Pumping Lemma
for
Context-free Languages

Take an **infinite** context-free language

Generates an infinite number of different strings

Example:

$$S \to ABE \mid bBd$$

$$A \to Aa \mid a$$

$$B \to bSD \mid cc$$

$$D \to Dd \mid d$$

$$E \to eE \mid e$$

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

A possible derivation:

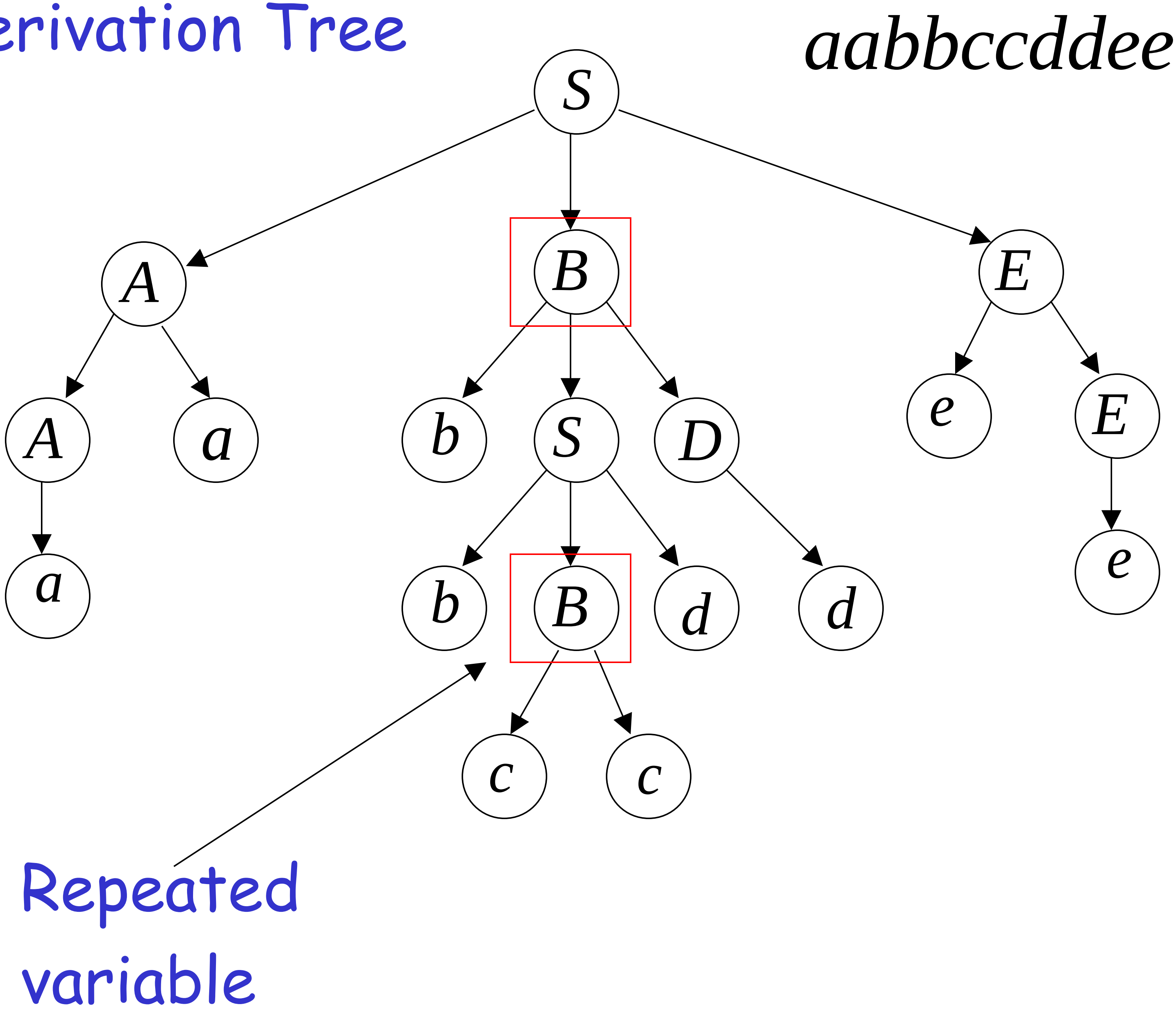
$$S \Rightarrow A\boxed{B}E \Rightarrow AaBE \Rightarrow aaBE$$

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

$$\Rightarrow aaabbccdDE \Rightarrow aabbccddE$$

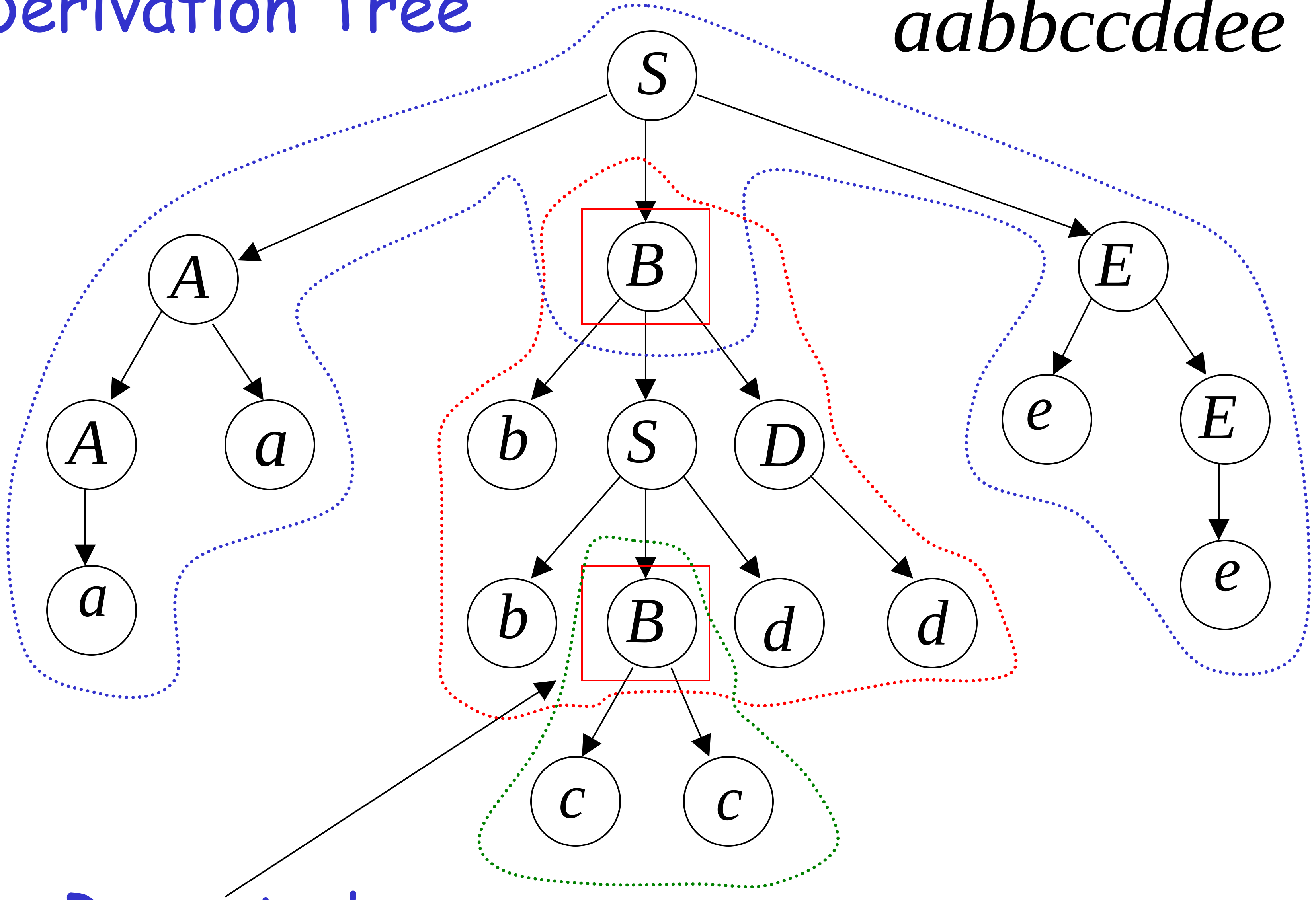
$$\Rightarrow aabbccddeE \Rightarrow aabbccddee$$

Derivation Tree



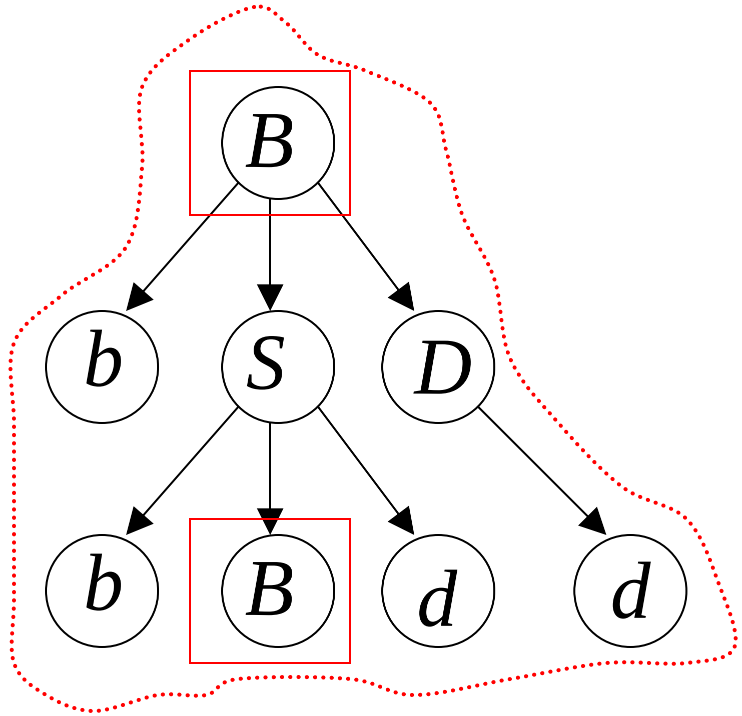
Derivation Tree

$aabbccddee$



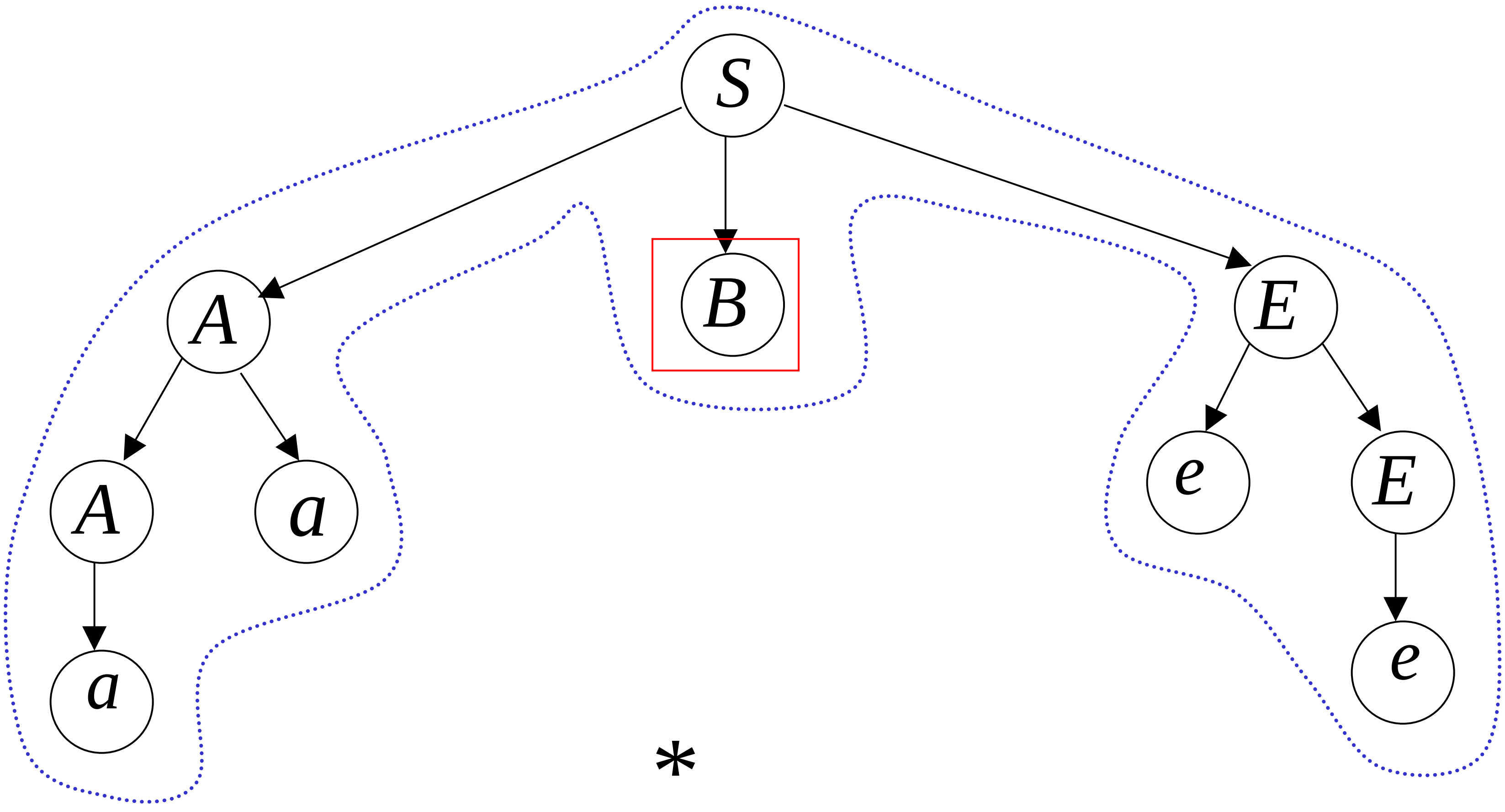
Repeated variable

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

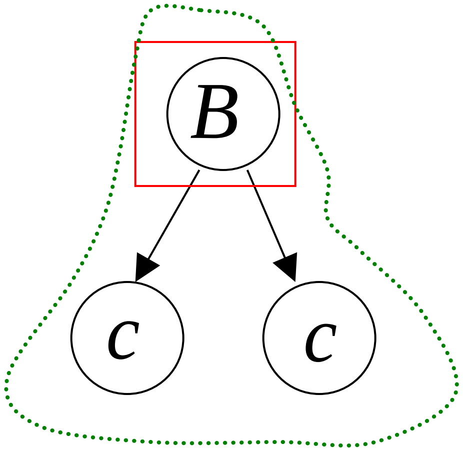


$$B \overset{*}{\Rightarrow} bbBdd$$

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

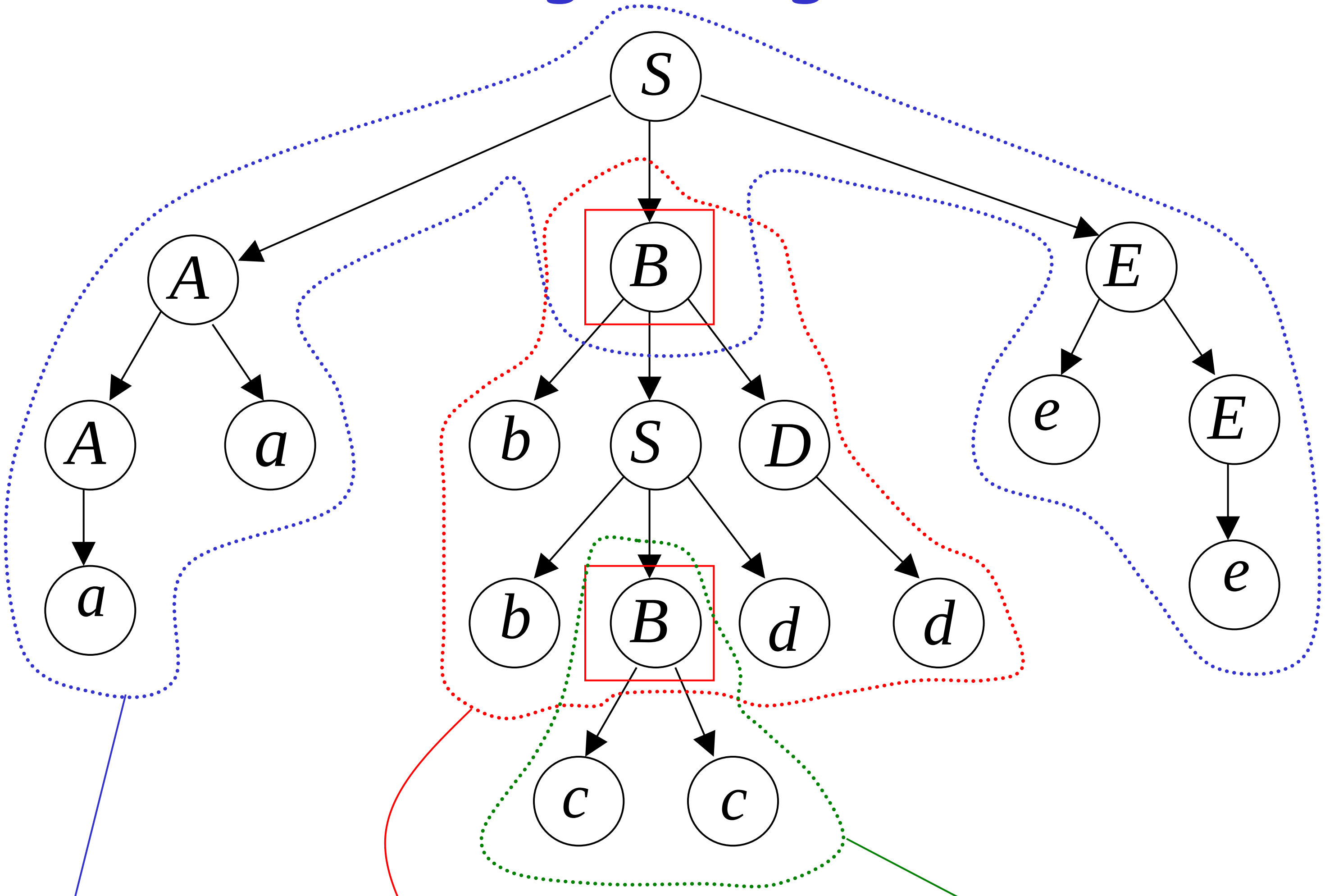


$$S \overset{*}{\Rightarrow} aaBee$$



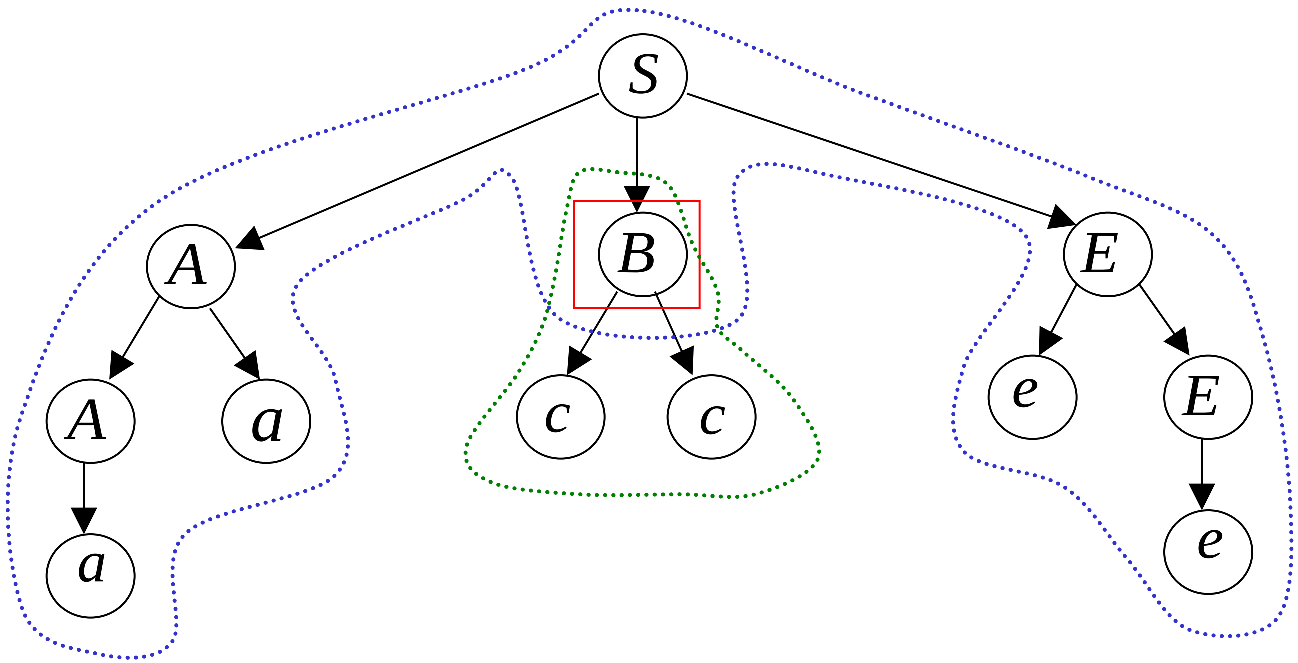
$$B \Rightarrow cc$$

Putting all together



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

We can remove the middle part



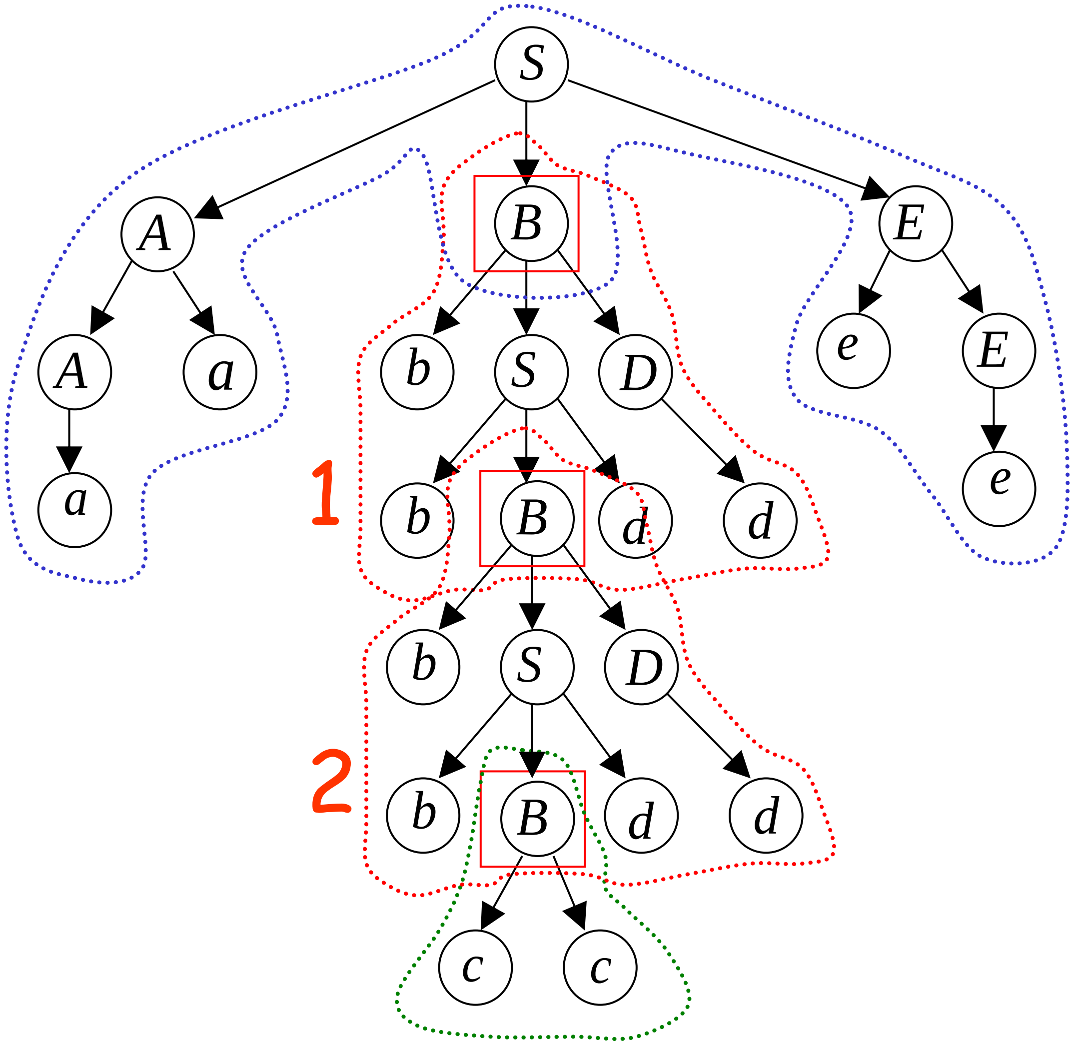
$$S \stackrel{*}{\Rightarrow} aa(bb)^0 cc(dd)^0 ee$$

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

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

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

We can repeated middle part two times



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

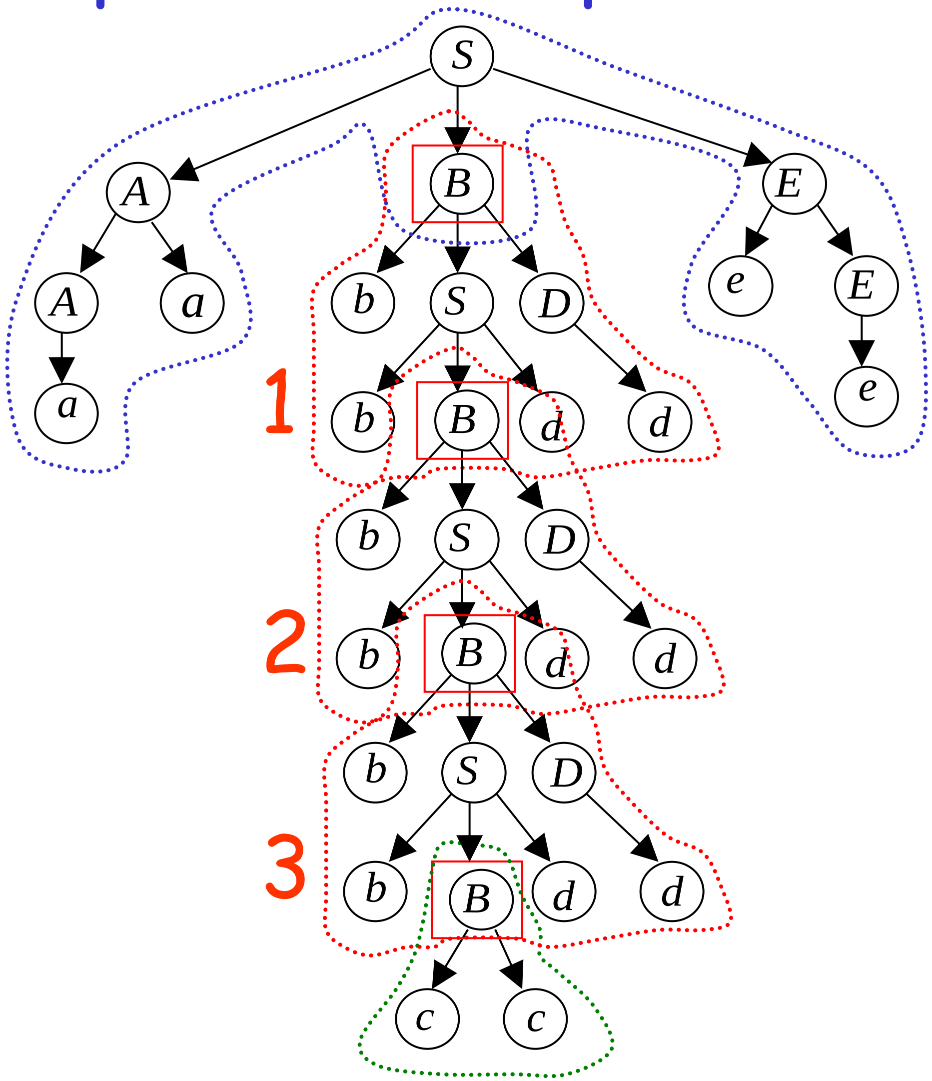
$$S \overset{*}{\Rightarrow} aaBee \qquad B \overset{*}{\Rightarrow} bbBdd \qquad B \Rightarrow cc$$

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

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

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

We can repeat middle part three times

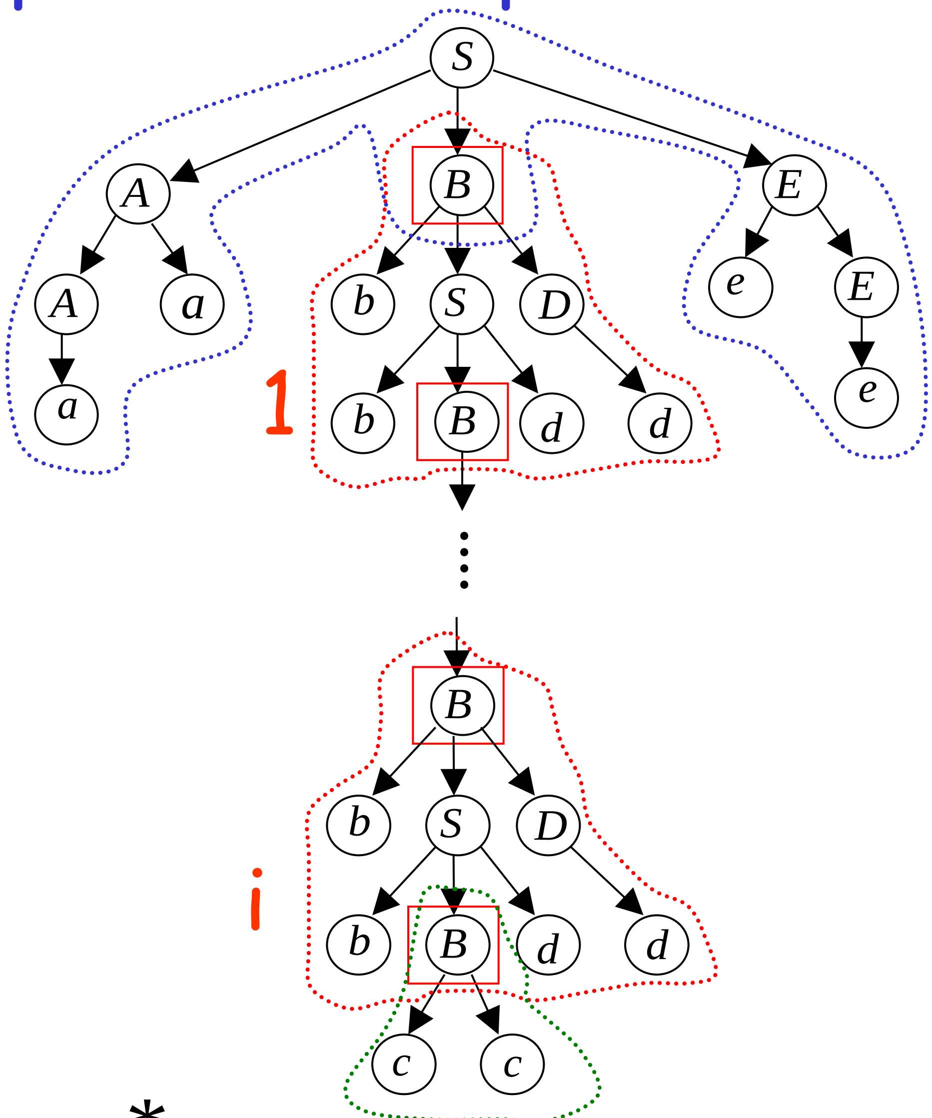


$$S \overset{*}{\Rightarrow} aa(bb)^3 cc(dd)^3 ee$$

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

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

Repeat middle part i times



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

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

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

For any $i \geq 0$

From Grammar

$$S \rightarrow ABE \mid bBd$$

$$A \rightarrow Aa \mid a$$

$$B \rightarrow bSD \mid cc$$

$$D \rightarrow Dd \mid d$$

$$E \rightarrow eE \mid e$$

and given string

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

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

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

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

$$r = 5$$

$$t = 3$$

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

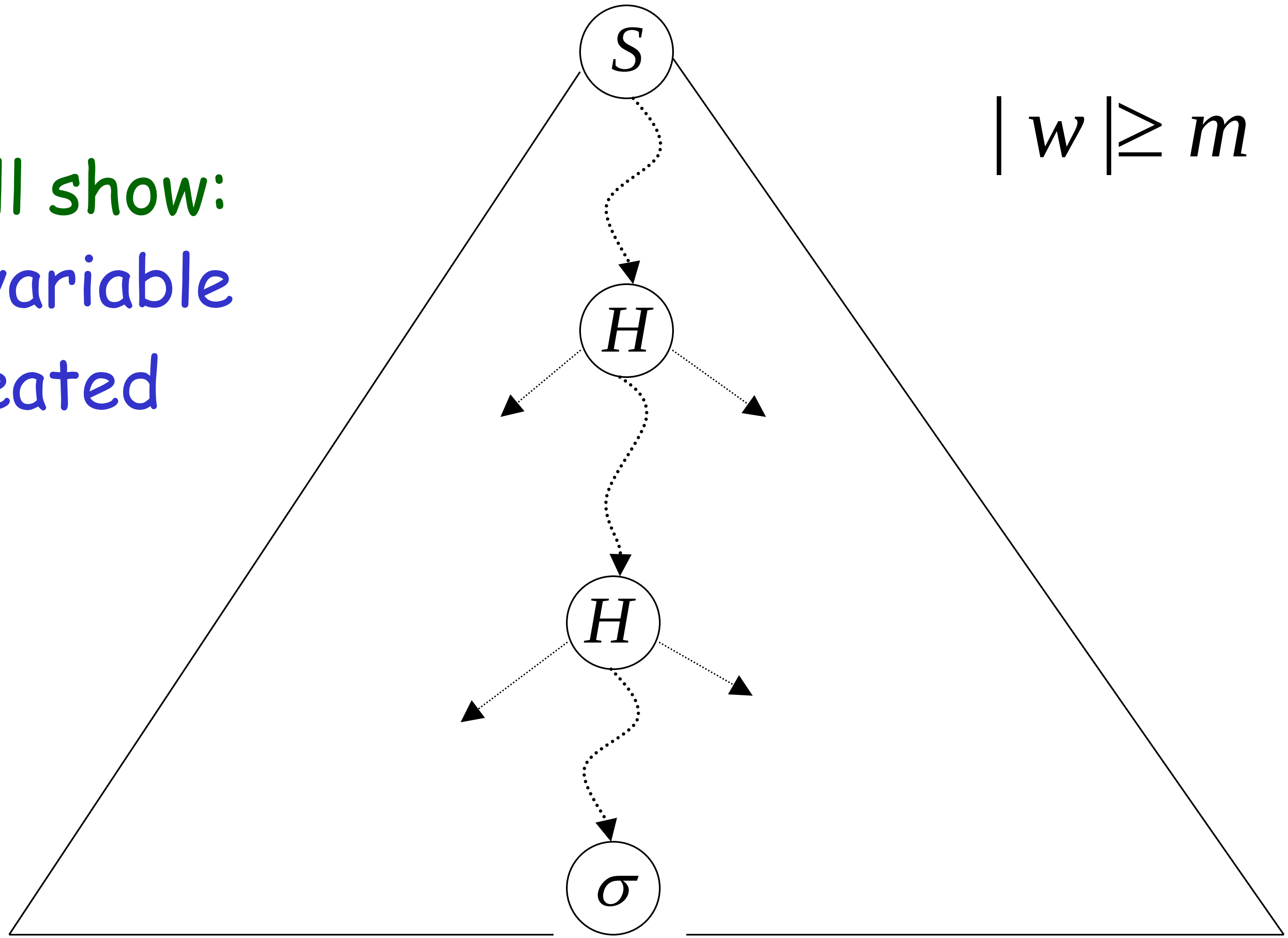
Proof:

Proof by contradiction

Derivation tree of w

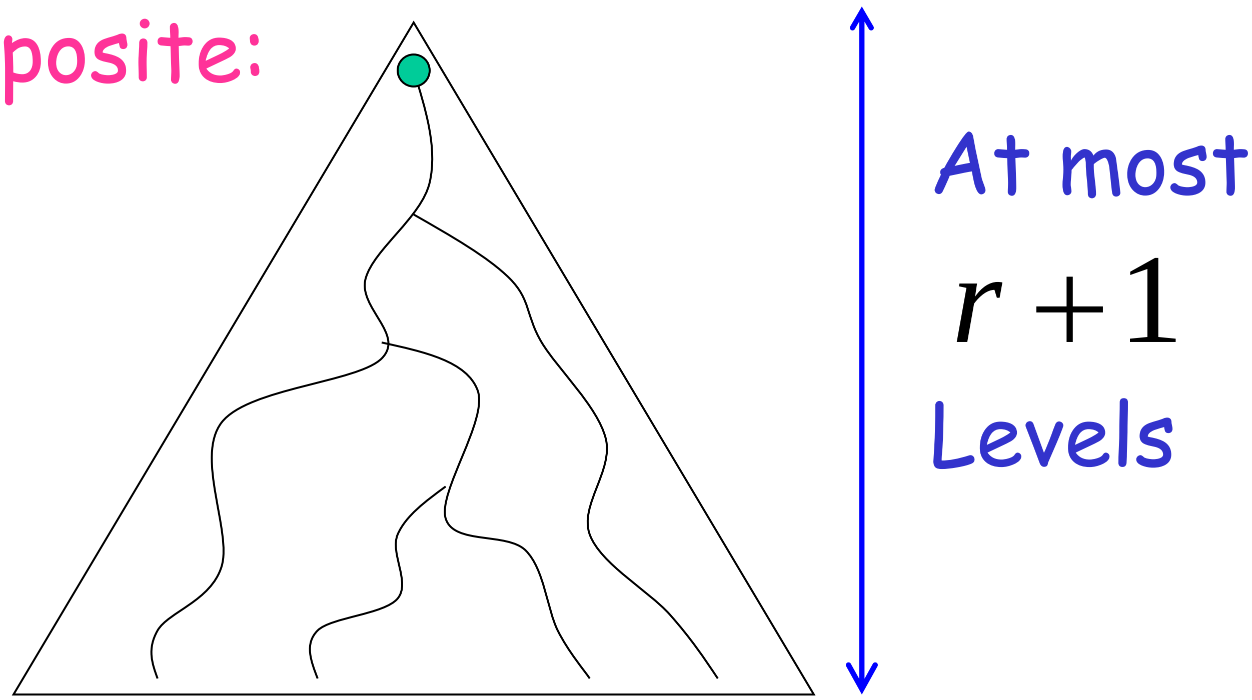
$|w| \geq m$

We will show:
some variable
is repeated

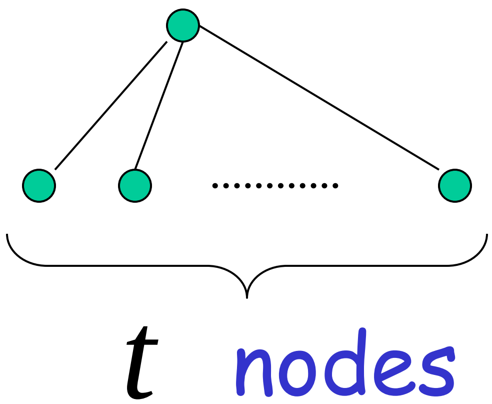


First we show that the tree of w has at least $r+2$ levels of nodes

Suppose the opposite:



Maximum number of nodes per level

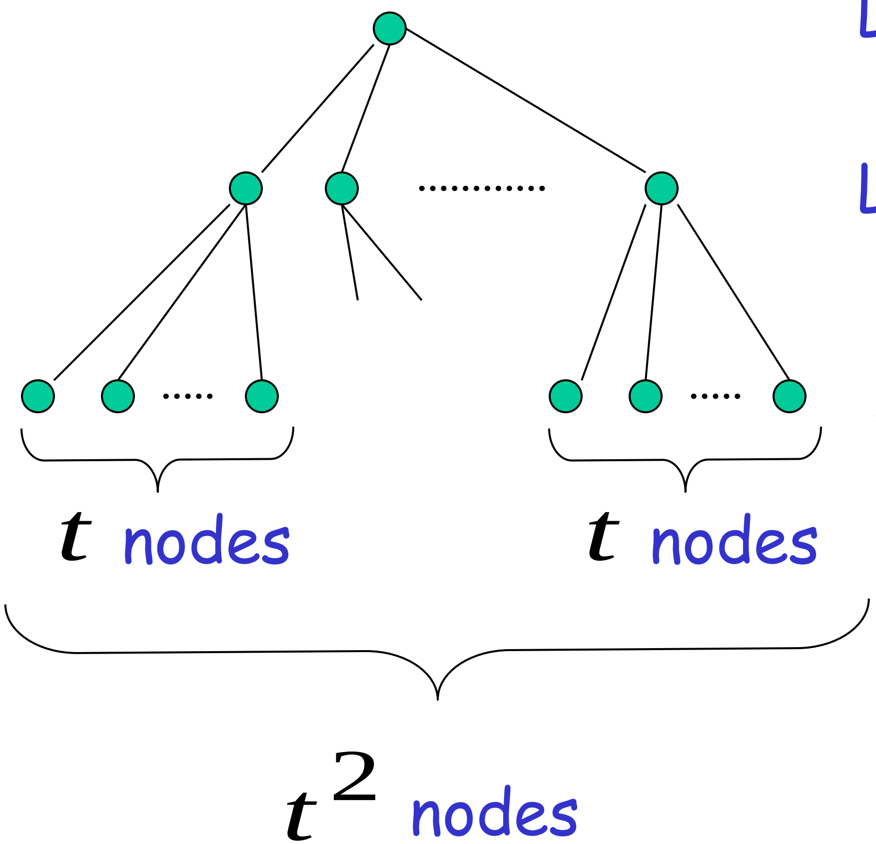


Level 0: 1 nodes

Level 1: t nodes

The maximum right-hand side of any production

Maximum number of nodes per level

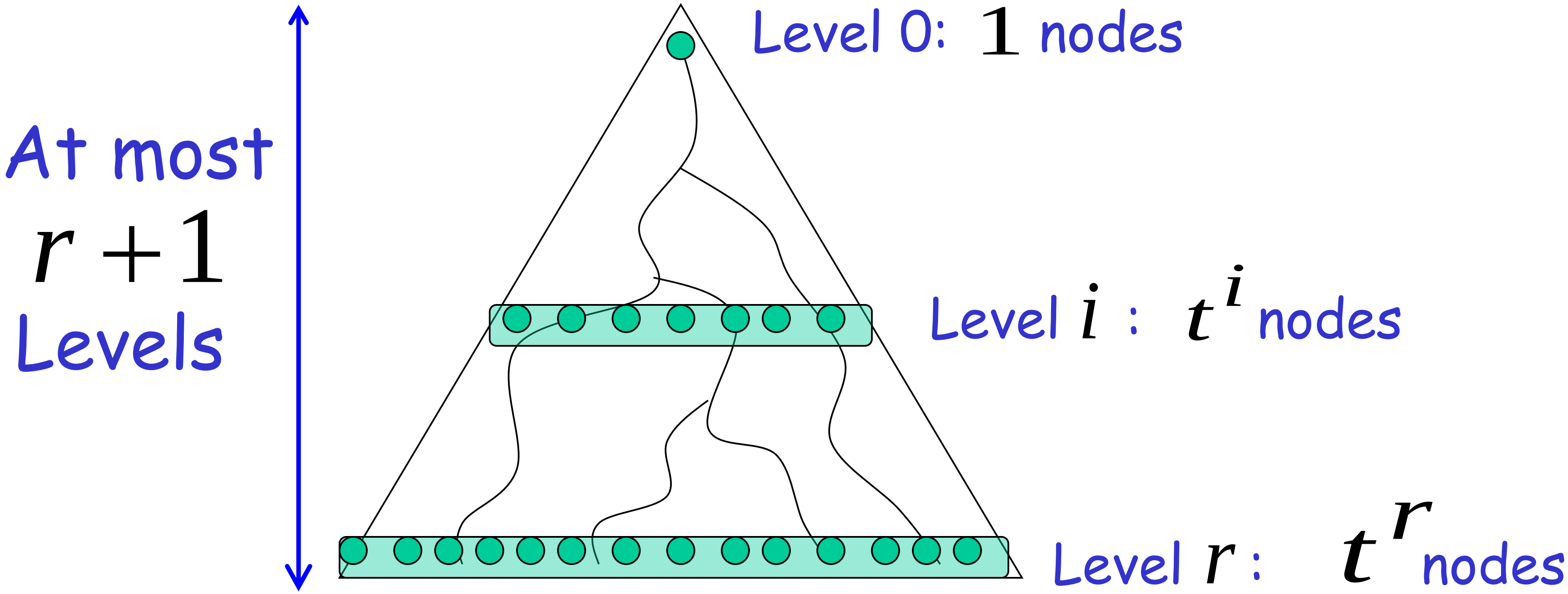


Level 0: 1 nodes

Level 1: t nodes

Level 2: t^2 nodes

Maximum number of nodes per level



Maximum possible string length

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

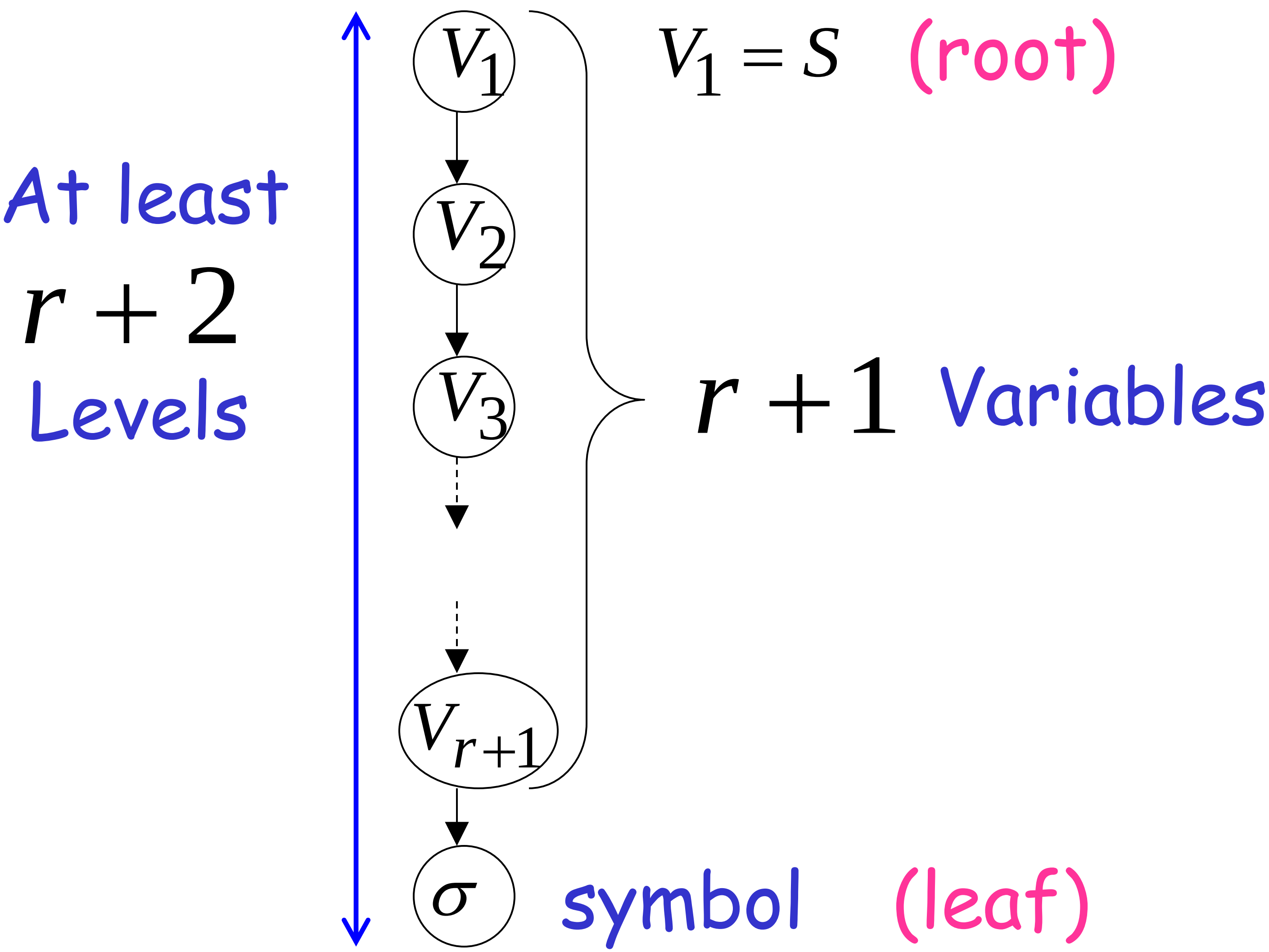
Therefore,
maximum length of string w : $|w| \leq t^r$

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

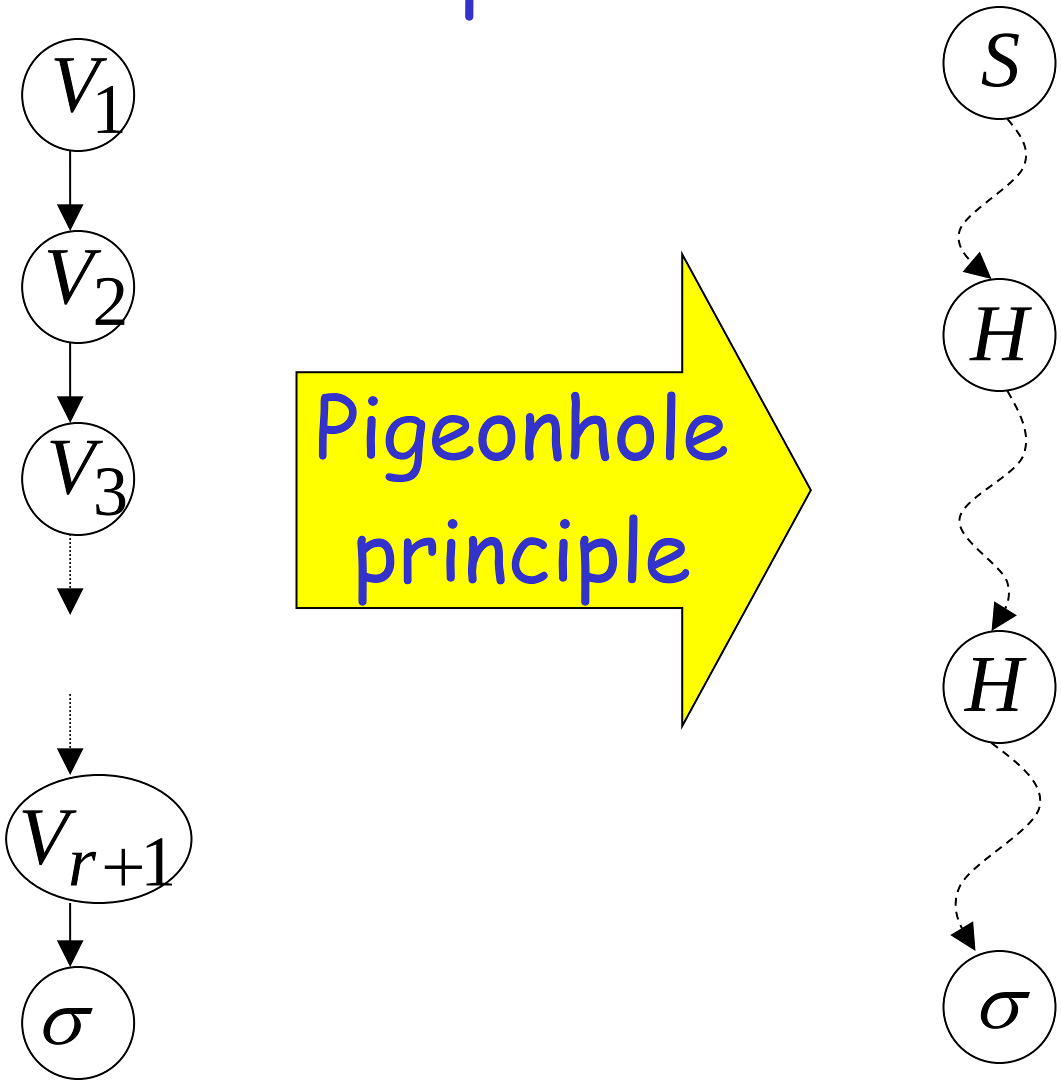
Contradiction!!!

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

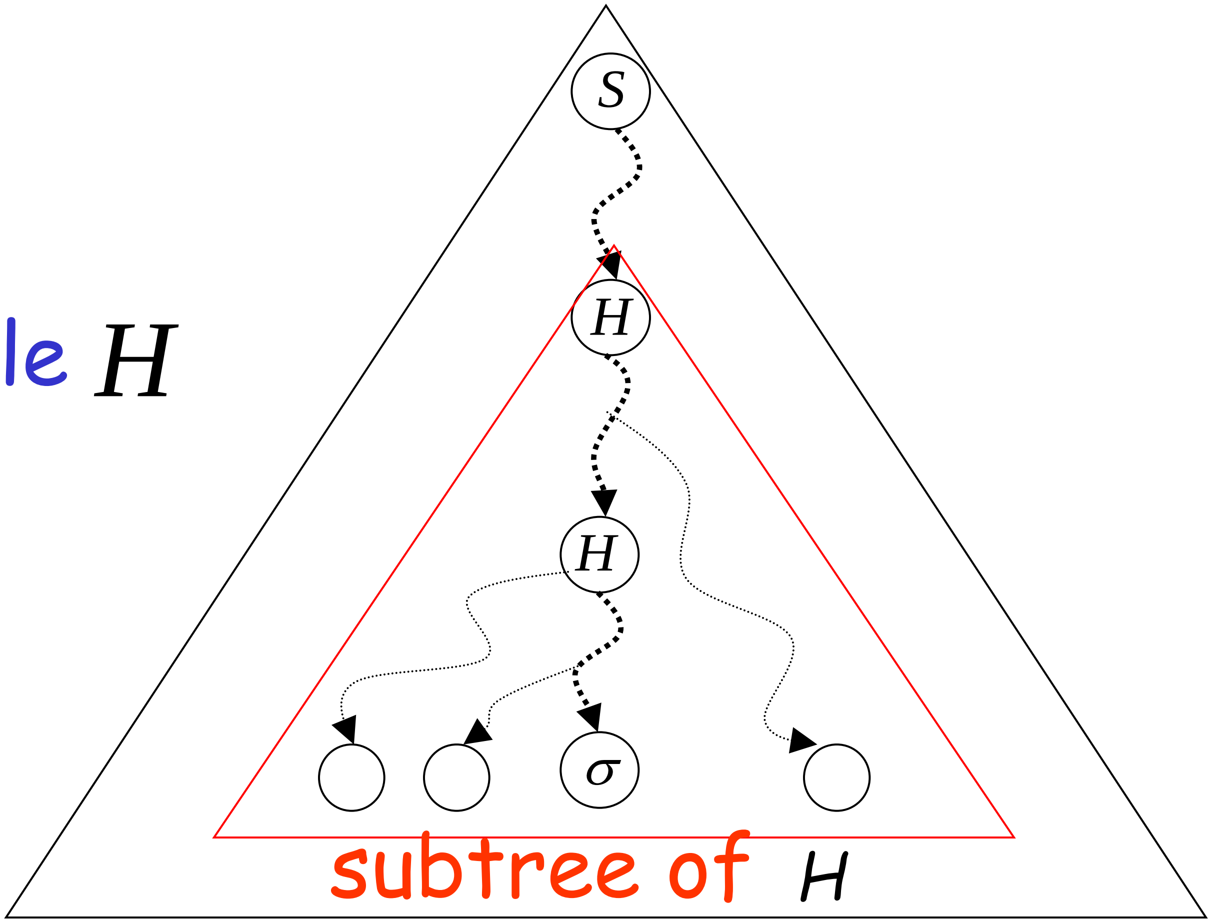


END OF CLAIM PROOF

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

From claim:

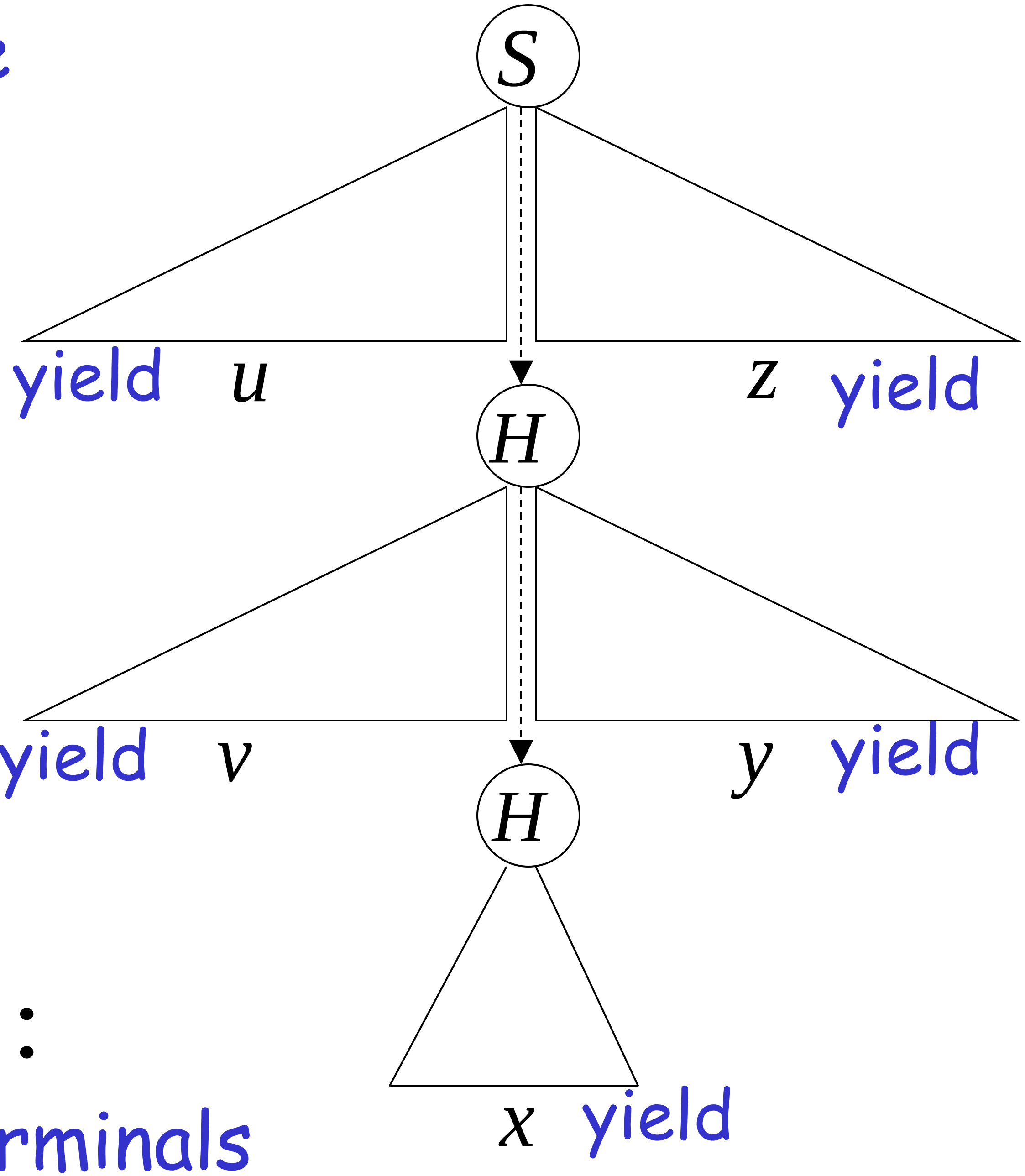
some variable H is repeated



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

We can write

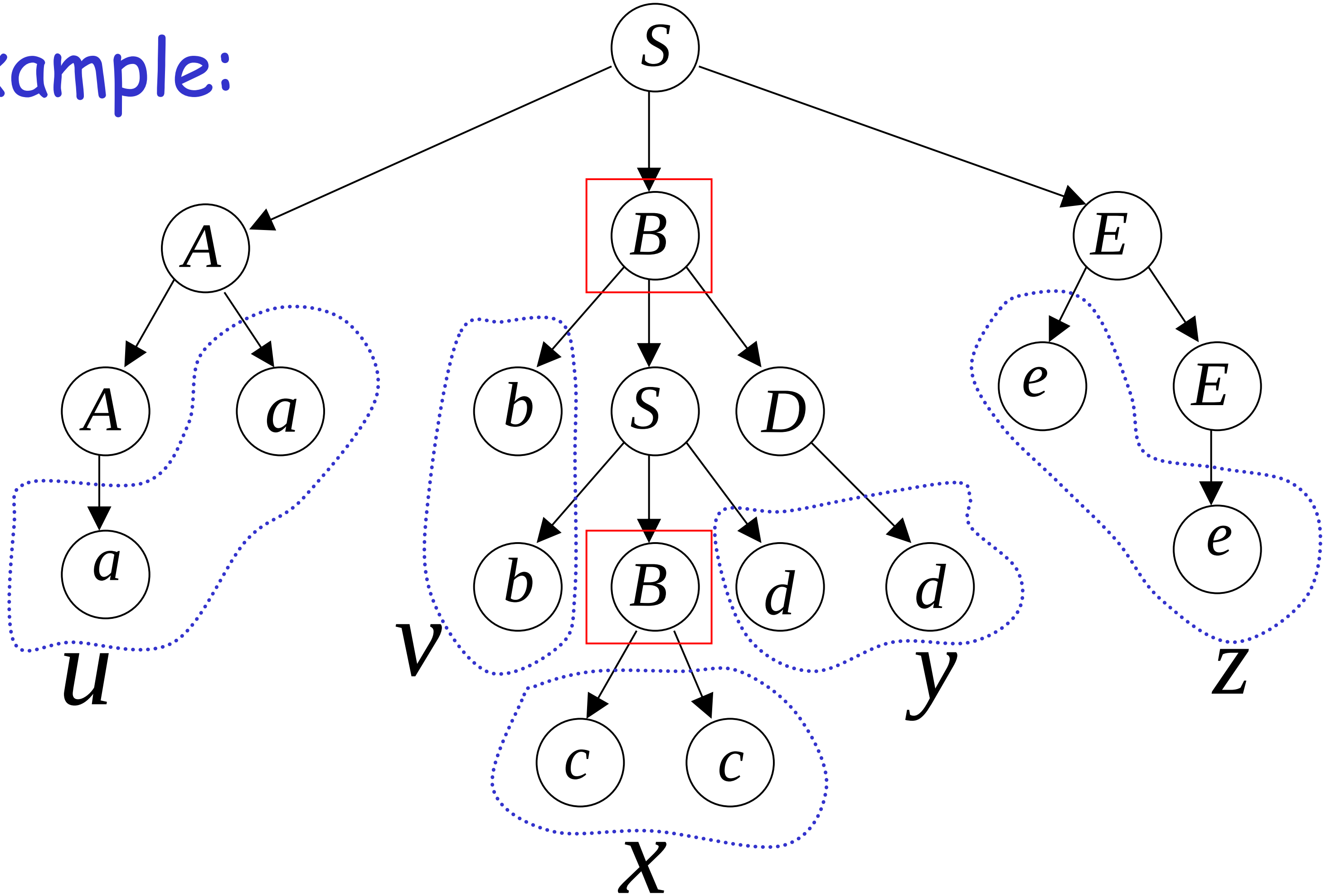
$w = uvxyz$



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

Strings of terminals

Example:



$u = aa$

$v = bb$

$x = cc$

$y = dd$

$z = ee$

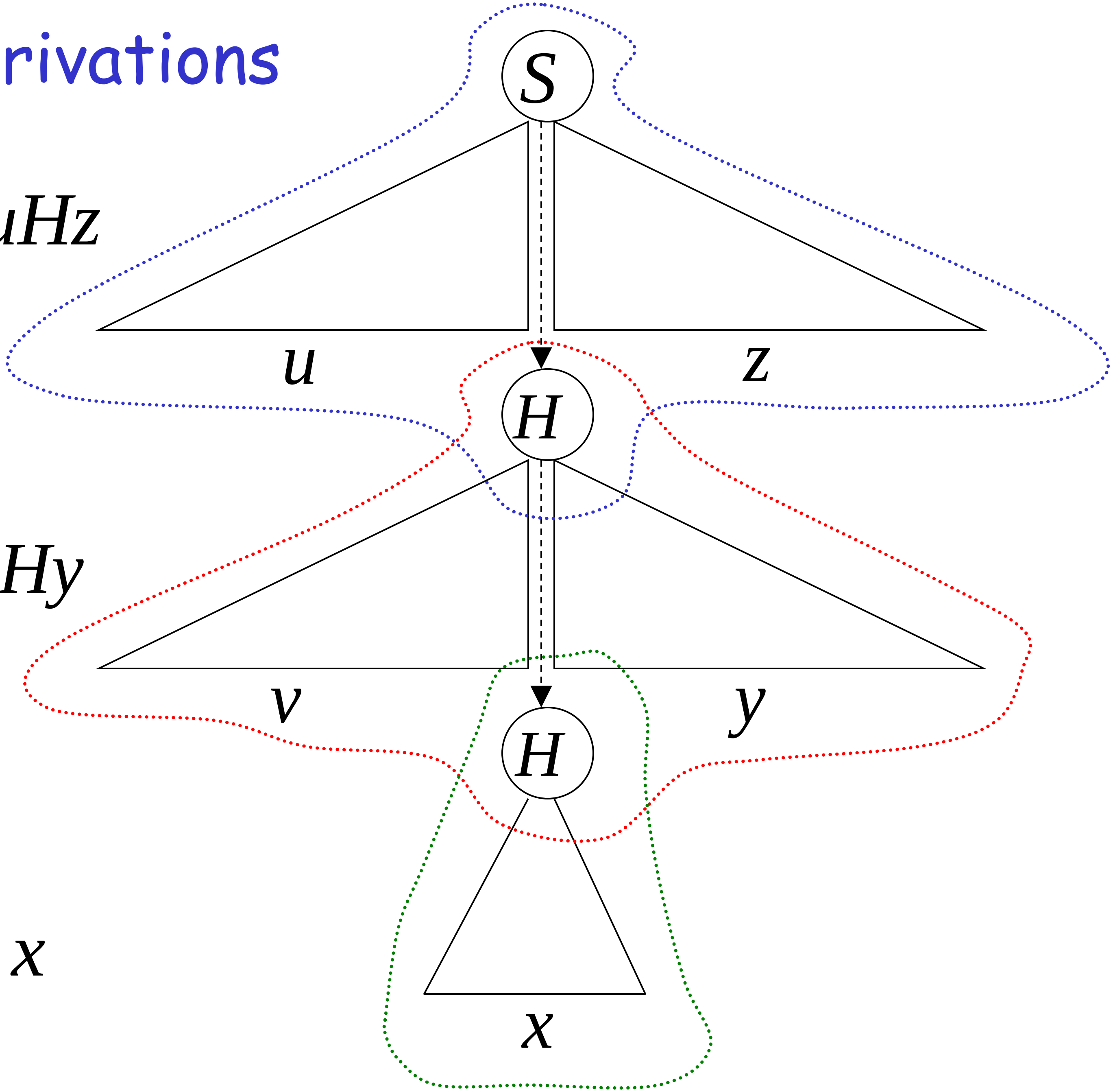
B correspond s to H

Possible derivations

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

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

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



Example:

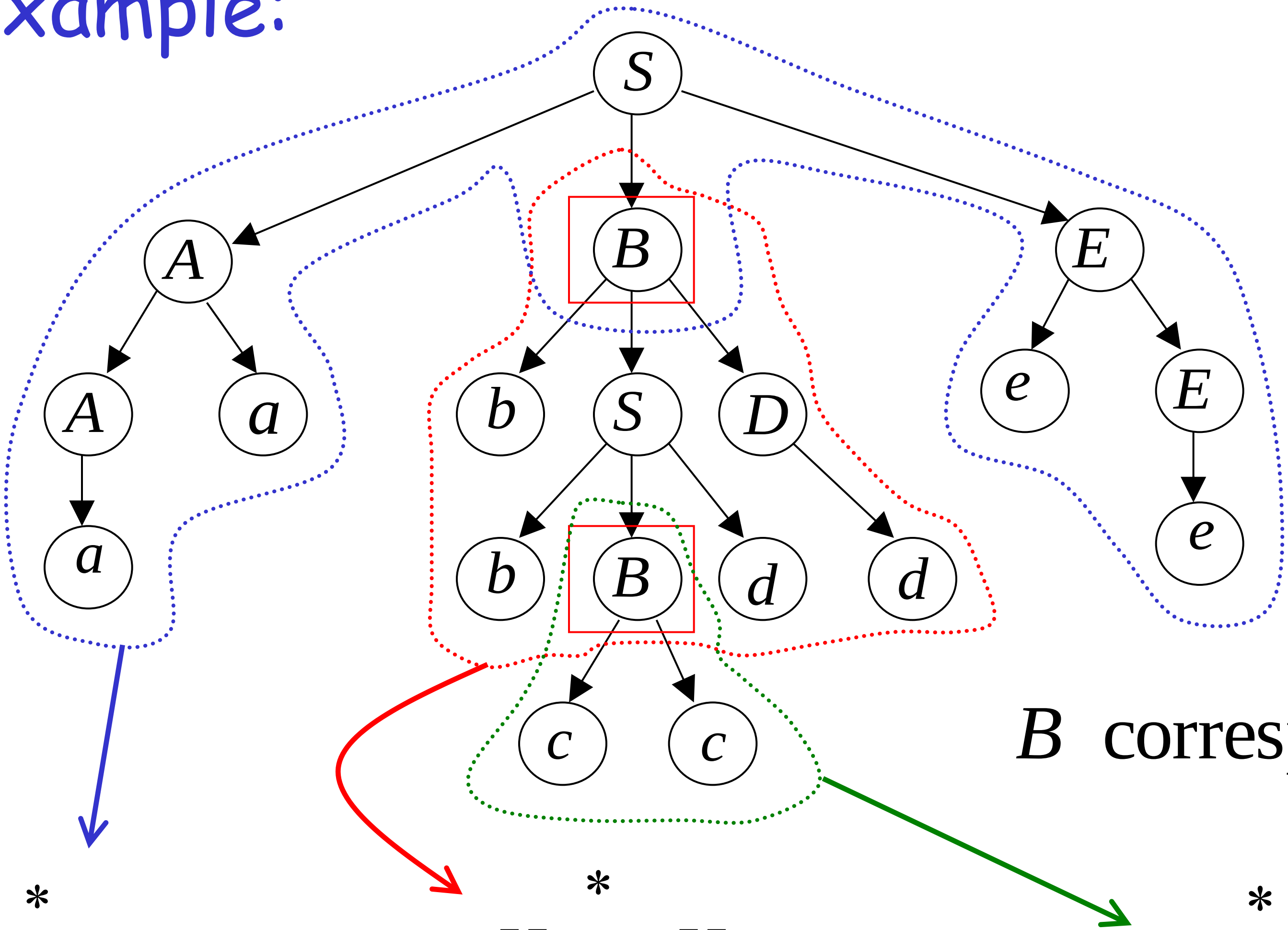
$u = aa$

$v = bb$

$x = cc$

$y = dd$

$z = ee$



B corresponds to H

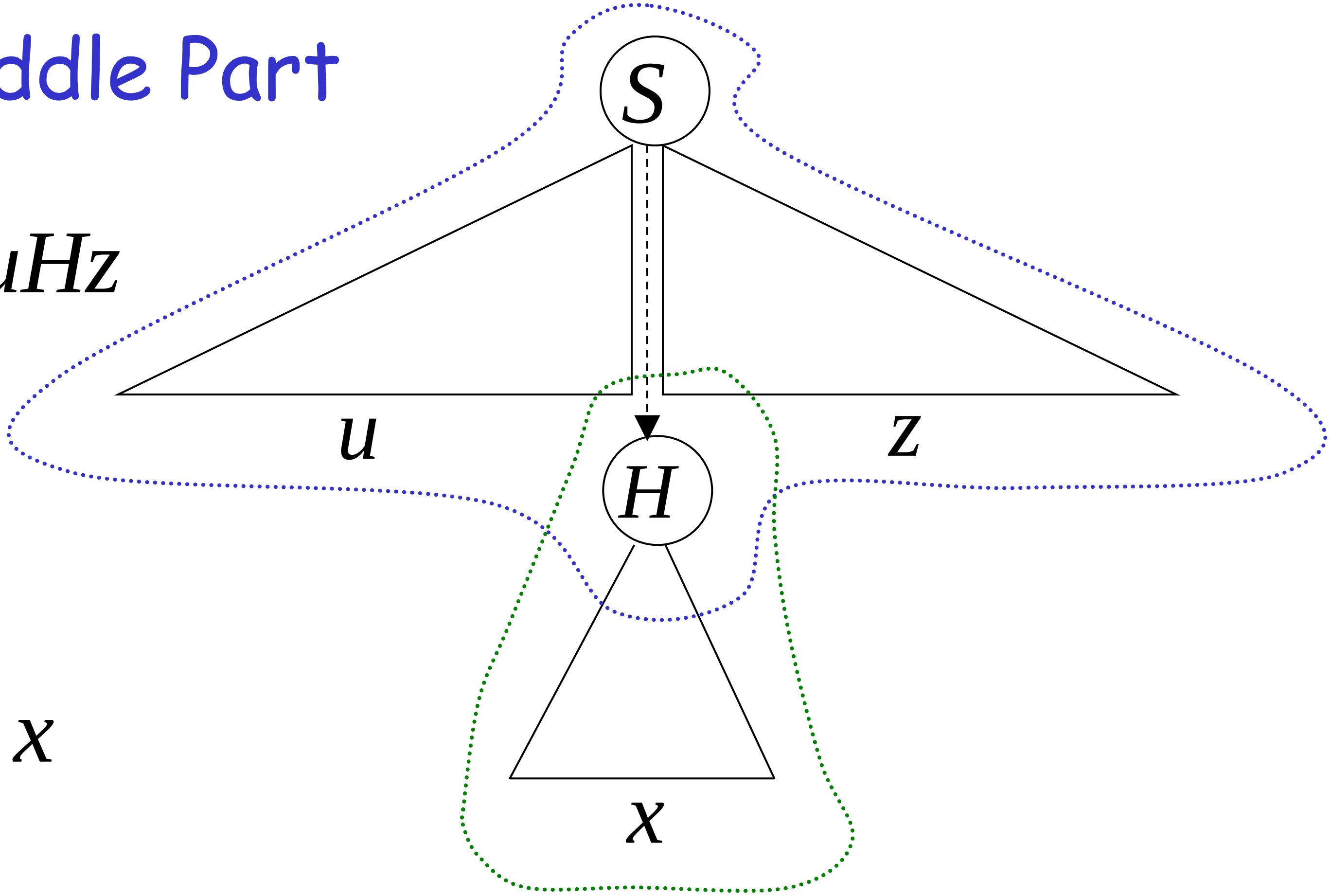
$S \overset{*}{\Rightarrow} uHz$ $\quad$ $H \overset{*}{\Rightarrow} vHy$ $\quad$ $H \overset{*}{\Rightarrow} x$

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

Remove Middle Part

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

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



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

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

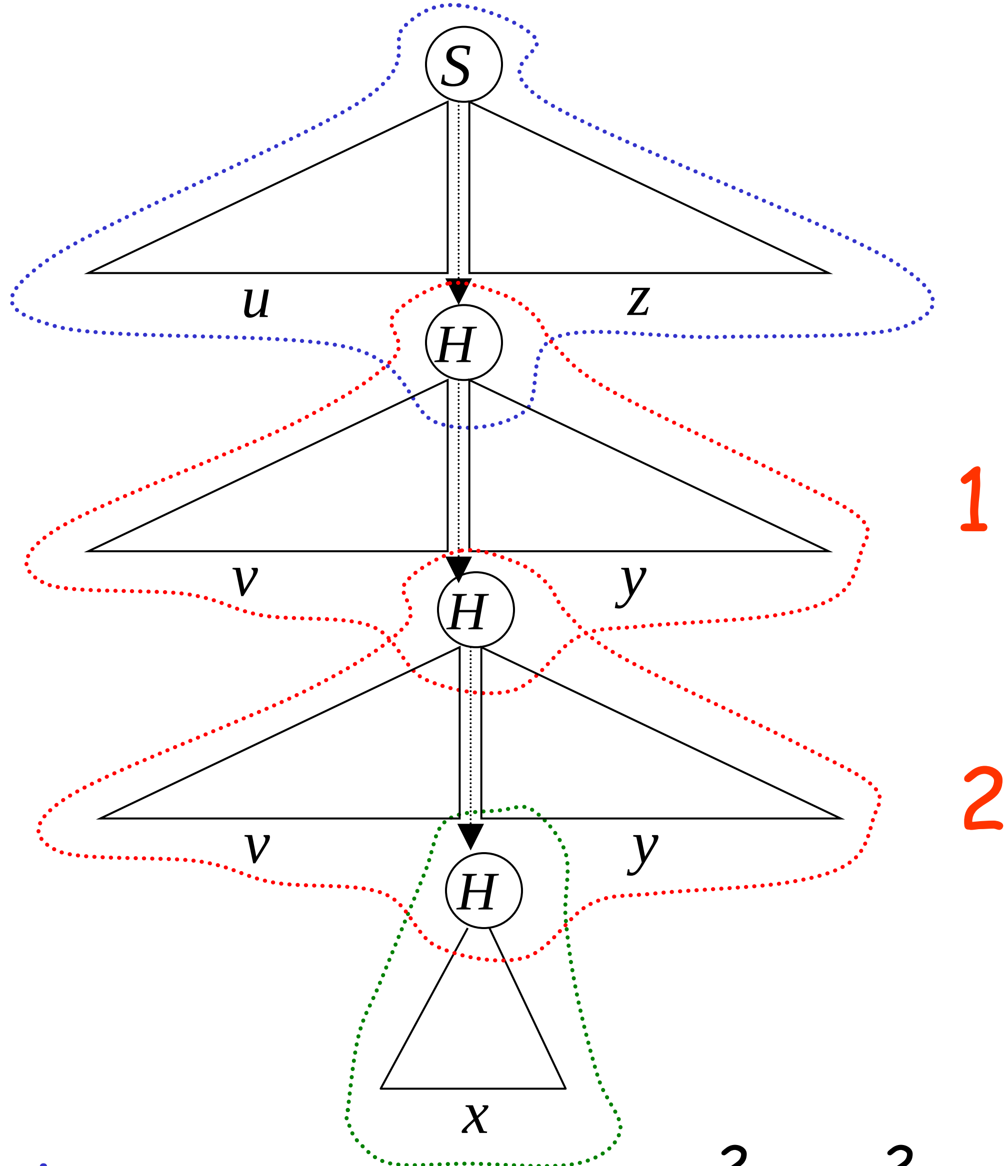
Repeat Middle part two times

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

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

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

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



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

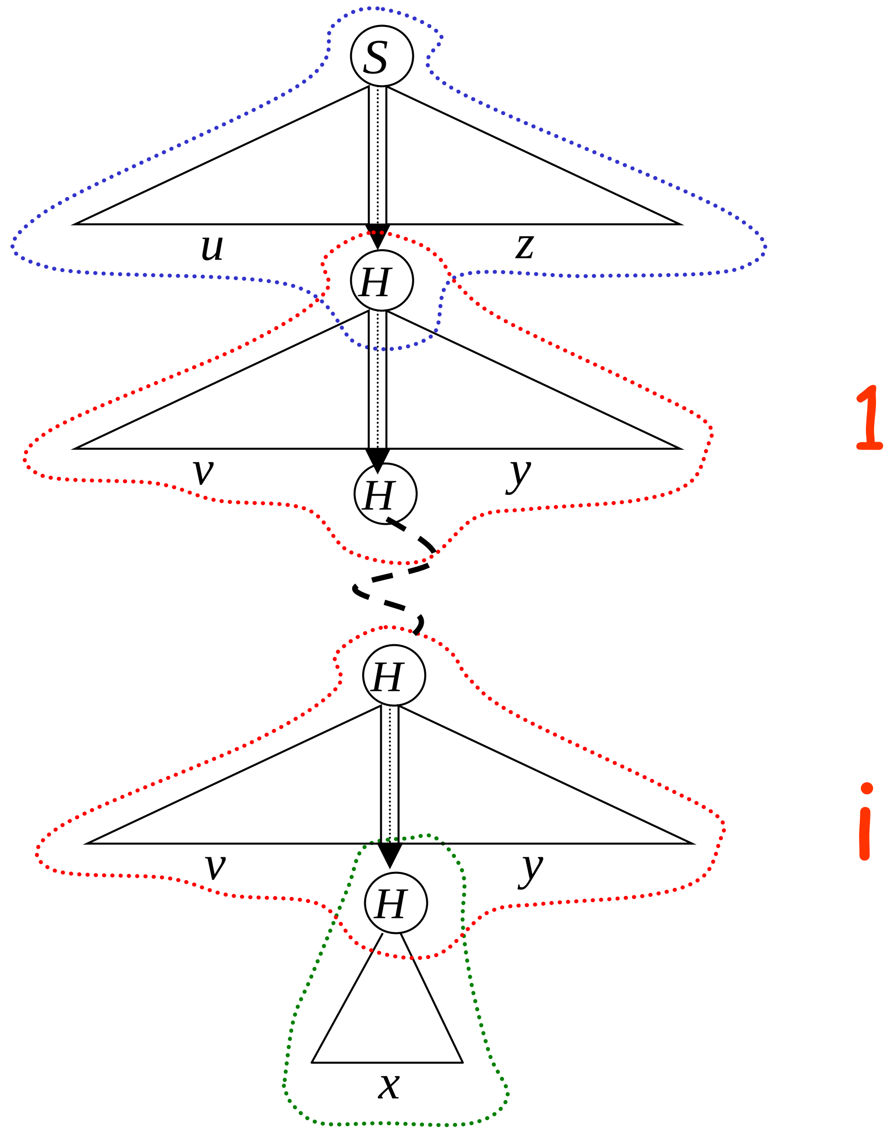
$$S \stackrel{*}{\Rightarrow} uHz$$

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

$$\vdots$$

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

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



Yield: $uv^i xy^i z$

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

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

$$\stackrel{*}{\Rightarrow} \ldots$$

$$\stackrel{*}{\Rightarrow} uv^iHy^iz \stackrel{*}{\Rightarrow} uv^ixy^iz \in L(G)$$

Therefore,

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

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

then we also know: $uv^i xy^i z \quad \in L(G)$

For all $i \geq 0$

since

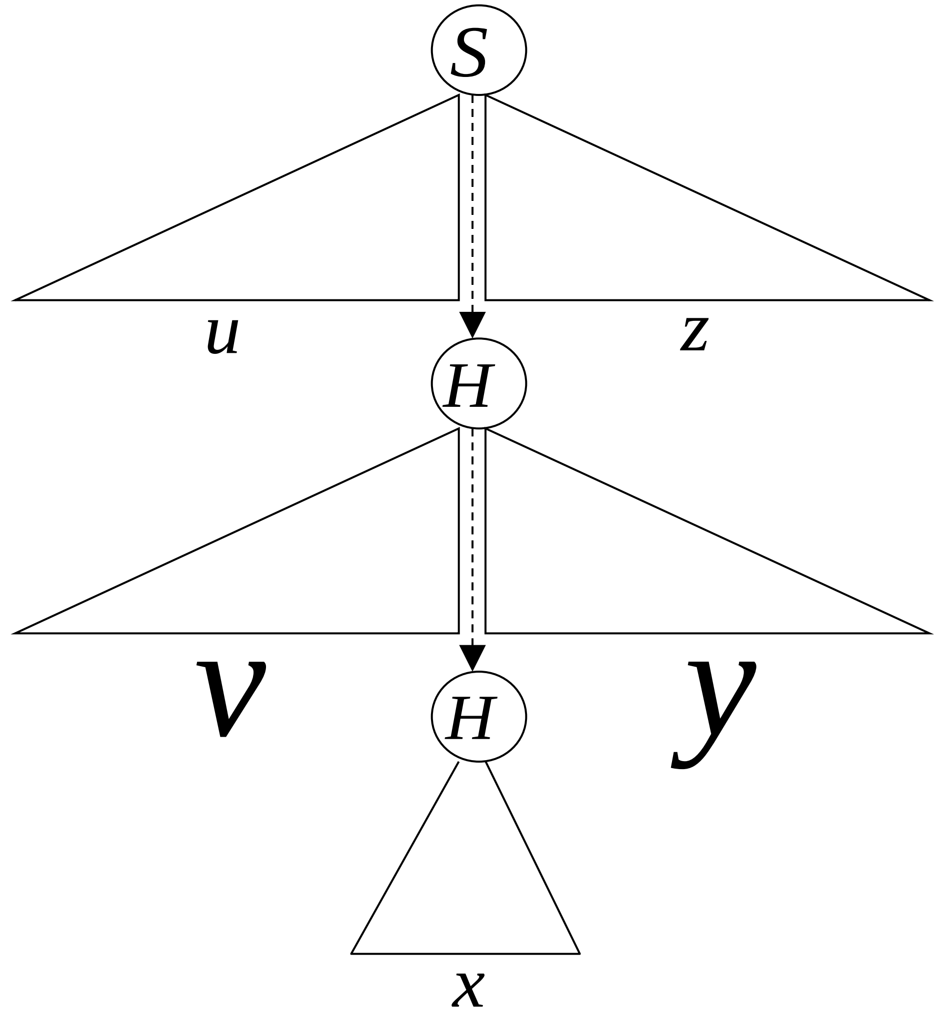
$L(G) = L - \{\lambda\}$

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

Observation 1:

$$|vy| \geq 1$$

Since G has no unit and λ-productions

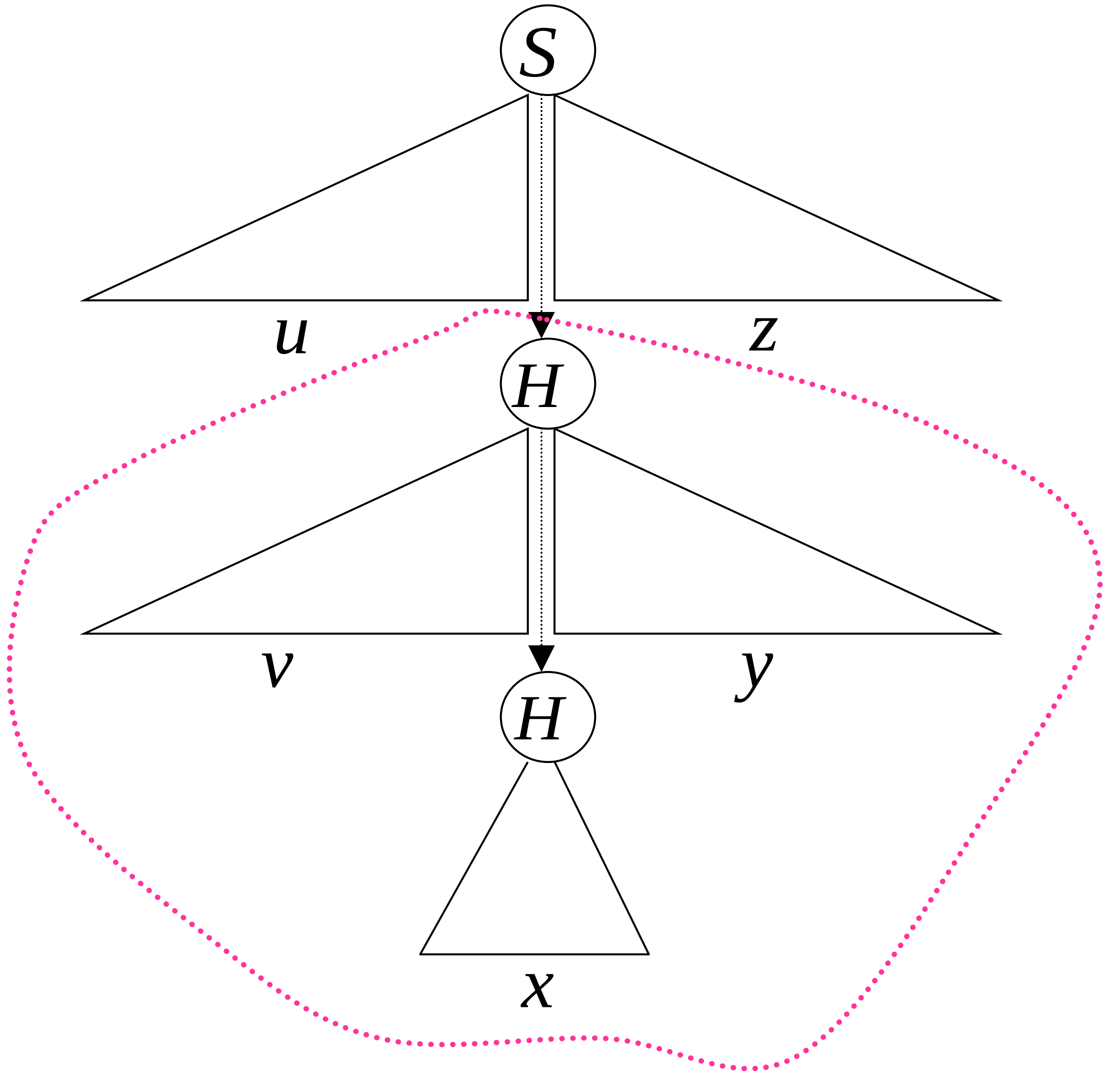


At least one of v or y is not λ

Observation 2:

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

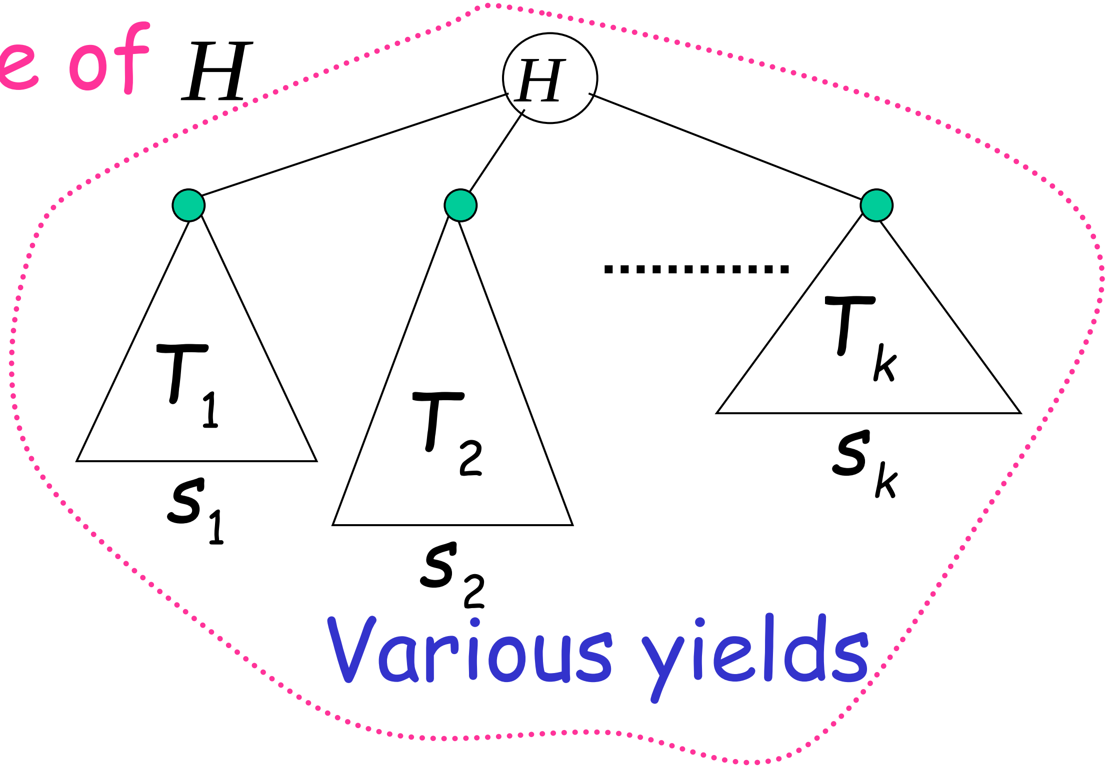
since in subtree only variable H is repeated



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 $|w| \geq m$

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

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

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

From pumping lemma:

we can write: $w = uvxyz$

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

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

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

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

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

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

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

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

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

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

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

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

where z spans the final $a\ldots a$, the b's and the c's.

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

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

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

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

$$\overbrace{\underbrace{a \ldots a}_{u}\underbrace{a \ldots a}_{v^2}\underbrace{a \ldots a}_{x}\underbrace{a \ldots a}_{y^2}a \ldots a}^{m+k_1+k_2}\ \overbrace{bbb \ldots bbb}^{m}\ \overbrace{ccc \ldots ccc}^{m}$$

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

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

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

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

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

$$k_1 + k_2 \geq 1$$

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

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

$$\underbrace{\overbrace{aaa\ldots aaa}^{m}\ b\ldots b}_{u}\underbrace{b\ldots b}_{vxy}\underbrace{b\ldots b\ \overbrace{ccc\ldots ccc}^{m}}_{z}$$

(the b's form a block of length m)

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

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

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

Case 3: vxy is in c^m

Similar to case 1

$$\overbrace{aaa\ldots aaa}^{m}\ \overbrace{bbb\ldots bbb}^{m}\ \overbrace{c\ldots cc\ldots cc\ldots c}^{m}$$

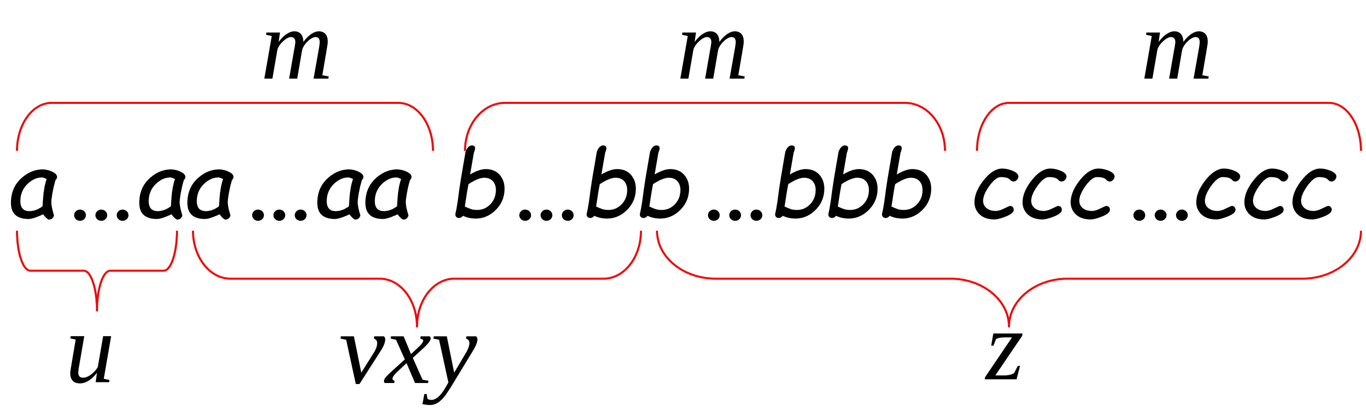
$$\underbrace{aaa\ldots aaa\ bbb\ldots bbb\ c\ldots c}_{u}\ \underbrace{c\ldots c}_{vxy}\ \underbrace{c\ldots c}_{z}$$

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

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

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

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



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

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

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

Sub-case 1: v contains only a
y contains only b

$$\overbrace{a\ldots a\,a\ldots a\,a\ldots a}^{m}\ \overbrace{b\ldots b\,b\ldots b\,b\ldots b}^{m}\ \overbrace{ccc\ldots ccc}^{m}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$k_1 + k_2 \geq 1$$

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

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

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

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

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

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

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

By assumption

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

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

$$\underbrace{a\ldots a}_{u}\underbrace{\overbrace{a\ldots a}^{k_1}\;\overbrace{b\ldots b}^{k_2}}_{v}\underbrace{b\ldots b}_{x}\underbrace{\overbrace{b\ldots b}^{k_3}}_{y}\underbrace{b\ldots b\;ccc\ldots ccc}_{z}$$

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

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

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

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

$$m \qquad\qquad m + k_3 \qquad\qquad m$$

$$\underbrace{a\ldots a}_{u}\,\underbrace{\overbrace{a\ldots a}^{k_1}\overbrace{b\ldots b}^{k_2}\overbrace{a\ldots a}^{k_1}\overbrace{b\ldots b}^{k_2}}_{v^2}\,\underbrace{b\ldots b}_{x}\,\underbrace{\overbrace{b\ldots b}^{2k_3}}_{y^2}\,\underbrace{b\ldots b\ ccc\ldots ccc}_{z}$$

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

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

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

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

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

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

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

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

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

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

Impossible!

$$\underbrace{aaa...}_{u}\!\!\overbrace{aaa...aaa}^{m}\;\overbrace{bbb...bbb}^{m}\;\overbrace{ccc...ccc}^{m}$$

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

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