Miscellaneous Problems on Syntax

* 1. A language is said to be *context-free* just if it is expressible by a context free grammar. For each of the following statements about languages over the alphabet $\{0, 1\}$, determine if it is true or false.

(a) Every word of a context-free language is of finite length.

(b) The language $\emptyset$ is context free.

(c) The language of all even length strings is context free.

(d) Every finite subset of $\{0, 1\}^*$ is context free.

* 2. For each of the following statements regarding context-free grammars and context-free languages over some alphabet Σ, determine if it is true or false.

(a) In a context-free grammar, there is exactly one rule for each nonterminal.

(b) The language Σ^* is context free.

(c) No context-free language can include the empty word.

(d) In a context-free grammar, there are always more terminal symbols than non-terminal symbols.

* 3. Consider the following grammar, with start symbol S:

$$\begin{aligned} S &\longrightarrow 0\ T\ 0 \mid 1\ T\ 1 \\ T &\longrightarrow 0\ T \mid 1\ T \mid 0 \mid 1 \end{aligned}$$

For each of the following words, give a derivation to show that it is in the language of this grammar.

(a) 0110

(b) 00110

(c) 11001

* 4. For each of the following CFG over $\{a, b, c\}$, with start symbol S, give (I) one word that is in the language and (II) one word that is not.

Both words should be over the alphabet $\{a, b, c\}$. Label the two words with (I) and (II) so that it is clear which is claimed to be in and which is claimed to be not in.

(a)

$$\begin{array}{rcl} S & \longrightarrow & XXX \\ X & \longrightarrow & a \mid b \end{array}$$

(b)

$$\begin{array}{rcl} S & \longrightarrow & T\ S \mid \epsilon \\ T & \longrightarrow & A\ b\ A\ b\ c \\ A & \longrightarrow & a\ A \mid \epsilon \end{array}$$

(c)

$$\begin{array}{rcl} S & \longrightarrow & AC \mid BC \\ A & \longrightarrow & a \mid a\ A \\ B & \longrightarrow & b \mid b\ B \\ C & \longrightarrow & c \mid c\ C \end{array}$$

(d)

$$S \longrightarrow a\ S\ a \mid b\ S\ \mid c$$

* 5. Consider the grammar for Lisp, given below with start symbol S.

$$\begin{array}{rcl} S & \longrightarrow & A \mid (\ E\ ) \\ E & \longrightarrow & S\ E \mid \epsilon \\ A & \longrightarrow & \mathsf{id} \mid \mathsf{num} \end{array}$$

This grammar is over the 4 terminal symbols:

( ) id num

(a) Compute the nullable, first and follow maps for this grammar.

(b) Construct the parse table for this grammar.

(c) Is this grammar LL(1)?

** 6. For each of the following languages over $\{0, 1\}$, construct a CFG to express it.

(a) $\{uv^n \mid u \in \{0\}^*,\ v = 11,\ n \in \mathbb{N}\}$

(b) $\{w \mid w \text{ starts with } 1\}$

(c) $\{0u1v0 \mid u \text{ is } v \text{ reversed}\}$

** 7. For each of the following languages over $\{a, b\}$, construct a CFG to express it.

(a) $\{w \mid \text{in } w \text{ every 'a' is followed immediately by a 'b'}\}$

(b) $\{w \mid \text{the number of occurrences of a in } w \text{ is a multiple of } 3\}$

(c) $\{w \mid w \text{ does } \textit{not} \text{ contain substring "ab"}\}$

** 8. Design CFG to express the following sets of strings over the alphabet of ASCII characters. Note: (a) you will find it convenient use some abbreviation (like $\cdots$) to help present the expressions compactly and (b) this would not be given as an exam question without specifying the shape of the strings in each part more precisely.

(a) Valid Bristol University usernames (two lowercase letters followed by five digits)

(b) Valid 24 hour clock times in format HH:MM

(c) Valid IPv4 addresses written in decimal

** 9. Construct a context-free grammar to recognise Haskell floating point literals, e.g. 2.99, 23.09e+34, 0.12e−200, 1.4e1.

A general description is as follows. A *decimal literal* is a non-empty sequence of digits (0–9). A *floating point literal* is either:

- a decimal literal followed by a decimal point followed by a decimal literal, optionally followed by an exponent
- or, a decimal literal followed by an exponent.

An exponent is the character e; optionally followed by the character $+$ or the character $-$; followed in all cases by a decimal literal.

** 10. For each of the following, give an equivalent grammar which is LL(1).

(a)

$$\begin{array}{rcl} S & \longrightarrow & S \wedge S \mid G \Rightarrow \mathsf{prop} \\ G & \longrightarrow & G \wedge G \mid \mathsf{prop} \end{array}$$

(b)

$$S \longrightarrow \mathsf{int} \mid \mathsf{string} \mid S \Rightarrow S \mid S \times S \mid (\ S\ )$$

** 11. Consider the following grammar, with start symbol *DeclList*:

$$\begin{array}{rcl} \mathit{DeclList} & \longrightarrow & \mathit{DeclList}\ ;\ \mathit{Decl} \mid \epsilon \\ \mathit{Decl} & \longrightarrow & \mathit{IdList}\ :\ \mathit{Type} \\ \mathit{IdList} & \longrightarrow & \mathit{IdList}\ ,\ \mathsf{id} \mid \mathsf{id} \\ \mathit{Type} & \longrightarrow & \mathsf{ty} \mid \mathit{Type}\ \mathsf{tymod} \end{array}$$

This grammar is over the six terminal symbols:

;   :   ,   id   ty   tymod

(a) Give an equivalent grammar which is LL(1).

(b) Demonstrate that your grammar is LL(1) by constructing the parse table.

*** 12. Construct a CFG expressing the language of bit strings (strings over $\{0, 1\}$) that represent numbers written in binary that are divisible by three. For example, 10010 should be derivable because it represents the decimal number 18 written in binary and this number is divisible by 3. However, 101 should not be derivable, because this is the binary representation of the number 5, which is not divisible by 3.