Non-Context-Free Languages

COMPSCI 3331

Outline

- Non-context-free Languages.
- Pumping Lemma for CFLs.
- Heuristics for CFLs and non-CFLs.
- Closure Properties of CFLs.

Non-Context-free Languages

- Not every language is a CFL.
- In PDAs, the stack can only be used to remember the most **recent** thing.
  - e.g., $L = \{a^n b^m a^n b^m \ : \ n, m \geq 0\}$.
- Once stack contents are used, there is no way to 'remember' them again.
  - e.g., $L = \{a^n b^n c^n \ : \ n \geq 0\}$.
- Need a formal method for proving languages not to be CFLs.

Pumping Lemma for CFLs

- Pumping lemma for regular languages: used DFA.
- Can prove a pumping lemma for CFLs using PDAs.
- Using CFGs is easier.

Pumping Lemma for CFLs

Theorem. Let L be a CFL. Then there exists a constant $n \geq 0$ such that for all $z \in L$ with $|z| \geq n$, there exists $u, v, w, x, y \in \Sigma^*$ such that $z = uvwxy$ with

- $|vwx| \leq n$;
- $vx \neq \varepsilon$;
- for all $i \geq 0$, $uv^iwx^iy \in L$.

$z = u\underbrace{vwx}_{\leq n}y$

Pumping Lemma for CFLs

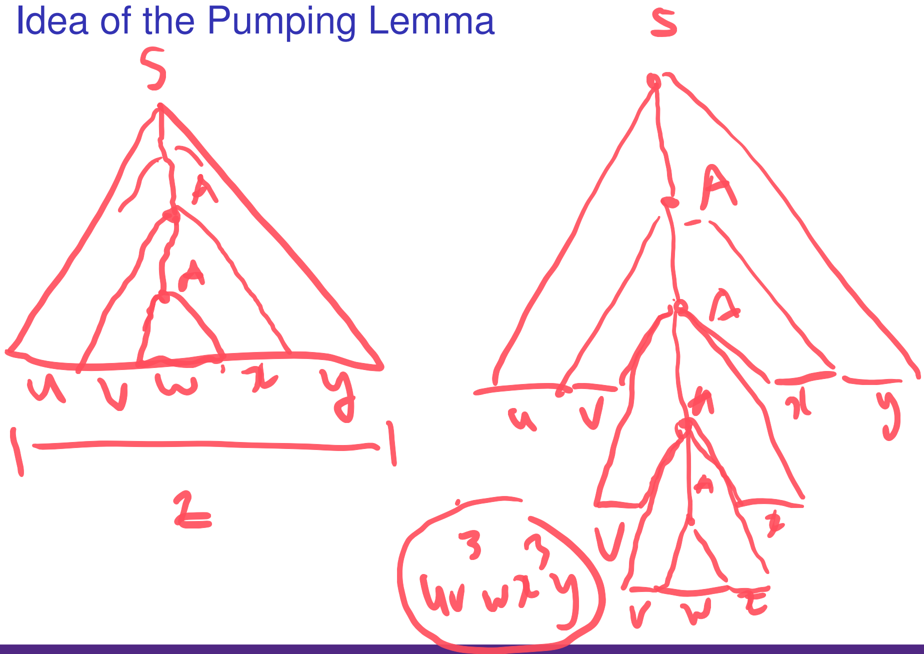
- Notice that *vwx* aren't at the start of the word anymore!
- Can only use the pumping lemma to prove that languages **aren't** CFLs

Let L be a language such that for all $n \geq 0$, there exists a $z \in L$ with $|z| \geq n$, such that for all ways of writing $z = uvwxy$ such that

- $|vwx| \leq n$;
- $vx \neq \varepsilon$,

if there always exists an $i \geq 0$ such that $uv^i wx^i y \notin L$, then the language L is not a CFL.

Idea of the Pumping Lemma



Heuristics for Non-CFLs

- The stack can only recall the most recent thing it stored.
- Once something is removed from the stack, it cannot be recovered!

Typical non-context-free “forms”

- Multiple agreement: $L = \{a^n b^n c^n \ : \ n \geq 0\}$.
- Cross agreement: $L = \{a^n b^m a^n b^m \ : \ n, m \geq 0\}$.
- Repetition: $L = \{ww \ : \ w \in \Sigma^*\}$.

Do not confuse ..

- Multiple agreement ($\{a^n b^n c^n : n \geq 0\}$) with independent agreement ($\{a^n b^n a^m b^m : n, m \geq 0\}$).
- Cross agreement ($\{a^n b^m a^n b^m : n, m \geq 0\}$) with nested agreement ($\{a^n b^m b^m a^n : n, m \geq 0\}$).
- Repetition ($\{ww : w \in \Sigma^*\}$) with reversed-repetition ($\{ww^R : w \in \Sigma^*\}$).

$a^n b^m c^m d^n$

Closure Properties for CFLs

Theorem. Let $L_1, L_2 \subseteq \Sigma^*$ be CFLs. Then $L_1 \cup L_2$, $L_1 L_2$ and L_1^* are CFLs.

Proof Let $G_i = (V_i, \Sigma, P_i, S_i)$ be CFGs such that $L(G_i) = L_i$ for $i = 1, 2$.

- $G_{\text{union}} = (V_1 \cup V_2 \cup \{S\}, \Sigma, P, S)$ where $P = P_1 \cup P_2 \cup \{S \to S_1 | S_2\}$.
- $G_{\text{cat}} = (V_1 \cup V_2 \cup \{S\}, \Sigma, P, S)$ where $P = P_1 \cup P_2 \cup \{S \to S_1 S_2\}$.
- $G_{\text{star}} = (V_1 \cup \{S\}, \Sigma, P, S)$ where $P = P_1 \cup \{S \to \varepsilon | S_1 S\}$.

Corollary. Every regular language is also a CFL.

Non-closure Properties for CFLs

Theorem. The CFLs are not closed under intersection.

Proof. Consider the following languages:

$$\begin{aligned} L_1 &= \{a^n b^n c^m \ : \ n, m \geq 0\} \\ L_2 &= \{a^m b^n c^n \ : \ n, m \geq 0\} \end{aligned}$$

Then L_1, L_2 are CFLs (exercise). However,

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

which is not a CFL.

Non-closure Properties for CFLs

Corollary. The CFLs are not closed under complement.

Proof. By contradiction. Assume that for all CFLs L, $\overline{L}$ is also a CFL. Consider that for all CFLs L_1, L_2 ...

1. $\overline{L_1}$, $\overline{L_2}$ are CFLs.
2. $\overline{L_1} \cup \overline{L_2}$ is a CFL.
3. $\overline{\overline{L_1} \cup \overline{L_2}}$ is a CFL
4. $L_1 \cap L_2$ is a CFL (de Morgan's law)

Closure properties using PDAs

Theorem. Let L be a CFL and R be a regular language. Then $R \cap L$ is a CFL.

Proof. Construction of a PDA.

$L(M_1) = L$

$L(M_2) = R$

$$\begin{aligned} M_1 &= (Q_1, \Sigma, \Gamma, \delta_1, q_1, Z_0, F_1) \\ M_2 &= (Q_2, \Sigma, \delta_2, q_2, F_2) \end{aligned}$$

PDA

DFA.

Assume M_1 accepts by final state. Then let

$$M = (Q_1 \times Q_2, \Sigma, \Gamma, \delta, [q_1, q_2], Z_0, F_1 \times F_2)$$

- This can be very helpful in showing non-closure of CFLs.
- Instead of showing L is not a CFL, pick R and show $L \cap R$ is not a CFL.

Conclusions

- Pumping lemma: showing that languages are not context-free.
- Closure properties: CFLs are closed under union, concatenation, Kleene closure.
- Non-closure properties: not closed under intersection, complement.
- Where to from here? Next step: Turing Machines.
- More powerful formalism than CFGs/PDAs.
- Next lecture: Determinism in PDAs.