CMPT 379 – Fall 2013 – Midterm

(1) (10pts) The following CFG describes regular expressions:

$$R \rightarrow R \text{ `|' } R \mid R\,R \mid R \text{ `*' } \mid \text{`(' } R \text{ `)' } \mid a \mid b$$

a. Convert this grammar into an unambiguous CFG that resolves ambiguity by assuming that Kleene closure, **`*'** has the highest priority, followed by concatenation, *RR*, followed by alternation, **`|'** . Also assume that each operation associates to the left, e.g. *RRR* should be treated as (*RR*)*R* and *R*|*R*|*R* should be treated as (*R*|*R*)|*R*. To make grading easier, for any new non-terminals that you introduce to solve this question, please use numeric subscripts, e.g. $R_1, R_2, R_3, \ldots$.

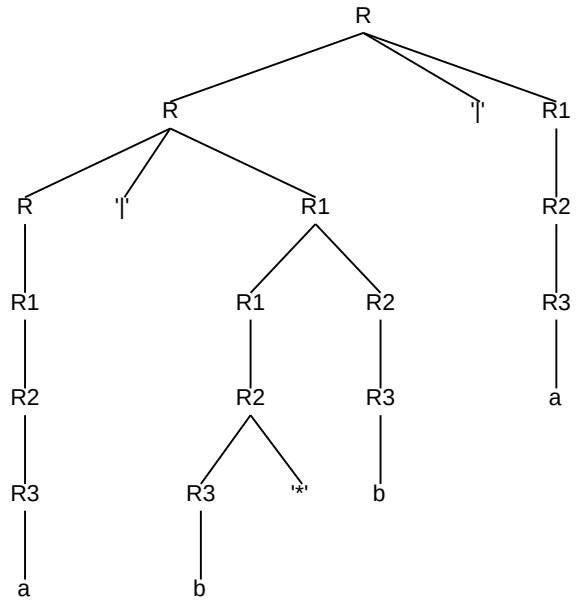
Answer:

$$\begin{aligned} R &\rightarrow R \text{ `|' } R_1 \mid R_1 \\ R_1 &\rightarrow R_1\,R_2 \mid R_2 \\ R_2 &\rightarrow R_3 \text{ `*' } \mid R_3 \\ R_3 &\rightarrow \text{`(' } R \text{ `)' } \mid a \mid b \end{aligned}$$

b. Use your unambiguous grammar to provide the parse tree for the regular expression `a|b*b|a`.

Answer:

```
R
R|R1
R|R1|R1
R1|R1|R1
R2|R1|R1
R3|R1|R1
a|R1|R1
a|R1 R2|R1
a|R2 R2|R1
a|R3 * R2|R1
a|b* R2|R1
a|b* R3|R1
a|b*b|R1
a|b*b|R2
a|b*b|R3
a|b*b|a
```



c. For the input string *abba* provide the lexemes that would be returned by a greedy longest match lexical analyzer assuming that the only token is defined by the (unambiguous) regular expression in Q(1b).

Answer: *a – bb – a* (note that the regexp can be reduced to *a*|*b*+).

(2) (10pts) Consider the augmented CFG G with S' as the start symbol:

$$
\begin{aligned}
S' &\rightarrow S && (1)\\
S &\rightarrow A\ a\ A\ b && (2)\\
S &\rightarrow B\ b\ B\ a && (3)\\
S &\rightarrow \epsilon && (4)\\
A &\rightarrow \epsilon && (5)\\
B &\rightarrow \epsilon && (6)
\end{aligned}
$$

a. Use the canonical LR(1) set-of-items construction and create an action/goto table for LR parsing for grammar G. Use the rule numbers that follow each rule in G above in your table. *Write down the table clearly and legibly*.

Answer:

$0:\ S' \rightarrow \bullet\, S,\ \$$
$\quad S \rightarrow \bullet\, AaAb,\ \$$
$\quad S \rightarrow \bullet\, BbBa,\ \$$
$\quad S \rightarrow \epsilon\bullet\ ,\ \$$
$\quad A \rightarrow \epsilon\bullet\ ,\ a$
$\quad B \rightarrow \epsilon\bullet\ ,\ b$

$1:\ S' \rightarrow S\bullet\ ,\ \$$

$2:\ S \rightarrow A \bullet\, aAb,\ \$$

$3:\ S \rightarrow B \bullet\, bBa,\ \$$

$4:\ S \rightarrow Aa \bullet\, Ab,\ \$$
$\quad A \rightarrow \epsilon\bullet\ ,\ b$

$5:\ S \rightarrow Bb \bullet\, Ba,\ \$$
$\quad B \rightarrow \epsilon\bullet\ ,\ a$

$6:\ S \rightarrow AaA \bullet\, b,\ \$$

$7:\ S \rightarrow BbB \bullet\, a,\ \$$

$8:\ S \rightarrow AaAb\bullet\ ,\ \$$

$9:\ S \rightarrow BbBa\bullet\ ,\ \$$

	a	b	$	S	A	B
0:	r5	r6	r4	1	2	3
1:			acc			
2:	s4					
3:		s5				
4:		r5			6	
5:	r6					7
6:		s8				
7:	s9					
8:			r2			
9:			r3			

b. Provide the LL(1) parsing table for G.

Answer:

	a	b	\$
S'	$S' \to S$	$S' \to S$	$S' \to S$
S	$S \to AaAb$	$S \to BbBa$	$S \to \epsilon$
A	$A \to \epsilon$	$A \to \epsilon$	
B	$B \to \epsilon$	$A \to \epsilon$	