# COL352 Lecture 22

## Contents



## 1 Recap

Proof that PDA = grammars completed.

## 2 Definitions

**Definition 1**

We define a simple PDA $P$ to be a PDA such that

1. $\Delta = \Delta_{push} \uplus \Delta_{pop}$, where
   (a) $\Delta_{push}$ contains transitions $(q, a, \epsilon, q', B)$ where $q, q' \in Q, a \in \Sigma_\epsilon, B \in \Gamma$ (i.e., not allowed to pop, must push), and
   (b) $\Delta_{pop}$ contains transitions $(q, a, B, q', \epsilon)$ where $q, q' \in Q, a \in \Sigma_\epsilon, B \in \Gamma$ (i.e., must pop, not allowed to push).
2. $|A| = 1$, i.e., unique accepting state.
3. If $x$ is accepted, then $x$ is accepted with an empty stack, i.e., $(q_{init}, x, \epsilon) \vdash^* (q_{acc}, \epsilon, \alpha)$ for some $\alpha \in \Sigma^*$ iff $q_{init}, x, \epsilon) \vdash^* (q_{acc}, \epsilon, \epsilon)$.

## 3 Content

**Question 1**

What kinds of closure properties can we think of for context free languages? Union? Intersection? Complementation? Concatenation? $*$?

*Answer.* Let $G = (N, \Sigma, R, S)$.
For $*$: $G' = (N \uplus \{T\}, \Sigma, R \cup \{T \to TT, T \to S, T \to \epsilon\})$ generates $L^*$.
Let $G = (N, \Sigma, R, S)$.
For $\cup$ : $G_1 = (N_1, \Sigma, R_1, S_1)$ generates $L_1$ and $G_2 = (N_2, \Sigma, R_2, S_2)$ generates $L_2$, then $(N_1 \uplus N_2 \uplus T, \Sigma, R_1 \cup R_2 \cup \{T \to S_1, T \to S_2\}, T)$ generates $L_1 \cup L_2$.

$(N_1 \uplus N_2 \uplus T, \Sigma, R_1 \cup R_2 \cup \{T \to S_1 S_2\}, T)$ generates $L_1 L_2$.

Not closed under intersection (would imply complementation by contradiction and contrapositive).

Let $L_1 = \{a^n b^n c^* \mid n \in \mathbb{N} \cup \{0\}\}$, and $L_2 = \{a^* b^n c^n \mid n \in \mathbb{N} \cup \{0\}\}$. Then $L_1 \cap L_2 = \{a^n b^n c^n \mid n \in \mathbb{N} \cup \{0\}\}$, which is probably not context free. We'll use a version of the pumping lemma for the DFA.

**Note 1**

Suppose $L$ is a context free language generated by a grammar $G$. Suppose $w \in L$ is a "long enough" string. Consider a smallest parse tree $T$ of $w$. Since $w$ is "long enough", $T$ is "tall enough". Look at the longest root-to-leaf path in $T$, say $p$.

Since $p$ is long enough, some two nodes on $P$ are labelled by the same non-terminal, say $A \in N$. Let $uw'z$ be $w$ such that the upper $A$ derives $w'$, and let $vxy$ be $w'$. Look at tree (very helpful). Break it into $S, A$'s tree (having $u, z$), $A, A$'s tree (having $v, y$) and $A$'s tree (having $x$).

By copy pasting the second tree into itself again and again, we can get $\forall i : uv^ixy^iz \in L$. This doesn't give us anything if $y = v = \epsilon$, so we enforce smallest tree constraints.

How tall is tall enough? If number of non-terminals is $k$, then we need height $\geq k + 1$.

How long is long enough? If $d = \max(\{|\alpha| \mid A \to \alpha \in R\})$, then we need $d^{k+1}$.

**Theorem 1**

**(Pumping lemma for Context-free languages)**

For every context-free language $L \subseteq \Sigma^*$, there exists $p \in \mathbb{N}$ such that for every $w \in L$ with $|w| \geq p$, there exist strings $u, v, x, y, z \in \Sigma^*$ such that

1. $w = uvxyz$
2. $|vy| > 0$
3. $|vxy| \leq p$
4. $\forall i \in \mathbb{N} \cup \{0\}$, we have $uv^ixy^iz \in L$.

*Proof.* $L = \mathcal{L}(G)$ for some grammar $G = (N, \Sigma, R, S)$. Let $k = |N|$, $d = \max(\{|\alpha| \mid A \to \alpha \in R\})$.

Choose $p = d^{k+1}$.

For any $w \in L$ with $|w| \geq p$, consider the smallest parse tree $T$ of $w$. Height of $T$ is at least $k + 1$. Let $P$ be a longest root-to-leaf path in $T$, so $P$ has $\geq k + 1$ non-leaf vertices. Consider the bottom-most $k + 1$ non-leaf vertices of $P$. There exist two vertices $l_1, l_2$ among those, whose label is the same non-terminal, say $A \in N$.

The yield of $l_1$ in this tree is some substring $w'$ of $w$, i.e., $w = uw'z$ for some $u, z \in \Sigma^*$.

The yield of $l_2$ in this tree is some substring $x$ of $w'$, i.e., $w' = vxy$ for some $v, y \in \Sigma^*$.

So, we have $w = uw'z = uvxyz$. $|vy| > 0$ by the minimality of $T$.

The height of the tree rooted at $l_1$ is at most $\leq k + 1$. Therefore, $|vxy| = |w'| \leq d^{k+1} = p$.

Observe that $S \overset{*}{\Longrightarrow} uAz$ (chop off the children of $l_1$ in $T$) and $A \overset{*}{\Longrightarrow} vAy$ (let $T'$ be the subtree of $T$ rooted at $l_1$, and chop off the children of $l_2$ in $T'$), and $A \overset{*}{\Longrightarrow} x$ (let $T''$ be the subtree of $T$ rooted at $l_2$).

$\forall i \in \mathbb{N} \cup \{0\}$, we have $S \overset{*}{\Longrightarrow} uAz \Longrightarrow uv^iAy^iz$ (by repeated application of the second relation), and using the last relation, we have $S \overset{*}{\Longrightarrow} uv^ixy^iz$. $\square$

The pumping game for context free languages $L \subseteq \Sigma^*$.

| Your opponent | You |
| --- | --- |
| $p \in \mathbb{N}$ | |
| | String $w$ with $|w| \geq p$, and $w \in L$ |
| $u, v, x, y, z \in \Sigma^*$ such that $w = uvxyz, |vy| > 0, |vxy| \leq p$ | |
| | $i \in \mathbb{N} \cup \{0\}$ |

You win if $uv^ixy^iz \notin L$, else your opponent wins.
The pumping lemma says that if $L$ is context-free, then your opponent has a winning strategy.
And if you have a winning strategy, then $L$ is not context-free.

**Example 1**

| Your opponent | You |
|---|---|
| $p$ | |
| | $a^p b^p c^p$ |
| $u, v, x, y, z \in \Sigma^*$ such that $w = uvxyz, \lvert vy \rvert > 0, \lvert vxy \rvert \leq p$ | so $vxy$ is either a substring of $a^p b^p$ or $b^p c^p$ |
| | Choose any $i \neq 1$ and reach a contradiction wrt number of $a$ |

**Claim 1.1**

$L = \{a^n b^n c^n \mid n \in \mathbb{N} \cup \{0\}\}$ is not context free.

*Proof.* As above. □

**Corollary 1**

The class of context-free languages is not closed under intersection.

*Proof.* Let $L_1 = \{a^m b^m c^n \mid m, n \in \mathbb{N} \cup \{0\}\}$ and $L_2 = \{a^m b^n c^n \mid m, n \in \mathbb{N} \cup \{0\}\}$. Both are context-free but $L_1 \cap L_2 = L$ is not context-free. □

**Corollary 2**

The class of context-free languages is not closed under complementation.

*Proof.* (Indirect): $L_1 \cap L_2 = (L_1^c \cup L_2^c)^c$, and we are done using the previous lemma.

(Direct): Directly show that $L^c$ is not context free for some language $L$. □

**Note 2**

Note that the language $\{w \cdot rev(w) \mid w \in \Sigma^*\}$ is context-free (can make a PDA for this).

**Claim 1.2**

The language $\{ww \mid w \in \Sigma^*\}$ is not context-free.

*Proof.*

| Your opponent | You |
|---|---|
| $p$ | |
| | $a^p b^p a^p b^p$ |
| $u, v, x, y, z \in \Sigma^*$ such that $w = uvxyz, \lvert vy \rvert > 0, \lvert vxy \rvert \leq p$ | so $vxy$ is either a substring of $a^p b^p$ or $b^p a^p$ |
| | Choose any $i \neq 1$ and reach a contradiction. |

□