Lecture 11

- HW2 due Monday Oct 16

Today: Practice Problems
(Pumping Lemma, PDA, CFg's)

Next Week
- Mon: Equivalence Between PDA and CFGs
- Wed: Review for Test 1

10/23
- Mon: Test 1 (in class)

Pumping Lemma for Context Free Languages

Lemma Let L be a CFL, $L \subseteq \Sigma^*$

Then there exists a number $p \geq 0$ such that for any string $w \in L$, $|w| \geq p$, $\exists$ strings $u, v, x, y, z \in \Sigma^*$ such that:

(0) $w = uvxyz$ and

(1) $|vxy| \leq p$ and $|vy| \geq 1$ and

(2) For every $i \geq 0$, $w^i = uv^i x y^i z$ is also in L.

Pumping Lemma for regular Languages: $\exists p \dots$ $|w| \geq p$

can write $w = xyz$, $|xy| \leq p$, $|y| \geq 1$

such that $\forall i$ $w \in L$ iff $w^i = xy^i z \in L$

Example 1 We will show $L = \{a^n b^n c^n \mid n \geq 0\}$ is not a CFL.

Let $w = a^{2p} b^{2p} c^{2p}$, where p is from the pumping lemma

then By pumping Lemma $w = uvxyz$, $|vxy| < p$, $|vy| \geq 1$

Case 1: v and y contain only a's. Then uv^2xy^2z will contain too many a's

Case 2: v and y contain only b's or only c's. Same arg as Case 1

Case 3 v and y together contain both a's and b's. then uv^2xy^2z will contain too few c's. Similar if v and y together contain b's + c's

There are no other cases since $|vxy| < p$ so vxy can't contain all three symbols a, b, c

Example 2 $L = \{ r\#s \mid r, s \in \{0,1\}^* \text{ and } r \text{ is a substring of } s \}$

Let p be pumping length; Let $w = 0^p 1^p \# 0^p 1^p \in L$.

then $w = uvxyz$, $|vxy| \leq p$, $|vy| \geq 1$

Case 1 v contains $\#$. Then $w^2 = uv^2xy^2z$ has 2 $\#$'s so $\notin L$

Case 2 y " $\#$. same argument as Case 1

Case 3 u contains $\#$. then $w^0 = uv^0xy^0z$ has fewer symbols after the $\#$ than before so $w^0 \notin L$

Case 4 z contains $\#$ then $w^2 = uv^2xy^2z$ has more than $2p$ symbols before "$\#$" and $2p$ after "$\#$" so $\notin L$

Case 5 x contains $\#$

Case (5a): $v \neq \varepsilon$

Case (5b): $v = \varepsilon$. then $y \neq \varepsilon$. So $w^0 = uv^0xy^0z$ has more symbols to left of $\#$ than to right

Example 3

$$L = \{ 0^n 1^m 0^n 1^m \mid n,m \geq 0 \}$$

Let p be pumping length, $w = 0^p 1^p 0^p 1^p$.

then $w = uvxyz$, $|vxy| \leq p$ $|vy| \geq 1$

Case 1 vxy in some 0-block or in some 1-block (vxy all 0's or all 1's)

$w^2 = uv^2xy^2z$ then w^2 has 3 blocks of length p and one block of length $> p$.

Case 2 v in two blocks, or y in two blocks,
then $w^2 = uv^2xy^2z$ has 6 blocks of alternating 0's and 1's.

Case 3 v in one block, y in next block.
then since either v or y is nonempty, either the 2 blocks of 0's in uv^2xy^2z have different lengths or the 2 blocks of 1's in uv^2xy^2z have different lengths

Example 4 $\quad L = \{ w \mid w = s \cdot s,\ s \in \{0,1\}^* \}$

1st try: $w = \underbrace{0^p 1}\underbrace{0^p 1} \in L$.

suppose $w = \underbrace{000\cdots}_{u}\underbrace{0}_{v}\underbrace{1}_{x}\underbrace{0}_{y}\underbrace{00\cdots 01}_{z}$

then $w^i \in L$ for all i, so this choice of w won't work to prove L not CFL

Example 4) $L = \{w \mid s \cdot s,\ s \in \{0,1\}^*\}$

2nd try: $w = 0^p 1^p 0^p 1^p$.

then w can be written as $w = uvxyz$, $|vxy| \le p$, $|vy| \ge 1$

write $w = abcd$, $a = 0^p$, $b = 1^p$, $c = 0^p$, $d = 1^p$

Case 1 vxy is a substring of a. Then $w^2 = uv^2xy^2z$ is of the form $a'bcd$, where $|a'| < p$ so Not in L

Same argument if vxy is a substring of c

Case 2 vxy is a substring of b. Then w^2 is of the form $ab'cd$ where $|b'| < p$, so Not in L. Same argument if vxy is a substring of d

Case 3 vxy is a substring of ab (containing at least one 0 + one 1) then $w^2 = uv^2xy^2z$ has too many groups of 1's

Similar arg if vxy a substring of bc, or of cd

Note this covers all cases since $|vxy| \le p$.

Another way to prove that a Language is not Context-free is by using closure properties.

1. Context free Languages are closed under union:

   If L_1 and L_2 are both context-free then so is $L = L_1 \cup L_2$

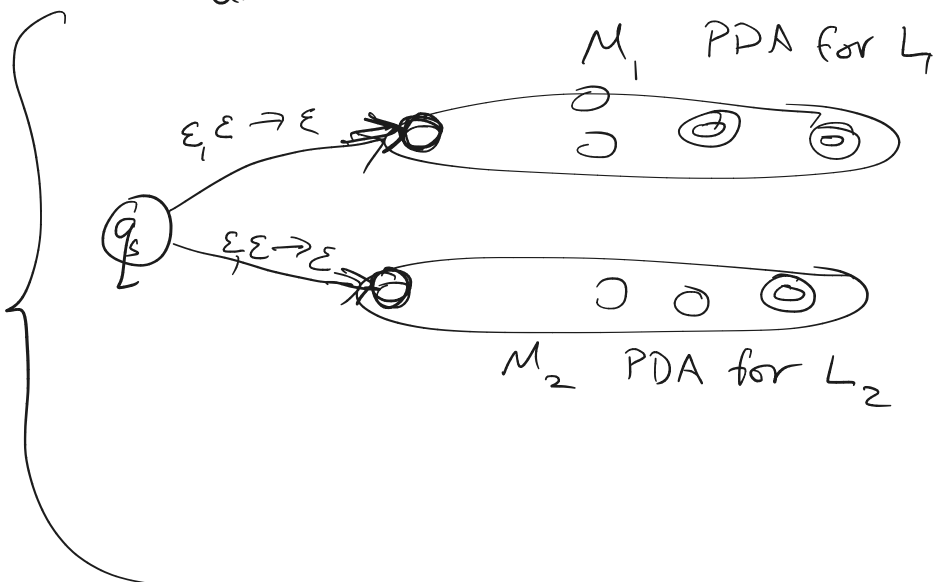
2. CFL's are closed under intersection with a regular Language: If L_1 is a CFL, and L_2 is a regular Language, then $L = L_1 \cap L_2$ is also CFL.

3. CFL's closed under concatenation

① For $L_1 \cup L_2$

Proof for PDA (closed under union)
essentially same as proof Reg Lang's closed under union.

M for L:



Q: Suppose L_1 is generated by CFG $G_1 = (V_1, \Sigma, R_1, S_1)$
L_2 " $G_2 = (V_2, \Sigma, R_2, S_2)$

How do we construct a CFG, G, for $L_1 \cup L_2$?

$G = (V, \Sigma, R, S)$

$V = V_1 \cup V_2 \cup \{S\}$

$R = \{S \to S_1 | S_2\} \cup R_1 \cup R_2$

BUT CFL's are <u>not</u> closed under:

- general intersection
  ($\exists$ CFL's L_1 and L_2, but $L = L_1 \cap L_2$ is NOT CFL)

- complement
  ($\exists L \subseteq \Sigma^*$ that is CFL, but $\overline{L} \notin$ CFL)

More Practice Problems on CFL's

(1.) Prove that $L = \{0^n \# 0^{2n} \# 0^{3n} \mid n \geq 0\}$ is Not a CFL

Let p be the pumping Length, and Let $w = 0^p \# 0^{2p} \# 0^{3p}$.
We will write $w = a \# b \# c$. Let $w = uvxyz$, $|vxy| \leq p$, $|vy| \geq 1$

Case (i) If vxy is a substring of a, then $w^0 = uv^0xy^0z$ is Not in ~~b~~ since $w^0 = a' \# b \# c$ where $|a'| < p$.

Case (ii) vxy is a substring of b. Then $w^2 = uv^2xy^2z$ since $w^2 = a \# b' \# c$, where b' has too many 0's

Case (iii) vxy is a substring of c. Again $w^2 \notin L$

Case (iv) vxy is a substring of $a \# b$. Then w^0 is Not of proper form (either is missing a #, or has too few 0's in a-part or b-part

Case (v) vxy is a substring of $b \# c$. Similar to (iv)

(2) $L = \{0^n 1^m 1^n 2^m \mid n, m \geq 0\}$.

Is L CFL?

we saw before

$L = 0^n 1^m 0^n 1^m$

is not CFL

$L = \{0^n 1^{n+m} 2^m \mid n, m \geq 0\}$

(2) $L = \{0^n 1^m 1^n 2^m \mid n, m \geq 0\}$.

Yes! L is a CFL $\qquad L = \{0^n \underline{1}^{m+n} 2^m\}$

Push 'X' when we see a 0

Then pop off 0 when we see a 1

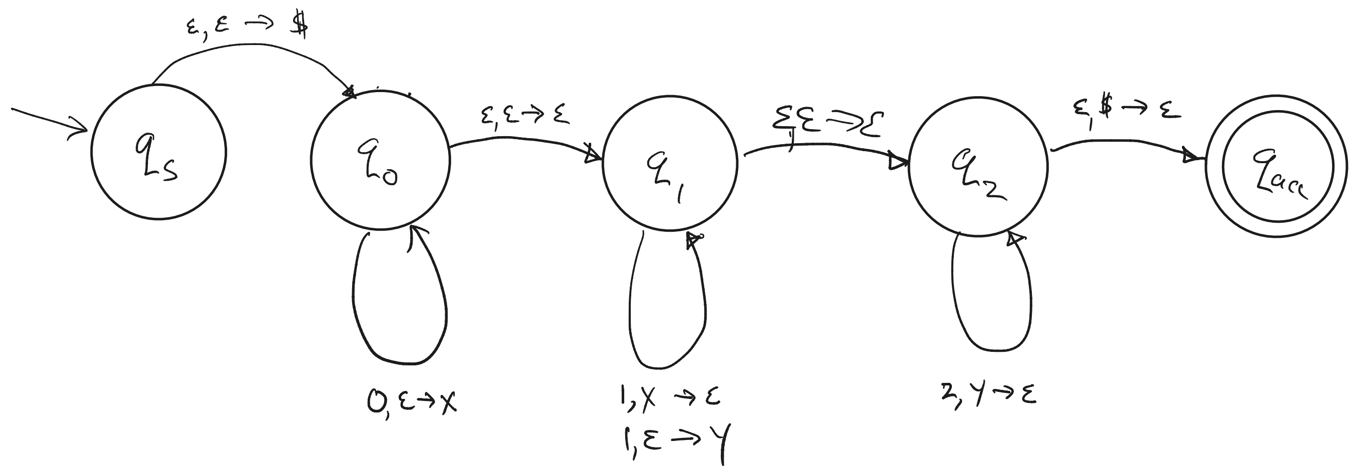
If we reach bottom of stack ('\$') when reading a 0,
start pushing y's (nondeterministically guess when to pop X's or when to push y's)

When reading 2's, pop them with Y symbols

If stack is empty ('\$) when we reach end of string, accept

(2) $L = \{0^n 1^m 1^n 2^m \mid n, m \geq 0\}$.

Yes! L is a CFL $\qquad L = \{0^n \underline{1}^{m+n} 2^m\}$



$$W = \underbrace{\underbrace{0\ 0\ 0}_{n}}_{\text{Push 'x' on '0'}}\ \underbrace{\underbrace{1\ 1\ 1}_{n}}_{\text{Pop 'x' on '1'}}\ \underbrace{\underbrace{1\ 1\ 1\ 1}_{m}}_{\text{Push 'y' on '1'}}\ \underbrace{\underbrace{2\ 2\ 2\ 2}_{m}}_{\text{pop 'y' on 2}}$$

(3) Is $L = \{0^n 1^m 2^n 1^m \mid m,n \geq 0\}$ a CFL? No!

$0^n 1^m 0^n 1^m$

Let p be pumping length. Let $w = 0^p 1^p 2^p 1^p$. $w \in L$.

By PL, can write $w = uvxyz$, $|vxy| \leq p$, $|vy| \geq 1$ and $\forall i\ uv^ixy^iz \in L$

Write $w = abcb'$

Case (1): vxy contained in a.

Then $w^0 = uv^0xy^0z = uxz$ has too few a's so $w^0 \notin L$

Similarly if vxy contained in b, or contained in c, or in b'

Case (2) vxy contained in ab (but not in a or in b)

$w^0 = uxy$. If $|v| \geq 1$ then $w^0 = 0^{n'} 1^{m} 2^n 0^m$ $n' < n$

If $|y| \geq 1$ then $w^0 = 0^{n'} 1^{m'} 2^n 0^m$ $m' < m$

Similar argument if vxy contained in bc or in cd

Another argument for why $L = 0^n 1^m 2^n 1^m$ is not a CFL:

Suppose that L is a CFL and let $G = (V, \Sigma, R, S)$ be a CFG generating L.

Then let $G' = (V, \Sigma, R', S)$ be obtained from G by replacing every occurrence of '2' in the rules of R by a '0' to get the new rules R'.

Can prove (by induction on $|w|$) that this gives is a 1-to-1, onto mapping from strings $w = 0^n 1^m 2^n 1^m$ generated by G to strings $w' = 0^n 1^m 0^n 1^m$ generated by G'.

$\therefore L(G') = \{0^n 1^m 0^n 1^m \mid n, m \geq 0\}$ is also a CFL.

But this contradicts fact that we already proved that $L(G')$ is not a CFL. ✓

t4) give a CFG for:

$L_1 = \{ w\#x \mid x = w^R y \text{, for some } y \in \{0,1\}^* \}$

$L_0 = \{ w\#x \mid x = w^R \}$

$L_2 = \{ w\#x \mid x \text{ contains the substring } w^R \}$

L_1 contains $\underbrace{0011}_{w} \# \overbrace{\underbrace{1100}_{w^R} \underbrace{1001}_{y}}^{x}$

L_2 contains $\underbrace{0011}_{w} \# \underbrace{00110}_{\text{any string}} \underbrace{1100}_{w^R} \underbrace{1101}_{\text{any string}}$

$$L_0 = \{ w \# w^R \mid w \in \{0,1\}^* \}$$

$$A \rightarrow 0A0 \mid 1A1 \mid \#$$

$$L_1 = \{ w \# w^R y \mid y \in \{0,1\}^* \}$$

$$A \rightarrow 0A0 \mid 1A1 \mid \#$$ generates $\{w \# w^R\}$

$$B \rightarrow \varepsilon \mid B0 \mid B1$$ generates all strings in $\{0,1\}^*$

$$S \rightarrow AB$$ generates all strings in L_1

$L_1 = \{ w \# w^R y \mid y \in \{0,1\}^* \}$

$A \to 0A0 \mid 1A1 \mid \#$

$B \to \varepsilon \mid B0 \mid B1$

$S \to AB$

$L_2 = \{ w \# y' w^R y \mid y, y' \in \{0,1\}^* \}$

$A' \Rightarrow 0A0 \mid 1A1 \mid \#B$

$B \to \varepsilon \mid B0 \mid B1$

$S \to A'B$

L_1:

$$\begin{cases} S_1 \to AB \\ A \to 0A0 \mid 1A1 \mid \# \\ B \to B0 \mid B1 \mid 0 \mid 1 \mid \varepsilon \end{cases}$$

generates $w\#w^R$

B generates $\{0,1\}^*$

L_2:

$$\begin{cases} S_1 \to A'B \\ A'' \to \overbrace{0A0} \mid 1A1 \mid \overbrace{\#B} \\ B \to \text{same} \end{cases}$$

generates $w\#\underbrace{(0+1)^*}w^R$

$= w\#Bw^R$

$\#$

$\underbrace{011}_{w}\#\underbrace{\text{arb}}\ \text{strng}\ \underbrace{110}_{w^R}$

(5) give a PDA for:

$L = \{ w \# x \mid w^R \text{ is a substring of } x \}$

(b) Describe the language generated by $\mathcal{G}$, in English.

$$\mathcal{G}: \quad \begin{aligned} S &\to TT \mid U \\ T &\to 111T1 \mid \# \\ U &\to 0U00 \mid \# \end{aligned}$$

T generates $\{1^{3m} \# 1^{m} \mid m \geq 0\}$

U generates $\{0^{n} \# 0^{2n} \mid n \geq 0\}$

S generates the language $L = L_1 \cup L_2$

$$L_1 = \{ \underbrace{1^{3m} \# 1^{m}} \; 1^{3n} \# 1^{n} \mid n, m \geq 0 \}$$ ← generated by TT

$$L_2 = \{0^{n} \# 0^{2n} \mid n \geq 0\}$$ ← strings generated by U