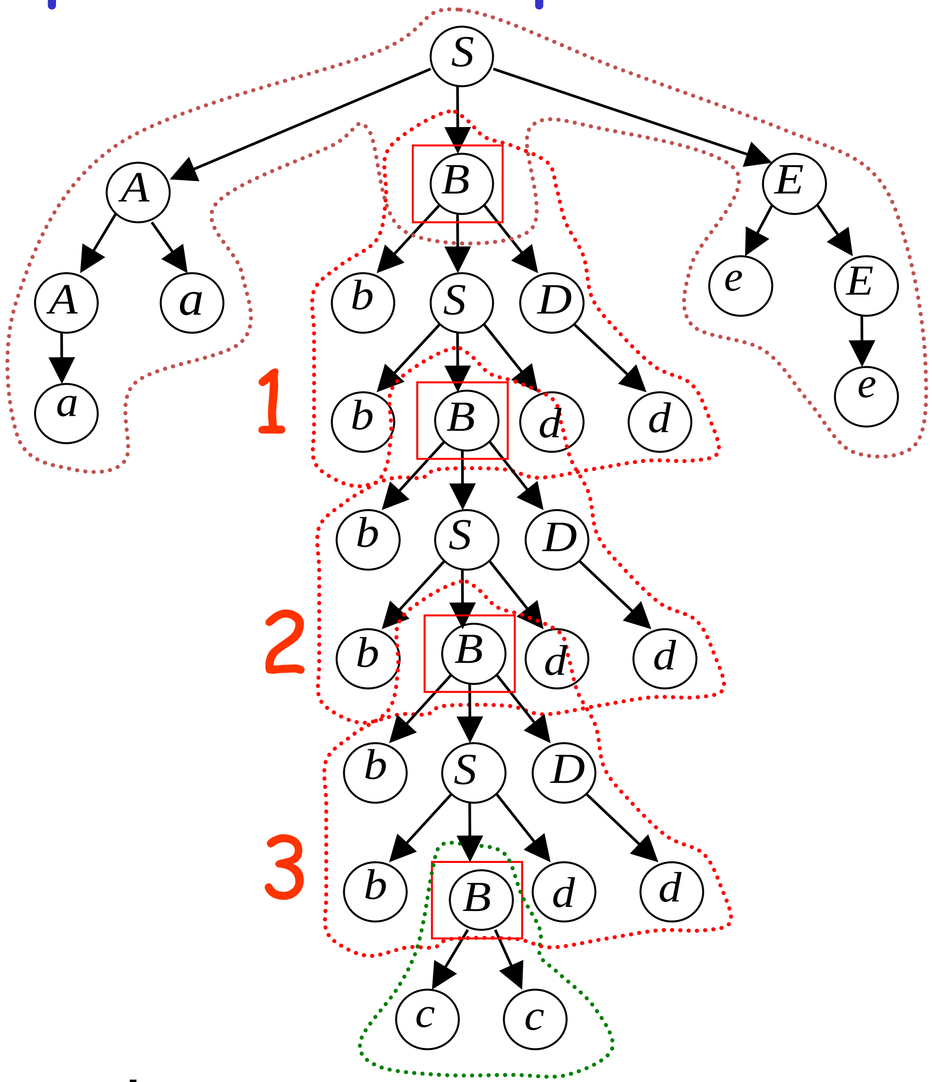
$$S \overset{*}{\Rightarrow} aaBee \qquad B \overset{*}{\Rightarrow} bbBdd \qquad B \Rightarrow cc$$

$$S \overset{*}{\Rightarrow} aaBee \overset{*}{\Rightarrow} aabbBddee$$

$$\overset{*}{\Rightarrow} aa(bb)^2 B(dd)^2 ee \overset{*}{\Rightarrow} aa(bb)^2 cc(dd)^2 ee$$

$$aa(bb)^2 cc(dd)^2 ee \quad \in L(G)$$

We can repeat middle part three times

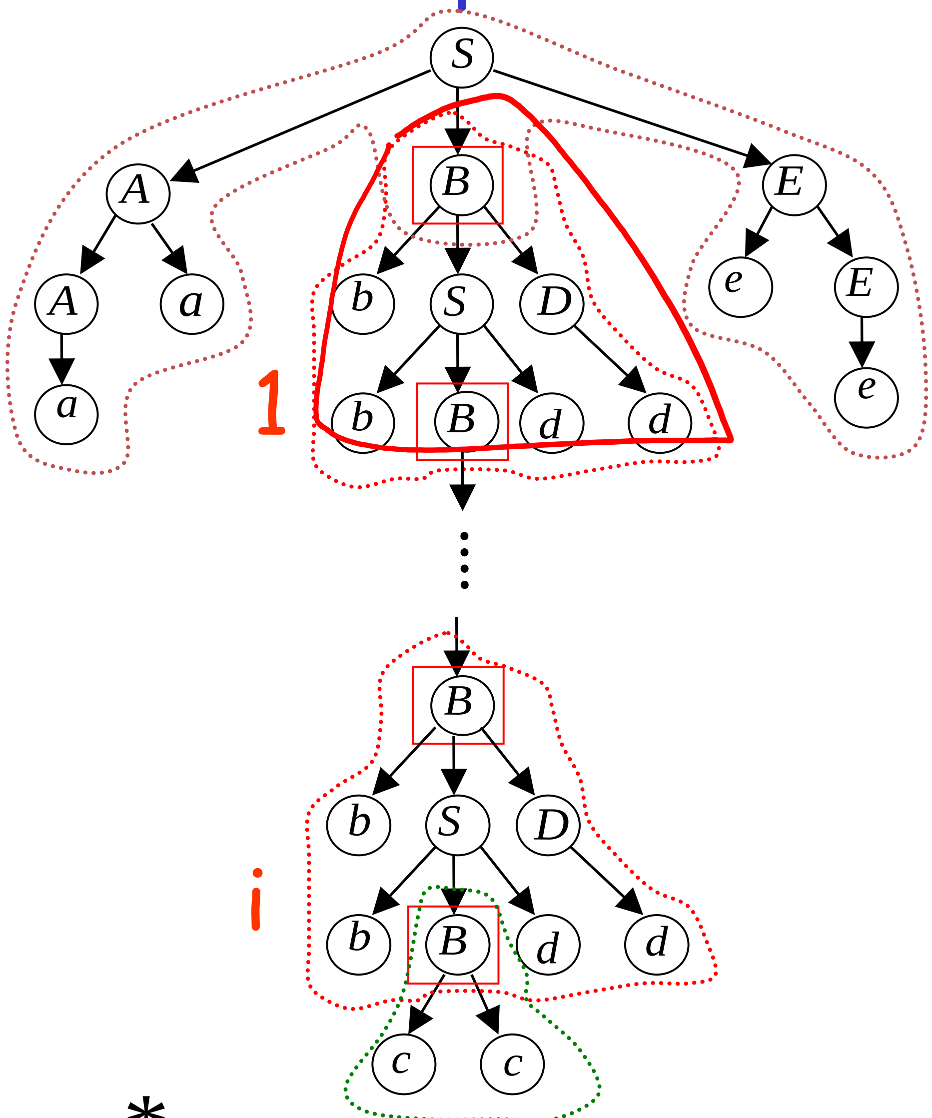


$$S \overset{*}{\Rightarrow} aa(bb)^3 cc(dd)^3 ee$$

$$S \overset{*}{\Rightarrow} aaBee \qquad B \overset{*}{\Rightarrow} bbBdd \qquad B \Rightarrow cc$$

$$S \overset{*}{\Rightarrow} aa(bb)^3cc(dd)^3ee \quad \in L(G)$$

Repeat middle part i times



$$S \overset{*}{\Rightarrow} aa(bb)^i cc(dd)^i ee$$

$$S \overset{*}{\Rightarrow} aaBee \qquad B \overset{*}{\Rightarrow} bbBdd \qquad B \Rightarrow cc$$

$$\Downarrow$$

$$S \overset{*}{\Rightarrow} \underline{aa(bb)^i cc(dd)^i ee \quad \in L(G)}$$

For any $i \geq 0$

From Grammar

$$S \to ABE \mid bBd$$

$$A \to Aa \mid a$$

$$B \to bSD \mid cc$$

$$D \to Dd \mid d$$

$$E \to eE \mid e$$

and given string

$$aabbccddee \quad \in L(G)$$

We inferred that a family of strings is in $L(G)$

$$S \stackrel{*}{\Rightarrow} \underline{aa(bb)^i cc(dd)^i} ee \quad \in L(G)$$ for any $i \geq 0$

Arbitrary Grammars

Consider now an arbitrary infinite context-free language L

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

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

(remove them)

Let r be the number of variables

Let t be the maximum right-hand size of any production

Example:

$$S \to ABE \mid bBd$$
$$A \to Aa \mid a$$
$$B \to bSD \mid cc$$
$$D \to Dd \mid d$$
$$E \to eE \mid e$$

$$r = 5$$

$$t = 3$$

S
A
B
D
E

Claim:

Take string $w \in L(G)$ with $|w| > t^r$.
Then in the derivation tree of w
there is a path from the root to a leaf
where a variable of G is repeated

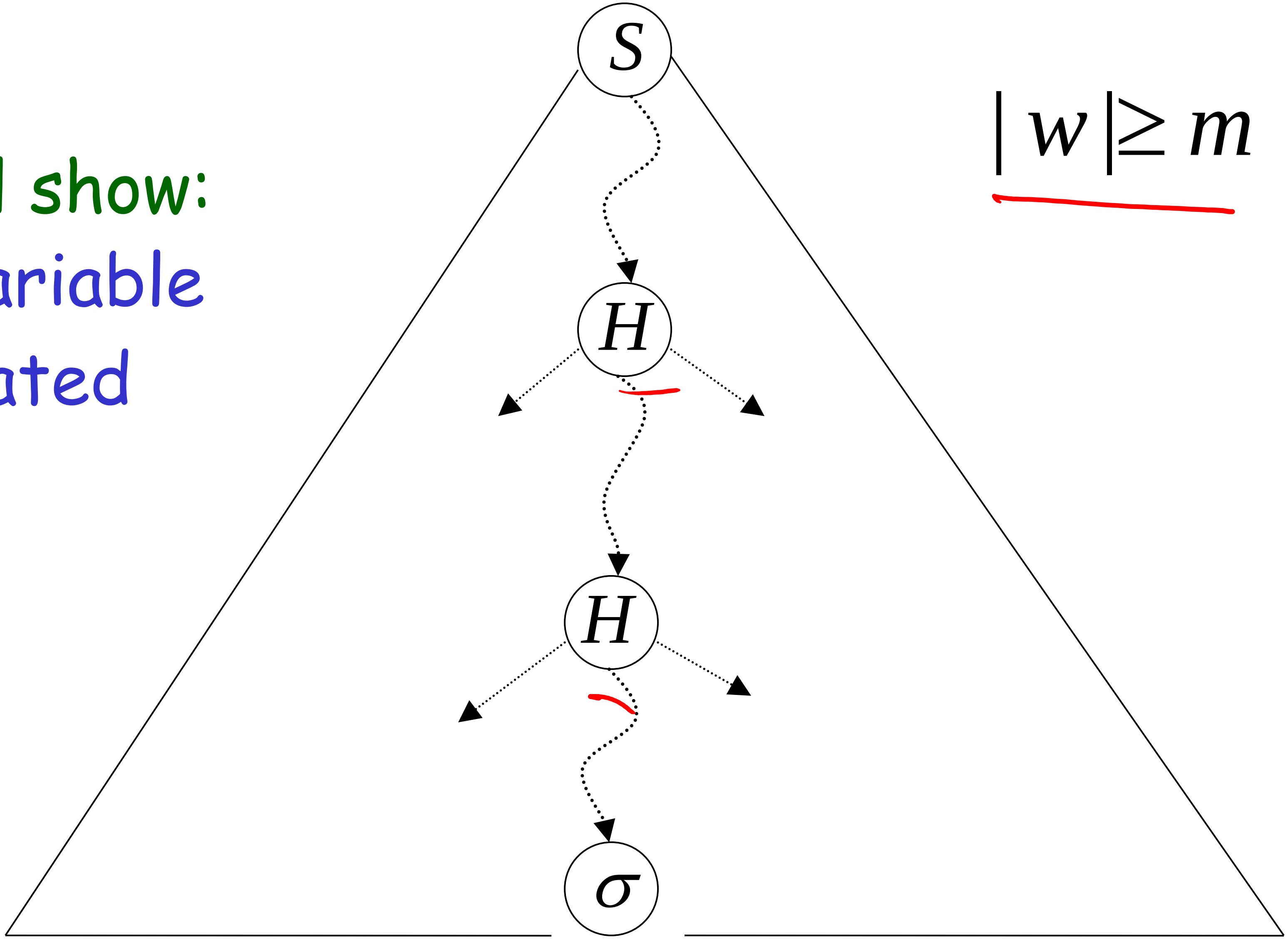
Proof:

Proof by contradiction

Derivation tree of w

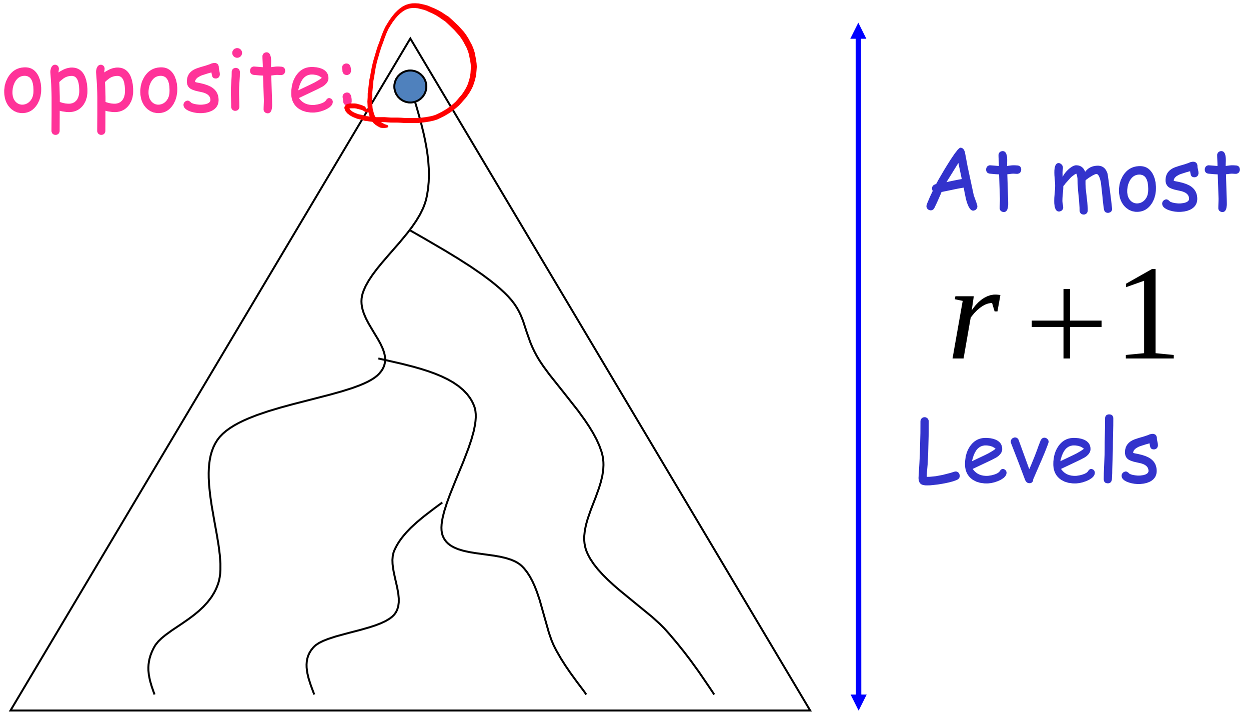
$|w| \geq m$

We will show:
some variable
is repeated

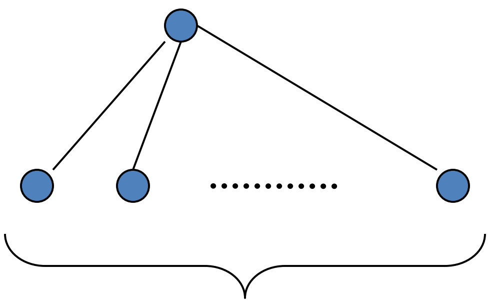


First we show that the tree of w has at least $r + 2$ levels of nodes

Suppose the opposite:



Maximum number of nodes per level



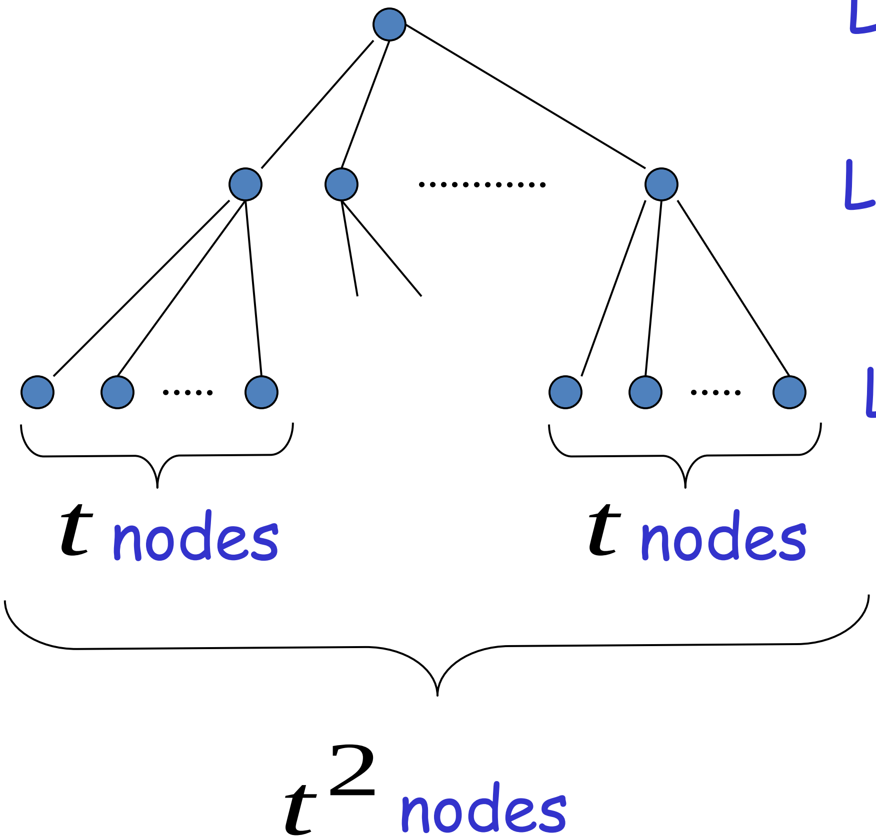
t nodes

Level 0: $\mathbf{1}$ nodes

Level 1: t nodes

The maximum right-hand side of any production

Maximum number of nodes per level

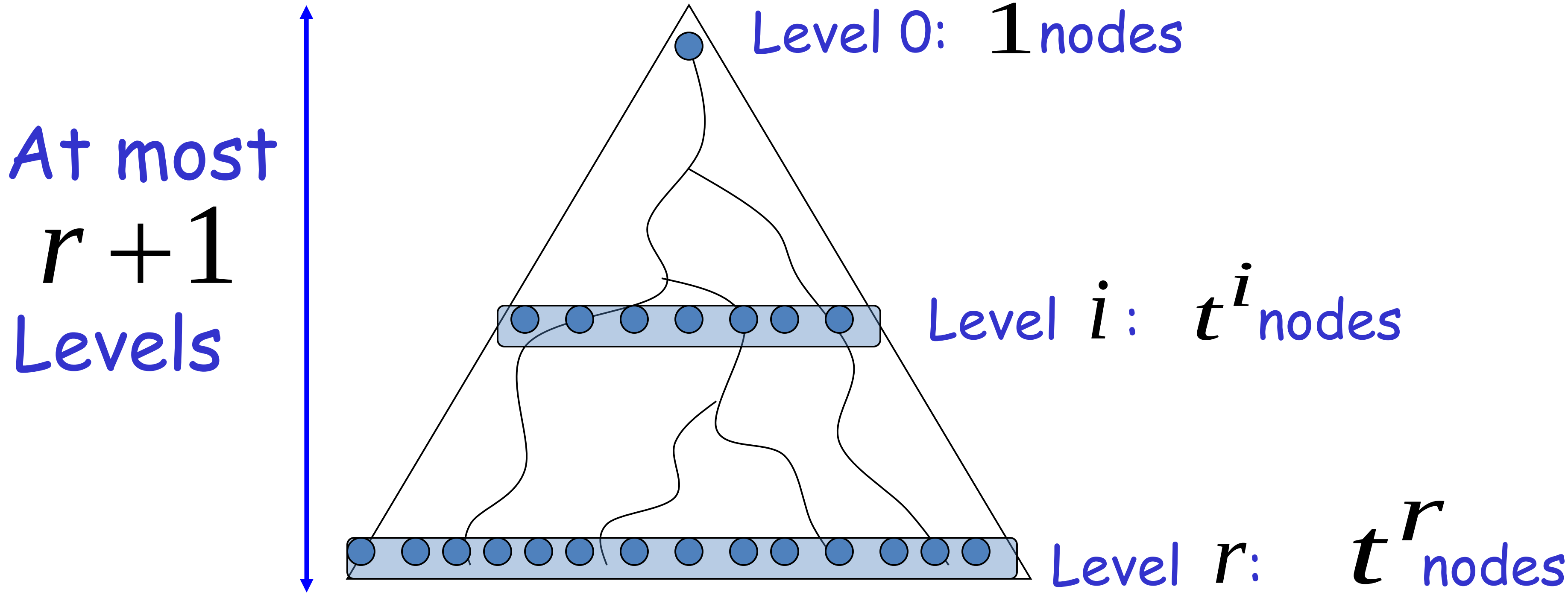


Level 0: 1 nodes

Level 1: t nodes

Level 2: t^2 nodes

Maximum number of nodes per level



Maximum possible string length

= max nodes at level r = $\boxed{t^r}$

Therefore,

maximum length of string w : $|w| \leq t^r$

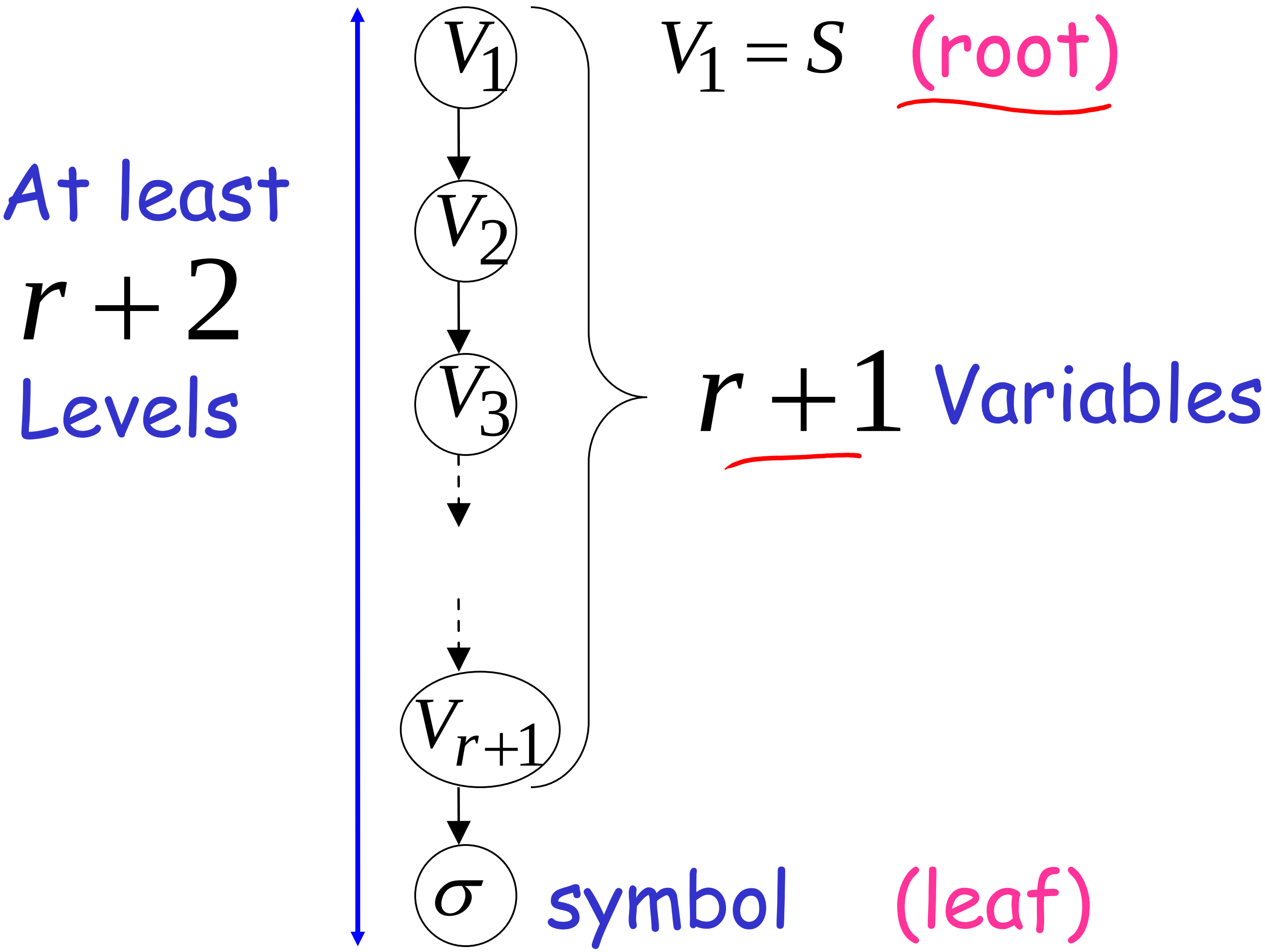
However we took, $|w| > t^r$

Contradiction!!!

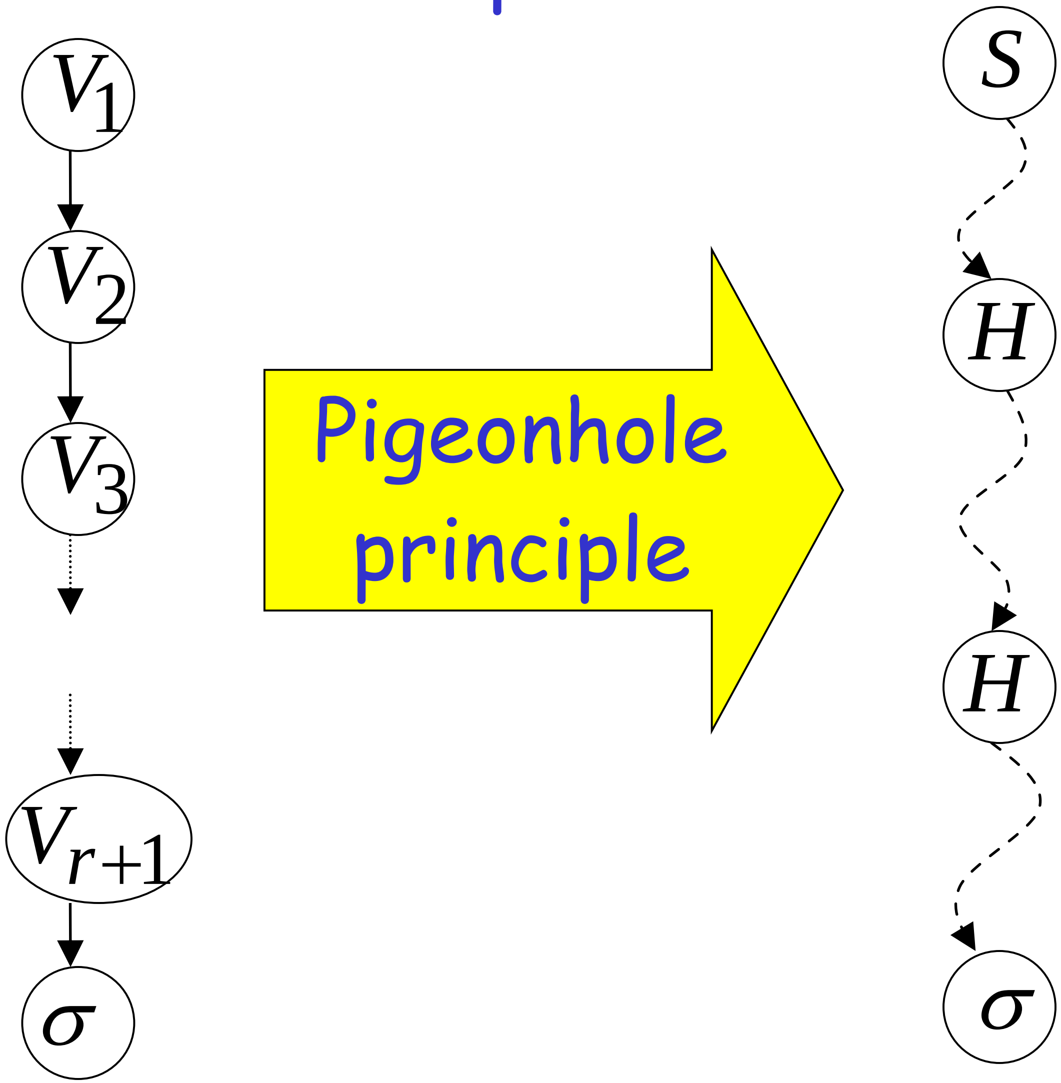
Therefore,

the tree must have at least $r+2$ levels

Thus, there is a path from the root to a leaf with at least $r+2$ nodes



Since there are at most r different variables some variable is repeated

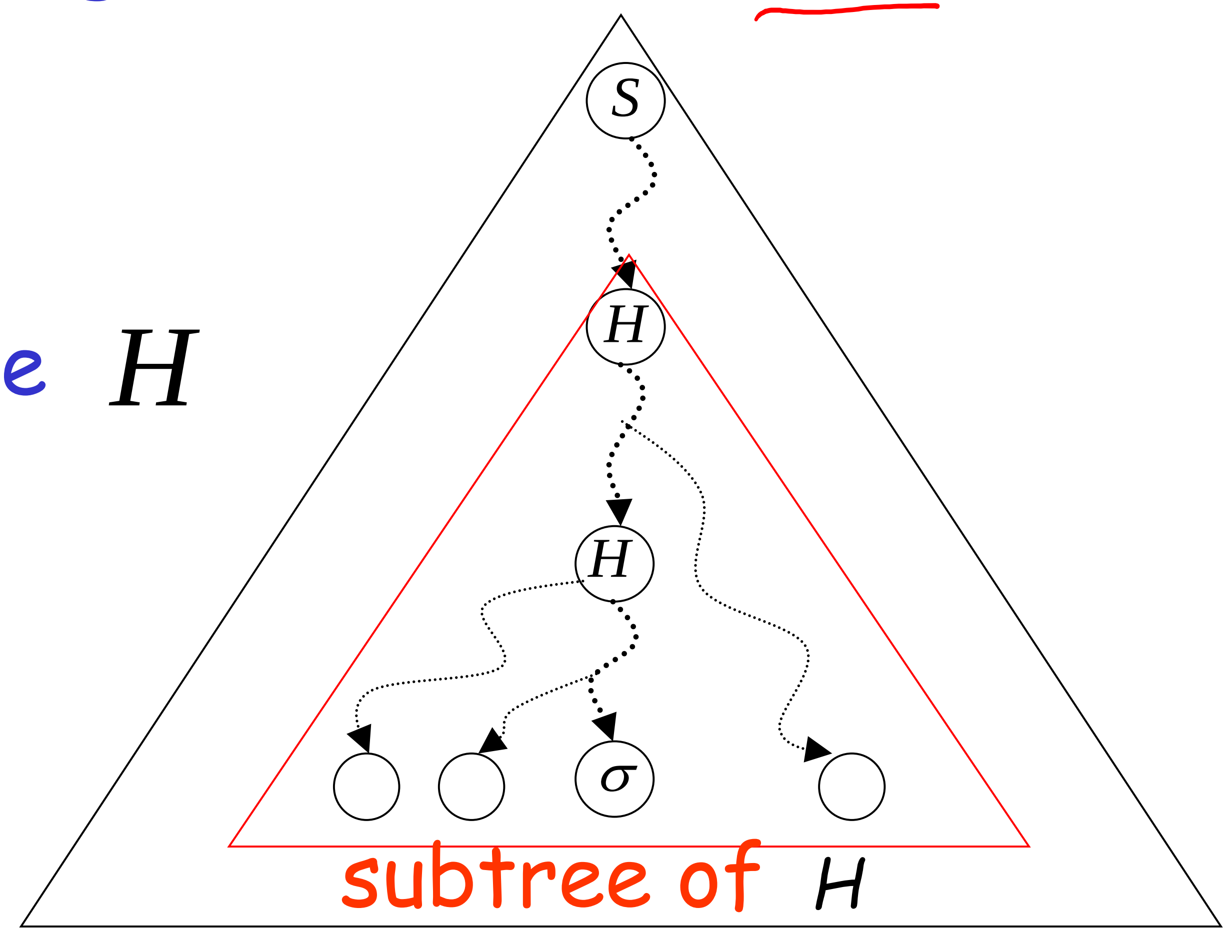


END OF CLAIM PROOF

Take now a string w with $|w| > t^r$

From claim:

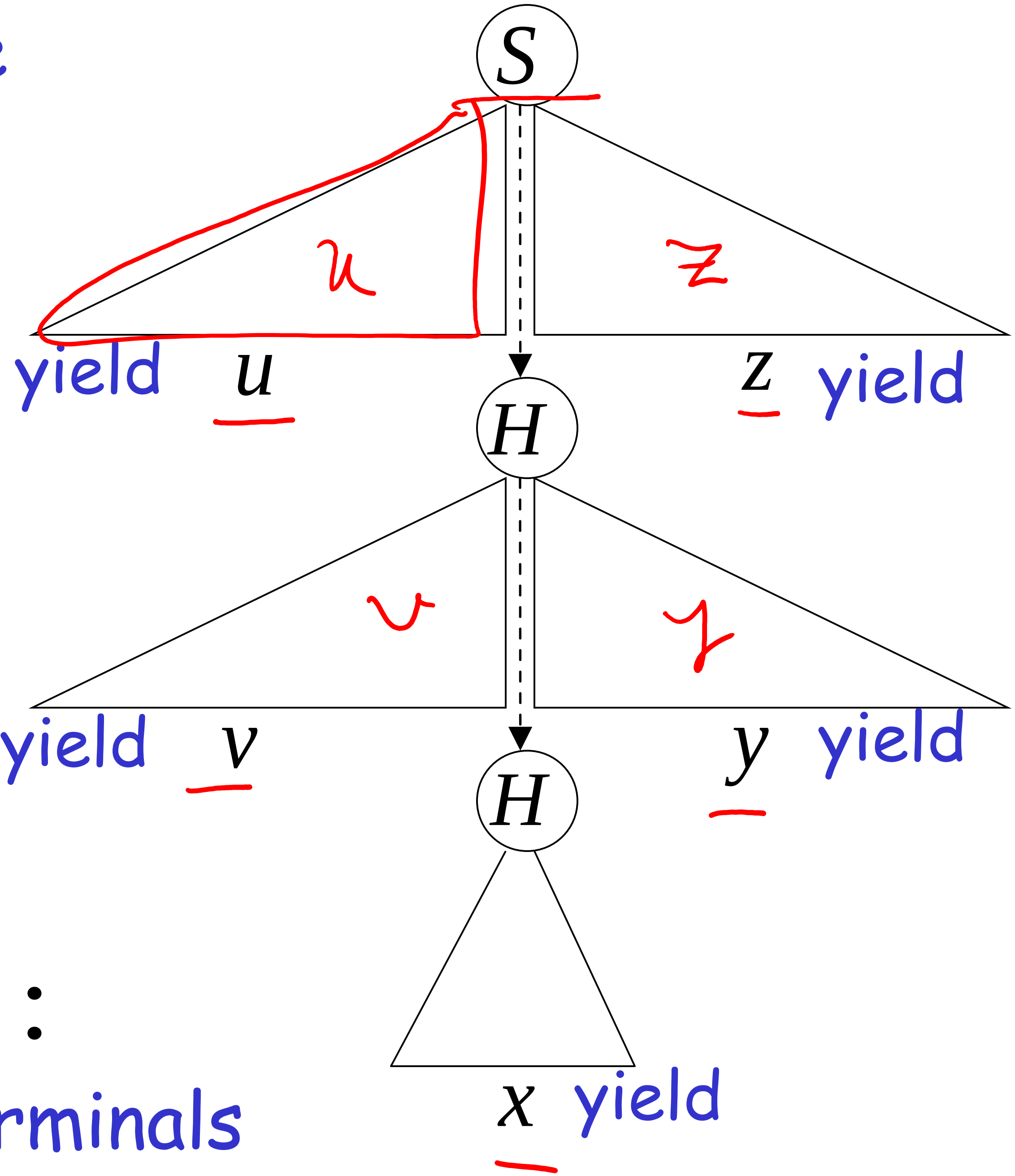
some variable H is repeated



Take H to be the deepest, so that only H is repeated in subtree

We can write

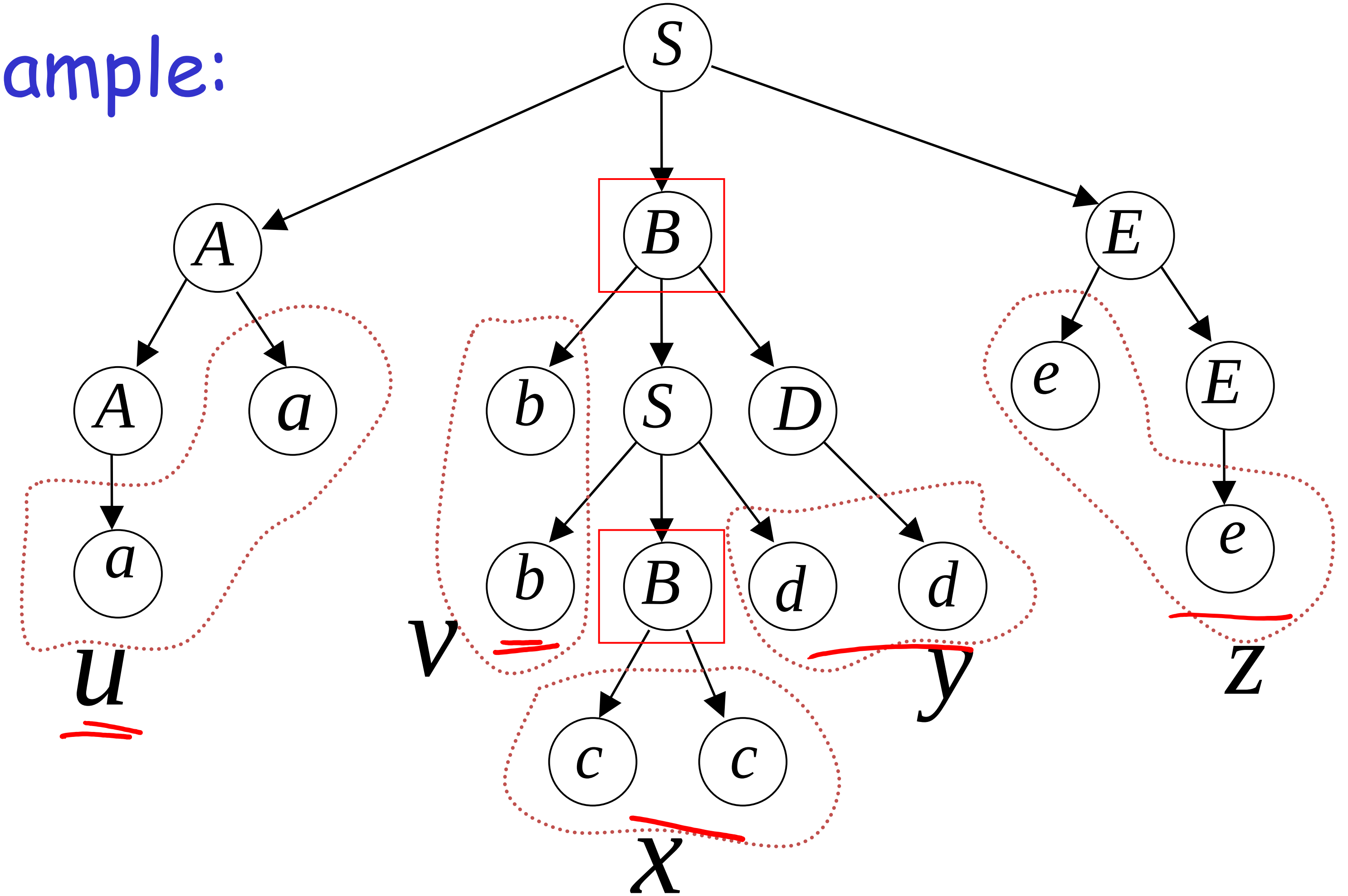
$$w = uvxyz$$



$$u, v, x, y, z:$$

Strings of terminals

Example:



$u = aa$

$v = bb$

$x = cc$

$y = dd$

$z = ee$

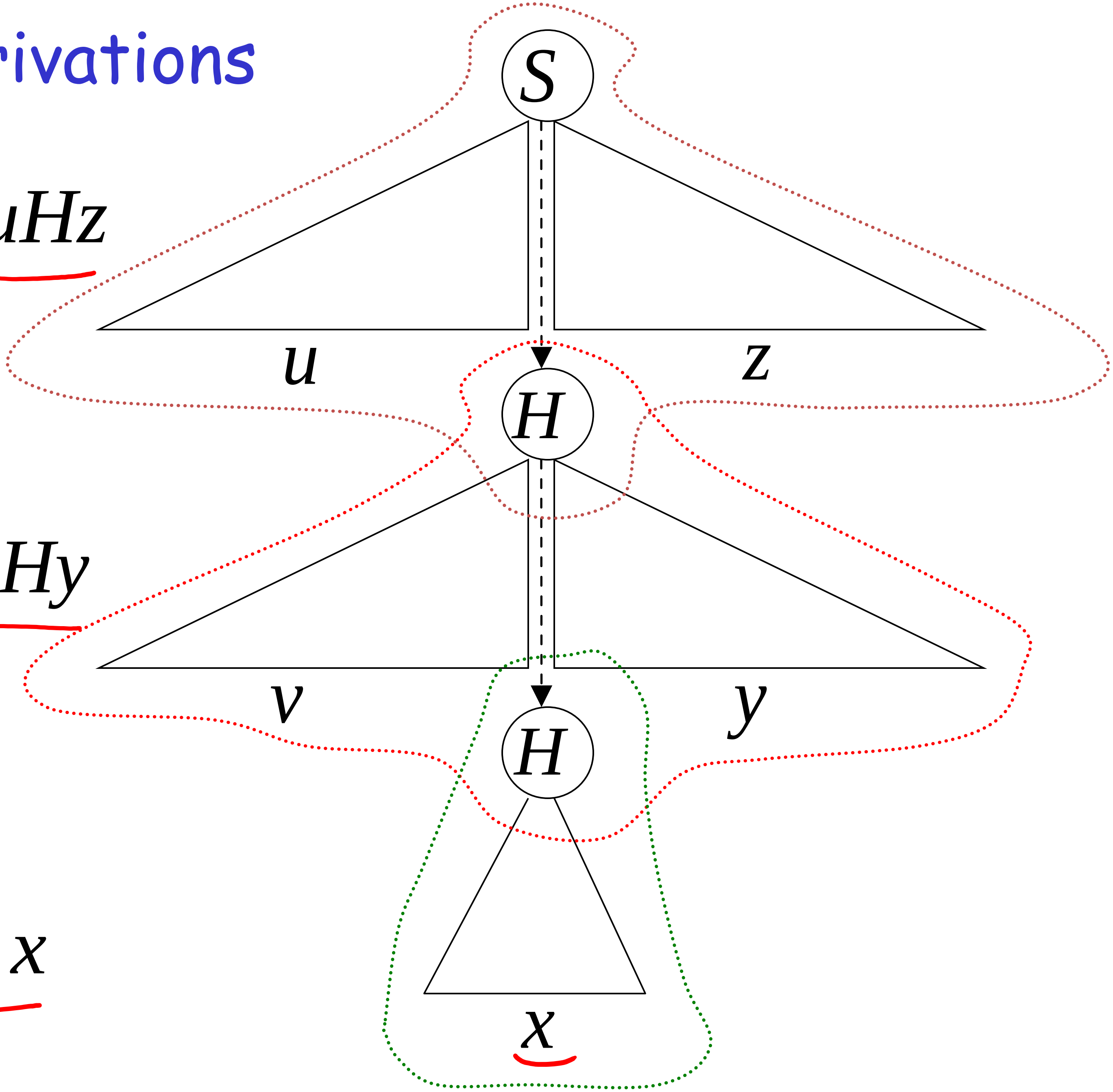
B corresponds to H

Possible derivations

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

$$H \overset{*}{\Rightarrow} vHy$$

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



Example:

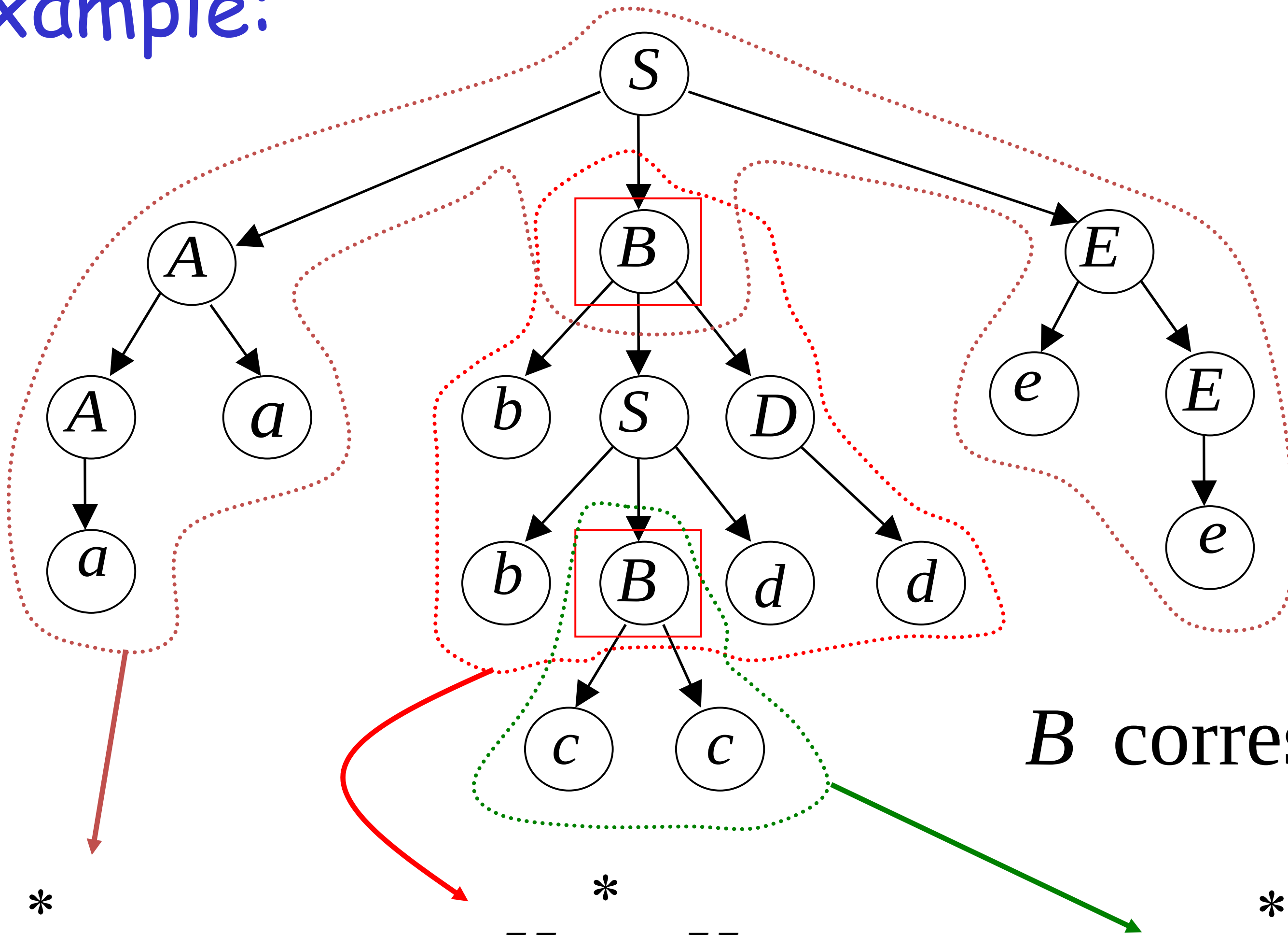
$u = aa$

$v = bb$

$x = cc$

$y = dd$

$z = ee$



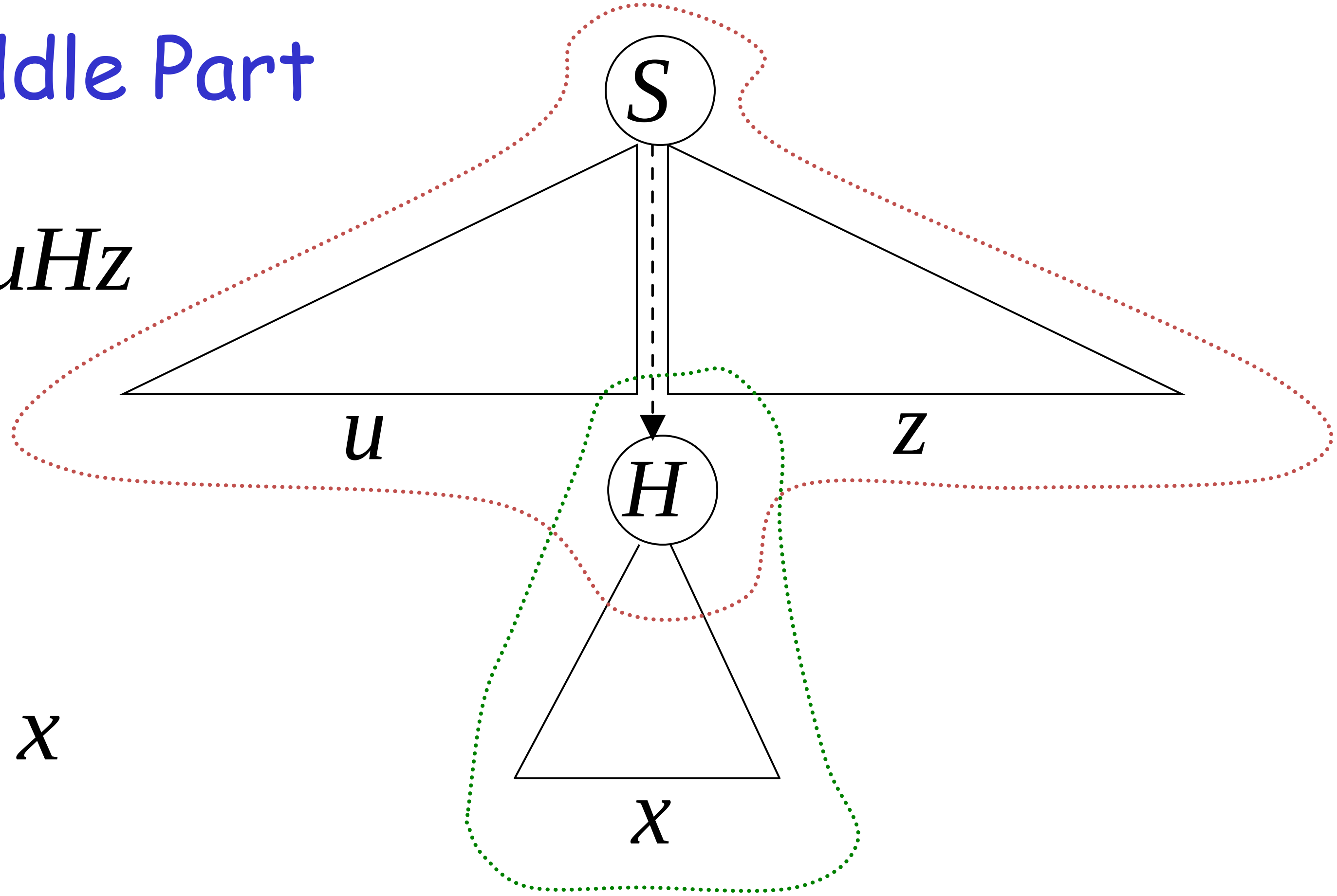
B corresponds to H

$$S \overset{*}{\Rightarrow} uHz \qquad H \overset{*}{\Rightarrow} vHy \qquad H \overset{*}{\Rightarrow} x$$

$$S \overset{*}{\Rightarrow} aaBee \qquad B \overset{*}{\Rightarrow} bbBdd \qquad B \Rightarrow cc$$

Remove Middle Part

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



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

Yield: $uxz = uv^0xy^0z$

$$S \stackrel{*}{\Rightarrow} uHz \stackrel{*}{\Rightarrow} uxz \quad = uv^0xy^0z \quad \in L(G)$$

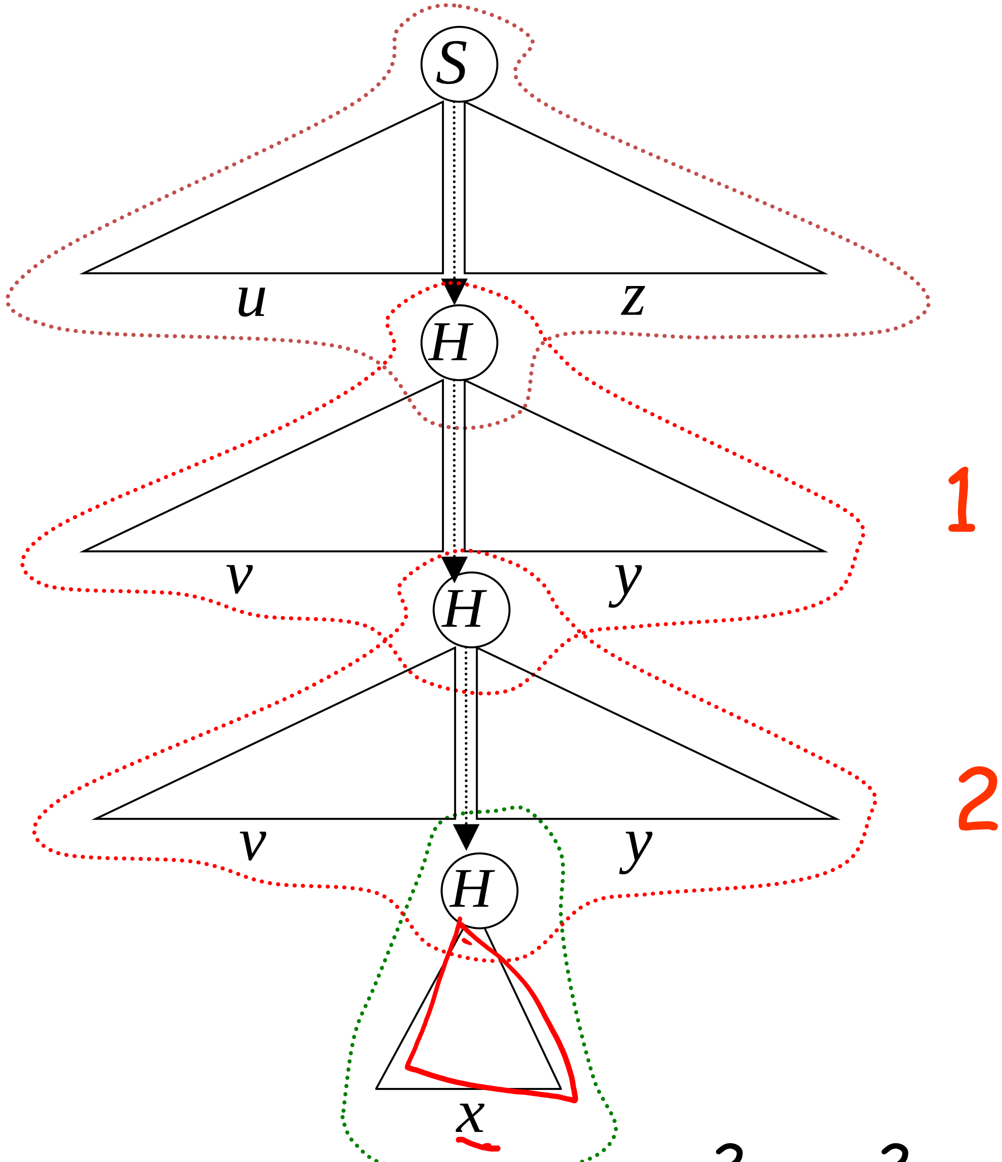
Repeat Middle part two times

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

$$H \overset{*}{\Rightarrow} vHy$$

$$H \overset{*}{\Rightarrow} vHy$$

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



Yield: $uvvxyyz = uv^2xy^2z$

$$S \overset{*}{\Rightarrow} uHz \qquad H \overset{*}{\Rightarrow} vHy \qquad H \overset{*}{\Rightarrow} x$$

$$S \overset{*}{\Rightarrow} uHz \overset{*}{\Rightarrow} uvHyz \overset{*}{\Rightarrow} uvvHyyz$$

$$\overset{*}{\Rightarrow} uvvxyyz \quad = uv^2xy^2z \quad \in L(G)$$

Repeat Middle part i times

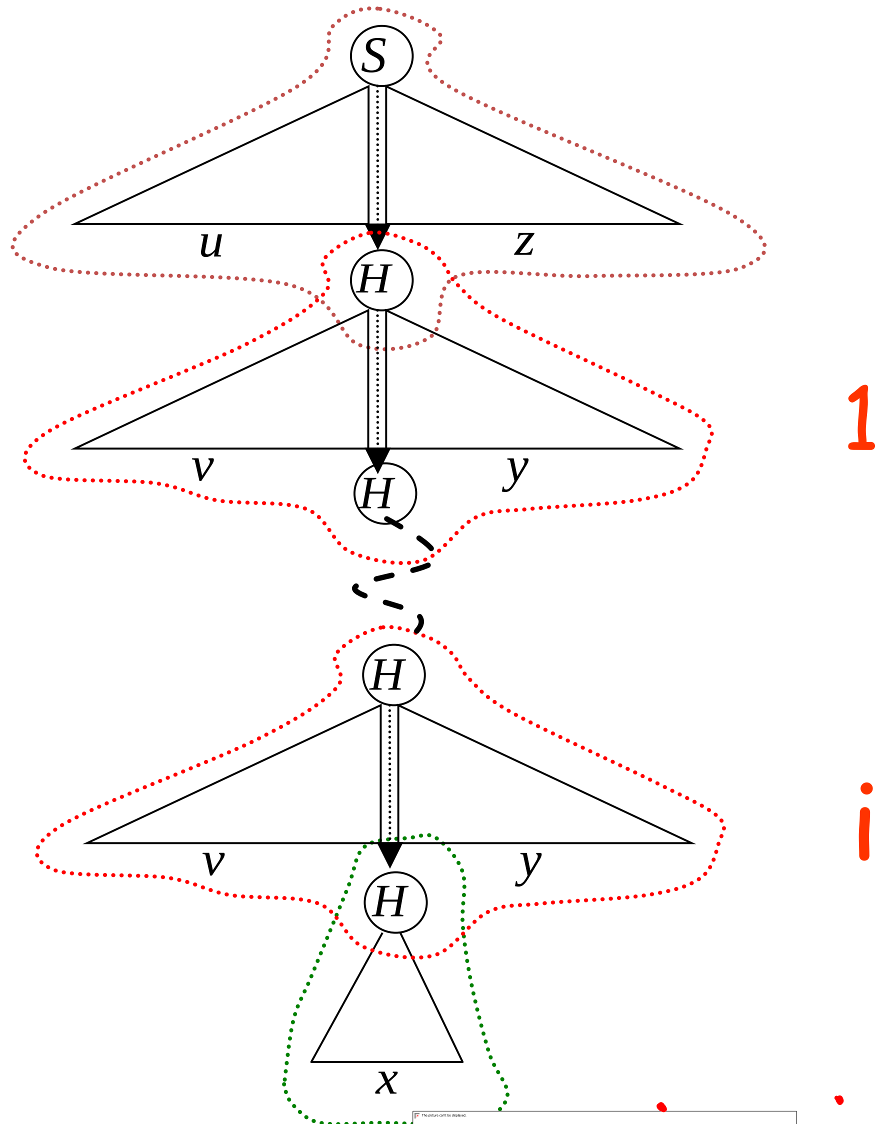
$$S \overset{*}{\Rightarrow} uHz$$

$$H \overset{*}{\Rightarrow} vHy$$

$$\vdots$$

$$H \overset{*}{\Rightarrow} vHy$$

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



Yield: $u v^i x y^i z$

$$S \overset{*}{\Rightarrow} uHz \qquad H \overset{*}{\Rightarrow} vHy \qquad H \overset{*}{\Rightarrow} x$$

$$S \overset{*}{\Rightarrow} uHz \overset{*}{\Rightarrow} uvHyz \overset{*}{\Rightarrow} uvvHyyz \overset{*}{\Rightarrow}$$

$$\overset{*}{\Rightarrow} \ldots$$

$$\overset{*}{\Rightarrow} uv^i H y^i z \overset{*}{\Rightarrow} \underline{uv^i x y^i z} \in \underline{L(G)}$$

Therefore,

$$|w| \geq t^r$$

If we know that: $w = uvxyz \quad \in L(G)$

then we know also know: $uv^ixy^iz \quad \in L(G)$

For all

since

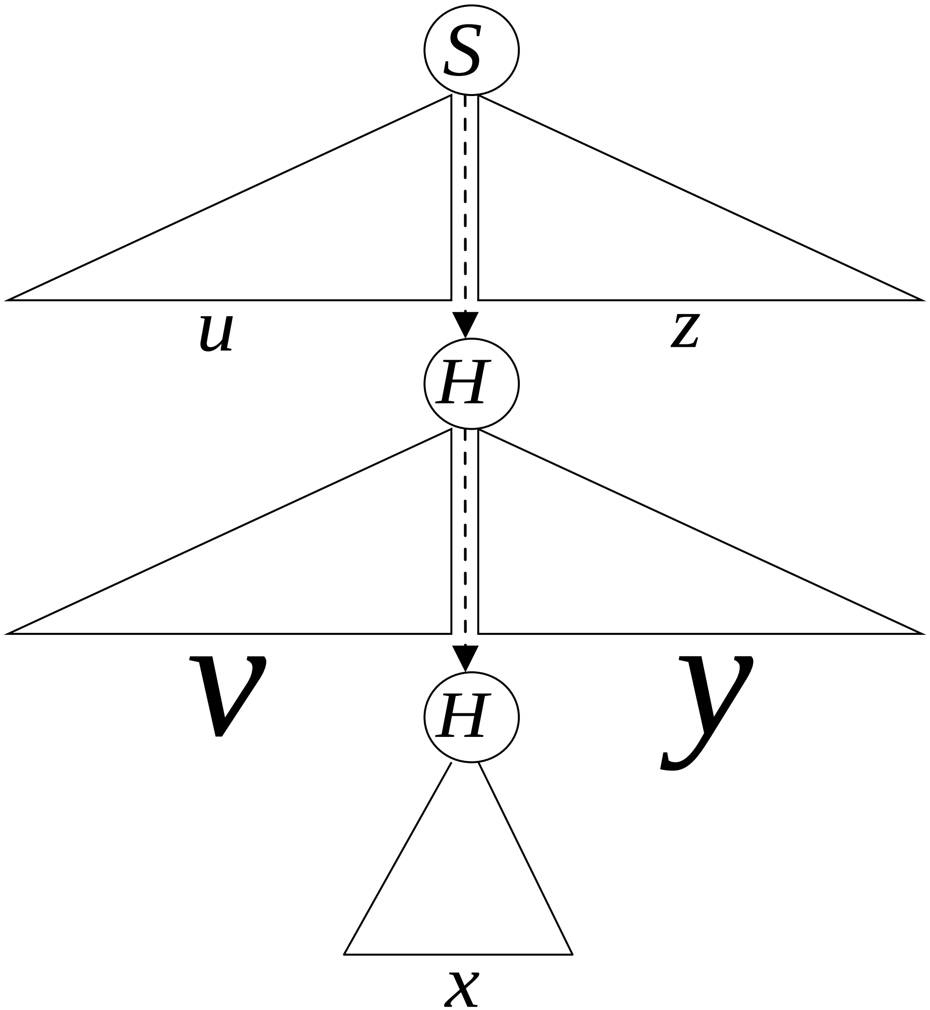
$L(G) = L - \{\lambda\}$

$$uv^ixy^iz \quad \in L$$

Observation 1:

$$|vy| \geq 1$$

Since G has no unit and λ-productions

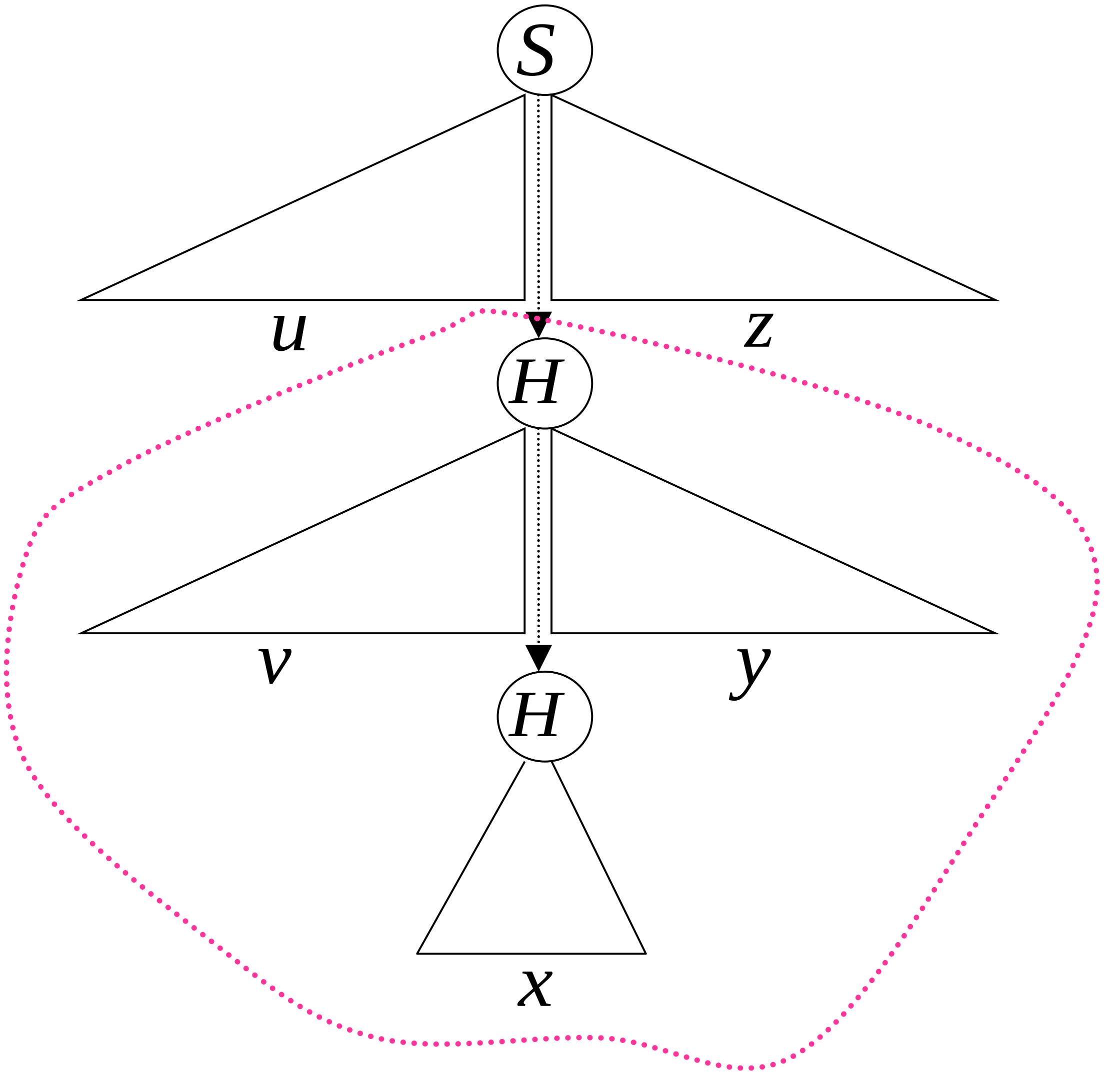


At least one of v or y is not λ

Observation 2:

$$|vxy| \leq t^{r+1}$$

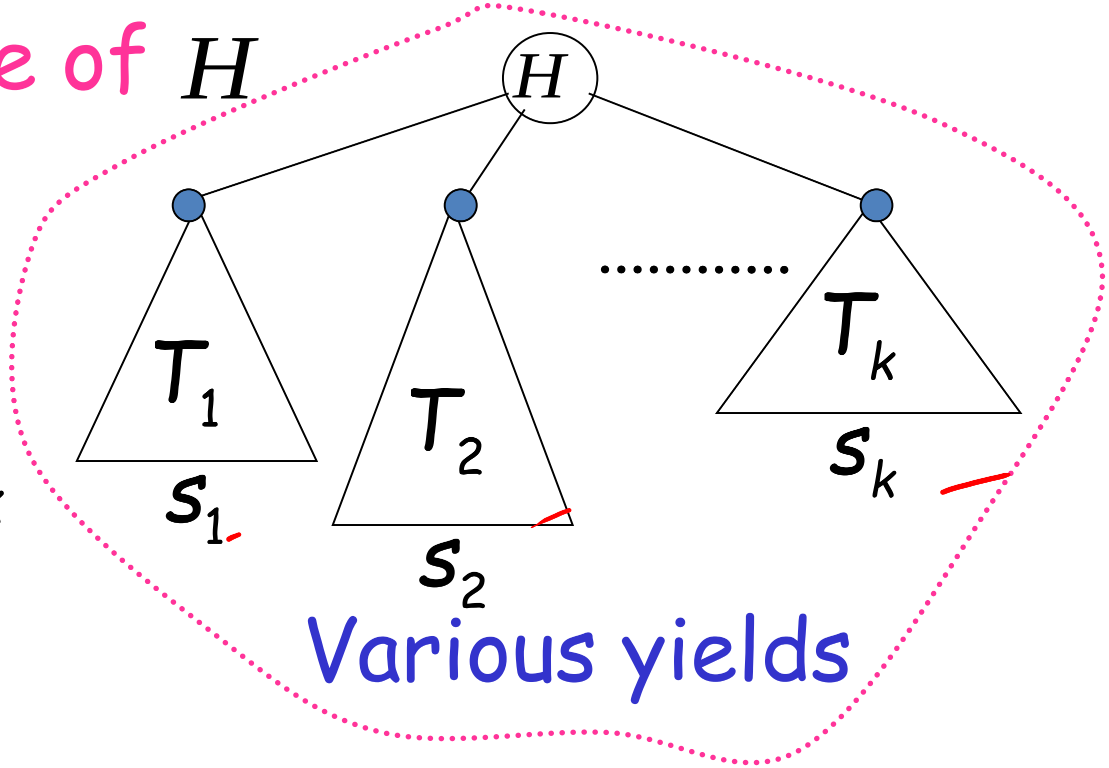
since in subtree only variable H is repeated



subtree of H

Explanation follows....

subtree of H



$$vxy = s_1 s_2 \cdots s_k$$

$|s_j| \leq t^r$ since no variable is repeated in T_j

$$|vxy| = \sum_{j=1}^{k} |s_j| \leq k \cdot t^r \leq t \cdot t^r = t^{r+1}$$

Maximum right-hand side of any production

Thus, if we choose critical length

$$m = t^{r+1} > t^r$$

then, we obtain the pumping lemma for context-free languages

The Pumping Lemma:

For any 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 that:

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

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

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

Let m be the critical length of the pumping lemma

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

$3m \geq m$

We pick: $w = \underline{a^m b^m c^m}$

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

$$w = \underline{a^m b^m c^m}$$

From pumping lemma:

we can write: $w = uvxyz$

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

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

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

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

Pumping Lemma says:

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

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

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

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

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

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

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

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

Case 1: vxy is in a^m

$$\overbrace{a \ldots a a \ldots a a \ldots a}^{m}\ \overbrace{bbb \ldots bbb}^{m}\ \overbrace{ccc \ldots ccc}^{m}$$

$$\underbrace{a \ldots a}_{u}\ \underbrace{a \ldots a}_{vxy}\ \underbrace{a \ldots a\ bbb \ldots bbb\ ccc \ldots ccc}_{z}$$

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

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

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

$$v = a^{k_1} \qquad y = a^{k_2} \qquad k_1 + k_2 \geq 1$$

$$\overbrace{\underbrace{a...a}_{u}\underbrace{a...a}_{v}\underbrace{a...a}_{x}\underbrace{a...a}_{y}a...a}^{m}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

(z spans from the last $a...a$ through $ccc...ccc$)

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

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

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

$$v = a^{k_1} \qquad y = a^{k_2} \qquad k_1 + k_2 \geq 1$$

$$\overbrace{a...aa...aa...aa...aa...a}^{m+k_1+k_2}\ \overbrace{bbb...bbb}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$\underbrace{a...a}_{u}\underbrace{a...a}_{v^2}\underbrace{a...a}_{x}\underbrace{a...a}_{y^2}\underbrace{a...a\ bbb...bbb\ ccc...ccc}_{z}$$

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

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

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

From Pumping Lemma: $\underline{uv^2xy^2}z \in L$

$\underline{i=2}$

$$k_1 + k_2 \geq 1$$

However: $u v^2 \underline{x y^2 z} = a^{m+k_1+k_2} \underline{b^m c^m} \notin L$

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 in b^m

Similar to case 1

$$\overbrace{aaa...aaa}^{m}\ \overbrace{b...bb...bb...b}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$\underbrace{aaa...aaa\ b...b}_{u}\underbrace{b...b}_{vxy}\underbrace{b...b\ ccc...ccc}_{z}$$

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

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

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

Case 3: vxy is in c^m

Similar to case 1

$$\underbrace{\overbrace{aaa\ldots aaa}^{m}\ \overbrace{bbb\ldots bbb}^{m}\ c\ldots c}_{u}\,\underbrace{c\ldots c}_{vxy}\,\underbrace{c\ldots c}_{z}$$

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

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

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

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

$$m \qquad\qquad m \qquad\qquad m$$

$$a...aa...aa\ b...bb...bbb\ ccc...ccc$$

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

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

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

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

Sub-case 1: v contains only a

y contains only b

$$\overbrace{a...aa...aa...a}^{m}\ \overbrace{b...bb...bb...b}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$\underbrace{a...a}_{u}\underbrace{a...a}_{v}\underbrace{a...a\ b...b}_{x}\underbrace{b...b}_{y}\underbrace{b...b\ ccc...ccc}_{z}$$

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

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

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

$$v = a^{k_1} \qquad y = b^{k_2} \qquad k_1 + k_2 \geq 1$$

$$\underbrace{\overbrace{a...a}^{u}\overbrace{a...a}^{v}a...a}_{m}\;\underbrace{b...b\,b...b\,b...b}_{m}\;\underbrace{ccc...ccc}_{m}$$

$$\underbrace{a...a}_{u}\,\underbrace{a...a}_{v}\,\underbrace{a...a\;b...b}_{x}\,\underbrace{b...b}_{y}\,\underbrace{b...b\;ccc...ccc}_{z}$$

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

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

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

$$v = a^{k_1} \qquad y = b^{k_2} \qquad k_1 + k_2 \geq 1$$

$$\overbrace{a...aa...aa...a}^{m+k_1}\;\overbrace{b...bb...bb...b}^{m+k_2}\;\overbrace{ccc...ccc}^{m}$$

$$\underbrace{a...a}_{u}\,\underbrace{a...a}_{v^2}\,\underbrace{a...a\;b...b}_{x}\,\underbrace{b...b}_{y^2}\,\underbrace{b...b\;ccc...ccc}_{z}$$

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

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

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

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

$$k_1 + k_2 \geq 1$$

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

Contradiction!!!

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

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

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

Sub-case 2: v contains a and b

y contains only b

$$\overbrace{a...aa...a}^{m}\ \overbrace{b...bb...bb...bb...b}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$\underbrace{a...a}_{u}\underbrace{a...a\ b...b}_{v}\underbrace{b...b}_{\underline{x}}\underbrace{b...b}_{y}\underbrace{b...b\ ccc...ccc}_{z}$$

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

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

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

By assumption

$$v = a^{k_1} b^{k_2} \qquad y = b^{k_3} \qquad k_1, k_2 \ge 1$$

$$\overbrace{a\ldots a\,\overbrace{a\ldots a}^{k_1}}^{m}\;\overbrace{\overbrace{b\ldots b}^{k_2}\,b\ldots b\,\overbrace{b\ldots b}^{k_3}\,b\ldots b}^{m}\;\overbrace{ccc\ldots ccc}^{m}$$

$$\underbrace{a\ldots a}_{u}\,\underbrace{a\ldots a\;b\ldots b}_{v}\,\underbrace{b\ldots b}_{x}\,\underbrace{b\ldots b}_{y}\,\underbrace{b\ldots b\;ccc\ldots ccc}_{z}$$

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

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

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

$$v = a^{k_1} b^{k_2} \qquad y = b^{k_3} \qquad k_1, k_2 \geq 1$$

$$\overbrace{a\ldots a\,\overbrace{a\ldots a}^{k_1}}^{m}\,\overbrace{b\ldots b}^{k_2}\,\overbrace{a\ldots a}^{k_1}\,\overbrace{\overbrace{b\ldots b}^{k_2}\,b\ldots b\,\overbrace{b\ldots b}^{2k_3}}^{m+k_3}\,b\ldots b\,\overbrace{ccc\ldots ccc}^{m}$$

$$\underbrace{a\ldots a}_{u}\,\underbrace{a\ldots ab\ldots ba\ldots ab\ldots b}_{v^2}\,\underbrace{b\ldots b}_{x}\,\underbrace{b\ldots b}_{y^2}\,\underbrace{b\ldots b\,ccc\ldots ccc}_{z}$$

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

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

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

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

$k_1, k_2 \geq 1$

However: $uv^2xy^2z = a^m b^{k_2} a^{k_1} b^{m+k_3} c^m \notin L$

Contradiction!!!

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

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

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

Sub-case 3: v contains only a

y contains a and b

Similar to sub-case 2

$$\overbrace{a...aa...aa...aa...a}^{m}\ \overbrace{b...bb...b}^{m}\ \overbrace{ccc...ccc}^{m}$$

$$\underbrace{a...a}_{u}\underbrace{a...a}_{v}\underbrace{a...a}_{x}\underbrace{a...a\ b...b}_{y}\underbrace{b...b\ ccc...ccc}_{z}$$

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

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

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

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

Similar to case 4

$$\underbrace{\overbrace{aaa...aaa}^{m}\,\overbrace{bbb..}^{m}}_{u}\underbrace{.bbb\,\overbrace{ccc...}^{m}}_{vxy}\underbrace{.ccc}_{z}$$

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

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

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

Case 6: vxy overlaps a^m, b^m and c^m

Impossible!

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

$$\underbrace{aaa...}_{u}\underbrace{aaa\ bbb...bbb\ ccc...}_{vxy}\underbrace{ccc}_{z}$$

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

Non-context free languages

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

$$\{ww : w \in \{a,b\}\}$$

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

Prove it !!!

Context-free languages

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

$$\{ww^R : w \in \{a,b\}^*\}$$

More Applications of The Pumping Lemma

$ww = \underline{a^n b^n a^n b^n}$

$w = a^n b^n$

he 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:

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

Non-context free languages

$$\{a^n b^n c^n : n \geq 0\} \qquad \{vv : v \in \{a,b\}\}$$

Context-free languages

$$\{a^n b^n : n \geq 0\} \qquad \{ww^R : w \in \{a,b\}^*\}$$

Theorem: The language

$$L = \{vv : v \in \{a,b\}^*\}$$

is **not** context free

Proof: Use the Pumping Lemma for context-free languages

$$L = \{vv : v \in \{a,b\}^*\}$$

Assume for contradiction that L is context-free

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

$$L = \{vv : v \in \{a, b\}^*\}$$

Pumping Lemma gives a magic number m such that:

Pick any string of L with length at least m

we pick: $a^m b^m a^m b^m \in L$

$$L = \{vv : v \in \{a,b\}^*\}$$

We can write: $a^m b^m a^m b^m = uvxyz$

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

Pumping Lemma says:

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

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

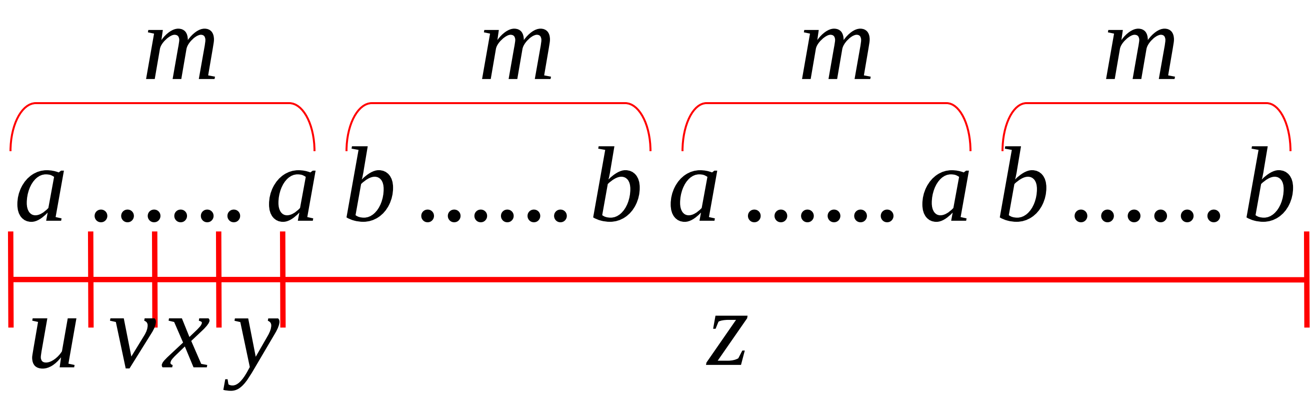
We examine **all** the possible locations of string vxy in $a^m b^m a^m b^m$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$$

Case 1: vxy is within the first a^m

$$v = a^{k_1} \qquad y = a^{k_2} \qquad k_1 + k_2 \ge 1$$



$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$$

$\boxed{\textbf{Case 1:} \quad vxy \text{ is within the first } a^m}$

$$v = a^{k_1} \qquad y = a^{k_2} \qquad k_1 + k_2 \ge 1$$

$$\overbrace{a\,\ldots\ldots\ldots\ldots\ldots\ldots\,a}^{m+k_1+k_2}\ \overbrace{b\,\ldots\ldots\,b}^{m}\ \overbrace{a\,\ldots\ldots\,a}^{m}\ \overbrace{b\,\ldots\ldots\,b}^{m}$$

$$u \quad v^2 \quad x \quad y^2 \quad z$$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$$

Case 1: vxy is within the first a^m

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

$$k_1 + k_2 \ge 1$$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 1: vxy is within the first a^m

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

However, from Pumping Lemma: $uv^2xy^2z \in L$

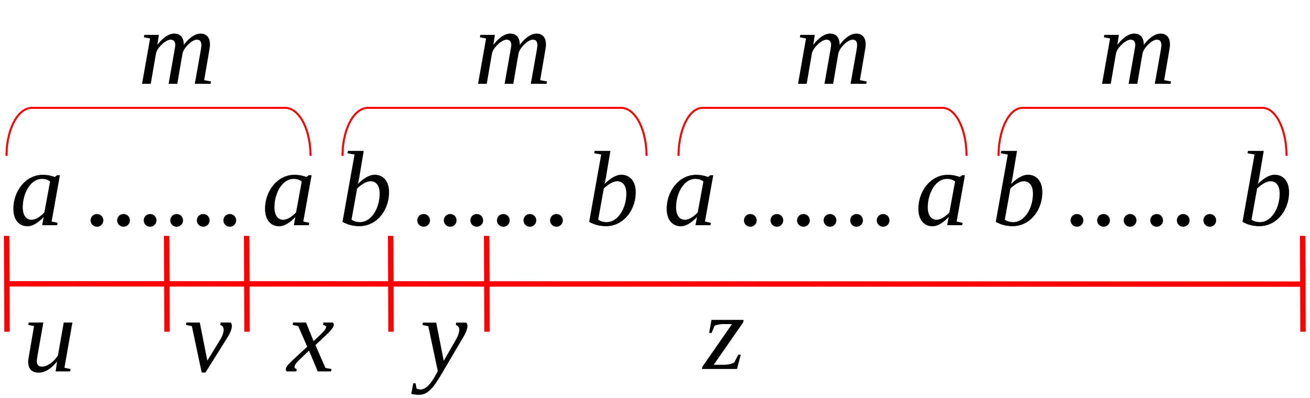
Contradiction!!!

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 2: v is in the first a^m
y is in the first b^m

$$v = a^{k_1} \qquad y = b^{k_2} \qquad k_1 + k_2 \geq 1$$



$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$$

Case 2: v is in the first a^m

y is in the first b^m

$$v = a^{k_1} \qquad y = b^{k_2} \qquad k_1 + k_2 \ge 1$$

$$\overbrace{a \ldots\ldots\ldots\ldots a}^{m+k_1}\ \overbrace{b \ldots\ldots\ldots\ldots b}^{m+k_2}\ \overbrace{a \ldots\ldots a}^{m}\ \overbrace{b \ldots\ldots b}^{m}$$

$$u \qquad v^2 \qquad x \qquad y^2 \qquad z$$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 2: v is in the first a^m
y is in the first b^m

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

$$k_1 + k_2 \geq 1$$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 2: v is in the first a^m

y is in the first b^m

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

However, from Pumping Lemma: $uv^2xy^2z \in L$

Contradiction!!!

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$$

Case 3: v overlaps the first $a^m b^m$
y is in the first b^m

$$v = a^{k_1} b^{k_2} \qquad y = b^{k_3} \qquad k_1, k_2 \ge 1$$

$$\overbrace{a\ldots\ldots a}^{m}\;\overbrace{b\ldots\ldots b}^{m}\;\overbrace{a\ldots\ldots a}^{m}\;\overbrace{b\ldots\ldots b}^{m}$$

$$u \qquad v \qquad x \quad y \qquad z$$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 3: v overlaps the first $a^m b^m$
y is in the first b^m

$$v = a^{k_1} b^{k_2} \qquad y = b^{k_3} \qquad k_1, k_2 \geq 1$$

$$\overbrace{a \ldots\ldots a}^{m}\ \overbrace{b \ldots b}^{k_2}\ \overbrace{a \ldots a}^{k_1}\ \overbrace{b \ldots\ldots\ldots b}^{m+k_3}\ \overbrace{a \ldots\ldots a}^{m}\ \overbrace{b \ldots\ldots b}^{m}$$

$$u \qquad v^2 \qquad x \qquad y^2 \qquad z$$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 3: v overlaps the first $a^m b^m$

y is in the first b^m

$$a^m b^{k_2} a^{k_1} b^{m+k_3} a^m b^m = uv^2xy^2z \quad \notin L$$

$$k_1, k_2 \geq 1$$

$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \le m \qquad |vy| \ge 1$$

Case 3: v overlaps the first $a^m b^m$

y is in the first b^m

$$a^m b^{k_2} a^{k_1} b^{k_3} a^m b^m = uv^2xy^2z \quad \notin L$$

However, from Pumping Lemma: $uv^2xy^2z \in L$

Contradiction!!!

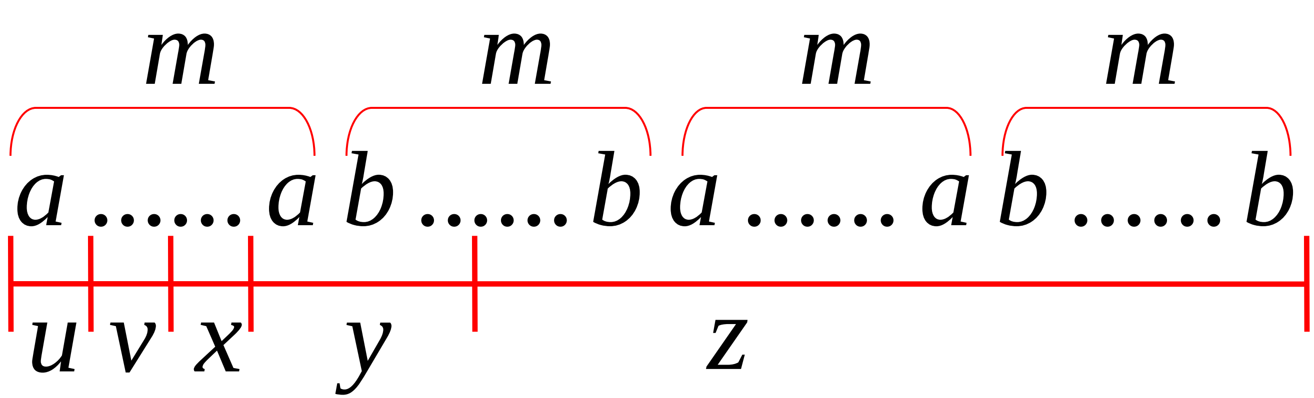
$$L = \{vv : v \in \{a,b\}^*\}$$

$$a^m b^m a^m b^m = uvxyz \qquad |vxy| \leq m \qquad |vy| \geq 1$$

Case 4: v in the first a^m

y Overlaps the first $a^m b^m$

Analysis is similar to case 3



Other cases: vxy is within $a^m \boxed{b^m} a^m b^m$

or

$a^m b^m \boxed{a^m} b^m$

or

$a^m b^m a^m \boxed{b^m}$

Analysis is similar to case 1:

$\boxed{a^m} b^m a^m b^m$

More cases: vxy overlaps $a^m \boxed{b^m a^m} b^m$

or

$a^m b^m \boxed{a^m b^m}$

Analysis is similar to cases 2,3,4:

$$\boxed{a^m b^m} a^m b^m$$

There are no other cases to consider

Since $|vxy| \leq m$, it is impossible vxy to overlap:

$$\boxed{a^m b^m a^m} b^m$$

nor

$$a^m \boxed{b^m a^m b^m}$$

nor

$$\boxed{a^m b^m a^m b^m}$$

In all cases we obtained a contradiction

Therefore: The original assumption that

$$L = \{vv : v \in \{a,b\}^*\}$$

is context-free must be wrong

Conclusion: L is not context-free