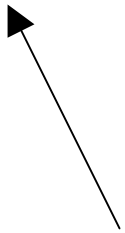
The Pumping Lemma for Context-Free Languages

Take an ***infinite*** context-free language



Generates an infinite number of different strings

Example:

$$S \to AB$$

$$A \to aBb$$

$$B \to Sb$$

$$B \to b$$

$$S \to AB$$

$$A \to aBb$$

$$B \to Sb$$

$$B \to b$$

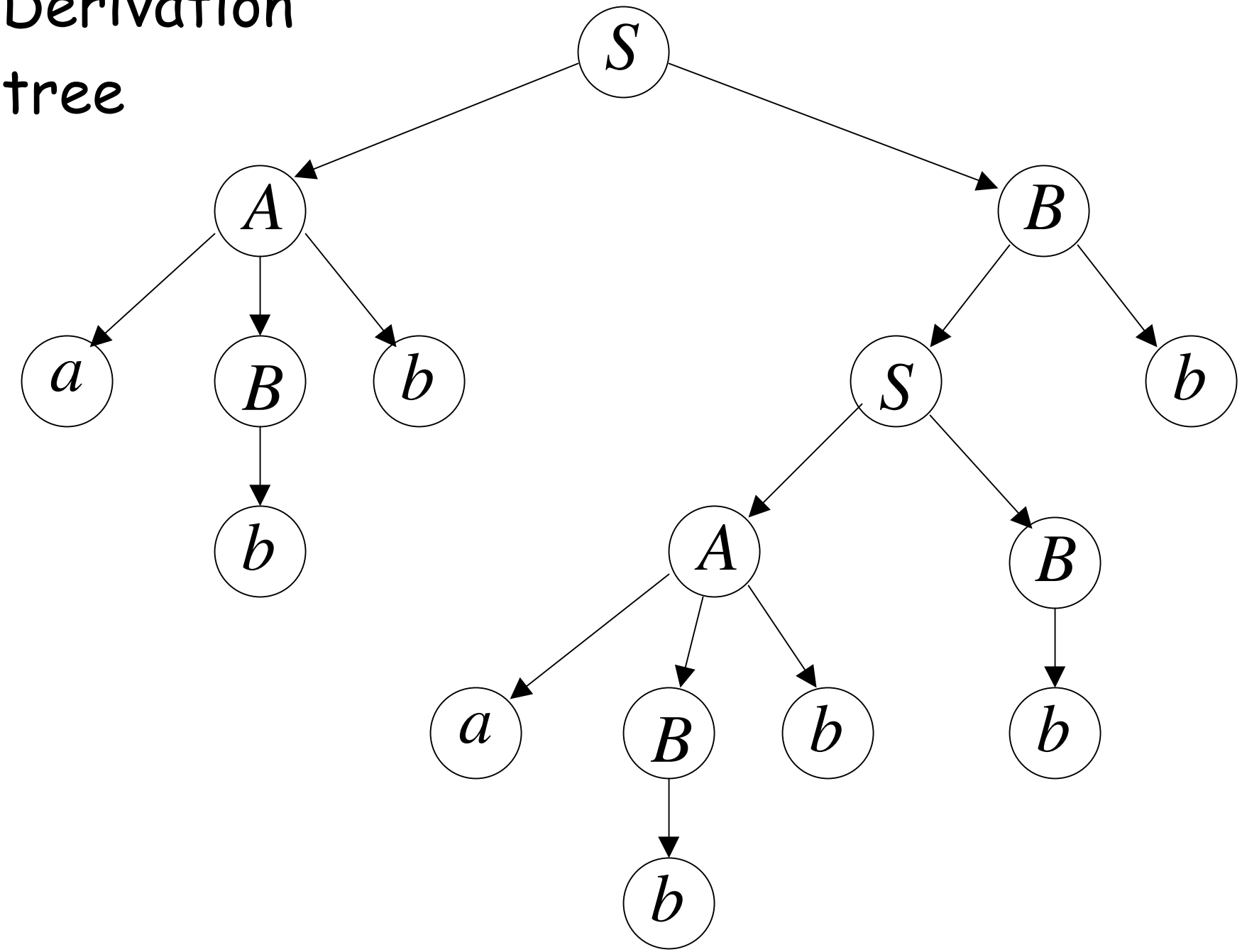
A derivation:

$$S \Rightarrow AB \Rightarrow aBbB \Rightarrow abbB \Rightarrow$$

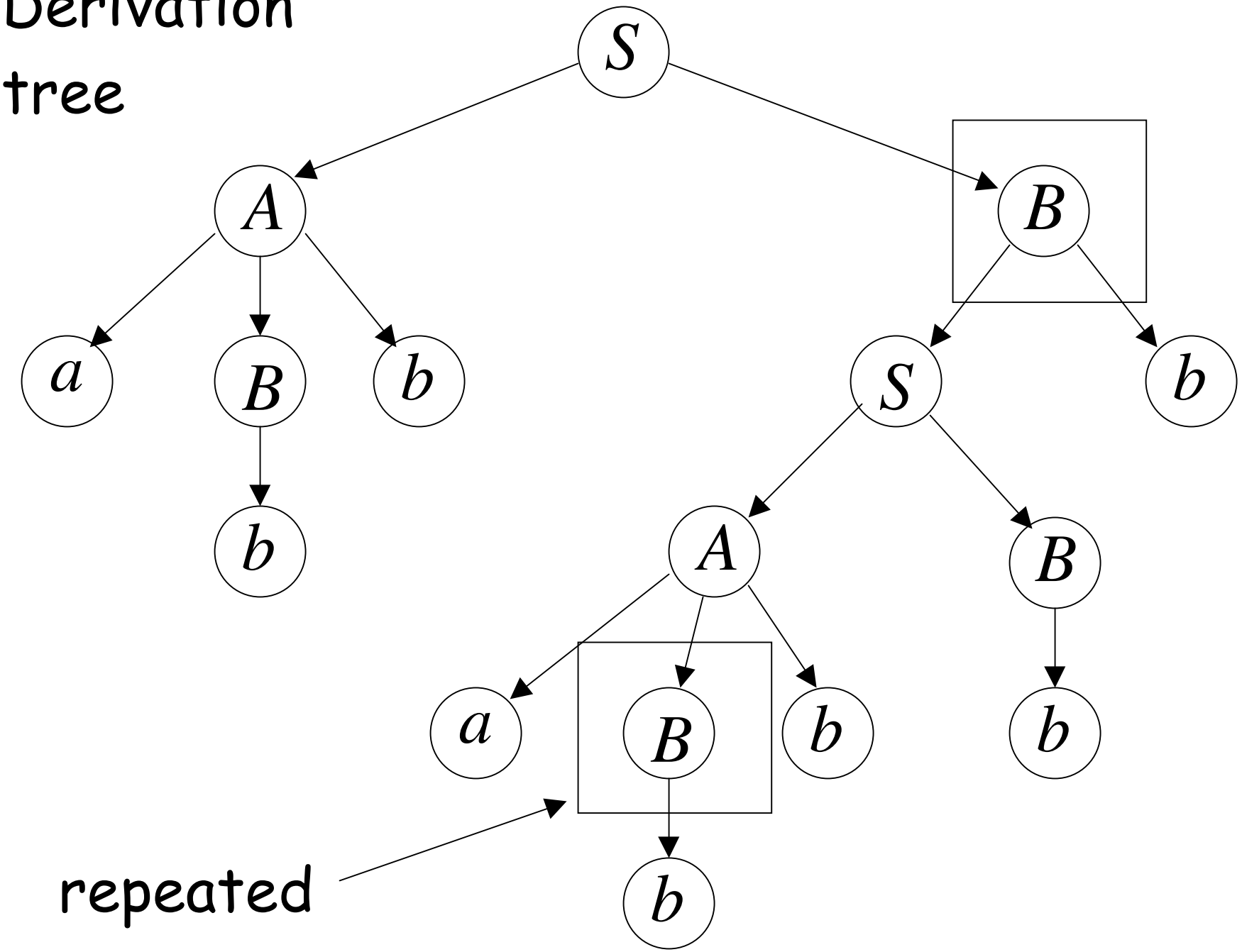
$$\Rightarrow abbSb \Rightarrow abbABb \Rightarrow abbaBbBb \Rightarrow$$

$$\Rightarrow abbabbBb \Rightarrow abbabbbb$$

Derivation tree

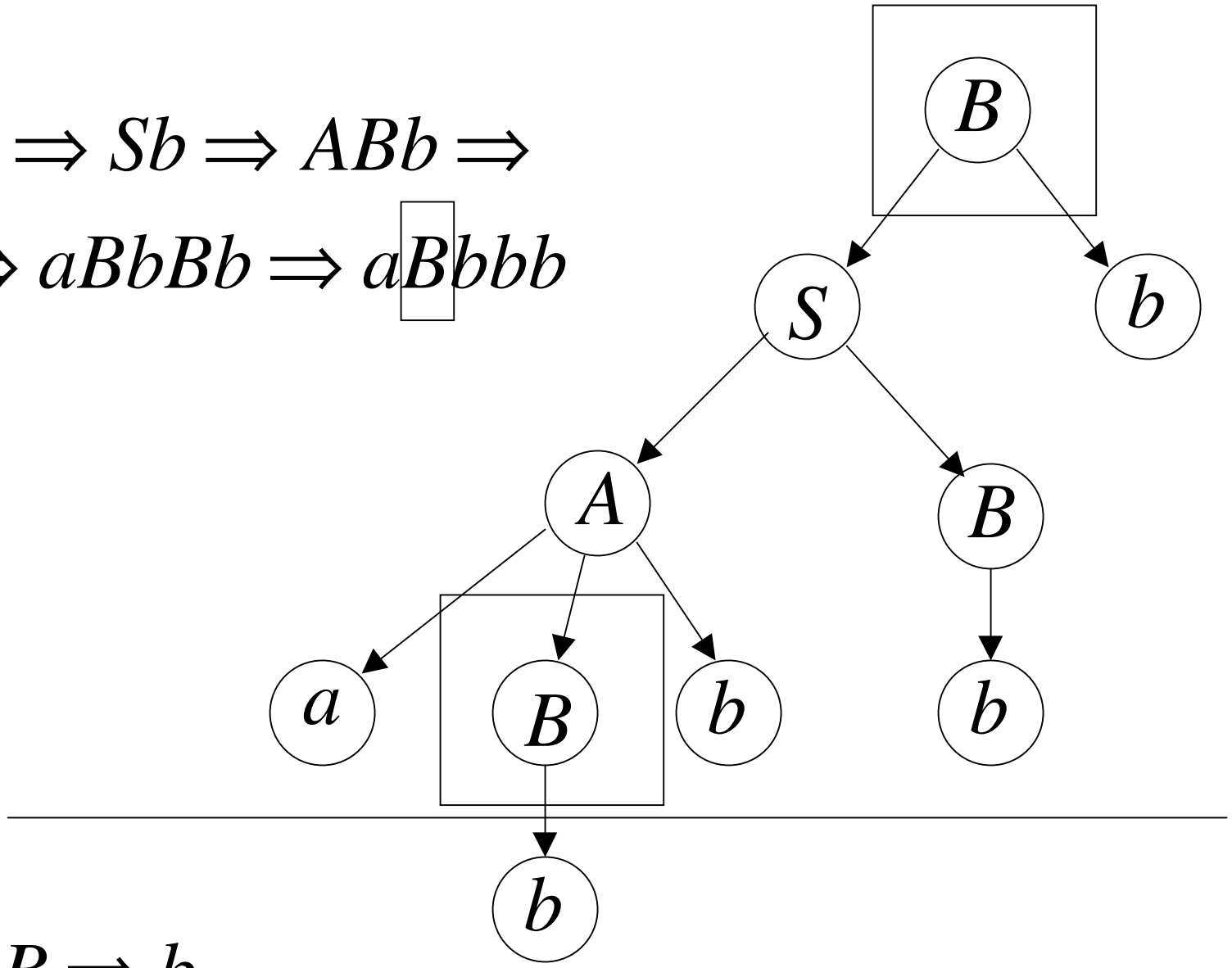


Derivation tree



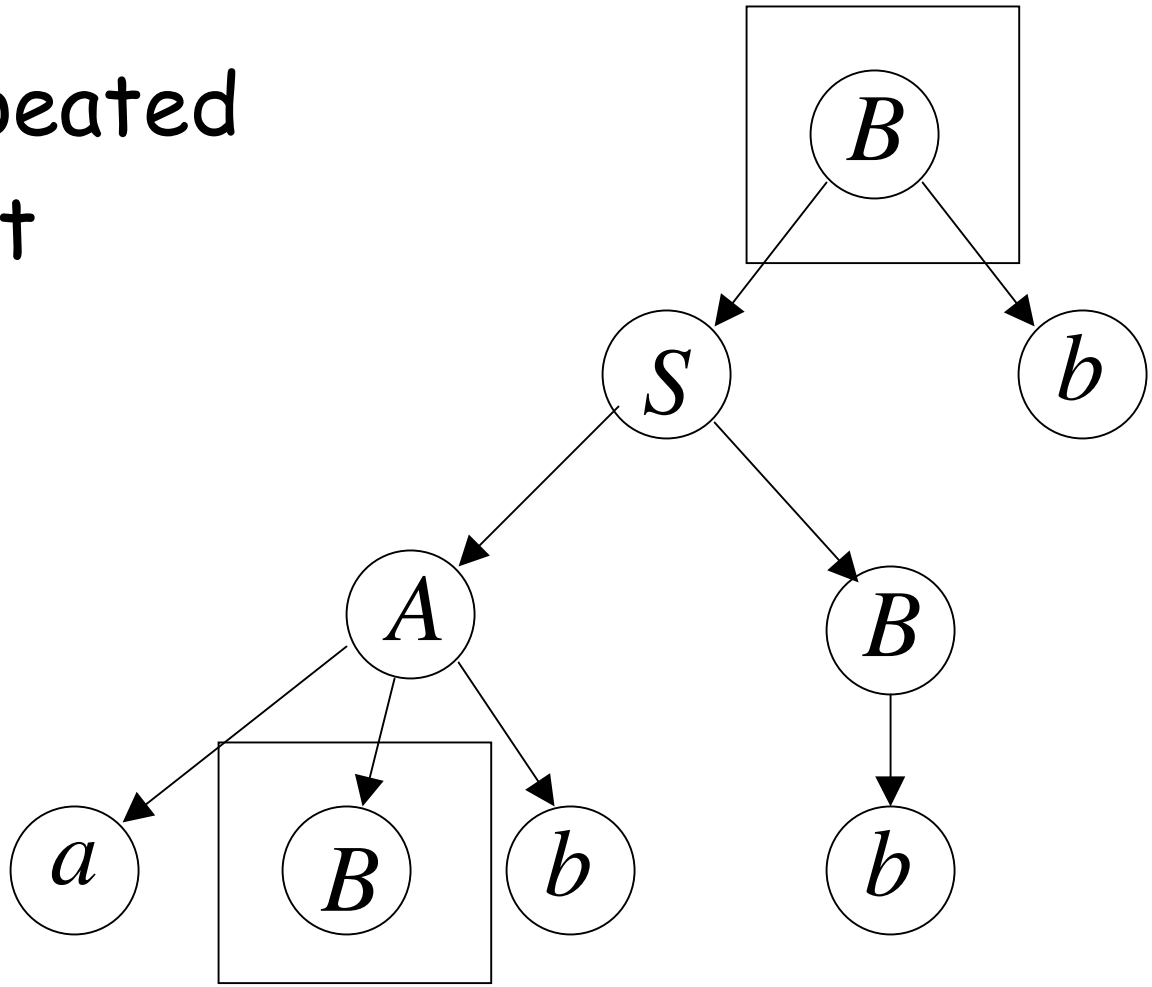
$B \Rightarrow Sb \Rightarrow ABb \Rightarrow$

$\Rightarrow aBbBb \Rightarrow aBbbb$



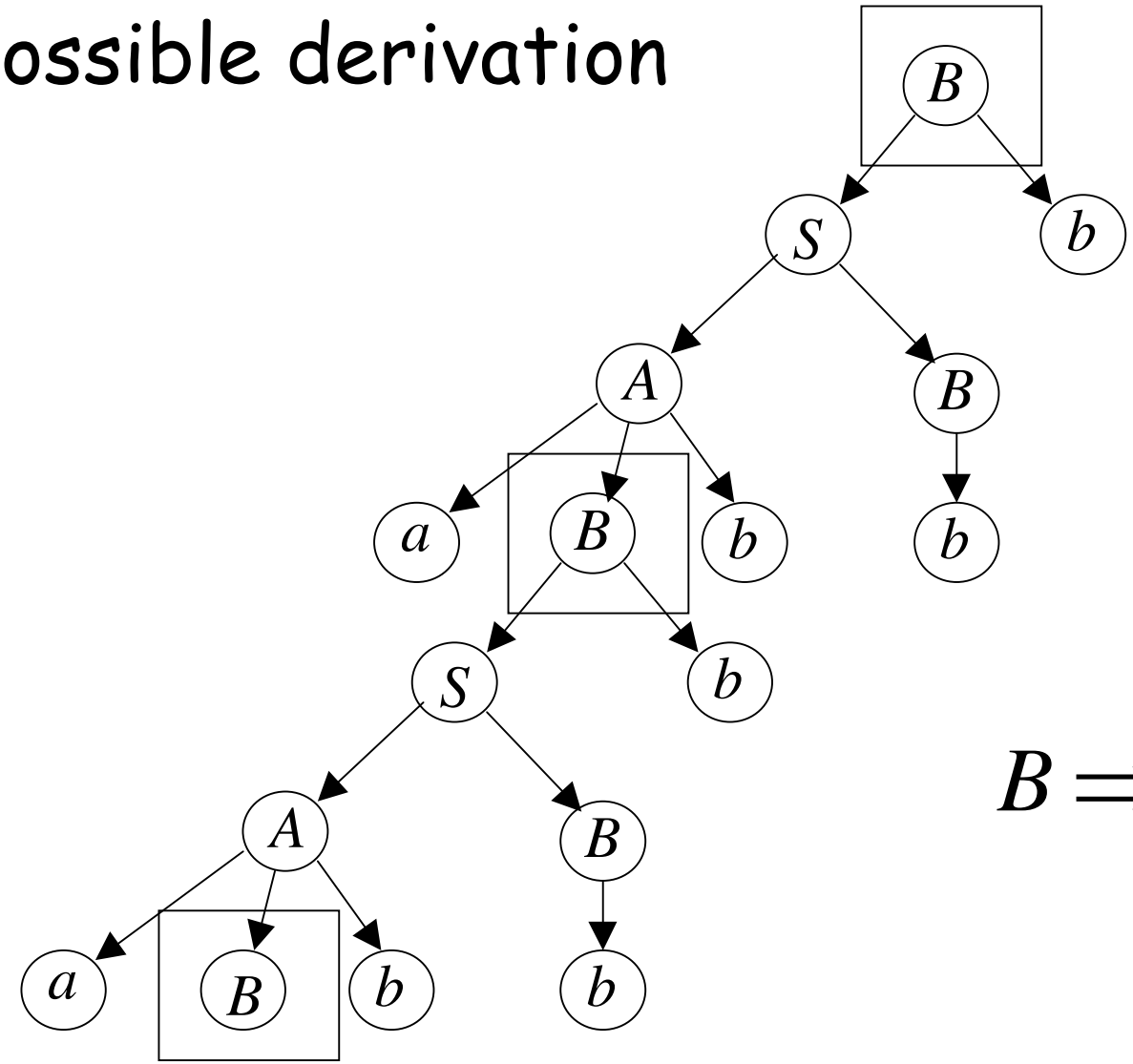
$B \Rightarrow b$

Repeated part



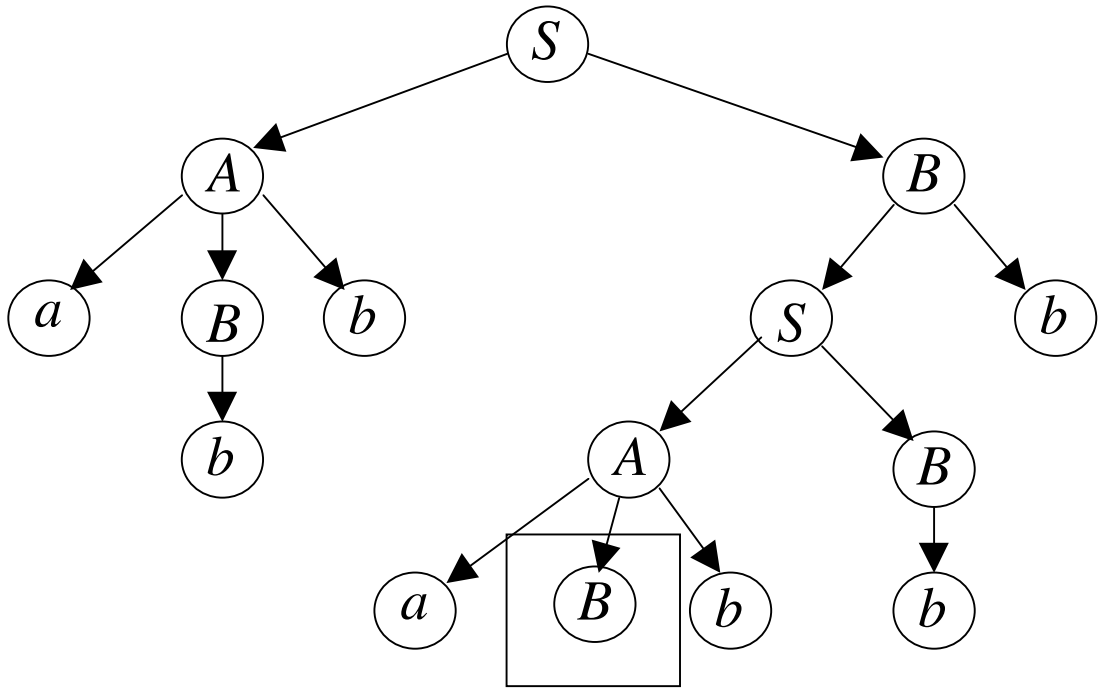
$$\boxed{B} \Rightarrow \ldots \Rightarrow a\boxed{B}bbb$$

A possible derivation



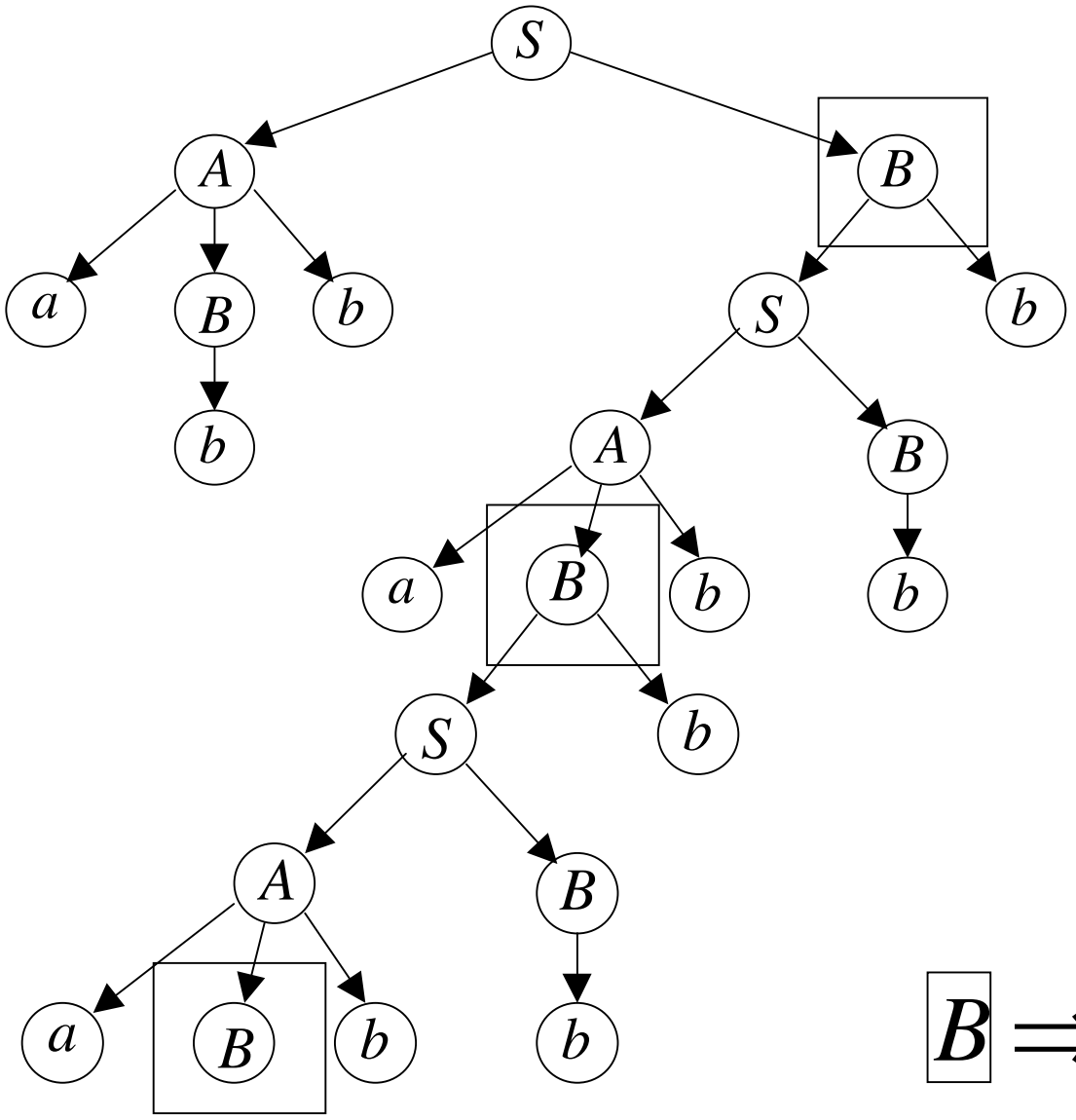
$$B \Rightarrow \ldots \Rightarrow aBbbb$$

$$B \Rightarrow \ldots \Rightarrow aBbbb \ldots \Rightarrow aaBbbbbbb$$



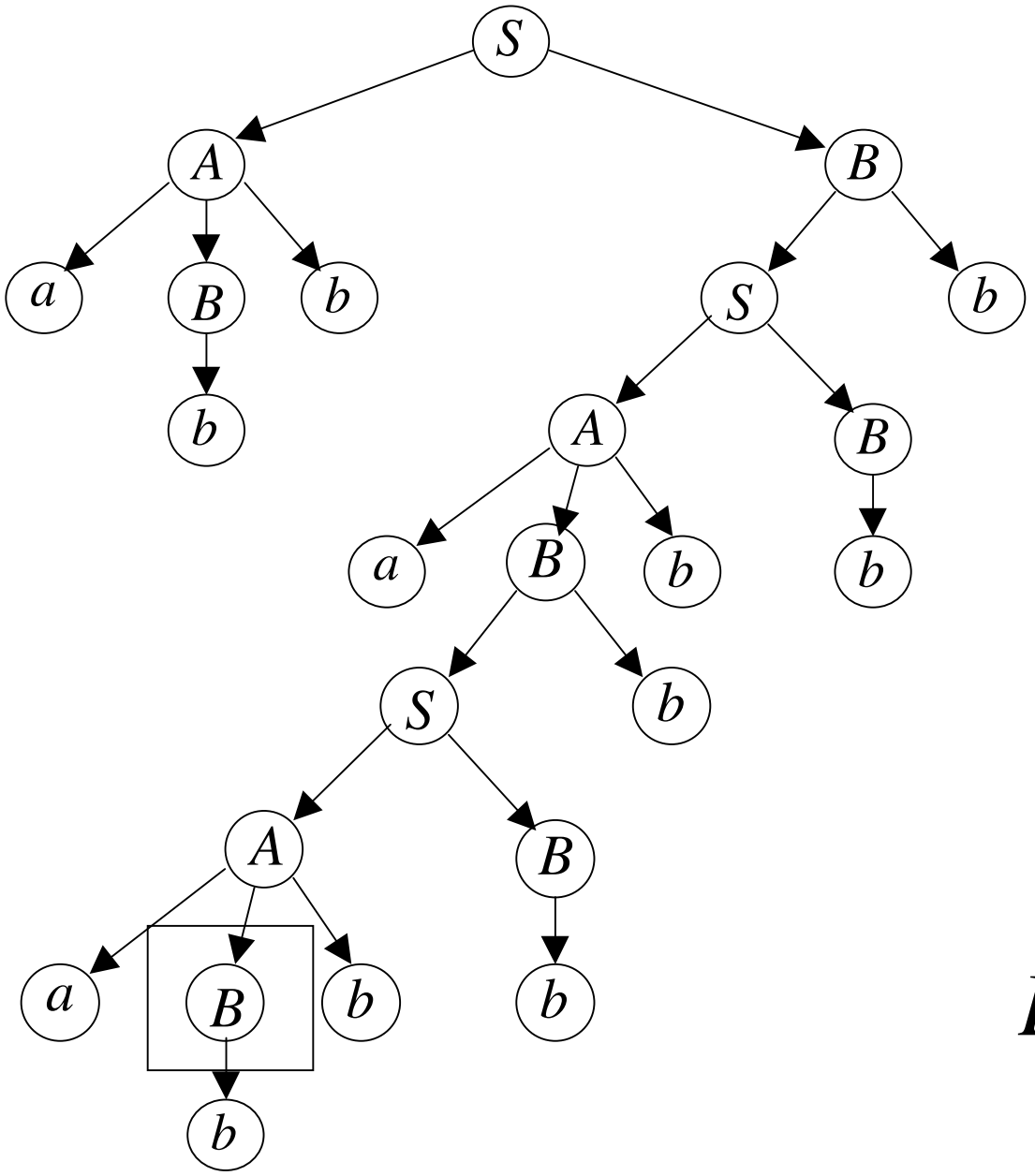
$$S \Rightarrow \ldots \Rightarrow abbaBbbb$$

$$B \Rightarrow \ldots \Rightarrow aBbbb$$



$$B \Rightarrow \ldots \Rightarrow aBbbb$$

$$S \Rightarrow \ldots \Rightarrow abbaBbbb \Rightarrow \ldots \Rightarrow abbaaBbbbbbb$$



$B \Rightarrow b$

$S \Rightarrow \ldots \Rightarrow abbaaBbbbbbb \Rightarrow abbaabbbbbbb$

$$S \Rightarrow \ldots \Rightarrow abbaabbbbbbb$$

Therefore, the string

$$abbaabbbbbbb$$

is generated by the grammar

We know $B \Rightarrow b$

$$B \Rightarrow \ldots \Rightarrow aBbbb$$

$$S \Rightarrow \ldots \Rightarrow abbaBbbb$$

This string is also generated:

$$S \Rightarrow \ldots \Rightarrow abbaBbbb \Rightarrow$$

$$\Rightarrow abba(a)B(bbb)bbb$$

$$\Rightarrow abba(a)^2 B(bbb)^2 bbb$$

$$\Rightarrow abba(a)^2 b(bbb)^2 bbb$$

We know $B \Rightarrow b$

$$B \Rightarrow \ldots \Rightarrow aBbbb$$

$$S \Rightarrow \ldots \Rightarrow abbaBbbb$$

This string is also generated:

$$S \Rightarrow \ldots \Rightarrow abbaBbbb \Rightarrow$$

$$\Rightarrow \ldots$$

$$\Rightarrow abba(a)^i B(bbb)^i bbb$$

$$\Rightarrow abba(a)^i b(bbb)^i bbb$$

Therefore, knowing that

$$abbabbbb$$

is generated, we also know that

$$abba(a)^i b(bbb)^i bbb$$

is generated

In general:

We are given an infinite context-free grammar G

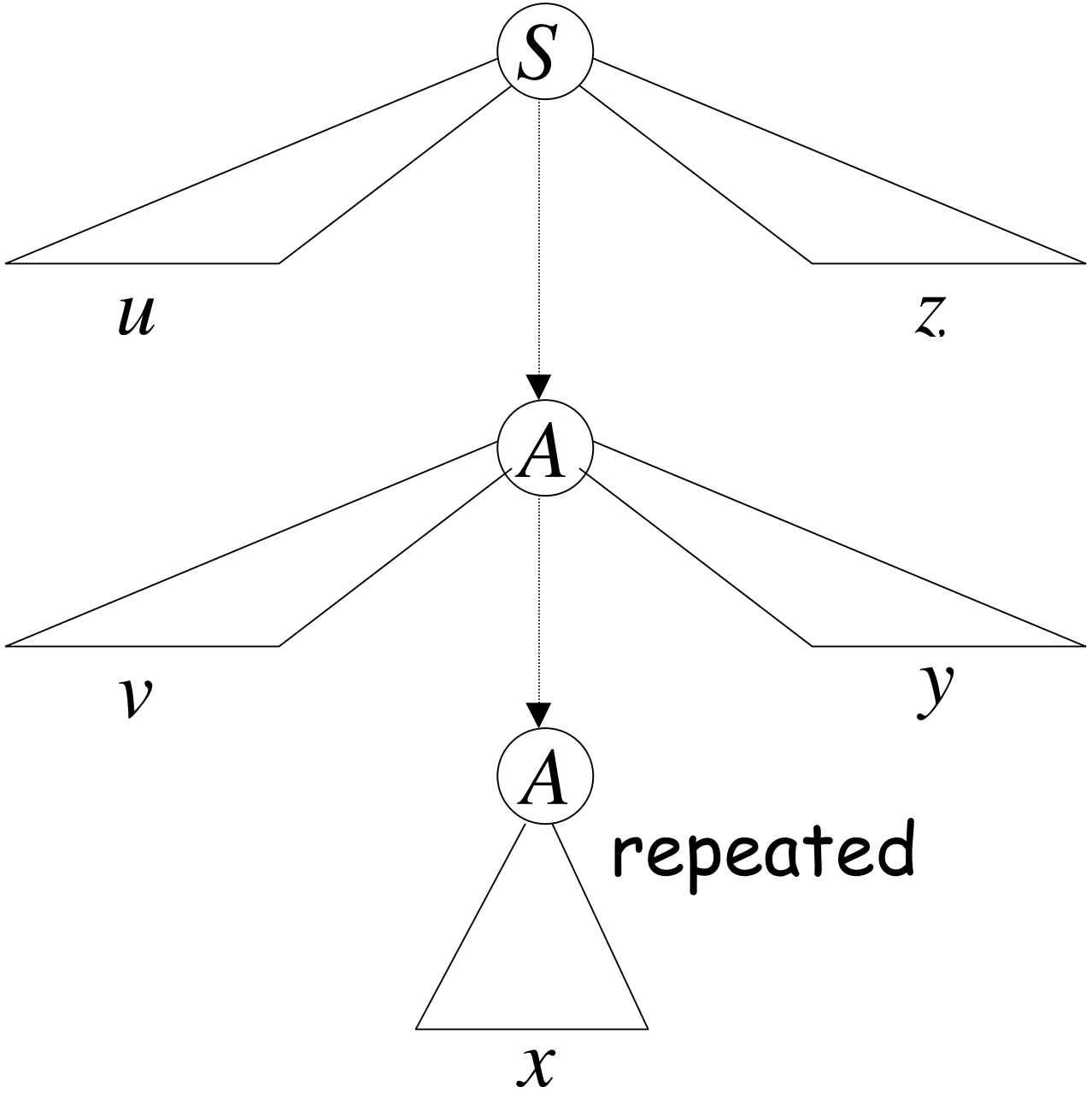
We take the derivation of a long enough string w

Take the length of w Bigger than

m = Productions * (largest production)

Some variable must be repeated
in the derivation

u, v, x, y, z: strings of terminals



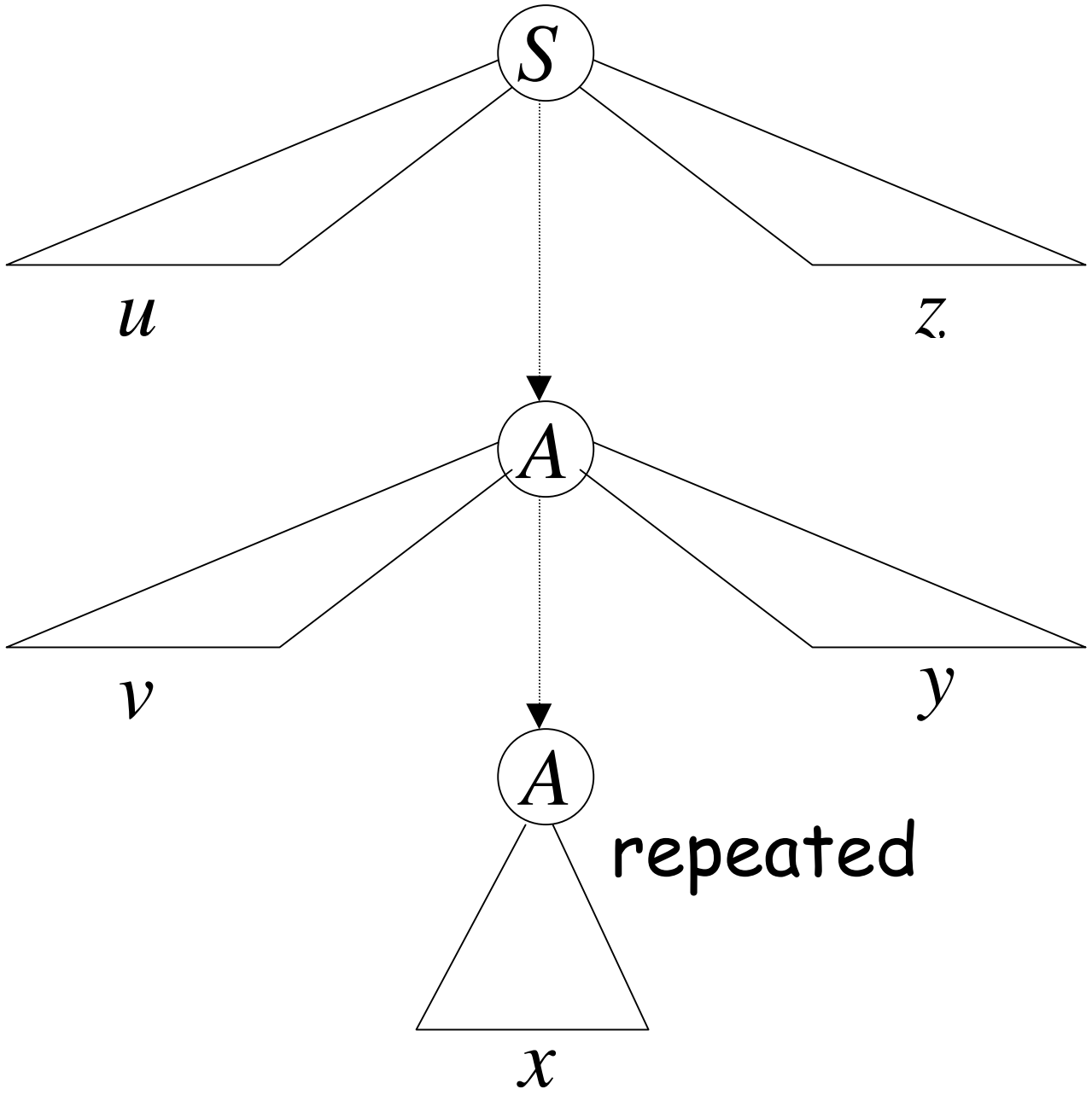
$w = uvxyz$

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:

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

$$uv^0xy^0z$$

We know:

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

This string is also generated:

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

(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:

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

$$uv^2xy^2z$$

We know:

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

This string is also generated:

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

$$\Rightarrow uvvvAyyyz \Rightarrow uvvvxyyyz$$

$$uv^3xy^3z$$

We know:

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

This string is also generated:

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

$$\Rightarrow uvvvAyyyz \Rightarrow \ldots$$

$$\Rightarrow uvvv \cdots vAy \cdots yyyz \Rightarrow$$

$$\Rightarrow uvvv \cdots vxy \cdots yyyz$$

$$uv^i xy^i z$$

Therefore, any string of the form

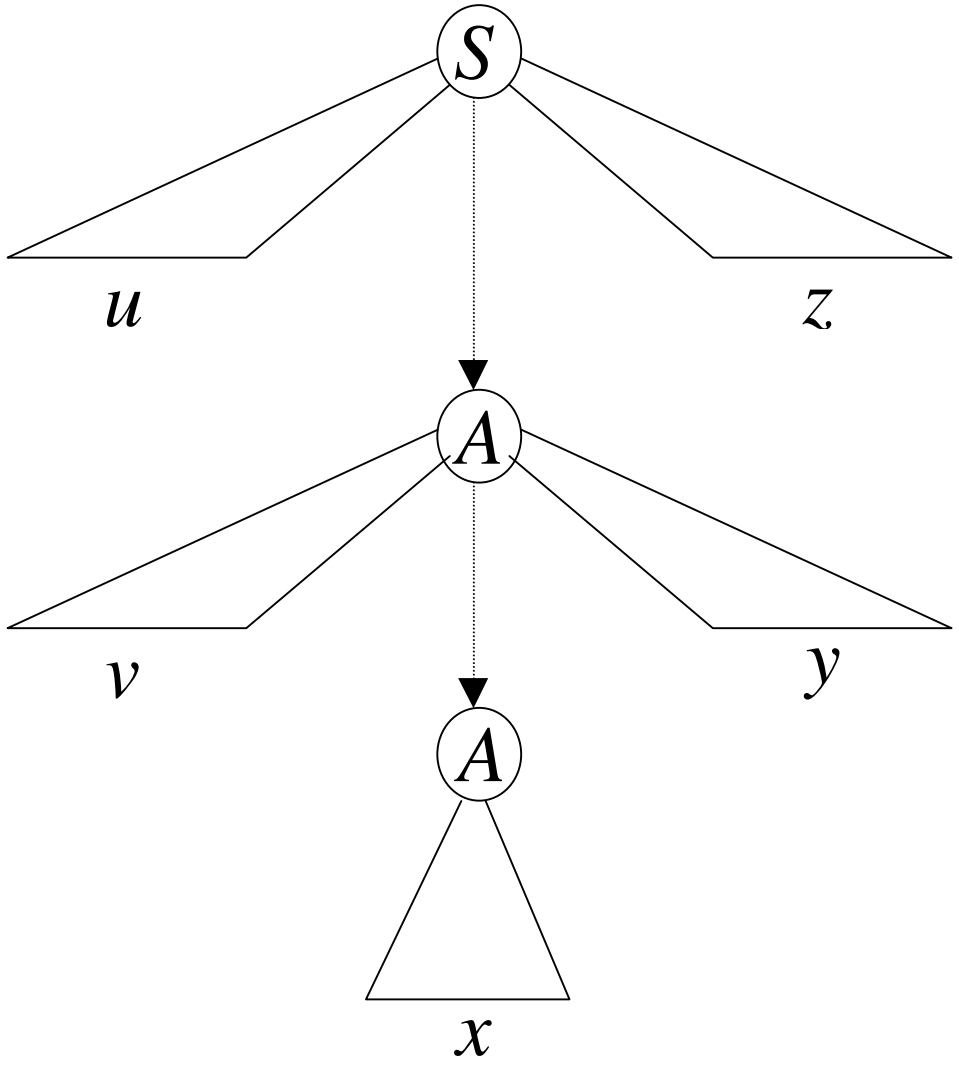
$$uv^i xy^i z \qquad\qquad i \geq 0$$

Is generated by the grammar G

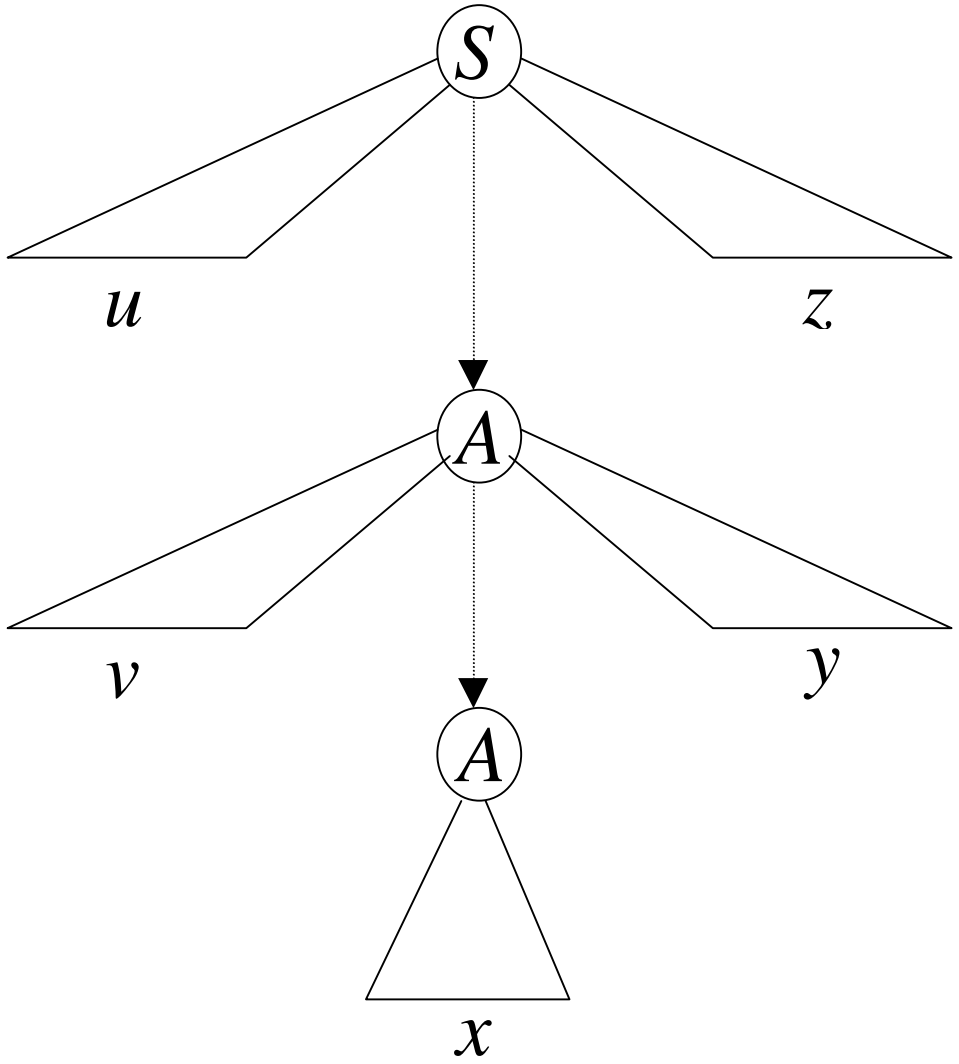
Therefore,

knowing that $uvxyz \in L(G)$

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



$$|vxy| \leq m$$



$$|vy| \geq 1$$

The pumping lemma:

For context-free language L

there exists an integer m

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

we can write $w = uvxyz$

with $|vxy| \leq m, \quad |vy| \geq 1$

Such that: $uv^i xy^i z \in L$