Please note that there is more than one way to answer most of these questions. The following only represents a sample solution.

(1) Let $R$ be the regular expression \((aa \cup bb)(a \cup bbb)^*(b \cup aaa)^*\). Produce a regular grammar that generates the language described by $R$.

**Answer:** A regular grammar that generates $L(R)$ is given by

$$G = (\{S, A, B, C, D, E, F, G, H\}, \{a, b\}, P, S)$$

where the production rules $P$ are given:

- \[ S \rightarrow aA \mid bB \]
- \[ A \rightarrow aC \]
- \[ B \rightarrow bC \]
- \[ C \rightarrow aC \mid bD \mid bF \mid aG \mid \epsilon \]
- \[ D \rightarrow bE \]
- \[ E \rightarrow bC \]
- \[ F \rightarrow bF \mid aG \mid \epsilon \]
- \[ G \rightarrow aH \]
- \[ H \rightarrow aF \]

(2) Let $L_1$ be the language described by the regular expression \((a \cup bbb \cup \epsilon)^*(b \cup aaa \cup \epsilon)^*\). Let $L$ be the language \(\{w \in \{a, b\}^* : w = x x^{rev} \text{ with } x \in L_1\}\). (Note that $x^{rev}$ means the reverse of $x$.) Show that the language $L$ is not regular.

**Answer:** For a contradiction, assume that $L$ is regular. Let $p$ be the pumping length given by the pumping lemma. Take $w = a^p bba^p$, then \(|w| > p\) and $w \in L$, since $a^p b \in L_1$. Then by the pumping lemma $w = xyz$ for some $x, y, z \in \Sigma^*$ with (1) \(|xy| \leq p\), (2) \(|y| > 0\) and (3) for any $i \geq 0$, $xy^i z \in L$. From (1) and (2), $y = 1^k$ for some $1 \leq k \leq p$. Take $i = 2$. However, $w_2 = xy^2 z = a^{p+k} bba^p \notin L$ since $w_2$ is not a palindrome, i.e. not the same forwards and backwards (it contains more $b$’s in the second half than the first half) and hence cannot be written as $ss^{rev}$ for any $s \in L_1$. However, this contradicts (3) above. Thus, $L$ cannot be regular.
(3) Produce a context-free grammar that generates the language $L$ from question 2.

**Answer:** A context-free grammar that generates $L$ is given by

$$G = (\{C, D, E, F, G, H\}, \{a, b\}, R, C)$$

where the production rules $R$ are given:

- $C \rightarrow aCa \mid bDb \mid bFb \mid aGa \mid \epsilon$
- $D \rightarrow bEb$
- $E \rightarrow bCb$
- $F \rightarrow bFb \mid aGa \mid \epsilon$
- $G \rightarrow aHa$
- $H \rightarrow aFa$

(4) Transform your context-free grammar from question 3 so that it is in Chomsky normal form. Explain your steps.

**Answer:** Step 1 is to add a new start state that yields the old start state of $G$ from problem (3):

- $S \rightarrow C$
- $C \rightarrow aCa \mid bDb \mid bFb \mid aGa \mid \epsilon$
- $D \rightarrow bEb$
- $E \rightarrow bCb$
- $F \rightarrow bFb \mid aGa \mid \epsilon$
- $G \rightarrow aHa$
- $H \rightarrow aFa$

Step 2 is to remove the rules that yield $\epsilon$, First we’ll remove $F \rightarrow \epsilon$ giving

- $S \rightarrow C$
- $C \rightarrow aCa \mid bDb \mid bFb \mid aGa \mid bb \mid \epsilon$
- $D \rightarrow bEb$
- $E \rightarrow bCb$
- $F \rightarrow bFb \mid aGa \mid bb$
- $G \rightarrow aHa$
- $H \rightarrow aFa \mid aa$

Next we’ll remove $C \rightarrow \epsilon$ giving:

- $S \rightarrow C \mid \epsilon$
- $C \rightarrow aCa \mid bDb \mid bFb \mid aGa \mid bb \mid aa$
- $D \rightarrow bEb$
- $E \rightarrow bCb \mid bb$
- $F \rightarrow bFb \mid aGa \mid bb$
- $G \rightarrow aHa$
- $H \rightarrow aFa \mid aa$

Now the only rule that can yield $\epsilon$ is the starting variable as required by CNF.
Step 3 is to remove all unit rules. The grammar only has one $S \to C$. Doing so gives

$$
S \to aCa | bDb | bFb | aGa | bb | aa | \epsilon \\
C \to aCa | bDb | bFb | aGa | bb | aa \\
D \to bEb \\
E \to bCb | bb \\
F \to bFb | aGa | bb \\
G \to aHa \\
H \to aFa | aa
$$

The last step is to add new variables and rules to convert all the remaining rules to have the required forms giving:

$$
S \to U_aC U_a | U_bD U_b | U_bF U_b | U_aG U_a | U_bU_b | U_aU_a | \epsilon \\
C \to U_aC U_a | U_bD U_b | U_bF U_b | U_aG U_a | U_bU_b | U_aU_a \\
D \to U_bE U_b \\
E \to U_bC U_b | U_bU_b \\
F \to U_bF U_b | U_aG U_a | U_bU_b \\
G \to U_aH U_a \\
H \to U_aF U_a | U_aU_a \\
U_a \to a \\
U_b \to b \\
U_aC \to U_aC \\
U_aF \to U_aF \\
U_aG \to U_aG \\
U_aH \to U_aH \\
U_bC \to U_bC \\
U_bD \to U_bD \\
U_bE \to U_bE \\
U_bF \to U_bF
$$

(5) **(not to be graded)** Sometimes students make the following argument.

(a) A language that contains only one string is a regular language.

(b) The union of any two regular languages is a regular language.

(c) Every language, whether it is infinite or not, is the union of languages each containing a single string.

(d) So every language is regular.

Is this a valid argument? Explain as precisely as you can why, or why not.

**Answer:** This is **NOT** a valid argument. While (a), (b) and (c) are valid statements, together they do not imply (d). As an counterexample consider the family of regular languages $L_i = \{a^ib^i\}$ for $i \geq 0$. Each language only contains one string and hence is regular as in (a). However, $\bigcup_{i=0}^{\infty} L_i = \{a^ib^i : i \geq 0\}$ a non-regular language. The main issue is that statement (b) only makes a statement about the union of two regular languages being regular. This can be extended by induction to state that the union of a finite number of regular languages are regular. However, it does not hold
in the limit, that is an infinite union of regular languages may not be regular, as was shown.