Formal Languages
The Pumping Lemma for CFLs

A Pumping Lemma for Context-Free Languages

Let L be an infinite context-free language. Then there exists some positive integer m such that any $\omega \in L$ with $|\omega| \geq m$ can be decomposed as

$$w = uvxyz, \tag{8.1}$$

with

$$|vxy| \leq m, \tag{8.2}$$

$$|vy| \geq 1, \tag{8.3}$$

such that

$$uv^i xy^i z \in L, \tag{8.4}$$

for all $i = 0,1, 2,\ldots$. This is known as the pumping lemma for context-free languages.

The Pumping Lemma - to identify non regular languages

- Given an infinite regular language L
- there exists an integer m
- for any string $w \in L$ with length $|w| \geq m$
- we can write $w = x\, y\, z$
- with $|x\, y| \leq m$ and $|y| \geq 1$
- such that: $x\, y^i\, z \in L \qquad i = 0, 1, 2, \ldots$

The Pumping Lemma:

For infinite context-free language L

there exists an integer m such that

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

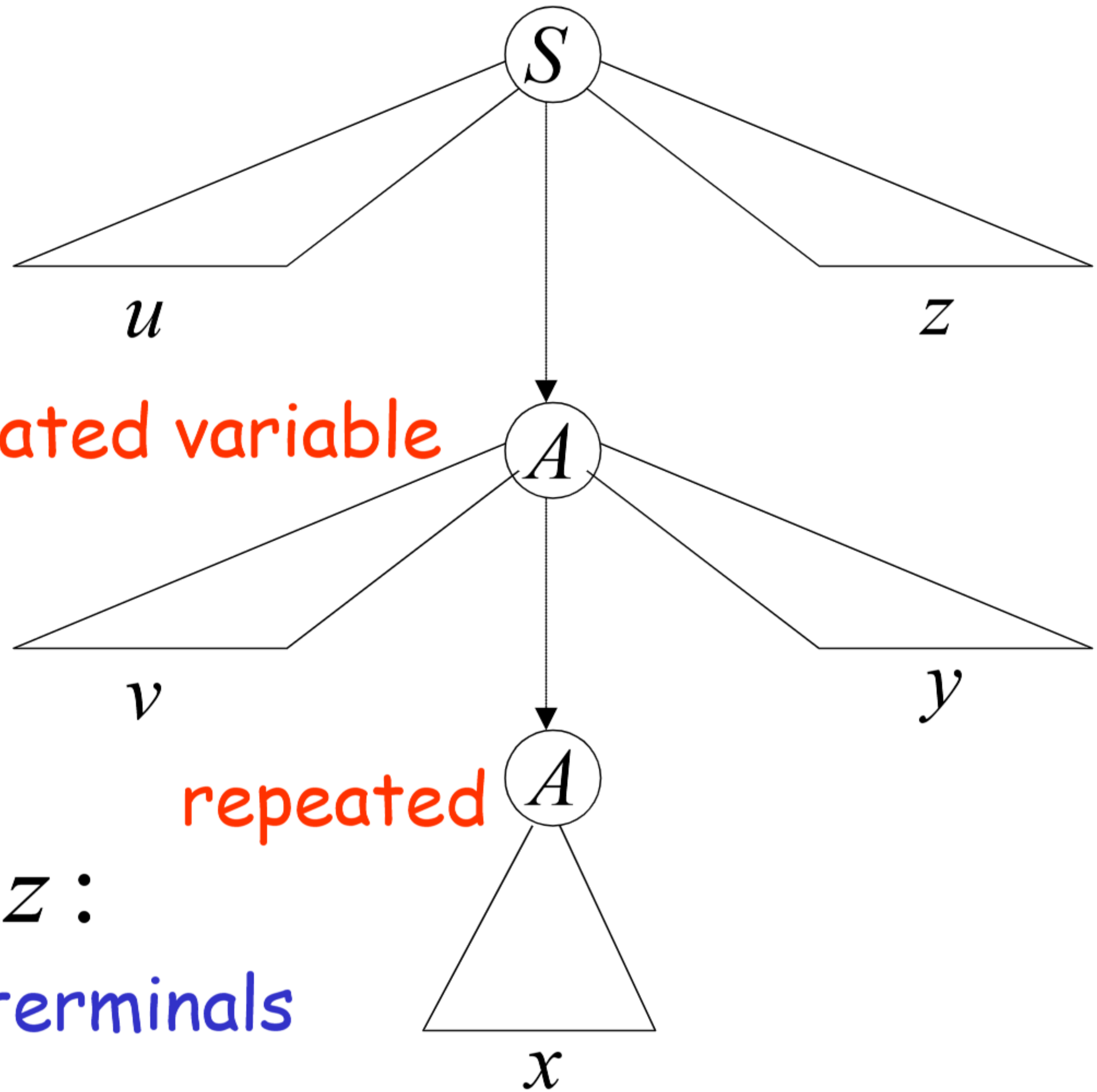
we can write $w = uvxyz$

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

and it must be:

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

Derivation tree of string w



$$w = uvxyz$$

u, v, x, y, z :

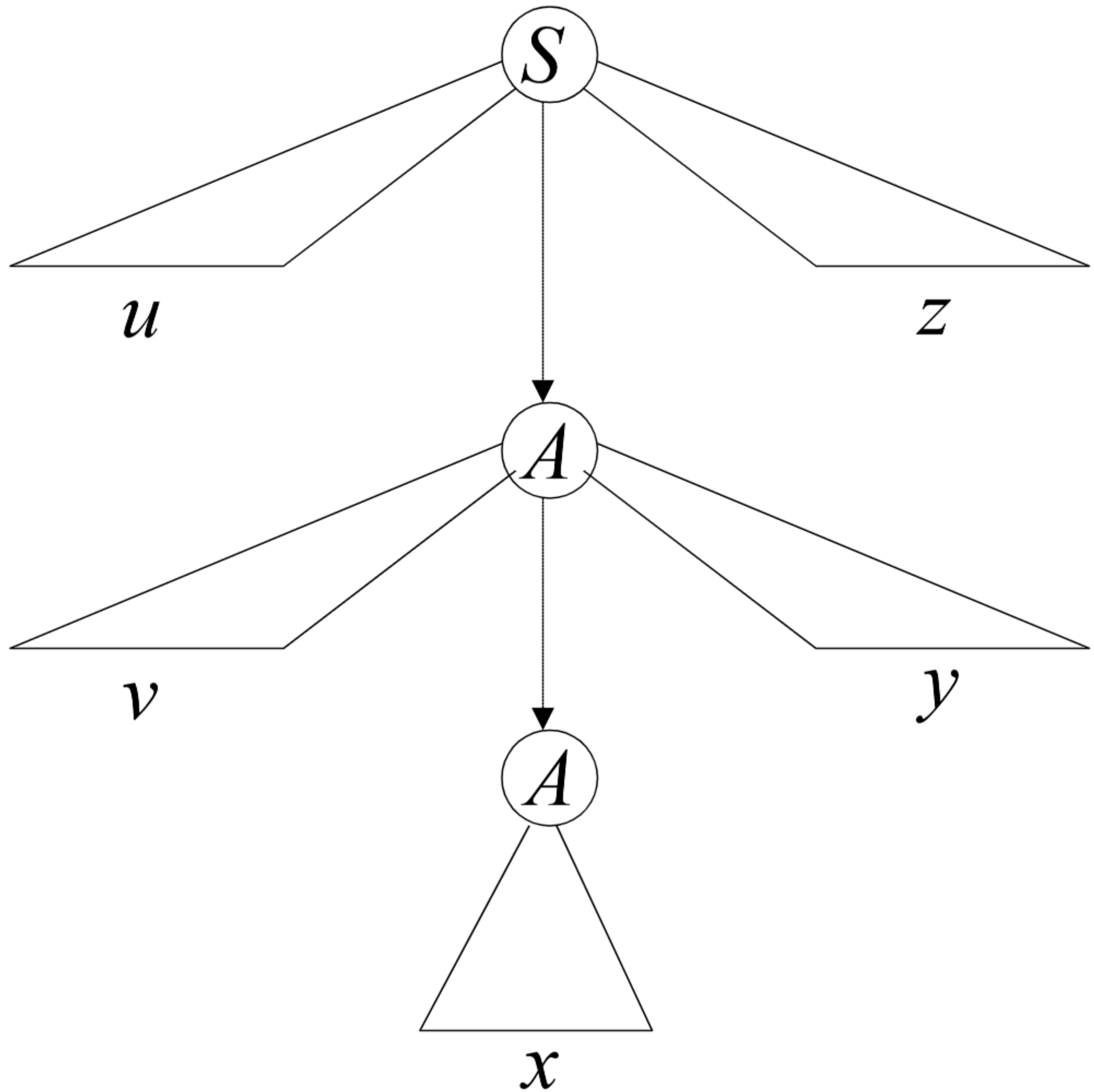
Strings of terminals

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:

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

$$uv^0xy^0z$$

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:

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

This string is also generated:

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

$$uv^2xy^2z$$

We know:

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

This string is also generated:

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

$$\overset{*}{\Rightarrow} uvvvAyyyz \overset{*}{\Rightarrow} uvvvxyyyz$$

$$uv^3xy^3z$$

We know:

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

This string is also generated:

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

$$\stackrel{*}{\Rightarrow} uvvvAyyyz \stackrel{*}{\Rightarrow} \dots$$

$$\stackrel{*}{\Rightarrow} uvvv\cdots vAy\cdots yyyz \stackrel{*}{\Rightarrow}$$

$$\stackrel{*}{\Rightarrow} uvvv\cdots vxy\cdots yyyz$$

$$uv^i xy^i z$$

Therefore, any string of the form

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

is generated by the grammar G

Therefore,

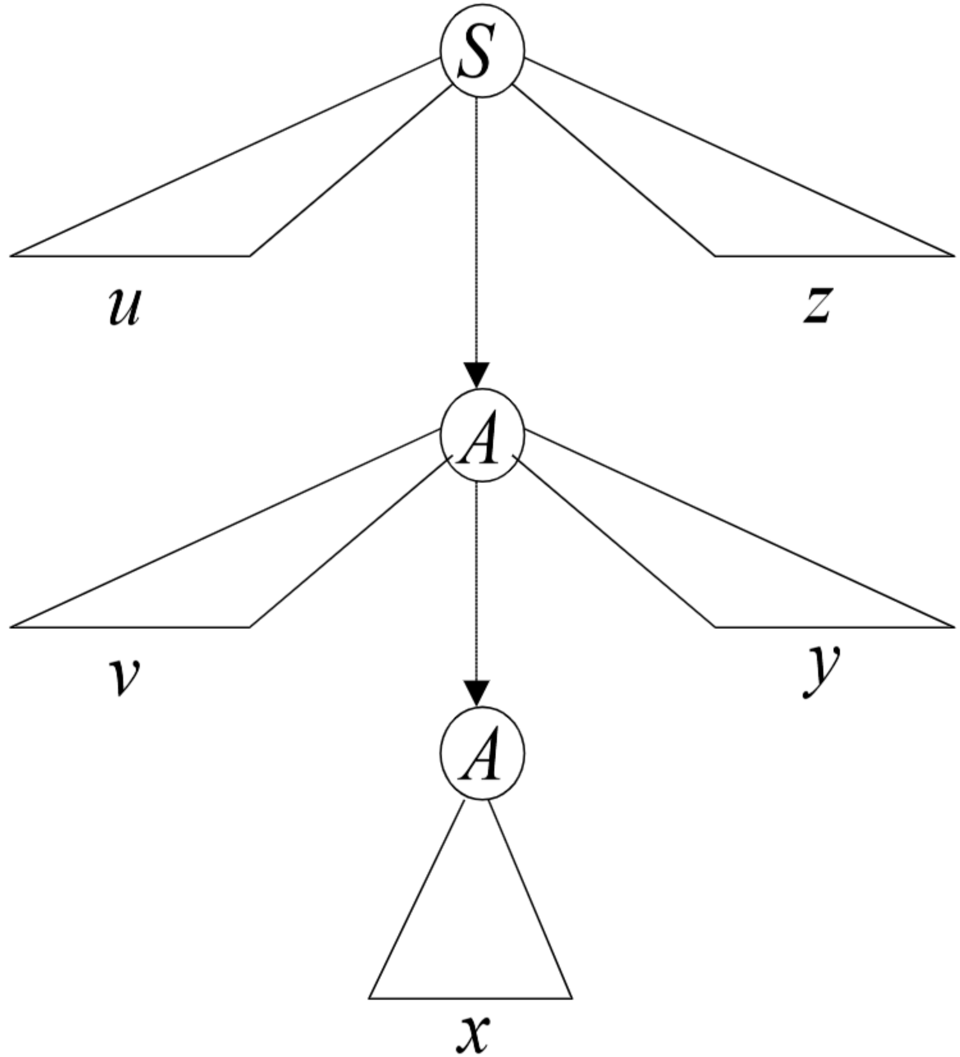
knowing that $uvxyz \in L(G)$

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

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

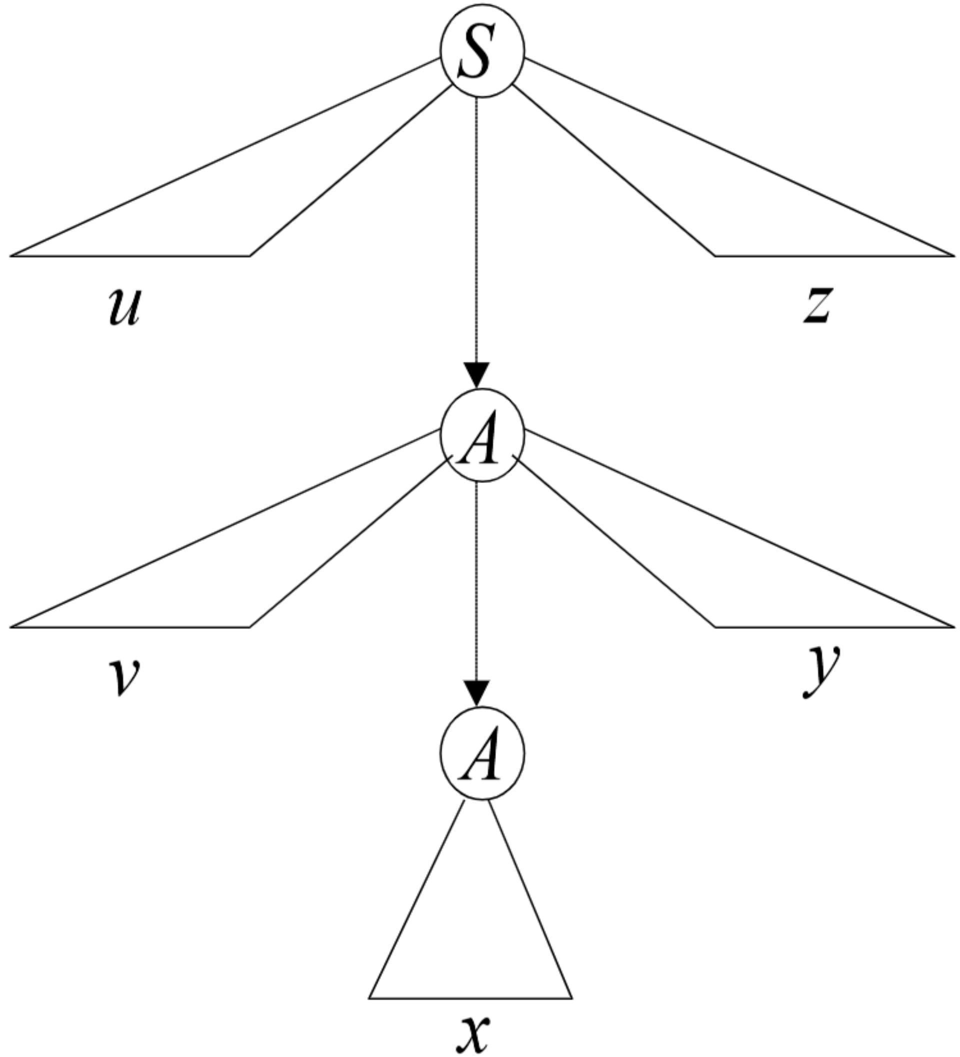
$$\Downarrow$$

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



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

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

with each of $aaa...aaa$, $bbb...bbb$, $ccc...ccc$ having 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 1: v and y only contain a

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

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

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

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

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

Case 1: Repeating v and y

$$k \geq 1$$

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

$$u \qquad v^2xy^2 \qquad z$$

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

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

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

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

$$k \geq 1$$

$$\overbrace{aaaaaa\ldots aaaaaa}^{m+k}\;\overbrace{bbb\ldots bbb}^{m}\;\overbrace{ccc\ldots ccc}^{m}$$

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

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

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

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

Case 1: 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!!!

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

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

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

Case 2: vxy is within b^m

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

(the middle m spans $bbb...bbb$)

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

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

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

Case 2: Same analysis as in case 1

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

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

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

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

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

Case 3: vxy is within c^m

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

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

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

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

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

Case 3: Same analysis as in case 1

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

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

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 possible locations of string vxy in $a^m b^m a^m b^m$

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

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

vxy is within the first a^m

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

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

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

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

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

$\boxed{vxy \text{ is within the first } a^m}$

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

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

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

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

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

vxy is within the first a^m

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

$$k_1 + k_2 \ge 1$$

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

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

vxy is within the first a^m

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

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

Contradiction!!!

Therefore: The original assumption that

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

is context-free must be wrong

Conclusion: L is not context-free

Closure Properties and Decision Algorithms for Context-Free Languages

The family of context-free languages is closed under union, concatenation, and star-closure.

Let L_1 and L_2 be two context-free languages generated by the context-free grammars $G_1 = (V_1, T_1, S_1, P_1)$ and $G_2 = (V_2, T_2, S_2, P_2)$, V1 and V2 are disjoint.

Consider now the language L (G3), generated by the grammar

$$G_3 = (V_1 \cup V_2 \cup \{S_3\}, T_1 \cup T_2, S_3, P_3),$$

where S_3 is a variable not in $V1 \cup V2$. The productions of G_3 are all the productions of G_1 and G_2, together with an alternative starting production that allows us to use one or the other grammars. More precisely

$$P_3 = P_1 \cup P_2 \cup \{S_3 \to S_1 | S_2\}.$$

G_3 is a context-free grammar, so that L (G_3) is a context-free language.

Concatenation

$$G_4 = (V_1 \cup V_2 \cup \{S_4\}, T_1 \cup T_2, S_4, P_4).$$

Here again S_4 is a new variable and

$$P_4 = P_1 \cup P_2 \cup \{S_4 \to S_1 S_2\}.$$

Then

$$L(G_4) = L(G_1) L(G_2)$$

Star Closure

Finally, consider $L(G_5)$ with

$$G_5 = (V_1 \cup \{S_5\}, T_1, S_5, P_5),$$

where S_5 is a new variable and

$$P_5 = P_1 \cup \{S_5 \to S_1 S_5 | \lambda\}.$$

Then

$$L(G_5) = L(G_1)^*.$$

CFL are not closed under Intersection and Complementation

The family of context-free languages is not closed under intersection and complementation.

Proof: Consider the two languages

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

or

and

$$L_2 = \{a^n b^m c^m : n \geq 0, m \geq 0\}.$$

There are several ways one can show that L_1 and L_2 are context-free. For instance, a grammar for L_1 is

$$\begin{aligned} S &\rightarrow S_1 S_2, \\ S_1 &\rightarrow a S_1 b | \lambda, \\ S_2 &\rightarrow c S_2 | \lambda. \end{aligned}$$

Alternatively, we note that L_1 is the concatenation of two context-free languages, so it is context-free by Theorem 8.3. But

$$L_1 \cap L_2 = \{a^n b^n c^n : n \geq 0\},$$

which we have already shown not to be context-free. Thus, the family of context-free languages is not closed under intersection.

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}.$$

If the family of context-free languages were closed under complementation, then the right side of the above expression would be a context-free language for any context-free L_1 and L_2. But this contradicts what we have just shown, that the intersection of two context-free languages is not necessarily context-free. Consequently, the family of context-free languages is not closed under complementation.■