CMSC 330 Practice Problem 4 Solutions

1. Context Free Grammars
   a. List the 4 components of a context free grammar.
      **Terminals, non-terminals, productions, start symbol**
   b. Describe the relationship between terminals, non-terminals, and productions.
      **Productions are rules for replacing a single non-terminal with a string of terminals and non-terminals**
   c. Define ambiguity.
      **Multiple left-most (or right-most) derivations for the same string**
   d. Describe the difference between scanning & parsing.
      **Scanning matches input to regular expressions to produce terminals, parsing matches terminals to grammars to create parse trees**
   e. Describe an abstract syntax tree (AST)
      **Compact representations of parse trees with only essential parts**

2. Describing Grammars
   a. Describe the language accepted by the following grammar:
      \[ S \rightarrow \text{abS} \mid a (ab)^n a \]
   b. Describe the language accepted by the following grammar:
      \[ S \rightarrow \text{aSb} \mid \varepsilon \]
      \[ a^n b^n, n \geq 0 \]
   c. Describe the language accepted by the following grammar:
      \[ S \rightarrow \text{bSb} \mid A \]
      \[ A \rightarrow \text{aA} \mid \varepsilon \]
      \[ b^n a^n b^n, n \geq 0 \]
   d. Describe the language accepted by the following grammar:
      \[ S \rightarrow \text{AS} \mid B \]
      \[ A \rightarrow \text{aAc} \mid \text{Aa} \mid \varepsilon \]
      \[ B \rightarrow \text{bBb} \mid \varepsilon \]
      **Strings of a & c with same or fewer c’s than a’s and no prefix has more c’s than a’s, followed by an even number of b’s**
   e. Describe the language accepted by the following grammar:
      \[ S \rightarrow \text{S and S} \mid \text{S or S} \mid (S) \mid \text{true} \mid \text{false} \]
      **Boolean expressions of true & false separated by and & or, with some expressions enclosed in parentheses**
   f. Which of the previous grammars are left recursive? **2d, 2e**
   g. Which of the previous grammars are right recursive? **2a, 2c, 2d, 2e**
   h. Which of the previous grammars are ambiguous? Provide proof.
      **Examples of multiple left-most derivations for the same string**
      **2d:**
      \[ S \Rightarrow \text{AS} \Rightarrow \text{AaS} \Rightarrow \text{aS} \Rightarrow \varepsilon \Rightarrow \text{a} \]
      \[ S \Rightarrow \text{AS} \Rightarrow \text{S \Rightarrow AS} \Rightarrow \text{AaS} \Rightarrow \text{aS} \Rightarrow \varepsilon \Rightarrow \text{a} \]
      **2e:**
      \[ S \Rightarrow \text{S and S \Rightarrow S and S and S \Rightarrow true and S and S} \]
      \[ \Rightarrow \text{true and true and S \Rightarrow true and true and true} \]
      \[ S \Rightarrow \text{S and S \Rightarrow true and S \Rightarrow true and S and S} \]
      \[ \Rightarrow \text{true and true and S \Rightarrow true and true and true} \]
3. Creating Grammars
   a. Write a grammar for $a^x b^y$, where $x = y$
      \[ S \rightarrow aSb | \varepsilon \]
   b. Write a grammar for $a^x b^y$, where $x > y$
      \[ S \rightarrow aL \quad L \rightarrow aL | aLb | \varepsilon \]
   c. Write a grammar for $a^x b^y$, where $x = 2y$
      \[ S \rightarrow aaSb | \varepsilon \]
   d. Write a grammar for $a^x b^y a^z$, where $z = x+y$
      \[ S \rightarrow aSa | aLa | aLb | \varepsilon \]
   e. Write a grammar for $a^x b^y a^z$, where $z = x-y$
      \[ S \rightarrow aSa | aLa | aLb | \varepsilon \]
   f. Write a grammar for all strings of $a$ and $b$ that are palindromes.
      \[ S \rightarrow aSa | bSb | L \quad L \rightarrow a | b | \varepsilon \]
   g. Write a grammar for all strings of $a$ and $b$ that include the substring $baa$.
      \[ S \rightarrow LbaaL \quad L \rightarrow aL | bL | \varepsilon \] // $L = any$
   h. Write a grammar for all strings of $a$ and $b$ with an odd number of $a$’s and an odd
      number of $b$’s.
      \[ S \rightarrow EaEbE | EbEaE \quad E \rightarrow EaEaE | EbEbE | \varepsilon | SS \quad // E = even #s \]
   i. Write a grammar for the “if” statement in OCaml
      \[ S \rightarrow if E then E else E | if E then E \quad E \rightarrow S | expr \]
   j. Write a grammar for all lists in OCaml
      \[ S \rightarrow [] | [E] | E::S \quad E \rightarrow elem | S \quad // Ignores types, allows lists of lists \]
   k. Which of your grammars are ambiguous? Can you come up with an
      unambiguous grammar that accepts the same language?
      Grammar for 3h is ambiguous. An unambiguous grammar must exist
      since the language can be recognized by a deterministic finite automaton,
      and DFA -> RE -> Regular Grammar.
      Grammar for 3i is ambiguous. Multiple derivations for “if expr then if
      expr then expr else expr”. It is possible to write an unambiguous
      grammar by restricting some $S$ so that no unbalanced if statement can be
      produced.

4. Derivations, Parse Trees, Precedence and Associativity
   For the following grammar: $S \rightarrow S$ and $S$ | $true$
   a. List 4 derivations for the string “true and true and true”.
      i. S => S and S => S and S and S => true and true and true
         and S => true and true and true
      ii. S => S and S => true and S => true and S and S => true and true
           and S => true and true and true
      iii. S => S and S => S and true => S and S and true => S and S and true
           => true and true and true
          => true and true and true
      v. S => S and S => S and S and S => true and S and S and S => true and true and true
b. Label each derivation as left-most, right-most, or neither.
   i and ii are left-most derivations, iii and iv are right-most derivations, remaining derivations are neither

c. List the parse tree for each derivation
   Tree 1 = ii, iii, x, xi, Tree 2 = rest

   ![Parse Trees](parse_trees.png)

   d. What is implied about the associativity of “and” for each parse tree?
   Tree 1 => and is right-associative, Tree 2 => and is left-associative

For the following grammar:  
S →  S and S | S or S | true

e. List all parse trees for the string “true and true or true”
f. What is implied about the precedence/associativity of “and” and “or” for each parse tree?

Tree 1 => or has higher precedence than and
Tree 2 => and has higher precedence than or

g. Rewrite the grammar so that “and” has higher precedence than “or” and is right associative

\[
S \rightarrow S \text{ or } S | L \quad // \text{ op closer to Start = lower precedence op}
L \rightarrow \text{ true and } L | \text{ true} \quad // \text{ right recursive = right associative}
\]

5. Left factoring

Rewrite the following grammars so they can be parsed by a predicative parser by applying left factoring where necessary

a. \[
S \rightarrow a \ b \ c | a \ c
\]

\[
\downarrow
\]

\[
S \rightarrow a \ L
L \rightarrow b \ c | c
\]

b. \[
S \rightarrow a \ a | a \ b | a
\]

\[
\downarrow
\]

\[
S \rightarrow a \ L
L \rightarrow a | b | \varepsilon
\]

c. \[
S \rightarrow a \ b \ A \ c | a \ b \ B \ a
\]

\[
\downarrow
\]

\[
S \rightarrow a \ b \ L
L \rightarrow A \ c | B \ a
\]

d. \[
S \rightarrow a \ a \ A | a \ a \ a \ B | a \ c
\]

\[
\downarrow
\]

\[
S \rightarrow a \ L
L \rightarrow a \ A | a \ a \ B | c
\]

\[
\downarrow
\]

\[
S \rightarrow a \ L
L \rightarrow a \ M | c
M \rightarrow A | a \ B
\]
6. Parsing

For the problem, assume the term “predictive parser” refers to a top-down, recursive descent, non-backtracking predictive parser.

a. Consider the following grammar: \( S \rightarrow S \) and \( S \) or \( S \) | \( (S) \) | \( \text{true} \) | \( \text{false} \)

i. Compute First sets for each production and nonterminal

\[
\text{First(true)} = \{ \text{“true”} \}
\]
\[
\text{First(false)} = \{ \text{“false”} \}
\]
\[
\text{First((S))} = \{ \text{“(“} \}
\]
\[
\text{First(S and S)} = \text{First(S or S)} = \text{First(S)} = \{ \text{“(“, “true”, “false”} \}
\]

ii. Explain why the grammar cannot be parsed by a predictive parser

**First sets of productions intersect, grammar is left recursive**

b. Consider the following grammar: \( S \rightarrow \text{abS} \mid \text{acS} \mid \text{c} \)

i. Compute First sets for each production and nonterminal

\[
\text{First(abS)} = \{ \text{a} \}
\]
\[
\text{First(acS)} = \{ \text{a} \}
\]
\[
\text{First(c)} = \{ \text{c} \}
\]
\[
\text{First(S)} = \{ \text{a, c} \}
\]

ii. Show why the grammar cannot be parsed by a predictive parser.

**First sets of productions overlap**

\[
\text{First(abS)} \cap \text{First(acS)} = \{ \text{a} \} \cap \{ \text{a} \} = \{ \text{a} \} \neq \{ \text{a}\}
\]

iii. Rewrite the grammar so it can be parsed by a predictive parser.

\[
S \rightarrow aL \mid c \quad L \rightarrow bS \mid cS
\]

iv. Write a predictive parser for the rewritten grammar.

```plaintext
parse_S() {
    if (lookahead == “a”) {
        match(“a”); // S → aL
        parse_L();
    } else if (lookahead == “c”) {
        match(“c”); // S → c
    } else error();
}
parse_L() {
    if (lookahead == “b”) {
        match(“b”); // L → bS
        parse_S();
    } else if (lookahead == “c”) {
        match(“c”); // L → cS
        parse_S();
    } else error();
}
```
c. Consider the following grammar: $S \rightarrow Sa \mid Sc \mid c$
   i. Show why the grammar cannot be parsed by a predictive parser.
      **First sets of productions intersect, grammar is left recursive**
   ii. Rewrite the grammar so it can be parsed by a predictive parser.
      $S \rightarrow cL$
      $L \rightarrow aL \mid cL \mid \varepsilon$
   iii. Write a recursive descent parser for your new grammar

```c
parse_S() {
    if (lookahead == "c") {
        match("c");  // S \rightarrow cL
        parse_L();
    } else error();
}
parse_L() {
    if (lookahead == "a") {
        match("a");  // L \rightarrow aL
        parse_L();
    } else if (lookahead == "c") {
        match("c");  // L \rightarrow cL
        parse_L();
    } else ;  // L \rightarrow \varepsilon
```