Intersection
of
Context-free Languages
and
Regular Languages

Intersection with Regular Languages

- Context-free languages are closed under **intersection with regular languages:**

  If L_1 is a context-free language and L_2 is a regular language, then the **intersection** of the languages L_1 and L_2:

$$L_1 \cap L_2$$

  is also context-free.

Intersection: Example

Machine M_1

NDPA for L_1
(context-free)

Machine M_2

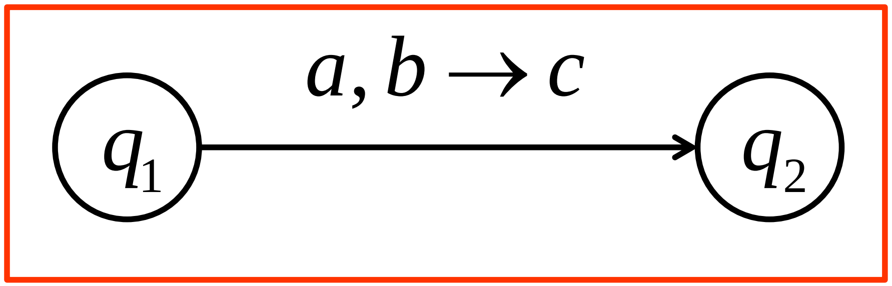
DFA for L_2
(regular language)

- Construct a new NPDA machine M that accepts $L_1 \cap L_2$.

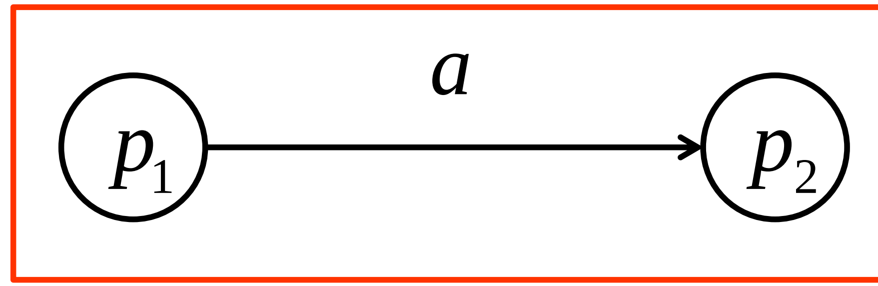
- M simulates in parallel M_1 and M_2.

Transition

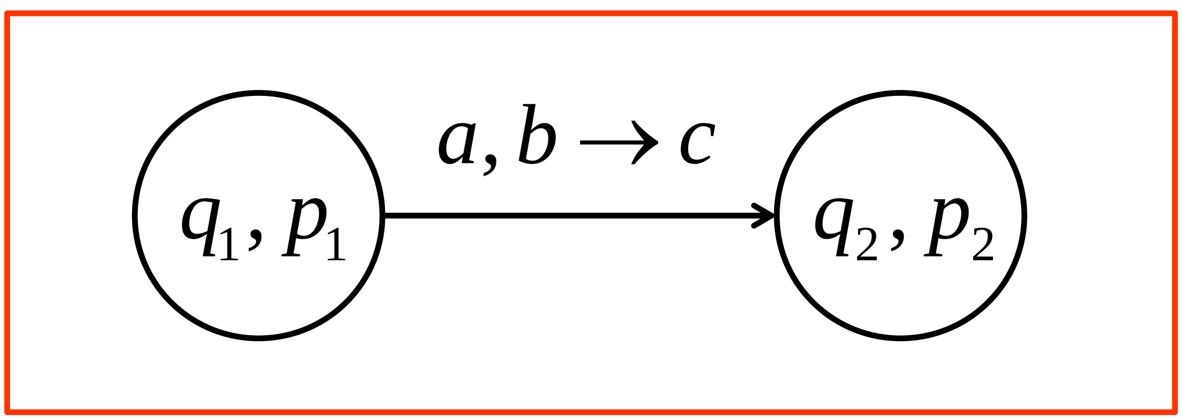
NDPA M_1



DFA M_2

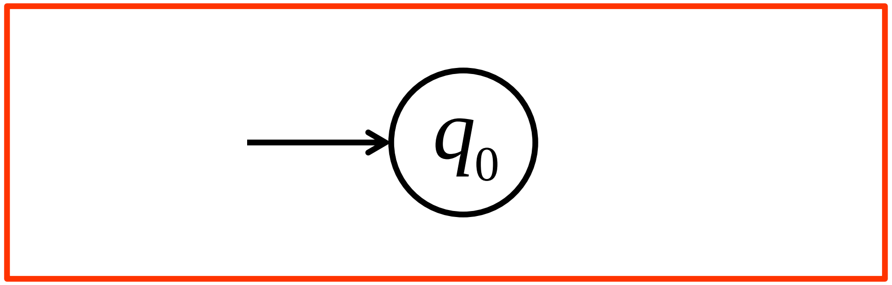


NDPA M_1

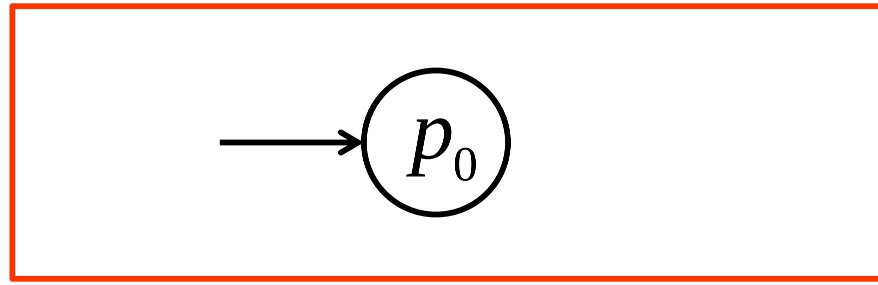


Initial State

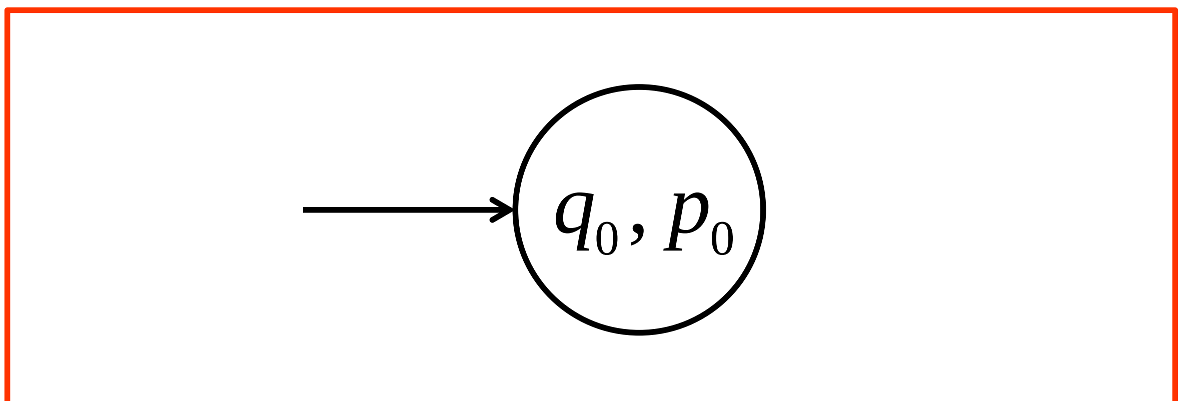
NDPA M_1



DFA M_2

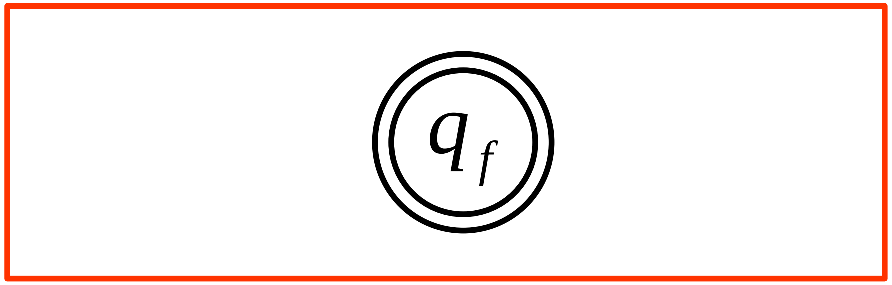


NDPA M_1

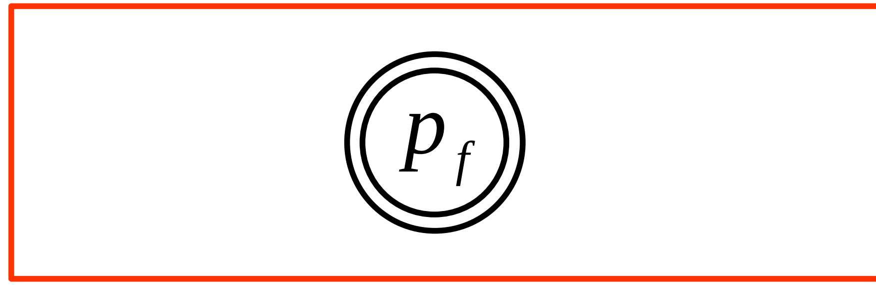


Final State

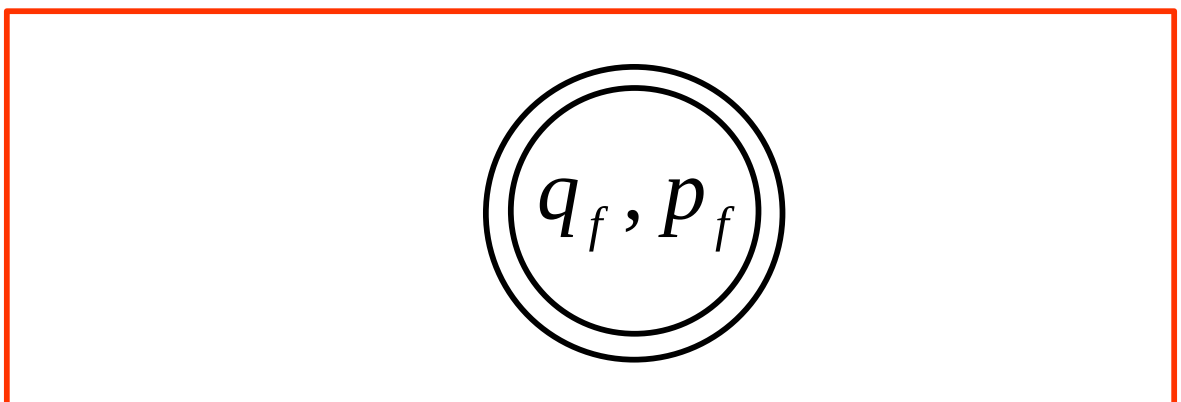
NDPA M_1



DFA M_2



NDPA M_1



Intersection with Regular Languages

- The NDPA M simulates in parallel the NDPA M_1 and the DFA M_2.

- The NDPA M accepts a string w if and only if
  - M_1 accepts the string w and
  - M_2 accepts the string w

- Therefore,

$$L_1 \cap L_2 = L(M) = L(M_1) \cap L(M_2)$$

is context-free.

Exercises

- Construct context-free grammars for the following languages:

1. $\{a^n b^{2n} c^m : n, m \geq 0\} \cup \{a^n b^m c^{2m} : n, m \geq 0\}$
2. $\{a^n b^{2n} c^m : n, m \geq 0\} \cdot \{a^n b^m c^{2m} : n, m \geq 0\}$
3. $\{a^n b^{2n} c^m : n, m \geq 0\}^*$
4. $\{a^n b^m c^{2m} : n, m \geq 0\}^*$

- Is $\{a^n b^{2n} c^m : n, m \geq 0\} \cap \{a^n b^m c^{2m} : n, m \geq 0\}$ a context-free language? Justify your answer.

Applications
of
Regular Closure

Intersection with Regular Languages

- Context-free languages are closed under **intersection with regular languages:**

  If L_1 is a context-free language and L_2 is a regular language, then the **intersection** of the languages L_1 and L_2:

  $$L_1 \cap L_2$$

  is also context-free.

An Application of Regular Closure

- Prove that

$$L = \{a^n b^n : n \neq 100\}$$

is context-free.

An Application of Regular Closure

- We know that

$$L_1 = \{a^n b^n : n \geq 0\}$$

is context-free.

An Application of Regular Closure

- We also know that

$$L_2 = \{a^{100}b^{100}\}$$

is regular (a finite language).

Then

$$\overline{L_2} = \{a, b\}^* - \{a^{100}b^{100}\}$$

is also regular (the complement).

An Application of Regular Closure

- Therefore,

$$L_1 \cap \overline{L_2}$$
$$= \{a^n b^n : n \geq 0\} \cap (\{a, b\}^* - \{a^{100} b^{100}\})$$
$$= \{a^n b^n : n \neq 100\} = L$$

is context-free (the intersection of a context-free language with a regular language).

Another Application of Regular Closure

- Prove that

  $$L = \{w \in \{a, b, c\}^* : n_a(w) = n_b(w) = n_c(w)\}$$

  is not context-free.

Another Application of Regular Closure

- Suppose that

  $$L = \{w \in \{a, b, c\}^*: n_a(w) = n_b(w) = n_c(w)\}$$

  is context-free.

  Then, according to the *regular closure*,

  $$L \cap (a^* b^* c^*) = \{a^n b^n c^n : n \geq 0\}$$

  is also context-free, which is the contradiction!

  Therefore, L is not context-free.

Decidable Properties
of
Context-Free Languages

Membership Problem

Membership Question:

for context-free grammar G, find if string

$$w \in L(G)$$

Membership Algorithms:

- Exhaustive search parser
- **CYK** parsing algorithm

Emptiness Problem

Emptiness Question:

for context-free grammar G, find if language

$$L(G) = \emptyset$$

Algorithm:

1. Remove useless variables
2. Check if the start symbol S is useless

Infiniteness Problem

Infiniteness Question:

for context-free grammar G, find if language

$L(G)$ is infinite

Algorithm:

1. Remove useless variables
2. Remove unit and λ-productions
3. Create dependency graph for variables
4. If there is a loop in the dependency graph then the language is infinite

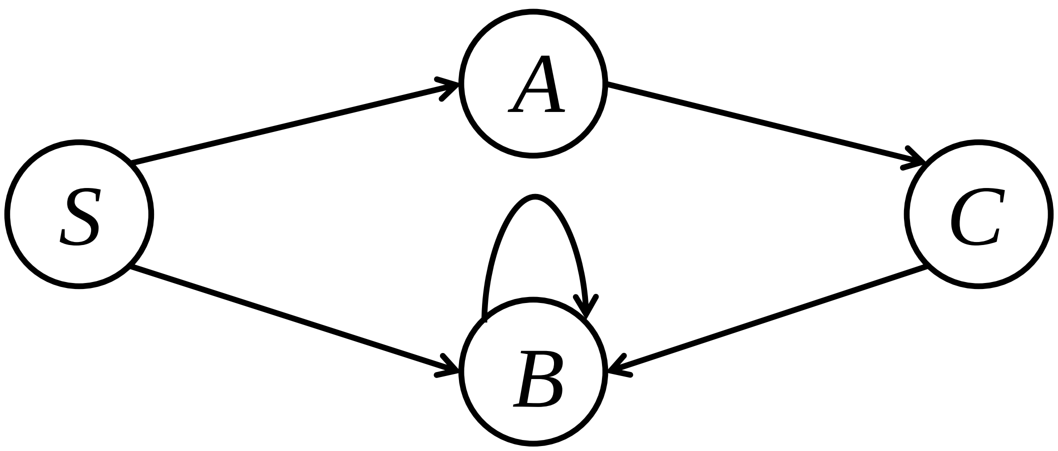
Infiniteness Problem

Example:

$$S \rightarrow AB$$
$$A \rightarrow aCb|a$$
$$B \rightarrow bB|bb$$
$$C \rightarrow cBS$$

Dependency graph:

Infinite language



Infiniteness Problem

Example:

$$S \rightarrow AB$$
$$A \rightarrow aCb|a$$
$$B \rightarrow bB|bb$$
$$C \rightarrow cBS$$

$$S \Rightarrow AB \Rightarrow aCbB \Rightarrow acBSbB \Rightarrow acbbSbbb$$

$$S \Rightarrow^* acbbSbbb \Rightarrow^* (acbb)^2S(bbb)^2$$
$$\Rightarrow^* (acbb)^iS(bbb)^i$$

The Pumping Lemma for Context-Free Languages

Pumping Lemma

- Consider an **infinite** context-free language (generates an infinite number of different strings)

- Example:

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

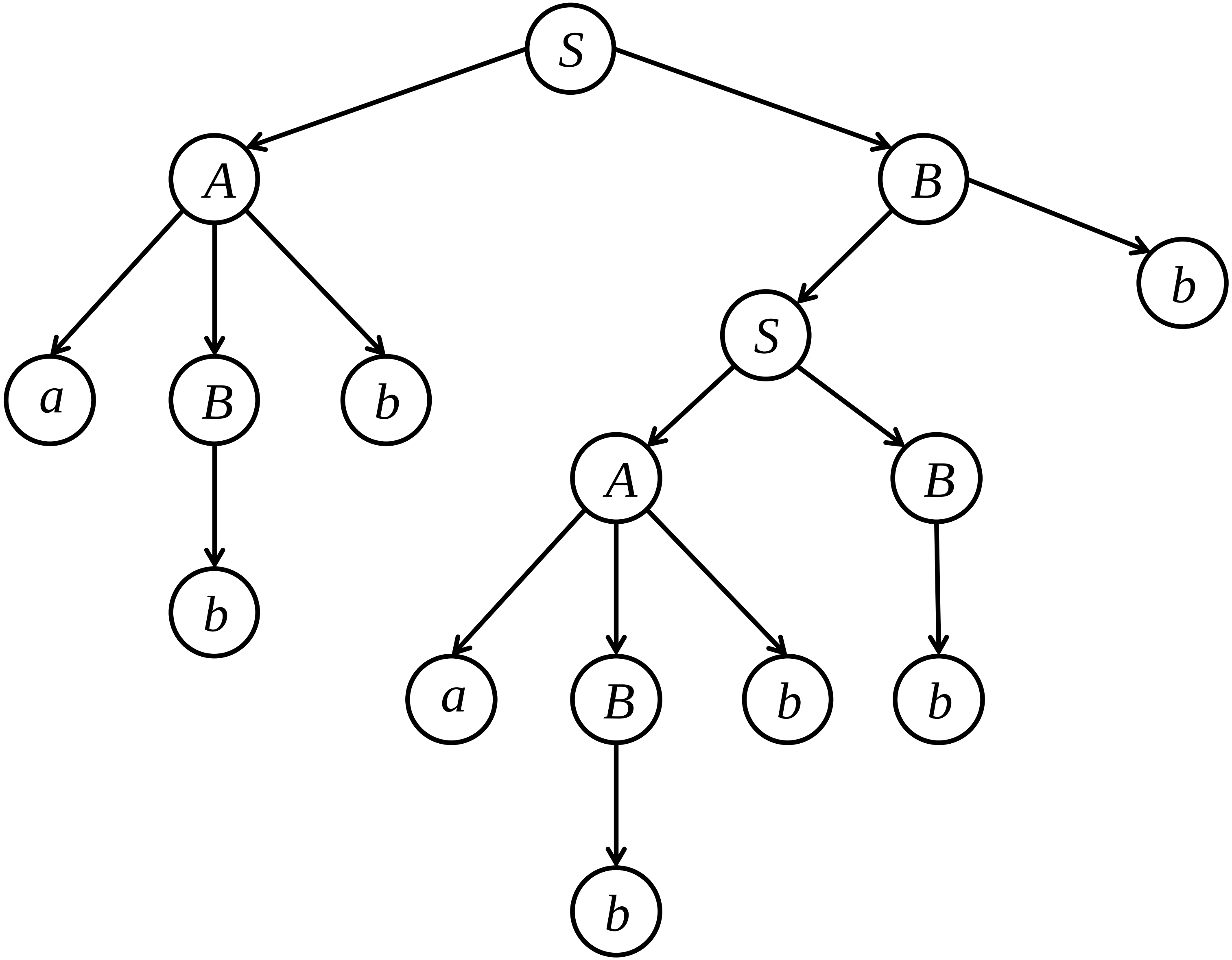
Pumping Lemma

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

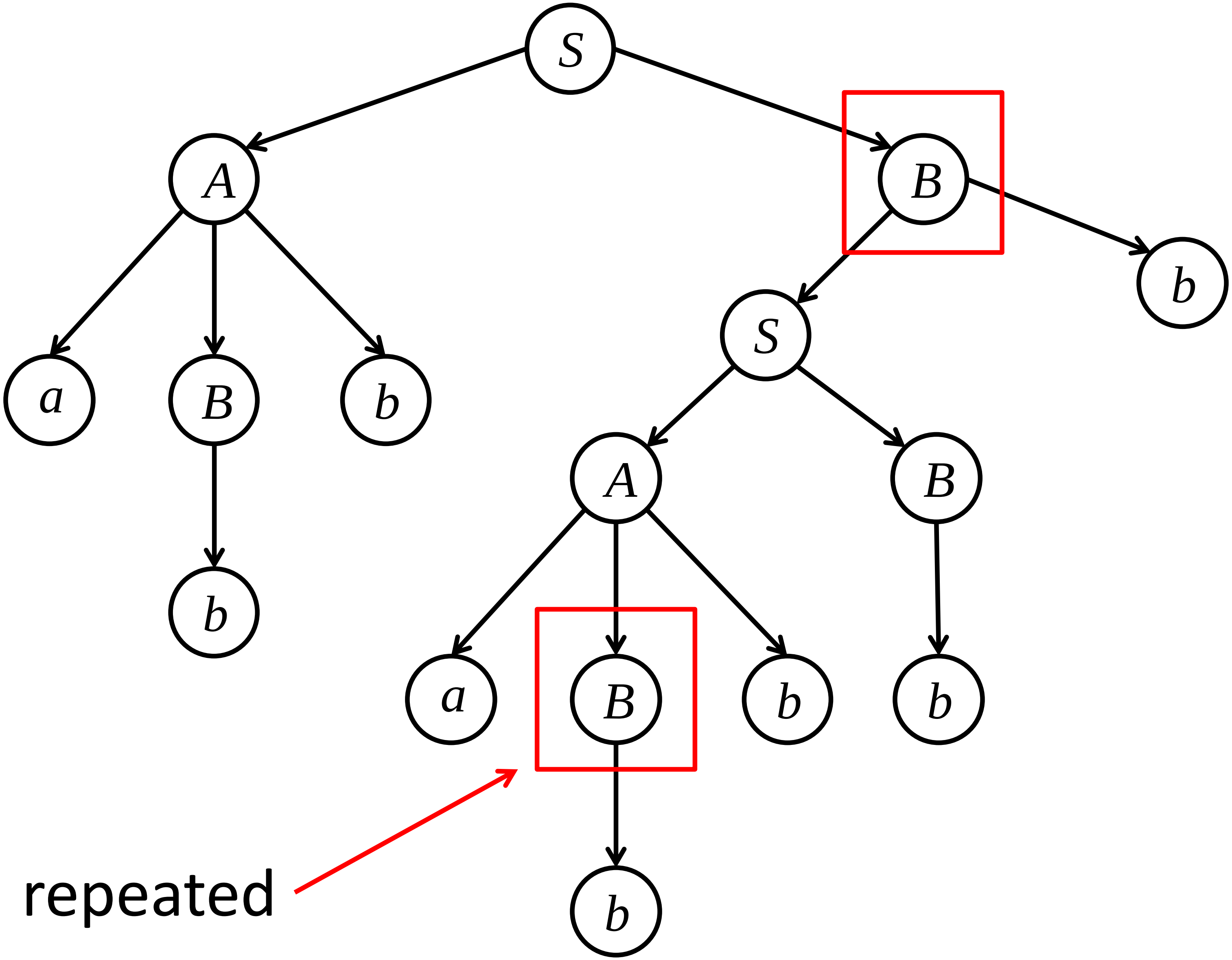
A derivation: Variables are repeated

$$S \Rightarrow AB \Rightarrow aBbB \Rightarrow abbB \Rightarrow abbSb \Rightarrow abbABb$$
$$\Rightarrow abbaBbBb \Rightarrow abbabbBb \Rightarrow abbabbbb$$

Pumping Lemma

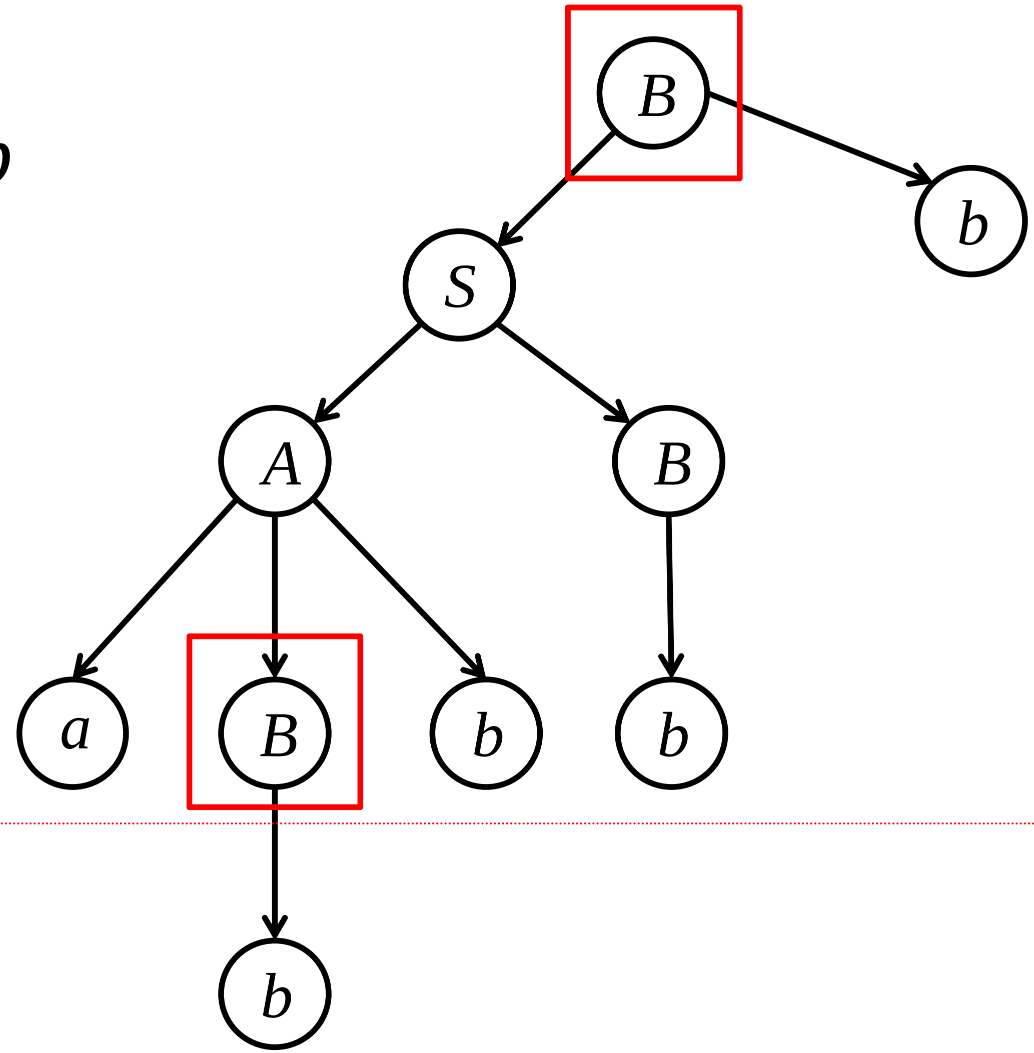


Pumping Lemma



Pumping Lemma

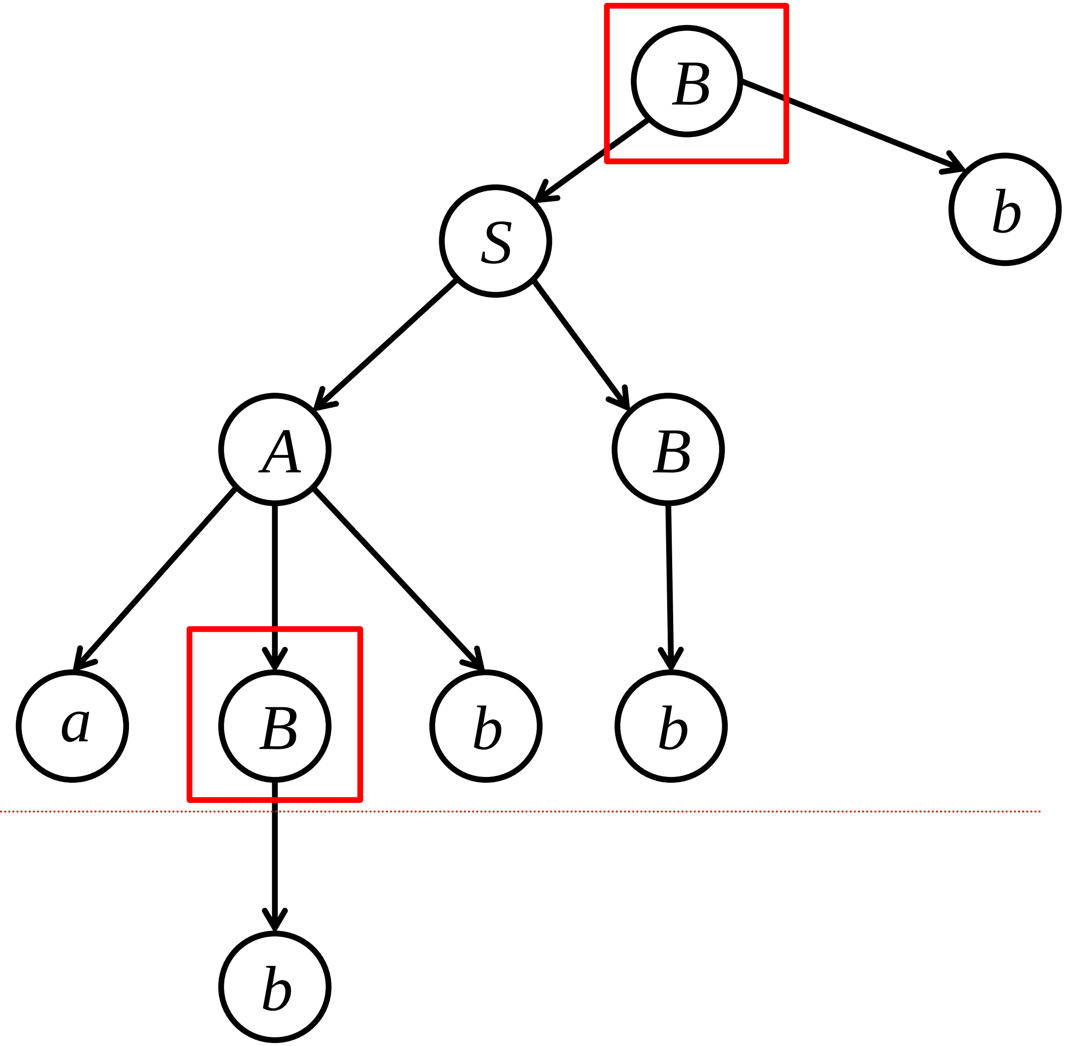
$$B \Rightarrow Sb \Rightarrow ABb$$
$$\Rightarrow aBbBb \Rightarrow aBbbb$$



$$B \Rightarrow b$$

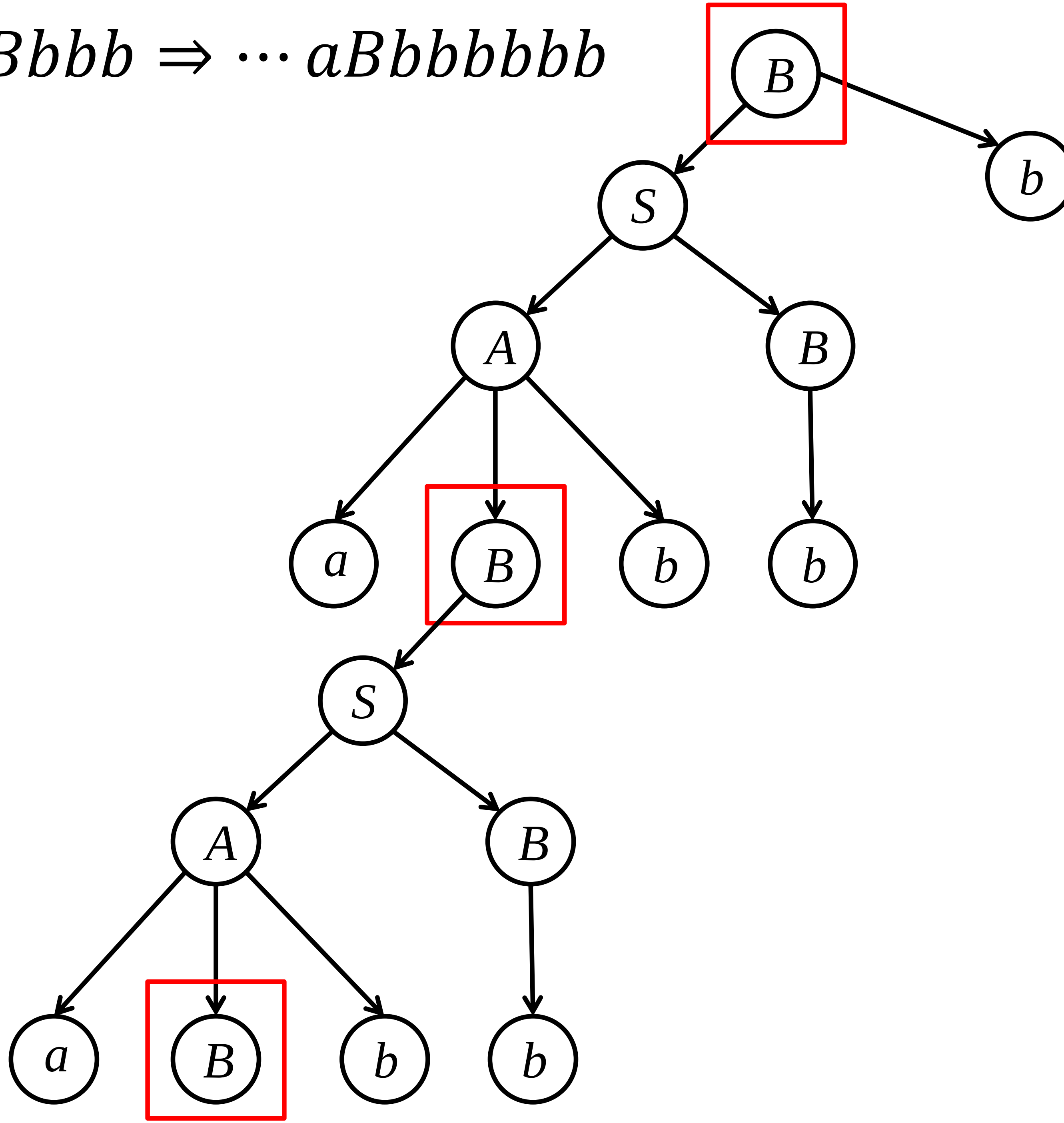
Repeated part:

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



$$B \Rightarrow b$$

$$\color{red}{B} \Rightarrow \cdots \Rightarrow aBbbb \Rightarrow \cdots aBbbbbbb$$



Pumping Lemma

Since

$$B \Rightarrow b$$

$$B \Rightarrow \cdots \Rightarrow aBbbb$$

$$S \Rightarrow \cdots \Rightarrow abbaBbbb$$

Then

$$S \Rightarrow \cdots \Rightarrow abbaBbbb \Rightarrow \cdots \Rightarrow abbaaBbbbbbb$$

$$\Rightarrow abbaabbbbbbb$$

Pumping Lemma

In general

$$B \Rightarrow b$$

$$B \Rightarrow \cdots \Rightarrow aBbbb$$

$$S \Rightarrow \cdots \Rightarrow abbaBbbb$$

Then

$$S \Rightarrow \cdots \Rightarrow abbaBbbb \Rightarrow \cdots$$

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

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

$$\Rightarrow^* abba(a)^3B(bbb)^3bbb \Rightarrow \cdots \Rightarrow$$

$$\Rightarrow abba(a)^iB(bbb)^ibb \Rightarrow abba(a)^ib(bbb)^ibbb$$

Pumping Lemma

In general:

We are given an infinite language generated by context-free grammar $G = (N, T, S, P)$.

Assume that G has no unit-productions and λ-productions.

Take a string $w \in L(G)$ with the length larger than m:

$$m > \#(\text{productions}) \times (\text{the largest right side of productions})$$

Pumping Lemma

- Take a string $w \in L(G)$ with the length larger than m such that

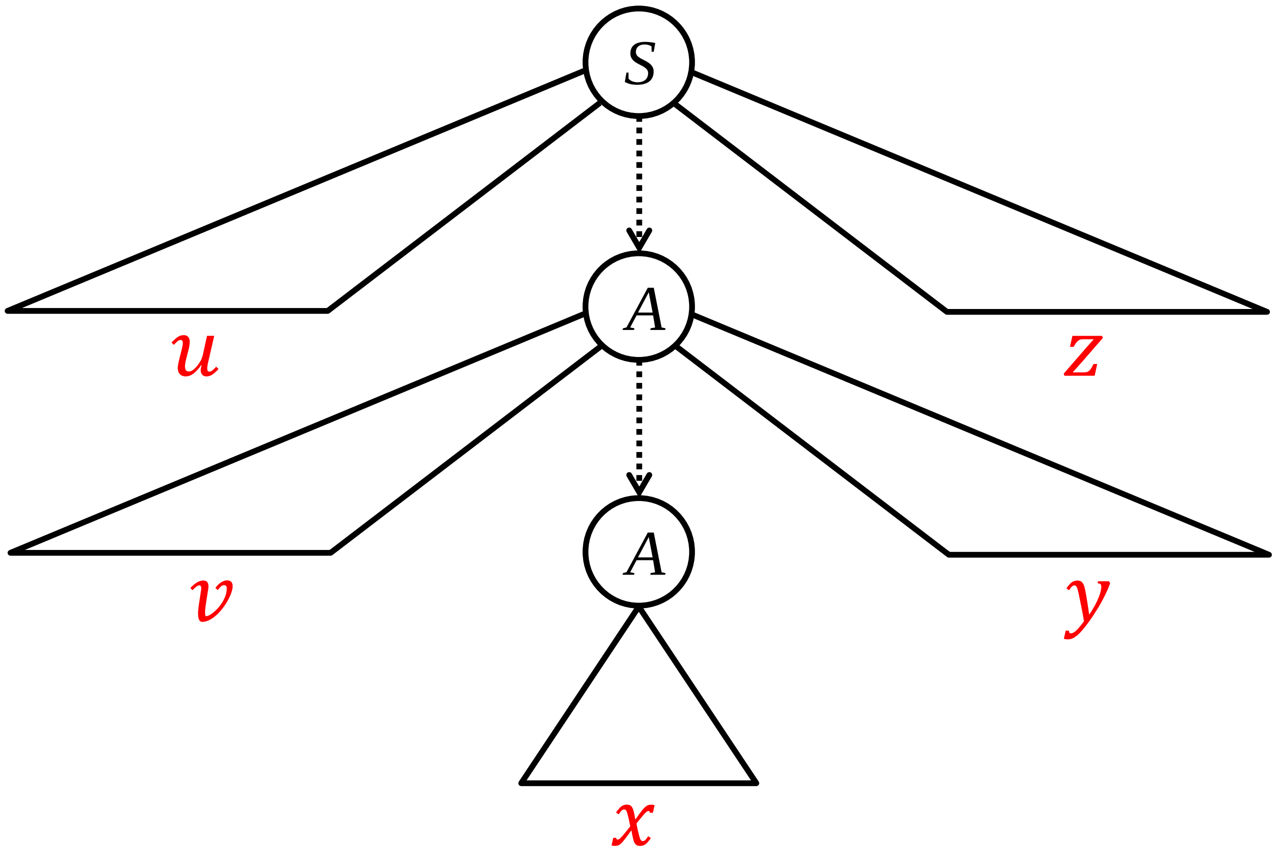
$$m > |P| \times (\text{largest right side of production})$$

- Consequence:

Some variable must be repeated in the derivation of w.

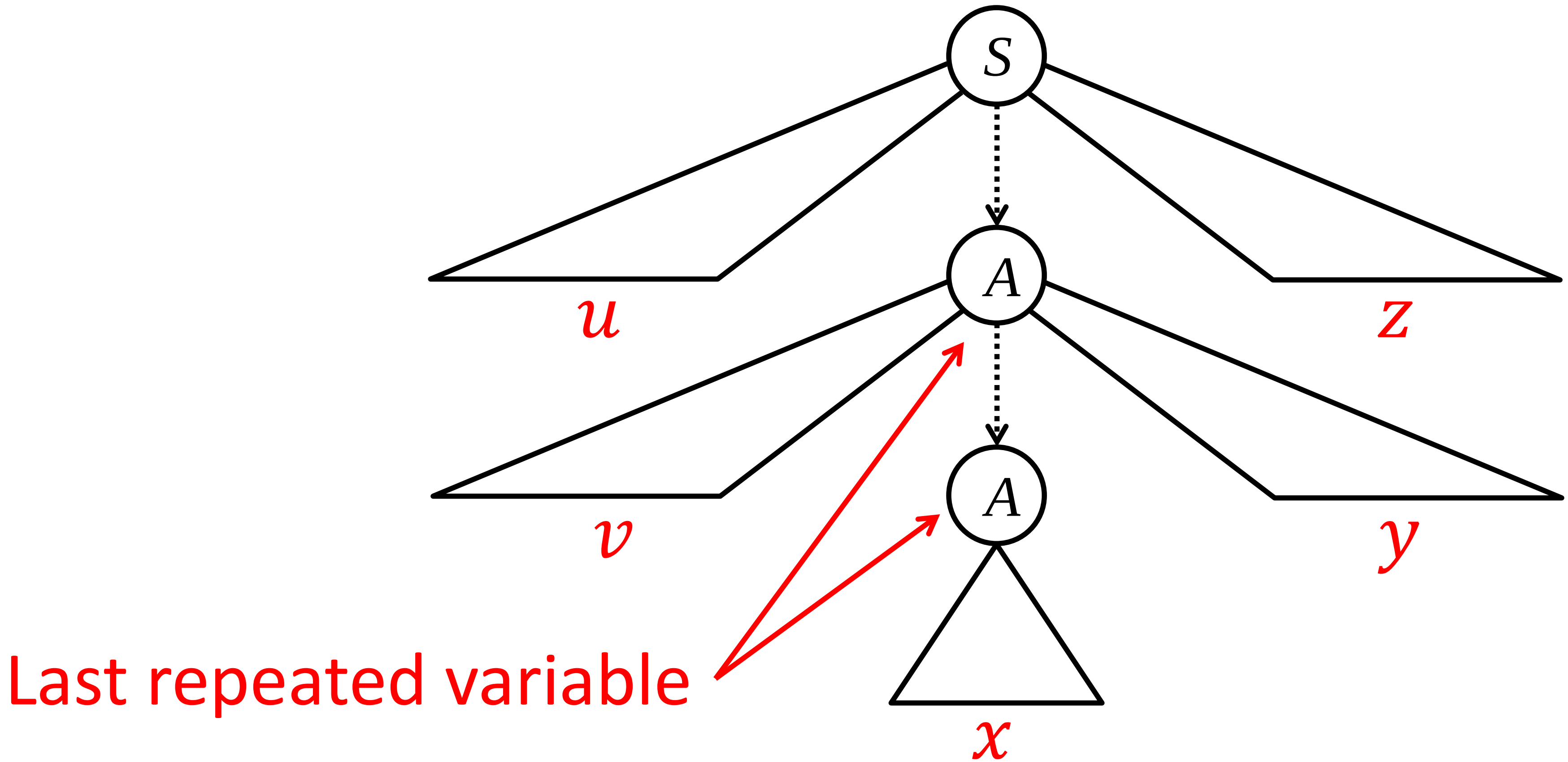
Pumping Lemma

$w = uvxyz \in L(G)$



Pumping Lemma

$w = uvxyz \in L(G)$



Pumping Lemma

Possible derivations:

$$S \Rightarrow^* uAz, \qquad A \Rightarrow^* vAy, \qquad A \Rightarrow^* x$$

Then the following string is also generated:

$$S \Rightarrow^* uAz \Rightarrow^* uxz = {\color{red} uv^0xy^0z}$$

Pumping Lemma

Possible derivations:

$$S \Rightarrow^* uAz, \qquad A \Rightarrow^* vAy, \qquad A \Rightarrow^* x$$

Then the following string is also generated:

$$S \Rightarrow^* uAz \Rightarrow^* uvxyz = uv^1xy^1z$$

Pumping Lemma

Possible derivations:

$$S \Rightarrow^* uAz, \qquad A \Rightarrow^* vAy, \qquad A \Rightarrow^* x$$

Then the following string is also generated:

$$\begin{aligned} S &\Rightarrow^* uAz \Rightarrow^* uvAyz \\ &\Rightarrow^* uvvAyyz \\ &\Rightarrow^* uvvxyyz = uv^2xy^2z \end{aligned}$$

Pumping Lemma

Possible derivations:

$$S \Rightarrow^* uAz, \qquad A \Rightarrow^* vAy, \qquad A \Rightarrow^* x$$

Then the following string is also generated:

$$\begin{aligned} S &\Rightarrow^* uAz \Rightarrow^* uvAyz \\ &\Rightarrow^* uvvAyyz \\ &\Rightarrow^* uvvvAyyyz \\ &\Rightarrow^* uvvvxyyyz = uv^3xy^3z \end{aligned}$$

Pumping Lemma

Possible derivations:

$$S \Rightarrow^* uAz, \qquad A \Rightarrow^* vAy, \qquad A \Rightarrow^* x$$

Then the following string is also generated:

$$\begin{aligned} S &\Rightarrow^* uAz \Rightarrow^* uvAyz \\ &\Rightarrow^* uvvAyyz \\ &\Rightarrow^* uvvvAyyyz \\ &\Rightarrow^* uvv \cdots vAyy \cdots yz \\ &\Rightarrow^* uvv \cdots vxyy \cdots yz = uv^ixy^iz \end{aligned}$$

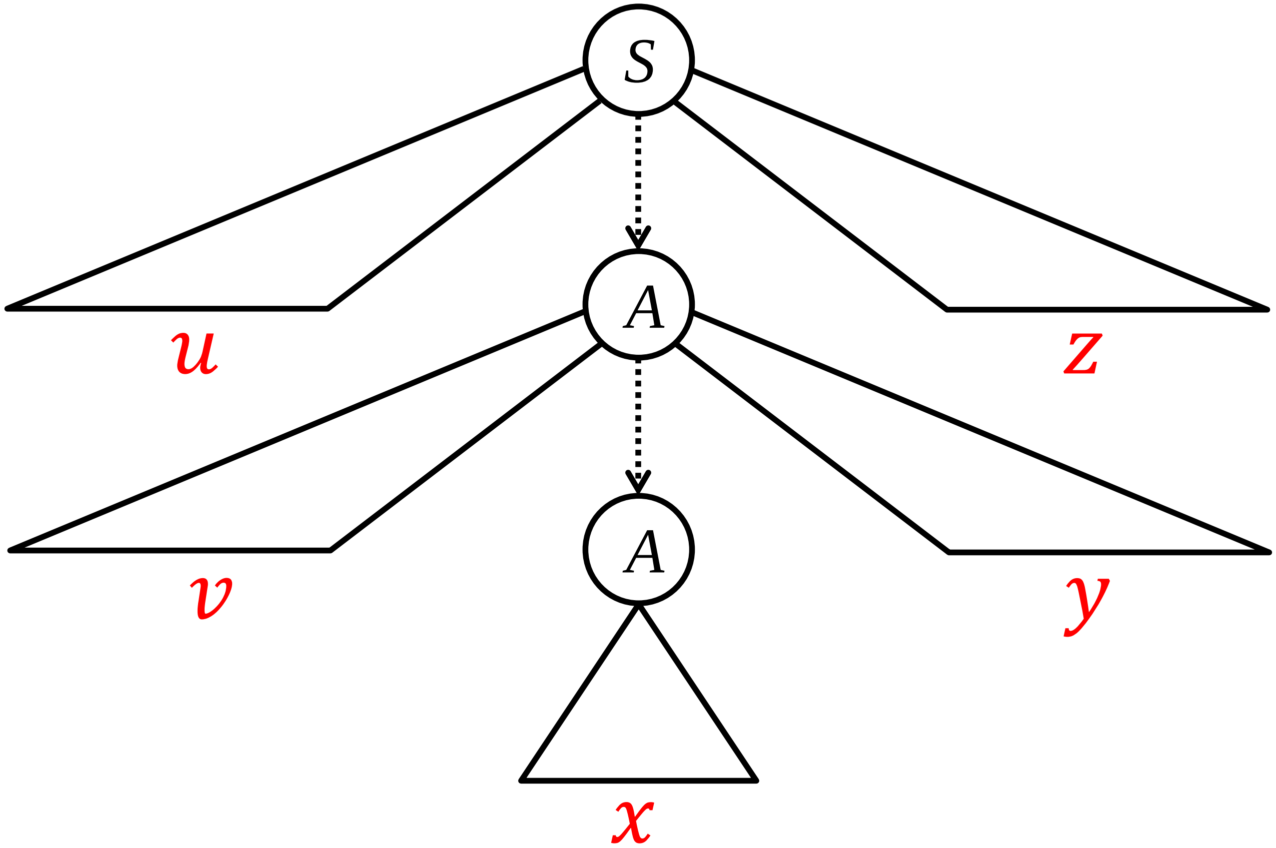
Pumping Lemma

- Therefore, any string of the form

$$uv^ixy^iz, i \geq 0$$

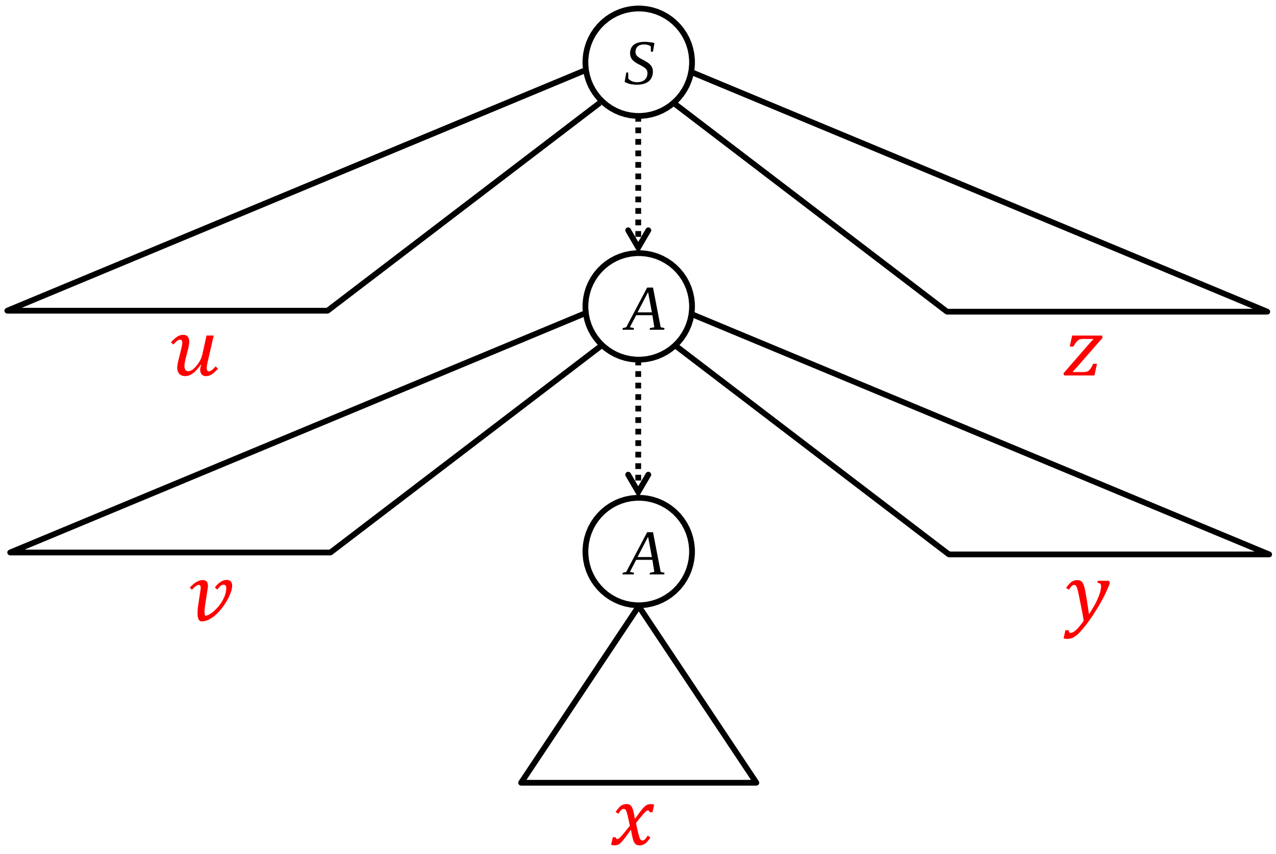
is generated by the grammar G.

Pumping Lemma: Observation



$|vxy| \leq m$, since A is the last repeated variable.

Pumping Lemma: Observation



$|vy| \geq 1$, since the are not unit and λ productions.

Pumping Lemma: Formal

Let L be an infinite context-free grammar. Then, there exists a positive integer m such that any string $w \in L$ with $|w| \geq m$, can be decomposed as

$$w = uvxyz,$$

with

$$|vxy| \leq m,$$

and

$$|vy| \geq 1,$$

such that

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

for all $i = 0, 1, 2, \ldots$

Applications
of
The Pumping Lemma

Application of Pumping Lemma

- The correct argument can be visualized as a **game** against an intelligent opponent.
- For regular languages, the substring xy whose length is bounded by m starts at the left end of w. Therefore the substring y that can be pumped is within m symbols of the beginning of w.
- For context-free languages, we only have a bound on $|vxy|$. The substring u that precedes vxy can be arbitrarily long. This gives additional freedom to the adversary.

Application of Pumping Lemma

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.

Application of Pumping Lemma

Proof:

Assume for contradiction that

is a context-free language.

Since L is infinite, we can apply the Pumping Lemma.

Application of Pumping Lemma

- Once the adversary has chosen m, we pick the string $w = a^m b^m c^m$ which is in L.
- If he chooses $|vxy|$ to contain only a's, then the pumped string will obviously not be in L.
- If he chooses a string containing an equal number of a's and b's, then the pumped string $a^k b^k c^m$ with $k \neq m$ can be generated, and again we have generated a string not in L.
- In fact, the only way the adversary could stop us from winning is to pick vxy so that vy has the same number of a's, b's and c's. But it is not possible because of restriction $|vxy| \leq m$.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

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

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 1: vxy is within a^n.

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

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

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 1: vxy is within a^n.

$$\overbrace{aaaaa\cdots aaaaa}^{m+2k}\ \overbrace{bbb\cdots bbb}^{m}\ \overbrace{ccc\cdots ccc}^{m}$$

$$\underbrace{a}_{u}\underbrace{aaaa\cdots aaa}_{v^2xy^2}\underbrace{a\ bbb\cdots bbb\ ccc\cdots ccc}_{z}$$

Repeat v and y.

Contradiction: $a^{m+k} b^m c^m \notin L$.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 2: vxy is within b^n.

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

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

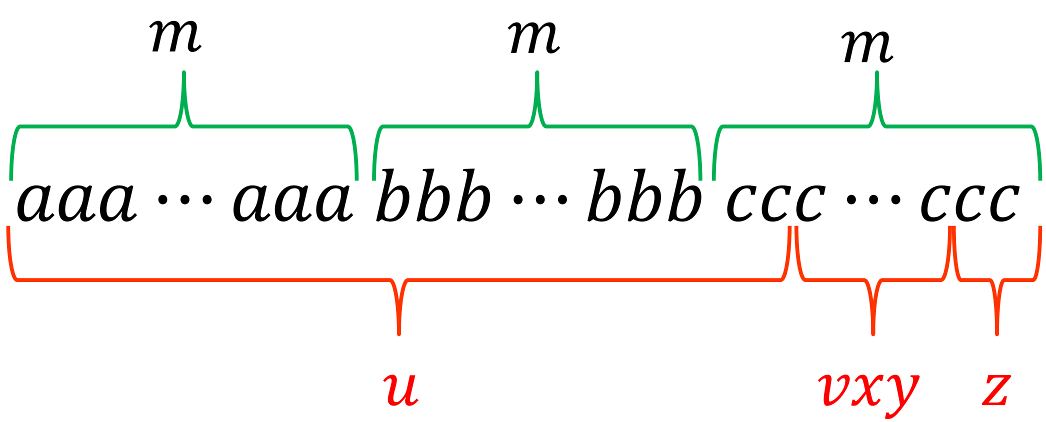
Similar to Case 1.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 3: vxy is within c^n.



Similar to Case 1.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 4: vxy is within $a^m b^m$.

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

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

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 4: vxy is within $a^m b^m$.

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

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

(a) v contains only a, and y contains only b.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 4: vxy is within $a^m b^m$.

$$\overbrace{aaa\cdots aaaaaa}^{m+k_1}\,\overbrace{bbbbbb\cdots bbb}^{m+k_2}\,\overbrace{ccc\cdots ccc}^{m}$$

$$\underbrace{aaa}_{u}\,\underbrace{\cdots aaaaaa\,bbbbbb}_{v^2xy^2}\,\underbrace{\cdots bbb\,ccc\cdots ccc}_{z}$$

Repeat v and y.

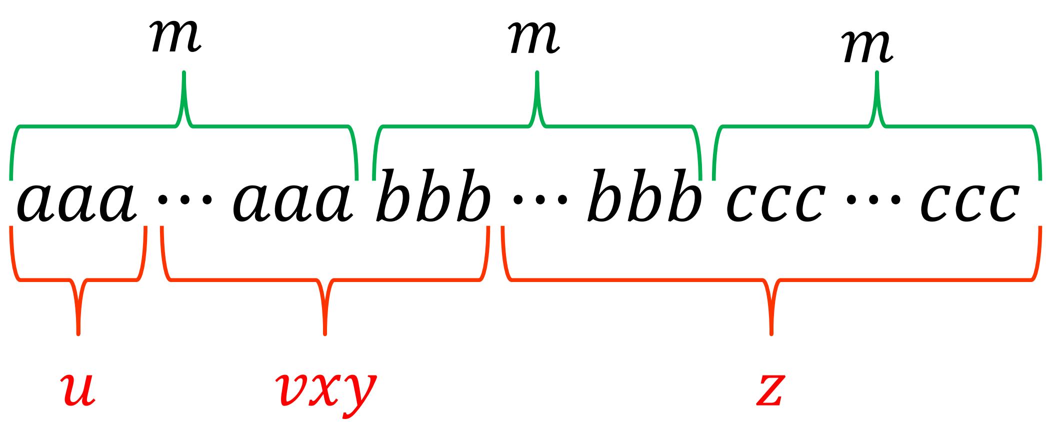
Contradiction: $a^{m+k_1} b^{m+k_2} c^m \notin L$.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 4: vxy is within $a^m b^m$.



(b) v contains a and b, and y contains only b.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 4: vxy is within $a^m b^m$.

$$\overbrace{aaa\cdots aaaaa}^{m}\,\overbrace{bb}^{k_1}\overbrace{aa}^{k_2}\overbrace{bb\ bbbbbb\cdots bbb}^{m+k}\,\overbrace{ccc\cdots ccc}^{m}$$

$$\underbrace{aaa}_{u}\underbrace{\cdots aaaaa\ bbaabb\ bbbbbb}_{vxy}\underbrace{\cdots bbb\ ccc\cdots ccc}_{z}$$

Repeat v and y. $k_1 + k_2 \geq 1$.

Contradiction: $a^m b^{k_1} a^{k_2} b^{m+k} c^m \notin L$.

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 4: vxy is within $a^m b^m$.

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

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

(b) v contains only a, and y contains a and b.

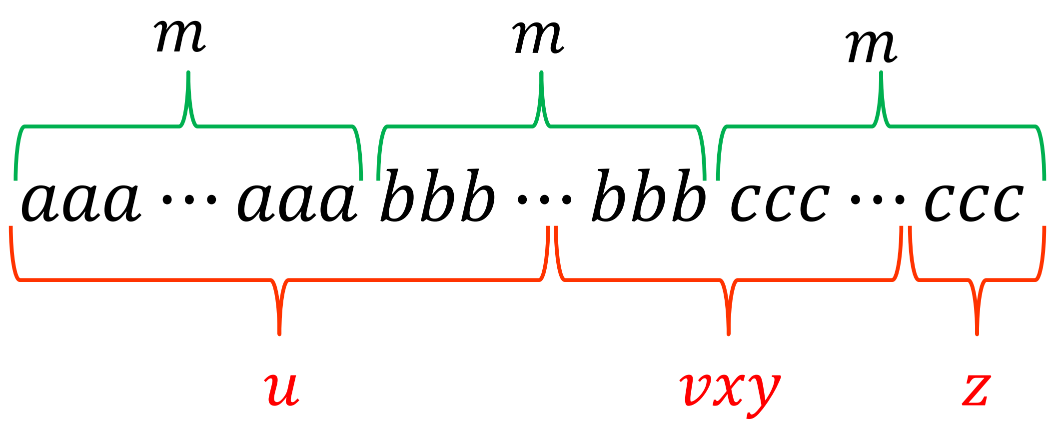
Similar to **Case 4** (a).

Application of Pumping Lemma

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

We can write $w = uvxyz$ with $|vxy| \leq m$ and $|vy| \geq 1$

Case 5: vxy is within $b^m c^m$.



Similar to **Case 4**.

Application of Pumping Lemma

- All possible cases are considered.
- Since $|vxy| \leq m$, string vxy cannot overlap a^m, b^m and c^m at the same time.
- In all cases, we obtain contradiction.
- Therefore, our initial assumption that

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

is context-free must be wrong.

- **Conclusion**: L is not context-free.

Exercises

- Prove that the following languages are not context-free:

1. $L = \{ww : w \in \{a, b\}^*\}$
2. $L = \{a^{n!} : n \geq 0\}$
3. $L = \{a^{n^2} b^n : n \geq 0\}$