Lab 11 is a non-programming homework. It is due in two weeks, on November 22 at 6:00pm. We will be covering all of the topics mentioned in this assignment over the next week and a half. Please turn in your answers as a PDF file.

Question 0. Use the Myhill-Nerode Theorem to prove that the language $\{w \mid w \in \Sigma^*, rev(w) = w\}$ of all **palindromes** is not regular.

$\{\varepsilon, a, aa, aaa, aaaa \ldots\}$

$w_i \in L$, $w_j \in L$

if $i=2$ $j=4$

$w_{24} = baa$

then

$w_i \cdot w_{ij} \in L$

$w_j \cdot w_{ij} \notin L$

Ex: $w_i = aa$ $w_j = aaaa$

$w_{ij} = baa$ $i=2$ $j=4$

$w_i \cdot w_{ij} = aa.baa$ ✓

$w_j \cdot w_{ij} = aaaa.baa$ X

Question 1. Consider the following context-free grammar G_1, with start symbol S:

$S \longrightarrow ABS \mid AB$

$A \longrightarrow aA \mid a$

$B \longrightarrow bA.$

Which of the following strings are in $L(G_1)$ and which are not? Provide derivations or parse trees for those that are in $L(G_1)$ and reasons for those that are not:

1. aabaab – not a part of $L(G_1)$ as a string cannot end on b since every b is followed by an a
2. aaaaba – S->AB->aAB->aaAb->aaaAB->aaaaB->aaaabA->aaaaba (Hence a part of $L(G_1)$)
3. aabbaa – not a part of $L(G_1)$ as you cannot repeat b's B->bA->baA
4. abaaba – S->ABS->aBS->abAS->abaS->abaAB->abaaB->abaabA->abaaba (Hence a part of $L(G_1)$)

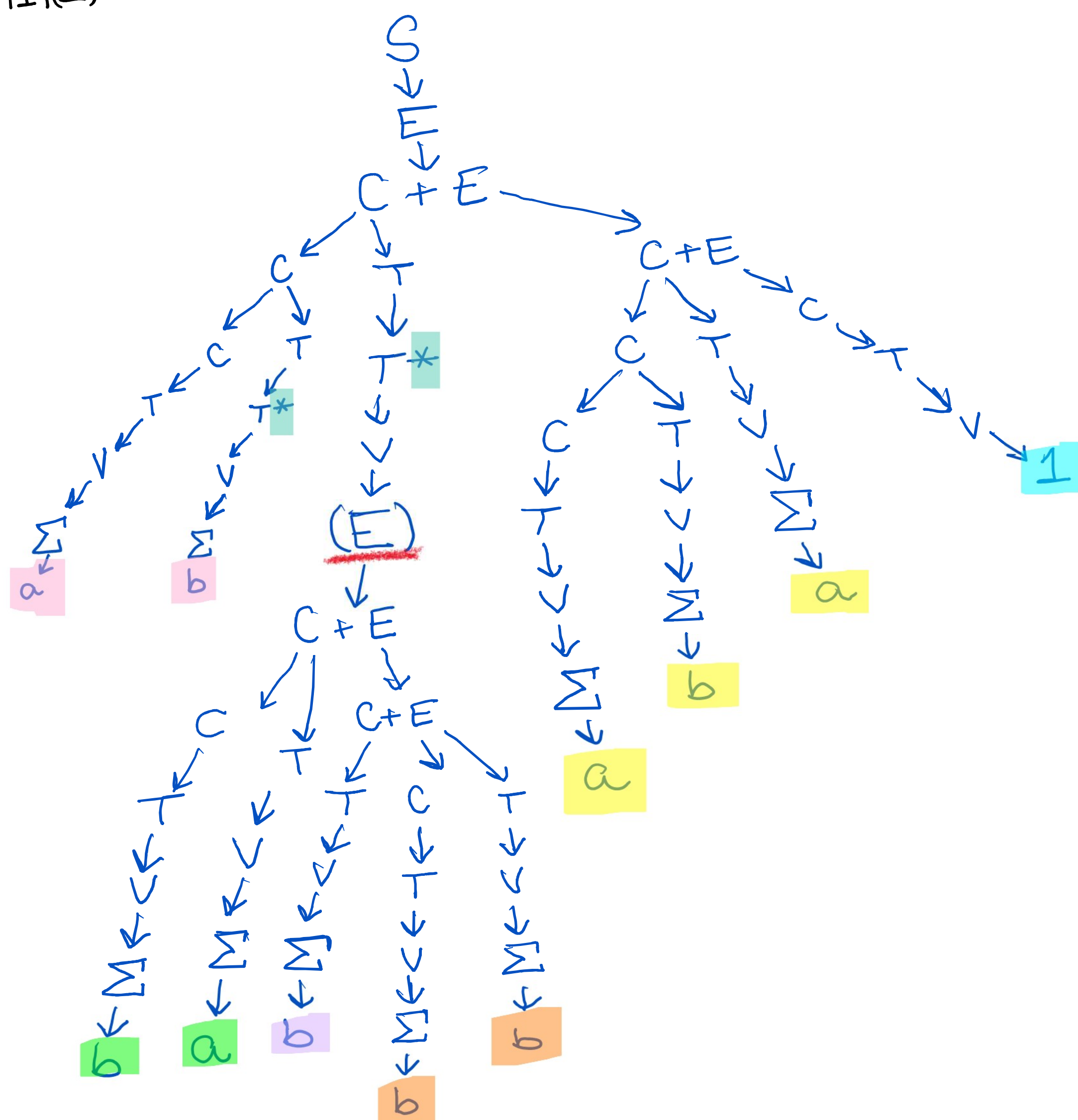
Question 2. Let Σ be a finite alphabet, and define $\Sigma'=\Sigma\cup\{0,1,+,*,(,)\}$. Find a context-free grammar G_2 with Σ' as the set of terminals which generates exactly the *regular expressions* over the alphabet Σ. Your grammar should be such that regular expressions are parsed *unambiguously*, where

- $*$ is postfix and has the highest precedence
- concatenation is left-associative and has the next highest precedence
- $+$ is right-associative and has the lowest precedence.

Show the derivation or parse tree of the string $r=ab*(ba+b+bb)*+aba+1$ in G_2.

$S \to E$
$E \to C \mid C+E$
$C \to CT \mid T$
$T \to T* \mid V$
$V \to \Sigma \mid 0 \mid 1 \mid (E)$

$S = ab*(ba+b+bb)*+aba+1$



Question 3. Find a context-free grammar G_3 that generates exactly the same strings as the regular expression r at the end of Question 2 (i.e., for which $L(G_3)=[[r]]$).

$S \to aba \mid M \mid \varepsilon$

$M \to aYN$

$Y \to bY\varepsilon$

$N \to baN \mid bN \mid bbN \mid \varepsilon$

$$\overbrace{ab^*\underbrace{(ba+b+bb)^*}_{N}}^{M} + aba + 1$$

Question 4. A CFG $G=(N,S,R)$ over Σ is in *Chomsky normal form* if every rule in R is of the form $A \longrightarrow B_1B_2$ or $A \longrightarrow a$, where $B_1,B_2 \in N$ and $a \in \Sigma$. Give a grammar in Chomsky normal form that generates the language $\{a^ib^ja^i \mid i \geq 0, j \geq 1\}$.

$\{a^i b^j a^i \mid i \geq 0, j \geq 1\}$

b, aba, aabaa, . . .

CFG :- $S \to aSa \mid B$

$B \to bB'$

$B' \to bB' \mid \varepsilon$

CNF : - No epsilon
- No unit rules
- Variable for letters

$S \to AY \mid XB' \mid b$

$B \to XB' \mid b$

$B' \to XB' \mid b$

$A \to a$

$X \to b$

$Y \to SA$

Question 5. Let $\Sigma=\{a,b\}$. Prove that the CFG G_5 with rules

$S \longrightarrow aSb|bSa|SS|\varepsilon$

generates the set of all strings in Σ_* with an equal number of as and bs. Do this by defining two recursive functions

$a,b:\Sigma_* \rightarrow N$

such that $a(w)$ is the number of as in w and $b(w)$ is the number of bs in w and finding and proving the appropriate condition on w and its prefixes, as we did with the balanced-parentheses grammar.

Lemma 1: $\forall w$, $w \in \{S, a, b\}^*$, if $S \to^* w$ then $a(w)=b(w)$

Lemma 2: $\forall w$, $w \in \{a,b\}^*$ if $a(w)=b(w)$ then $S \to^* w$

By string induction on $|w|$. Assume that (L_2) is true for all strings w $|w|<n$

$S \to aSb|bSa|SS|\varepsilon$ $\#a = \#b$

$S \to ab|ba|SS|\cancel{S}|\cancel{S}$

$S \to ab|ba|SS$

$S \to AB|BA|SS$

$A \to a$

$B \to b$

Due to rules of CNF both a and b occur the same amount of time.

Question 6. What set is generated by the following grammar?

$S \longrightarrow bS|Sa|aSb|\varepsilon$

Give a proof (in the same style as in Question 5).

bS → string starts with b

Sa → String ends with a

aSb → String starts with a ends with b

$S \to bS|Sa|aSb|\varepsilon$

$S \to b|a|ab$

$S \to B|A|AB$

$A \to a$

$B \to b$

Question 7 (optional extra credit - worth an additional 2 points). A CFG $G=(N,S,R)$ over Σ is in *Greibach normal form* if every rule in R is of the form $A \longrightarrow aB_1B_2 \cdots B_n$, where $a \in \Sigma$, each $B_i \in N$, and $n \geq 0$. (Note that the case $n=0$ means that the rule looks like $A \longrightarrow a$ for $a \in \Sigma$.) Find a CFG G_7 in Greibach normal form such that $L(G_7)=L(G_5)-\{\varepsilon\}$, i.e., that generates exactly the nonempty strings generated by the grammar in Question 5. Prove this. Note that, since G_7 will necessarily have multiple nonterminal symbols, you will have to find a condition for each and then prove all of these conditions simultaneously.