The Pumping Lemma for Context-Free Languages

Take an **infinite** context-free language

Generates an infinite number of different strings

Example:

$$S \rightarrow AB$$

$$A \rightarrow aBb$$

$$B \rightarrow Sb$$

$$B \rightarrow b$$

$$S \rightarrow AB$$

$$A \rightarrow aBb$$

$$B \rightarrow Sb$$

$$B \rightarrow b$$

In a derivation of a long string, variables are repeated

A derivation:

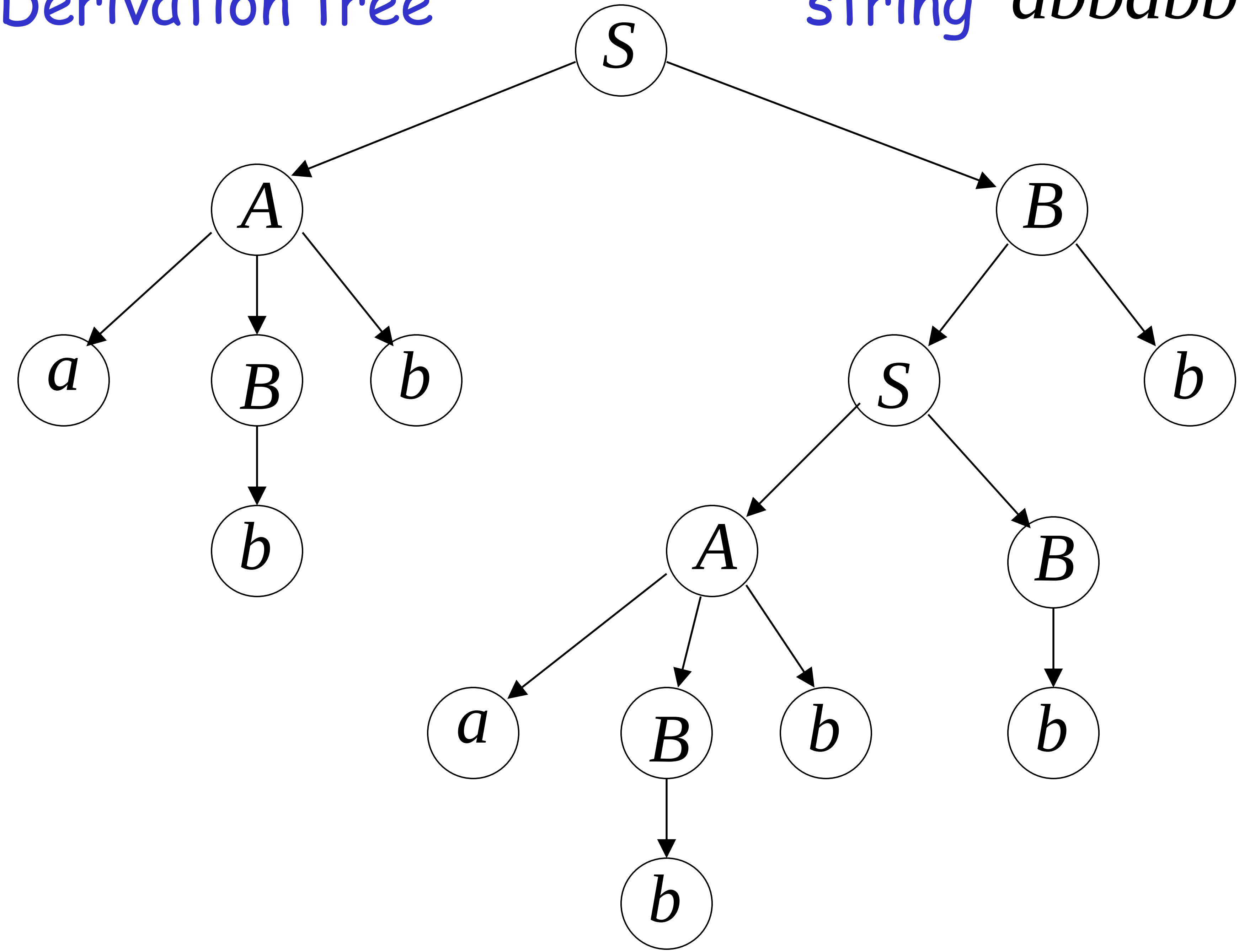
$$S \Rightarrow AB \Rightarrow aBbB \Rightarrow abb\underline{B}$$

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

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

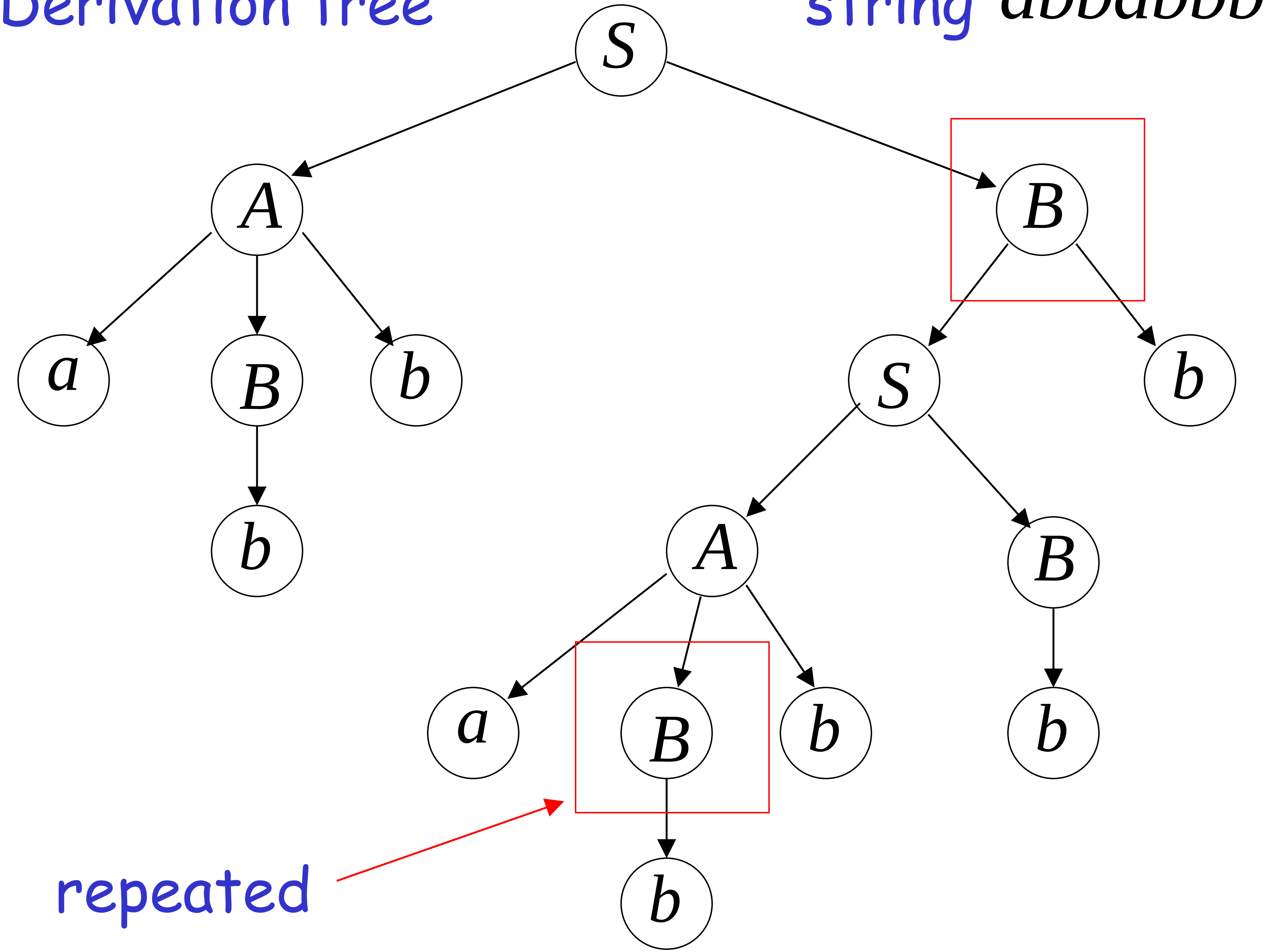
Derivation tree

string $abbabbbb$



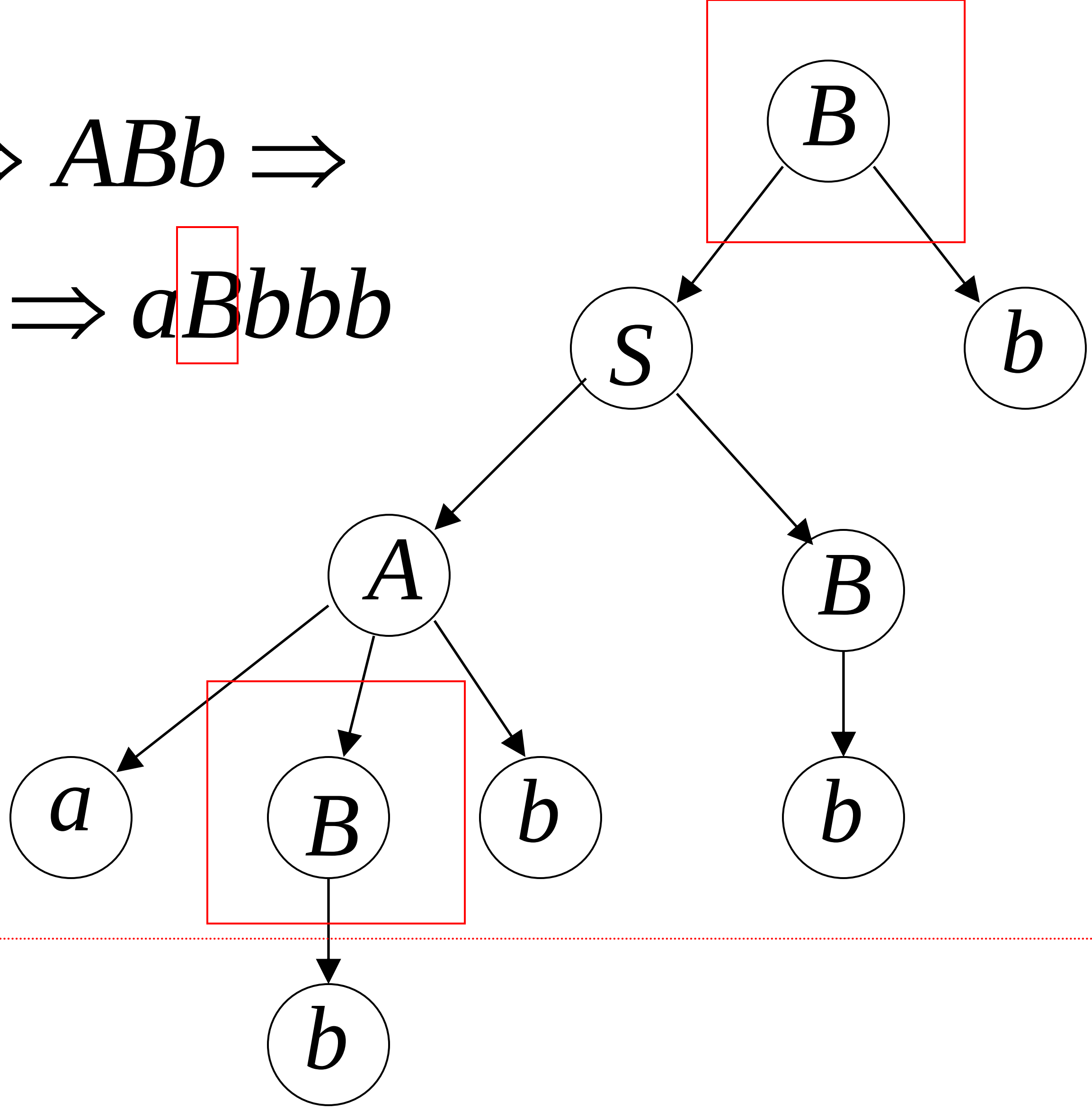
Derivation tree

string $abbabbbb$



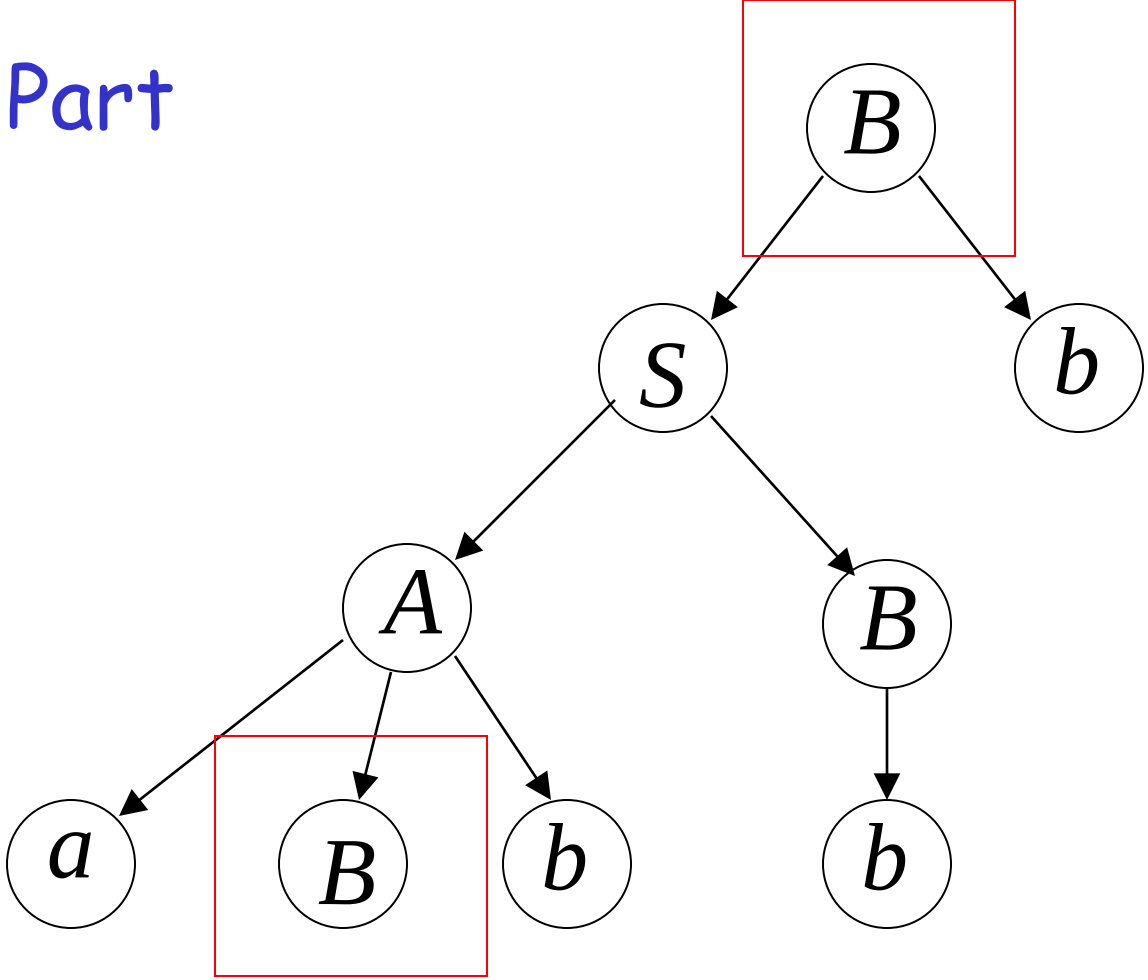
$$\boxed{B} \Rightarrow Sb \Rightarrow ABb \Rightarrow$$

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



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

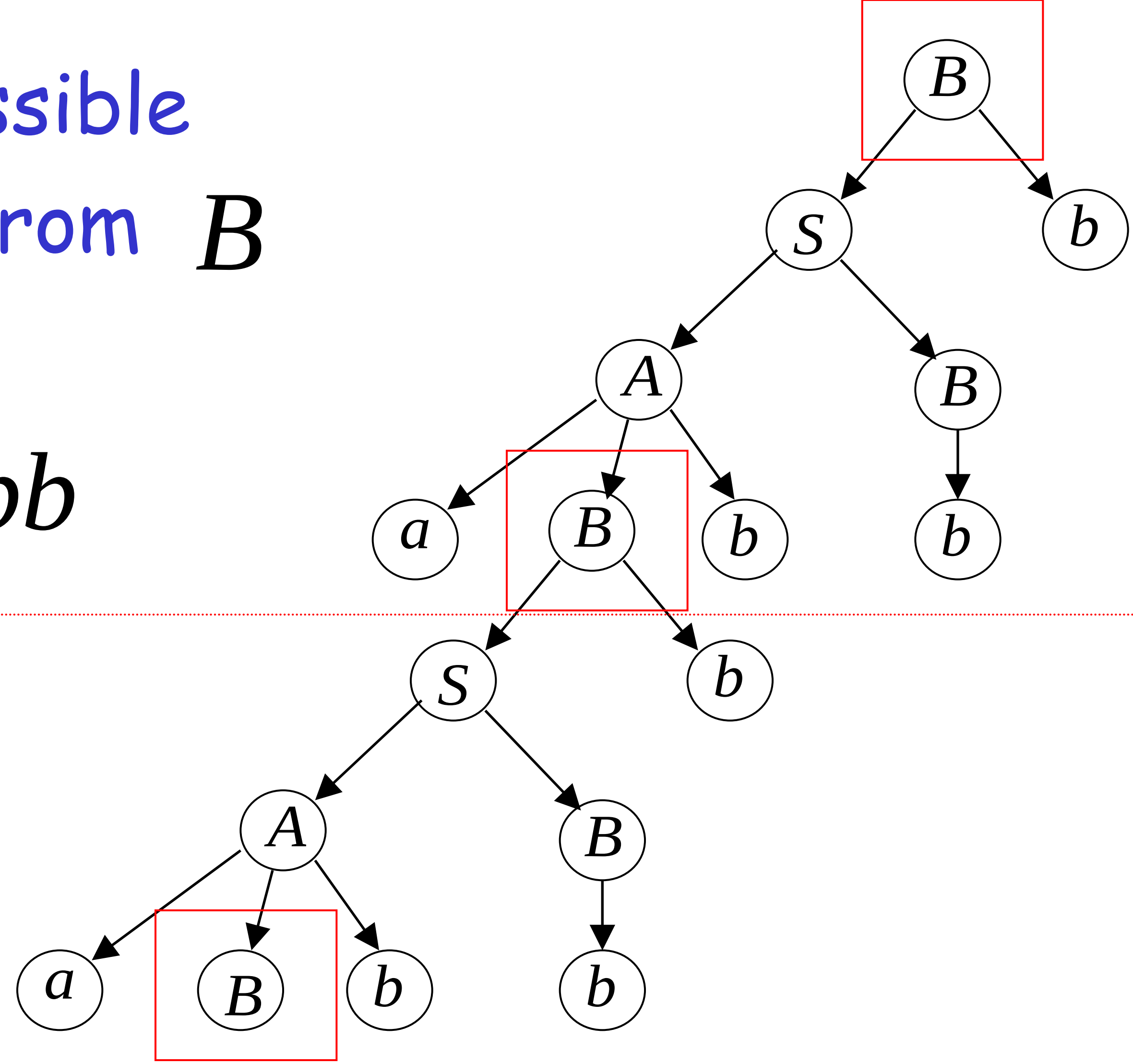
Repeated Part



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

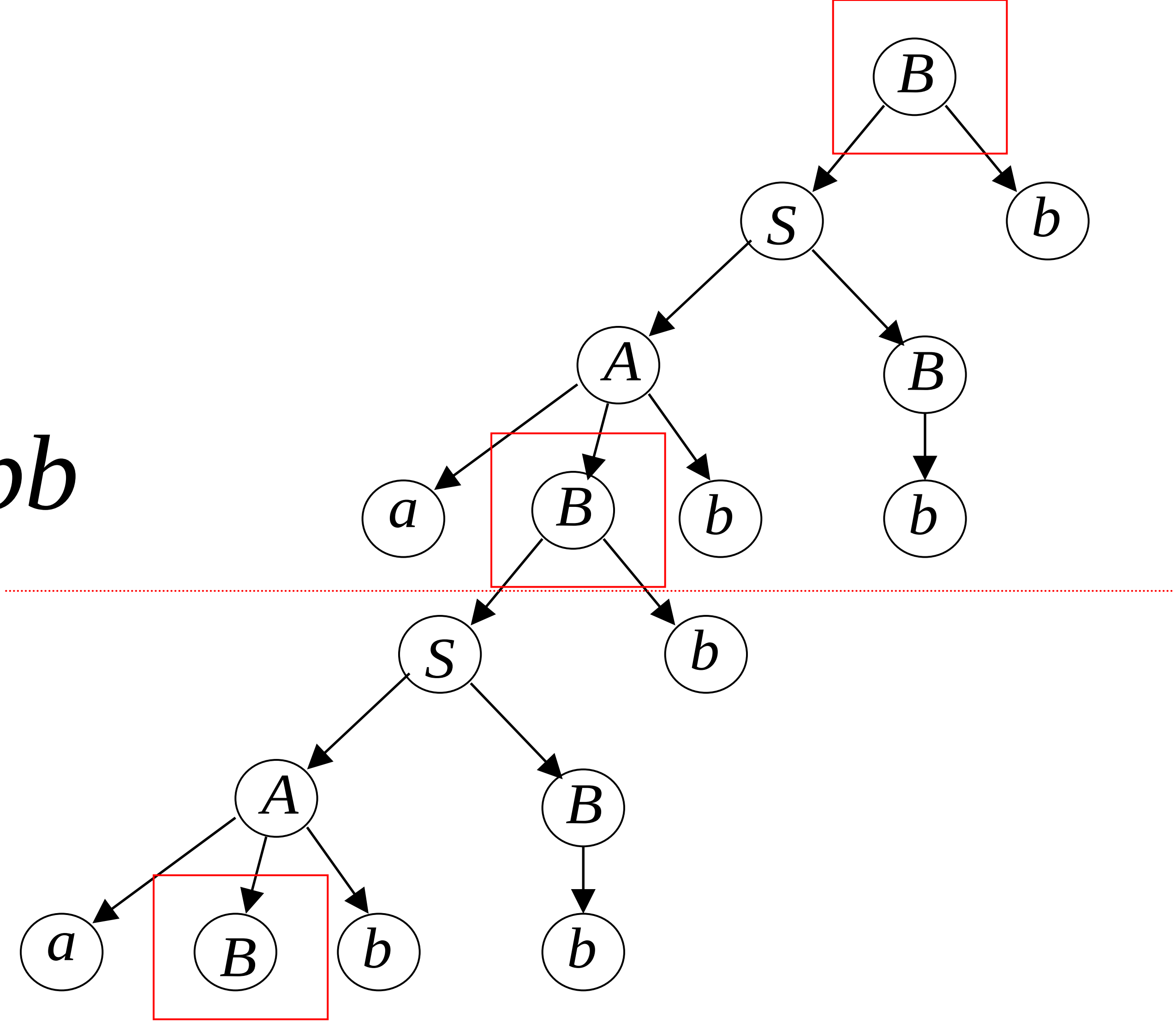
Another possible derivation from B

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



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

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

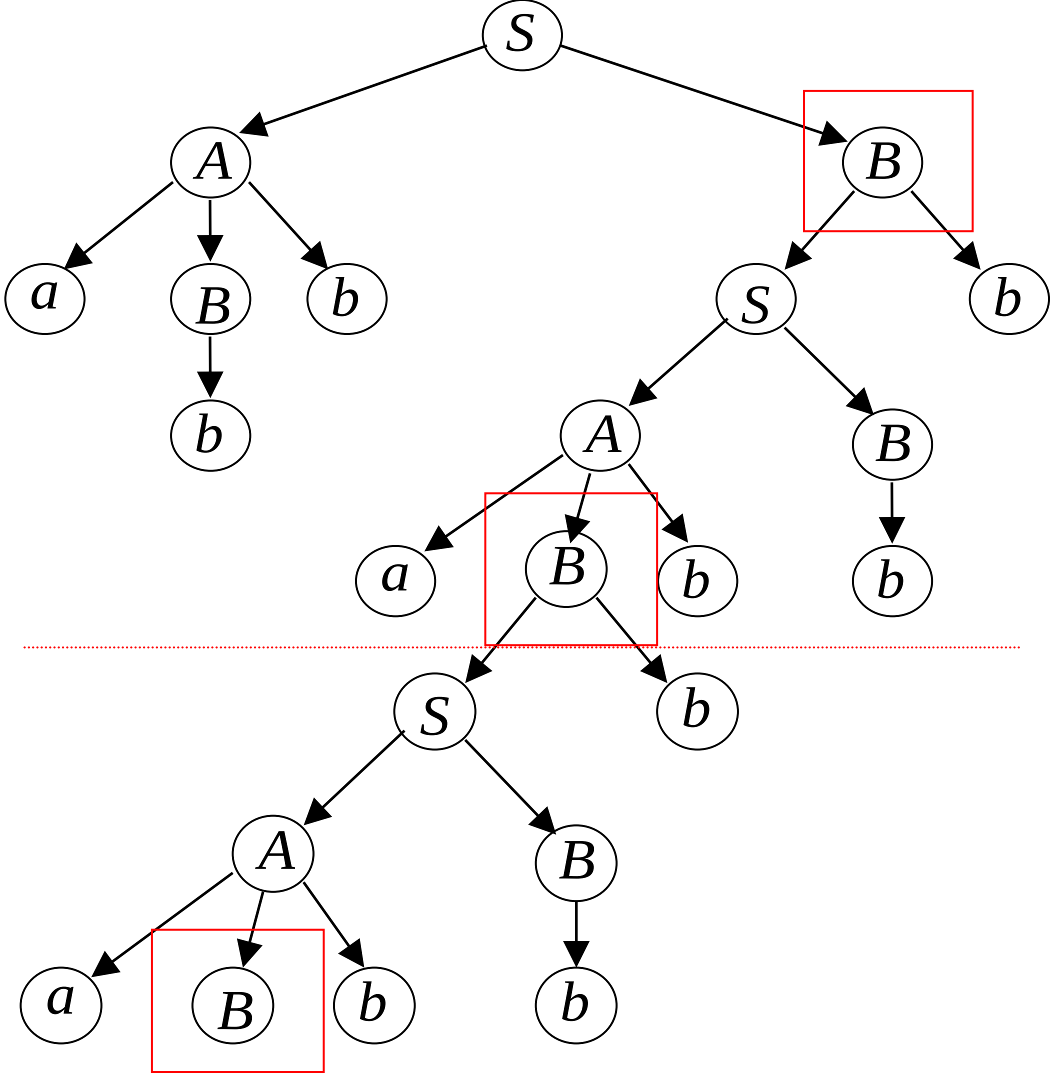


$$B \overset{*}{\Rightarrow} (a)B(bbb) \overset{*}{\Rightarrow} (a)^2 B(bbb)^2$$

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

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

$$B \Rightarrow b$$

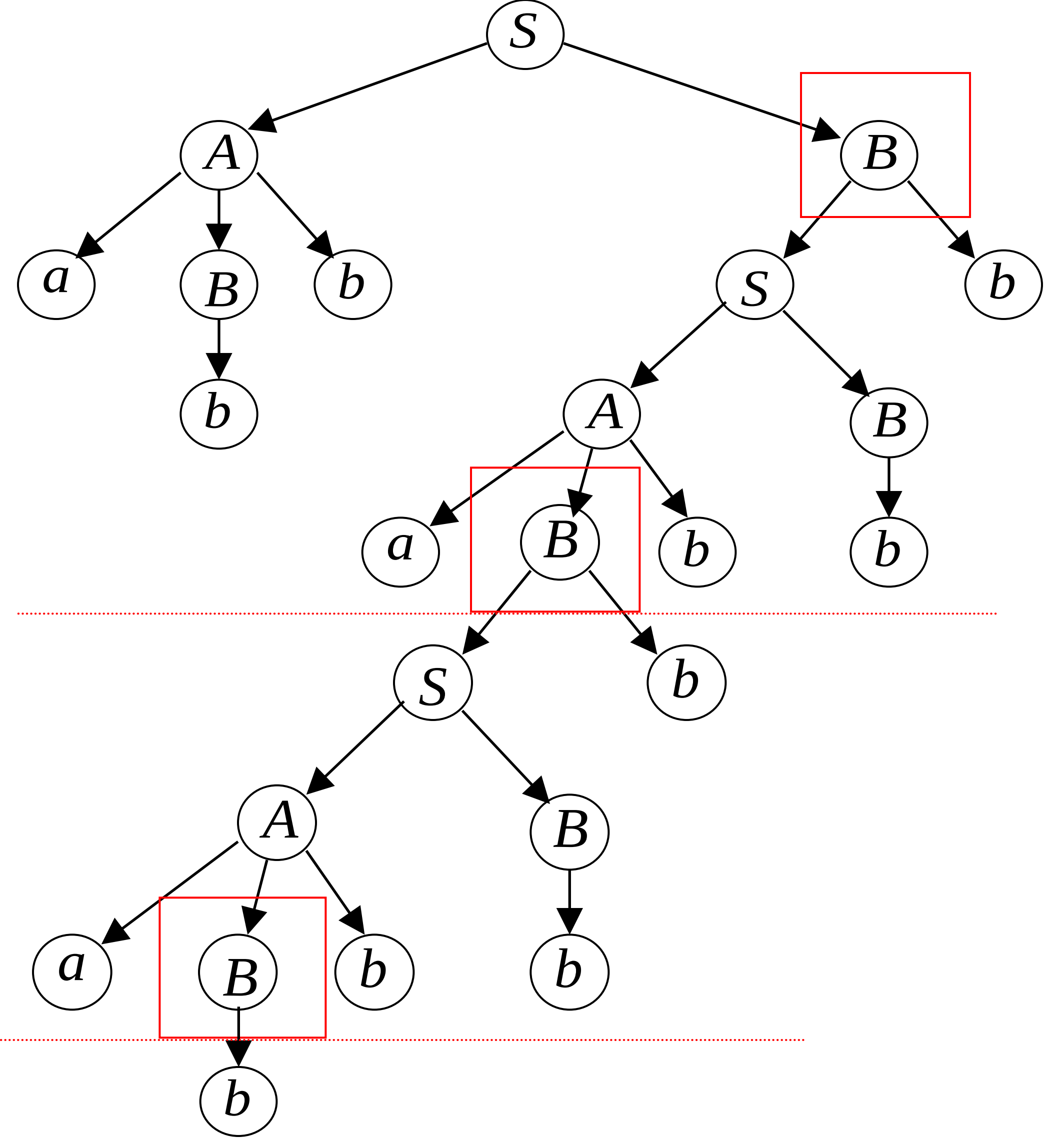


$$S \overset{*}{\Rightarrow} abb(a)B(bbb) \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2$$

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

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

$$B \Rightarrow b$$



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

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

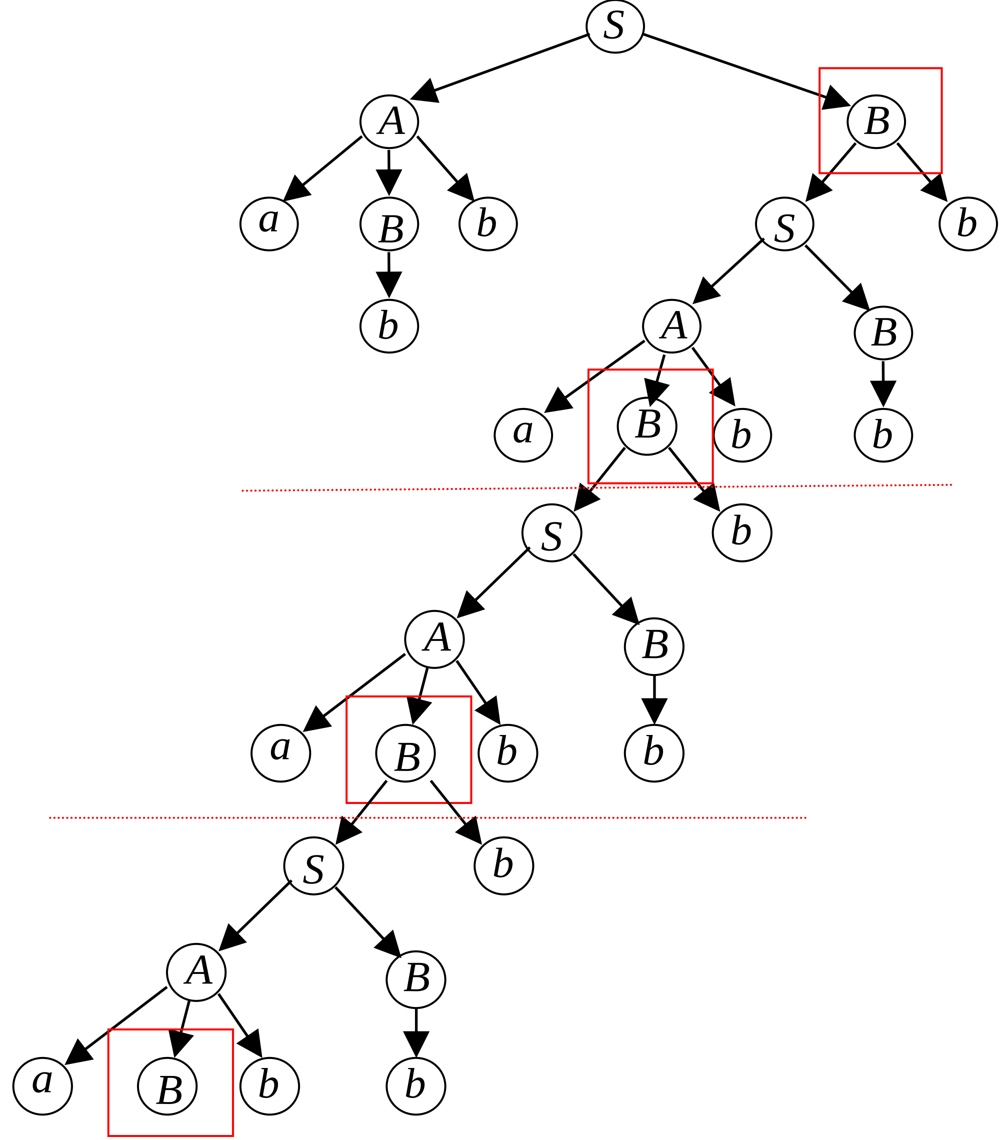
$$S \overset{*}{\Rightarrow} abb(a)^2 b(bbb)^2$$

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

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

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

$$B \Rightarrow b$$

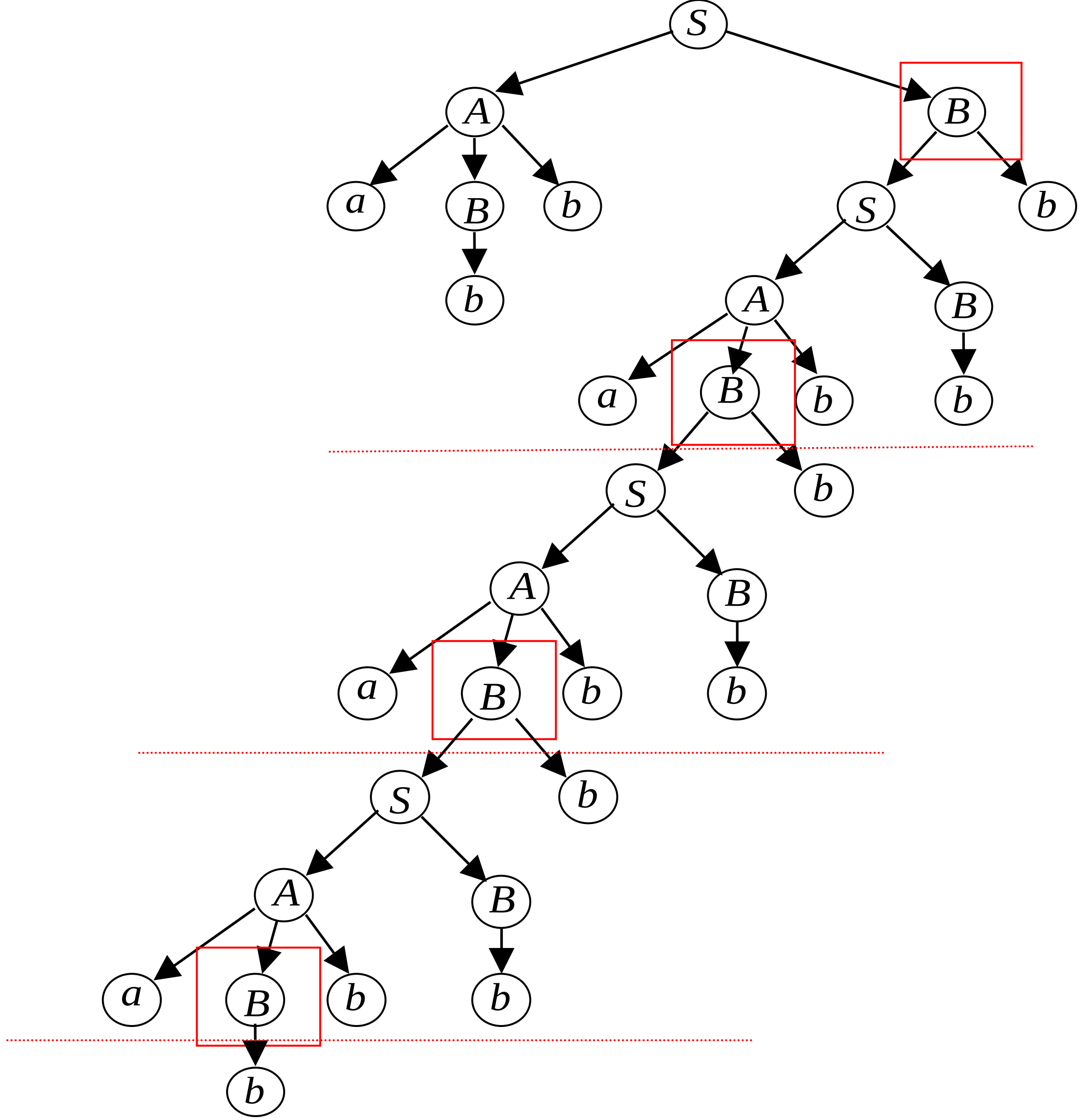


$$S \overset{*}{\Rightarrow} abb(a)^2 B(bbb)^2 \overset{*}{\Rightarrow} abb(a)^3 B(bbb)^3$$

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

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

$$B \Rightarrow b$$



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

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

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

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

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 it be f)

Let $m = p + 1$

Example $G:$ $S \to AB$

$A \to aBb$

$B \to Sb$

$B \to b$

$$p = 4 \times 3 = 12$$

$$m = p + 1 = 13$$

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

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

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

$$S = v_1$$

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

$$|v_{i+1}| < |v_i| + f$$

maximum right hand side
of any production

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

$$m \leq |w| \leq k \cdot f \quad \Rightarrow \quad p < k \cdot f$$

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

$$p < k \cdot f$$

$$\Downarrow$$

$$k > \frac{p}{f}$$

Number of productions in grammar

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

$k >$ 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

$$S \to r_1$$
$$A \to r_2$$
$$B \to r_2$$
$$\cdots$$

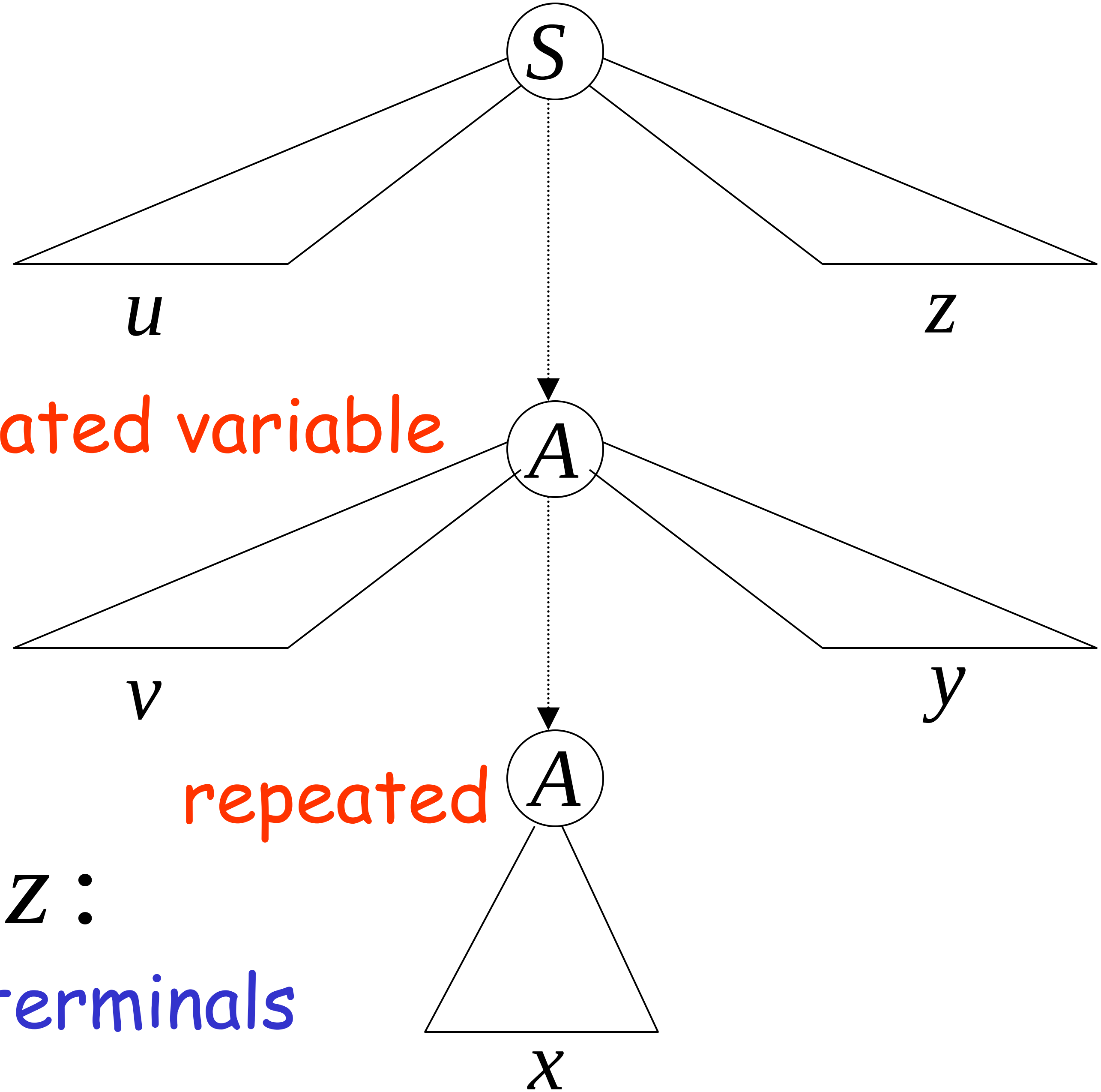
$w \in L(G)$ $\qquad$ $|w| \geq m$

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



$$w = uvxyz$$

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

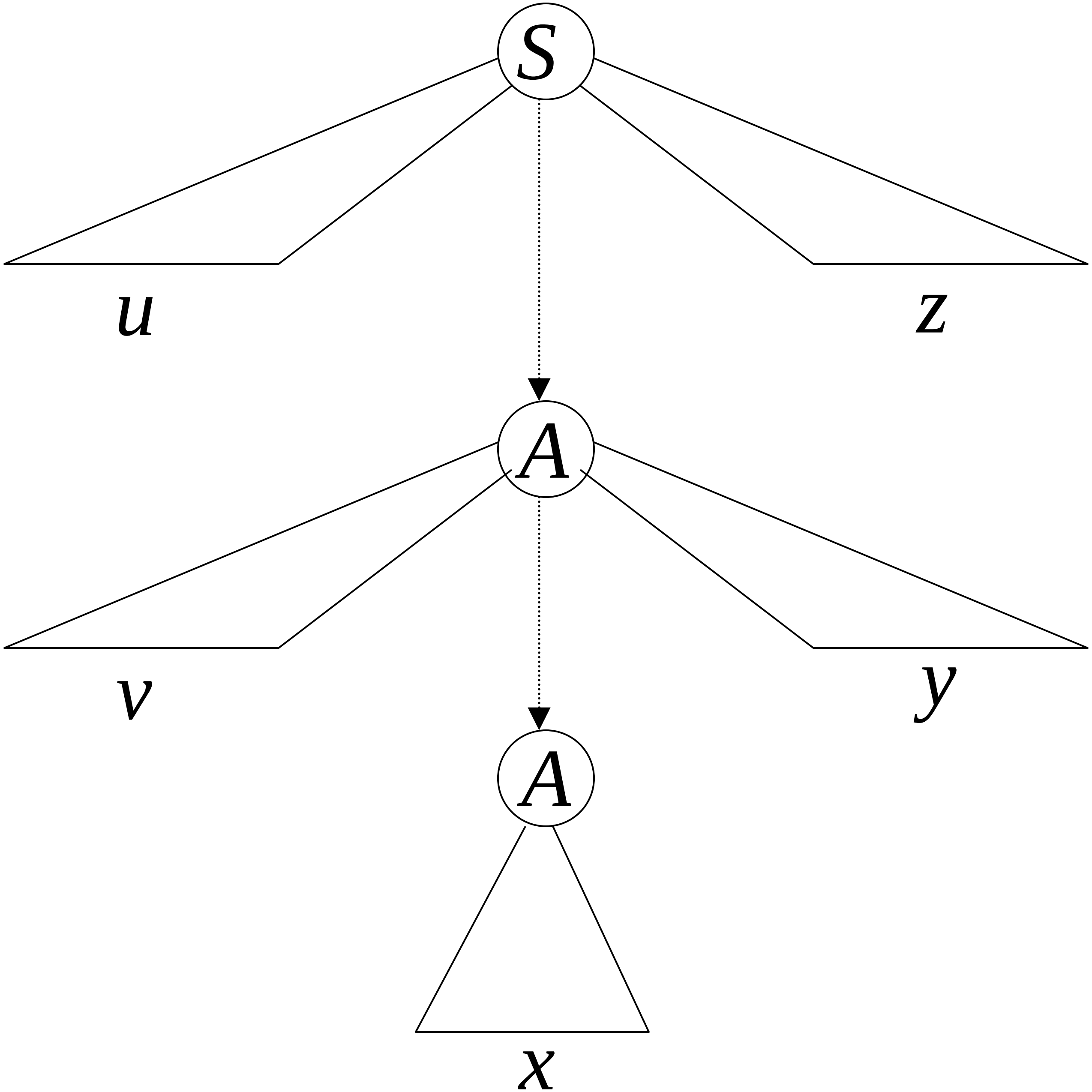
Strings of terminals

Possible derivations:

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

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

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



We know:

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

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:

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

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:

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

This string is also generated:

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

$$uv^i xy^i z$$

Therefore, any string of the form

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

is generated by the grammar G

Therefore,

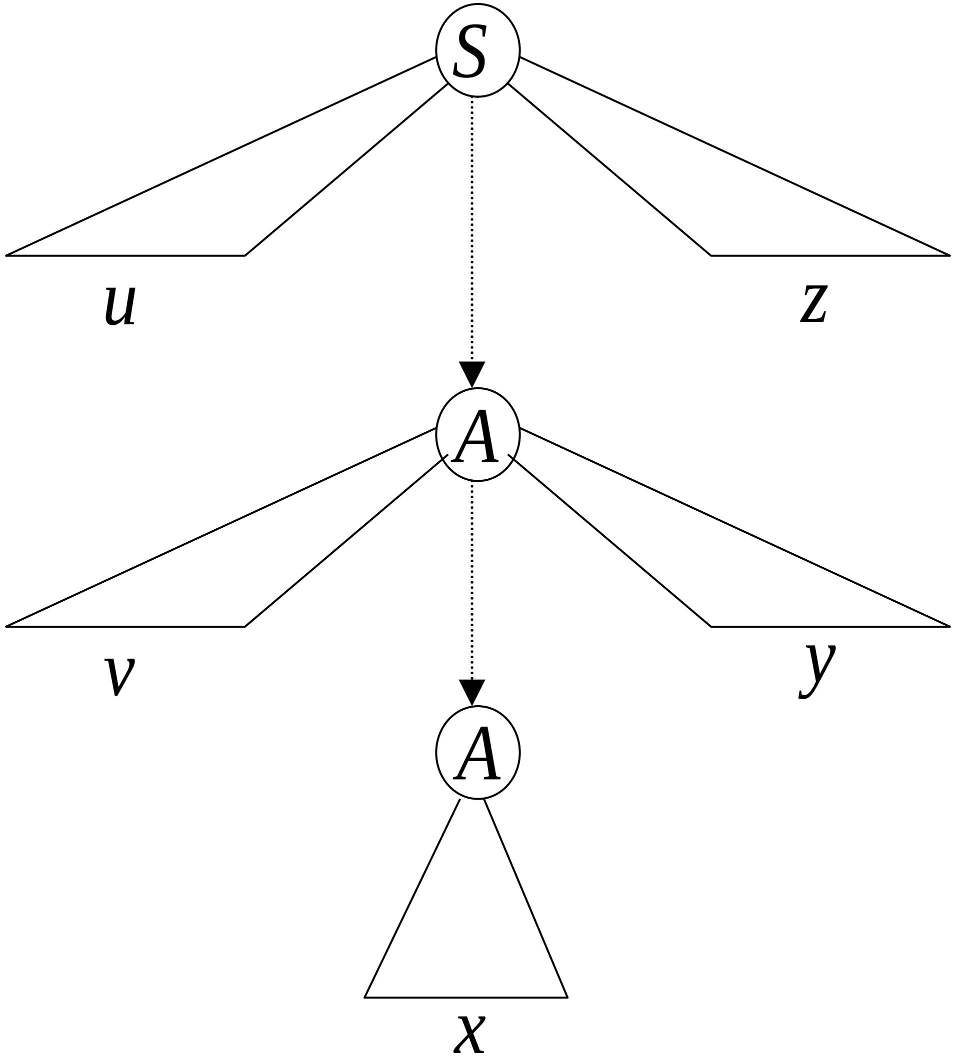
knowing that $uvxyz \in L(G)$

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

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

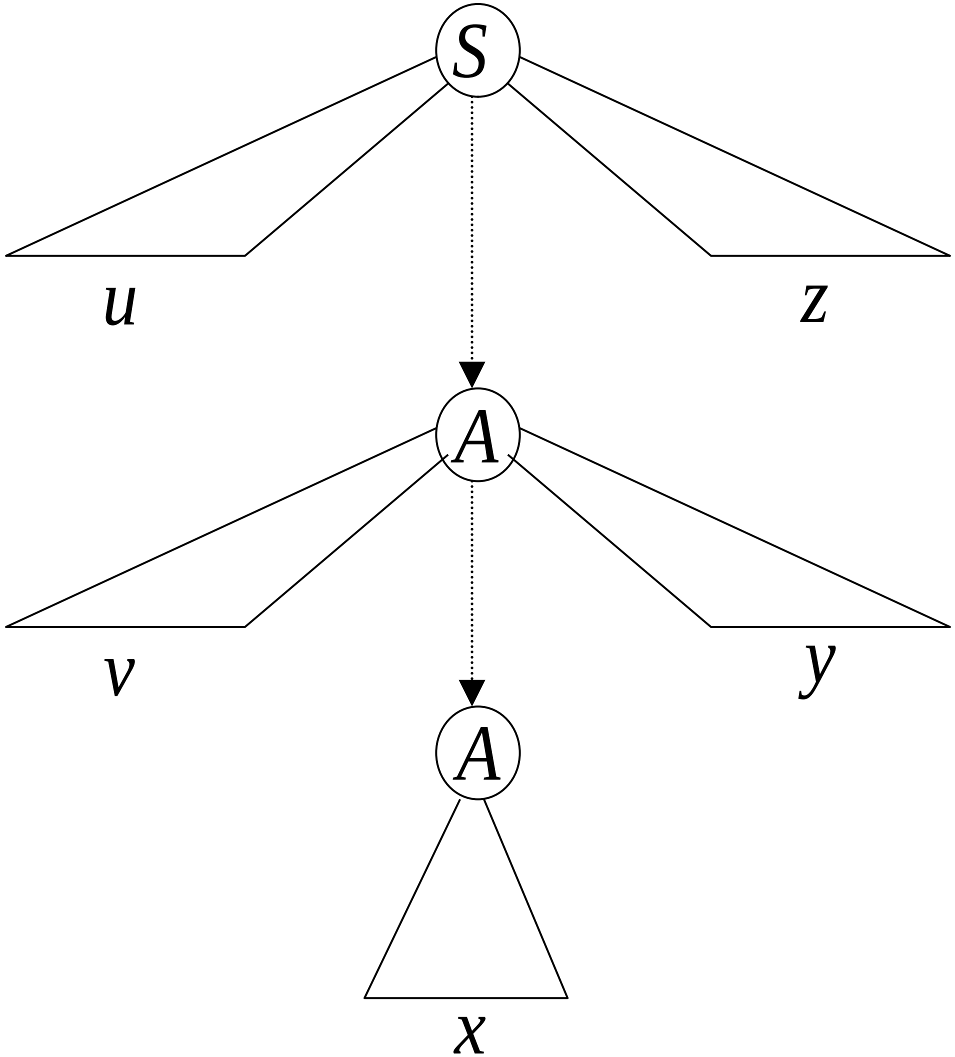
$$\Downarrow$$

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



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 $|vy| \geq 1$

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

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

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

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 within a^m

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

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

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

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

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

Case 1: v and y consist from only a

$$\underbrace{\overbrace{aa}^{}}_{u}\!\!\!\!\!\!\!\!\!\!\! \quad \overbrace{\underbrace{aa\ldots a}_{u\quad vxy}\ \underbrace{a}_{}}^{m}$$

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

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

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

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

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

Case 1: Repeating v and y

$$k \geq 1$$

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

$$\underbrace{a}_{u}\underbrace{aaaaa...aaaaa}_{v^2xy^2}\underbrace{a\,bbb...bbb\,ccc...ccc}_{z}$$

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

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

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

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

$$k \geq 1$$

$$\overbrace{\underbrace{a}_{u}\underbrace{aaaaa...aaaaa}_{v^2xy^2}a}^{m+k}\underbrace{\overbrace{bbb...bbb}^{m}\;\overbrace{ccc...ccc}^{m}}_{z}$$

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

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

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

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!!!

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

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

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

$\boxed{\textbf{Case 2:} \quad vxy \text{ is within } b^m}$

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

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

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

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

Case 2: Similar analysis with case 1

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

$$\underbrace{aaa\ldots aaa\,b}_{u}\,\underbrace{bb\ldots bb}_{vxy}\,\underbrace{b\,ccc\ldots ccc}_{z}$$

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

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

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

Case 3: vxy is within c^m

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

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

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

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

Case 3: Similar analysis with case 1

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

$$\underbrace{aaa...aaa\,bbb...bbb\,c}_{u}\,\underbrace{cc...c}_{vxy}\,\underbrace{cc}_{z}$$

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

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

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

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

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

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

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

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

Case 4: Possibility 1: v contains only a

y contains only b

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

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

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

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

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

Case 4: Possibility 1: v contains only a

y contains only b

$$k_1 + k_2 \geq 1$$

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

$$\underbrace{aaa}_{u}\underbrace{\ldots aaaaaaa\ bbbbbbb}_{v^2xy^2}\underbrace{\ldots bbb\ ccc\ldots ccc}_{z}$$

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

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

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

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

$$k_1 + k_2 \geq 1$$

$$\underbrace{aaa}_{u}\underbrace{\overbrace{\ldots aaaaaaa}^{m+k_1}\,\overbrace{bbbbbbb}^{m+k_2}}_{v^2xy^2}\underbrace{\overbrace{\ldots bbb}^{}\,\overbrace{ccc\ldots ccc}^{m}}_{z}$$

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

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

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

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

$$k_1 + k_2 \geq 1$$

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

Contradiction!!!

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

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

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

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

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

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

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

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

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

Case 4: Possibility 2: v contains a and b

y contains only b

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

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

$$\underbrace{aaa}_{u}\underbrace{\ldots aaaaabbaabb\ bbbbbbb}_{v^2xy^2}\underbrace{\ldots bbb\ ccc\ldots ccc}_{z}$$

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

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

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

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

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

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

$$\underbrace{aaa}_{u}\underbrace{\ldots aaaaabbaabb\ bbbbbbb}_{v^2xy^2}\underbrace{\ldots bbb\ ccc\ldots ccc}_{z}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Case 5: Similar analysis with case 4

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

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

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