Pumping Lemma for Context-Free Languages

Let $L = L(G),\ G$ a CFG.

Then $\exists k > 0$ s.t. $\forall w \in L$ with $|w| \geq k$,

w can be **decomposed** as:

$$w = uvxyz$$

where $$vy \neq \lambda$$

and $$|vxy| \leq k$$

s.t. $\forall i \geq 0$,

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

$Proof$: By **Chomsky's Theorem** (p. 6-6) we can construct a CFG G^C in **CNF** for $L \backslash \{\lambda\}$.

Let $n = \#$ of non-terminals in G^C.

Let $k = 2^{n+1}$.

Let $w \in L,\ |w| \geq k$.

Let PT be a parse tree in G^C for w.

Then (from p. 6-13) PT has **depth** $\geq n + 1$

and the **longest path** in PT has length $\geq n + 1$.

Therefore, it contains $\geq n + 1$ occurences of non-terminals.

Therefore by the PHP, some **non-terminal** X of G^C is **repeated** on this path.

Let T = subtree of PT rooted at an **upper** occ. of X

Let t = subtree of PT rooted at a **lower** occ. of X

Then we can write:

$$w = uvxyz$$

where

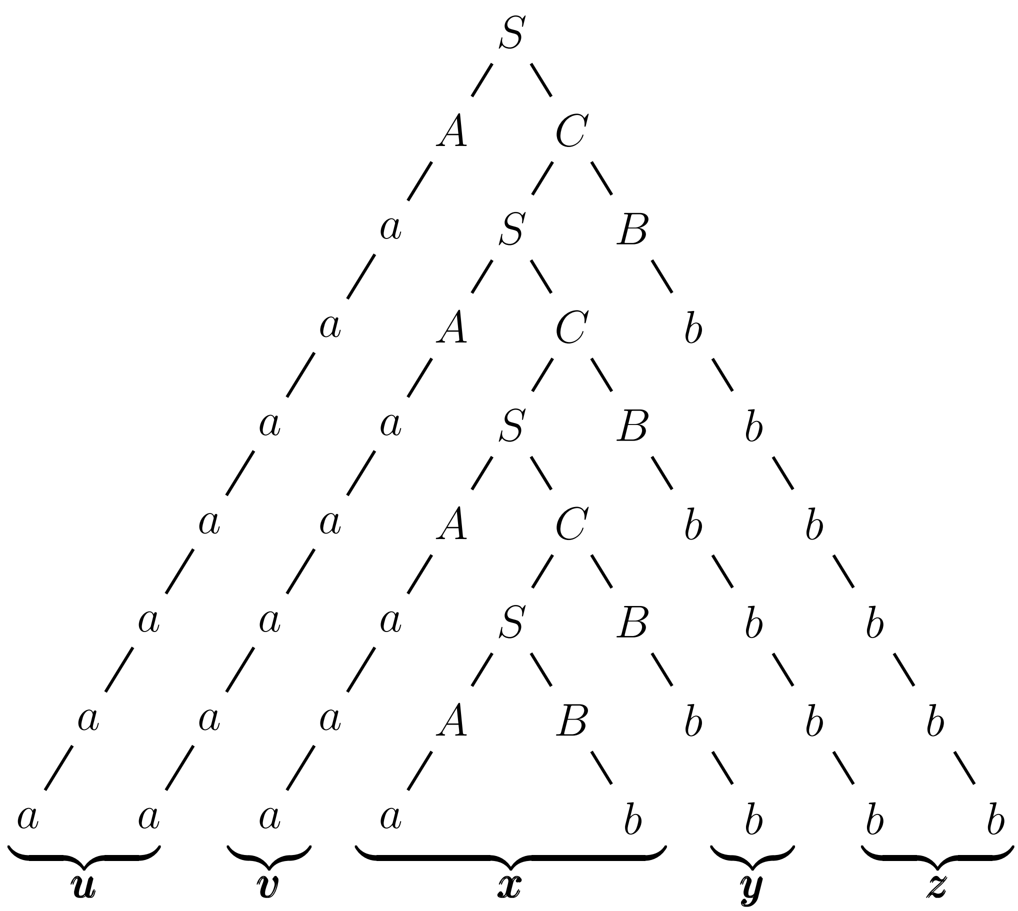
x is the substring generated by t (i.e. lower occ. of X)

vxy is the substring generated by T (i.e. upper occ. of X)

[E.g. for our example (p.) $n = 4$, $k = 2^5$ we can take $w = a^k b^4$]

In this way, get **parse trees in G^C** for
$w_i = uv^i xy^i z \quad (i = 0, 1, 2, ...)$
which are all distinct, since $vy \neq \lambda$

Example: **Extended parse tree in G^C for $w = a^4b^4$**



We can take

$$\boldsymbol{u = aa}$$

$$\boldsymbol{v = a}$$

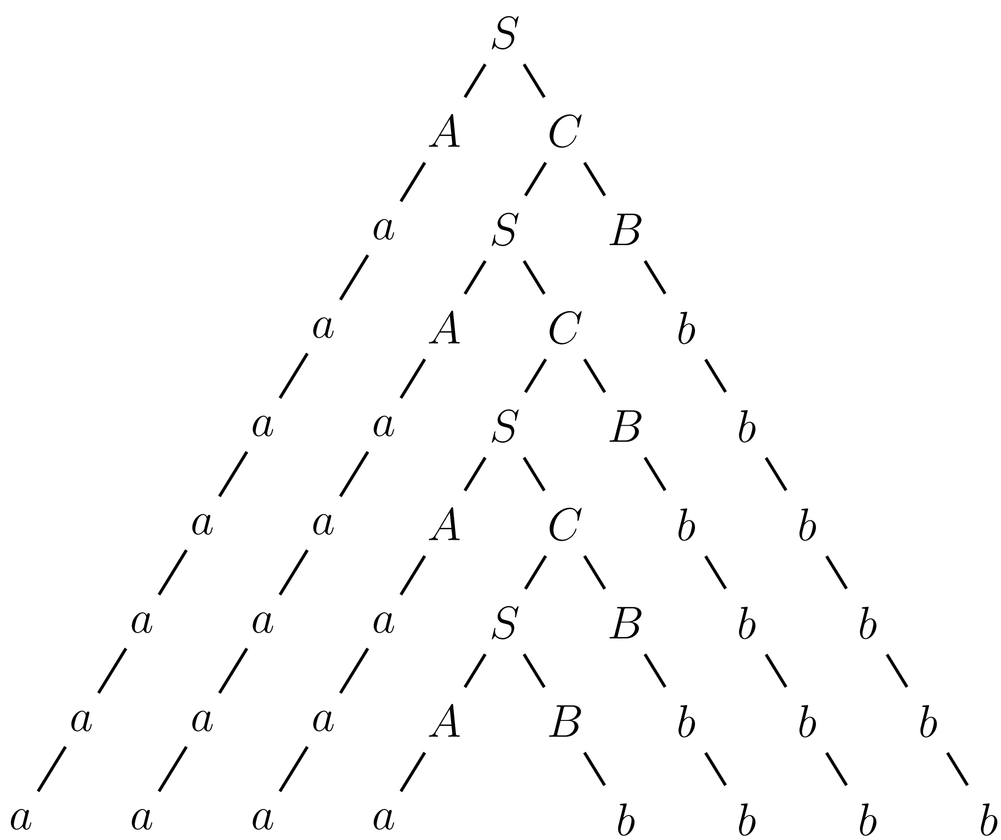
$$\boldsymbol{x = ab}$$

$$\boldsymbol{y = b}$$

$$\boldsymbol{z = bb}$$

Then get words in $\boldsymbol{L}$:

$$\boldsymbol{w_i = uv^i xy^i z} \quad (\boldsymbol{i} = 0, 1, 2, ...)$$



$\boldsymbol{w_2 = uv^2xy^2z}$ formed from $\boldsymbol{w}$ by replacing $\boldsymbol{t}$ by $\boldsymbol{T}$

$\boldsymbol{w_3 = uv^3xy^3z}$ formed from $\boldsymbol{w_2}$ by replacing $\boldsymbol{t}$ by $\boldsymbol{T}$

$\boldsymbol{w_0 = uv^0xy^0z}$ formed from $\boldsymbol{w}$ by replacing $\boldsymbol{T}$ by $\boldsymbol{t}$, etc.

We have been using the set $\{a^n b^n \mid n \geq 0\}$ with the grammar G^C as a running example.

Returning to the general case:

Recall $k = 2^{n+1}$ where $n = \#$ of non-terminals.

Suppose $|w| \geq k$

Then by the *Conclusion* on p. 6-13:

A parse tree for w has w at level $\geq n+1$.

Then by PHP, some non-terminal X, occurs **twice on this path**.

Take the **two lowest occurrences** of any **repeating non-terminal X on this path**.

Let T be the tree under the higher X

Let t be the tree under lower X

Now the **height** of the **higher X** on this path is **at most $n+1$**,

by PHP (since otherwise there would be a lower repeating pair).

Hence $|vxy| \leq$ $\square$

Examples: (1) $L = \{a^n b^n a^n \mid n \geq 0\}$ is **not** a CFL.

$Proof$: By contradiction:

Suppose it is.

Let G^C be a Chomsky grammar that generates:

$$L \backslash \{\lambda\} = \{a^n b^n a^n \mid n \geq 0\}$$

Suppose G^C has n non-terminals.

Let $k = 2^{n+1}$.

Let $w = a^k b^k a^k \in L$

By PL for CFL's:

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

$$w = uvxyz,$$

$$vy \neq \lambda,$$

and

$$|vxy| \leq k$$

and $\forall i \geq 0$

$$w_i = uv^i xy^i z \in L \backslash \{\lambda\}.$$

Represent w as:

$$w = \underbrace{\overbrace{a \ldots a}^{k\ a's}}_{\textbf{Region 1}} \;\Big|\; \underbrace{\overbrace{b \ldots b}^{k\ b's}}_{\textbf{Region 2}} \;\Big|\; \underbrace{\overbrace{a \ldots a}^{k\ a's}}_{\textbf{Region 3}}$$

Boundary 1 Boundary 2

Now where is v?

Assume $v \neq \lambda$.

There are a **number of possibilities**:

Possibility 1: v lies in Region 1, so $v \preccurlyeq a^k$.

Since $|vxy| \leq k$,

vxy is **completely in Regions 1 and 2**.

Consider $w_2 = uv^2xy^2z \in L'$

In $w_2 = uv^2 \ldots$,

so Region 1 has $> k\ \ a's$.

But Region 3 has only $k\ \ a's$,

since $|vxy| \leq k$, and so vxy does not reach Region 3.

Therefore $w_2 \notin L$.

Possibility 2: v crosses Boundary 1.

so $v = a^i b^j$ for some $i, j > 0$.

Therefore

$$\begin{aligned} w_2 &= uv^2 ... \\ &= ...a^i b^j a^i b^j ... \end{aligned}$$

So w_2 **has** $>$ **3 regions!**

Therefore $w_2 \notin L$

Possibility 3: v lies in Region 2, so $v \preccurlyeq b^k$.

Therefore, $w_2 = uv^2 ...$

So Region 2 has $> k$ $b's$

but Region 1 has **only** k $a's$ since v^2 does not touch Region 1.

Therefore $w_2 \notin L$.

Possibility 4: v crosses Boundary 2, so $v = a^i b^j$ for some $i, j > 0$.

Then, (just as in **Possibility 2**)

w_2 **has** $>$ **3 regions!**

Therefore $w_2 \notin L$.

Possibility 5: v lies in Region 3, so $v \preccurlyeq a^k$.

Then, (just as in **Possibility 1**)

In $w_2 = uv^2...$

Region 3 has $> k\ \ a's$.

Further, since v lies in Region 3, vxy does not touch Region 1.

Therefore, in w_2, Region 1 has only $k\ \ a's$.

Hence $w_2 \notin L$.

Therefore, by all cases $w_2 \notin L$

<u>Note:</u> We have assumed $v \neq \lambda$.

But if $v = \lambda$, then $y \neq \lambda$.

So use **exactly** the same argument with y, proceeding $R \to L$ on w_2.

Other examples of Non-CFL's:

$\{a^n b^n c^n \mid n \geq 0\}$,

$\{a^n b^{2n} c^{3n} \mid n \geq 0\}$, etc., etc.

Similar proofs for both.

<u>BUT</u>

$\{a^n b^n a^k \mid n, k \geq 0\}$,

$\{a^n b^k a^k \mid n, k \geq 0\}$,

$\{a^n b^k c^{2k} \mid n, k \geq 0\}$, etc.

— with k "**independent** of n" — **ARE** CFL's

But, e.g.:

$\{a^n b^k a^n \mid k \geq n$ (**or** $k < n$)$\}$, etc., etc.

— where k "**depends on**" n — **are NOT** CFL's

Another Example of a non-CFL:

(2) $\Sigma = \{a, b\}$,

$L = \{uu \mid u \in \Sigma^*\}$

$\boldsymbol{Proof}$: Consider $u = a^k b^k$ for k "**big**".

Similar to (1).