**CSE396** **Prelim II Answer Key** **Spring 2016**

Closed books and laptops, one notes sheet allowed, closed neighbors, 75 minutes. Do ALL FIVE problems **on these exam sheets**. The exam totals 100 pts., subdivided as shown.

**(1) ($5 \times 5 = 25$ pts.)**

This problem had diagrams for Turing machines with these accompanying prose descriptions. The language is given after each machine.

1. One-tape TM $M_1$ = "Do zero or more *passes* and accept if and when [the] tape has no $a$'s or $b$'s left. Each *pass* X-es out one $a$ then one $b$ or vice-versa, rejecting if there is no opposite char to X out." The work alphabet $\Gamma$ adds '$X$' and the blank $B$ to the input alphabet $\Sigma = \{a, b\}$. *Answer:* This machine was drawn to accept on reading the blank right away so it accepts $\epsilon$ and gives $L(M_1) = \{x \in \{a,b\}^* : \#a(x) = \#b(x)\}$.
2. Two-tape TM $M_2$, arcs going to $q_{rej}$ not shown: "Push an '$a$' for each '$a$' read. On reading '$b$', switch to popping one $a$ for each $b$. Accept iff [the] stack empties on the last $b$." *Answer:* This machine does not accept $\epsilon$ since it needs reading at least one $b$, so $L(M_2) = \{a^n b^n : n \geq 1\}$.
3. Two-tape TM $M_3$, again cases of immediate rejection are not shown: "Similar to $M_2$, except also accept if no $b$'s are found." *Answer:* This machine does accept $\epsilon$ under the "no $b$'s found" clause, so $L(M_3) = L(M_2) \cup a^*$.
4. One-tape TM $M_4$ that emulates a DFA: [no prose description given or needed]. *Answer:* In the diagram $q_{acc}$ and $q_{rej}$ are add-ons, the latter subbing for a dead state. The only states that matter for the language can be labeled $s, f$ and coded by $(s,a,s), (s,b,f), (f,a,s)$ with $F = \{f\}$, formally as an NFA. The language is $L(M_4) = (a + ba)^* b$ = the set of strings with no $bb$ substring that also end in $b$. Note $\epsilon \notin L(M_4)$.
5. "None of the Above."

With this done, the problem was given each context-free grammar $G$ below, say which if any Turing machine $M_i$ at right makes $L(G) = L(M_i)$.

(a) $S \rightarrow aSb \mid ab$ *Answer:* Generates $\{\, a^n b^n : n \geq 1 \,\}$, so $L(M_2)$.

(b) $S \rightarrow aSb \mid \epsilon$ *Answer:* Generates $\{\, a^n b^n : n \geq 0 \,\}$ which includes $\epsilon$, but the two machines that accept $\epsilon$ do not yield this language, so 5, "None of the Above."

(c) $S \rightarrow aS \mid baS \mid b$ *Answer:* Generates $(a + ba)^* b$, so $L(M_4)$.

(d) $S \rightarrow aSb \mid bSa \mid SS \mid \epsilon$ *Answer:* This was the CFG from lecture for $\{\, x : \#a(x) = \#b(x) \,\}$, so $L(M_1)$.

(e) $S \rightarrow aSb \mid ab \mid A$,
$A \rightarrow aA \mid \epsilon$ *Answer:* Includes $L(M_2) \cup a^*$, so includes $L(M_3)$ (and in particular generates $\epsilon$), but by self-recursion before transiting to $A$, it also includes $\{\, a^m b^n : m > n \,\}$. So another "None of the Above" (but partial credit given for $M_3$).

**(2) (6 × 3 = 18 pts.)** *True/False:* Please write out `true` or `false` in full, no justifications needed. The space below can also be used as scratch for Problem (1).

(a) For all context-free languages $A$, $B$, and $C$, the language $A^* \cup B \cdot C$ is also context-free. *True:* the CFLs are closed under the regular operations.

(b) If a CFG $G$ is unambiguous, and $G'$ is obtained by deleting one or more rules of $G$, then $G'$ is unambiguous. *True:* if a string $x$ has two distinct parse trees in $G'$, then it already had them in $G$.

(c) For every NFA $N$ there is a CFG $G$ such that $L(G) = L(N)$. *True* because every regular language is a CFL.

(d) If a CFG $G$ has a variable $A$ such that $A \Longrightarrow^* \epsilon$, then $\epsilon \in L(G)$. *False:* $A$ is nullable but the start symbol $S$ need not be.

(e) For every Turing machine $M$ there is a CFG $G$ such that $L(G) = L(M)$. *False:* some TMs $M$ accept languages like $\{ a^n b^n a^n \}$ that are not CFLs.

(f) The language of odd-length palindromes, i.e., $\{wcw^R : w \in \Sigma^*, \ c \in \Sigma\}$, is context-free. *True:* $S \rightarrow aSa \mid bSb \mid a \mid b$.

**(3) (6 + 3 + 3 + 12 = 24 pts.)**

Let $E$ be the language of all strings over $\Sigma = \{a, b\}$ that do not have the substring $bb$, and let $G$ be the following context-free grammar:

$$\begin{aligned} S &\rightarrow \epsilon \mid b \mid BS \mid SA \\ A &\rightarrow aS \mid AA \\ B &\rightarrow a \mid bAaB \end{aligned}$$

(a) Show that the string $babab$ is ambiguous in the grammar $G$, by giving two different parse trees. *Answer:* Start $S \rightarrow SA$ in both cases. In the first, derive the $S$ to $b$ and $A$ to $aS$. This gives $baS$, which repeats to give $babaS$ and finally $babab$. The second derives the $A$ to $AA$ instead, giving $bAA$. The next level of the tree has $baSaS$, and deriving both $S$-es to $b$ completes the tree.

(b) Is any other variable besides $S$ nullable? Give one(s) if so. *Answer:* No; in particular, $A$ is not nullable since it eventually must derive at least one $a$.

(c) Do any unit rules occur during the conversion to Chomsky normal form? Give one(s) if so—but do not do any more of the conversion. *Answer:* Yes—the rules $S \rightarrow B$ and $S \rightarrow A$ occur. The rule $A \rightarrow a$ is also added, but it does not count as a unit rule.

(d) Prove by structural induction that $L(G) \subseteq E$. *Hint:* Ask yourself what additional properties, besides not allowing a $bb$ substring themselves, must the variables $A$ and $B$ maintain?

*Answer:* Define $P_S \equiv$ "Every $x$ that I derive has no substring $b$," which is just the "vanilla" statement of membership in $E$. The properties you want for $P_A$ and $P_B$ both *include* $P_S$, and *add* the clause "and if $A \Longrightarrow^* x$ then $x$ begins with $a$" for $P_A$ and "if $B \Longrightarrow^* x$ then $x$ ends in $a$" to $P_B$.

The key points were that these stronger properties for $A$ and $B$ were needed to prevent $bb$ occurring "at boundaries" and that we need to uphold these extra clauses in the rules for $A$ and $B$ as well (though not immediately within the rules for $S$).

- $S \to \epsilon \mid b$: These rules uphold $P_S$ immediately. (Note also that the latter means a string derived from $S$ can both begin and end with $b$, which is why the stronger properties for the other5 variables need to be leaned on.)
- $S \to BS$: Suppose $S \Longrightarrow^* x$ using this rule first. Then $x =: yz$ where $B \Longrightarrow^* y$ and $S \Longrightarrow^* z$. By IH $P_B, P_S$ on the right-hand side, neither $y$ nor $z$ has a $bb$ substring *and* $y$ ends in $a$, so no $bb$ appears where they concatenate either. So $x$ has no $bb$, which upholds $P_S$ on the left-hand side.
- $S \to SA$: Suppose $S \Longrightarrow^* x$ utrf. Then $x =: yz$ where $S \Longrightarrow^* y$ this time and $A \Longrightarrow^* z$. By IH $P_S, P_A$ on RHS, $y$ and $z$ each have no internal $bb$ *and* $z$ begins with $a$. So $x$ has no $bb$ at the boundary either, which upholds $P_S$ on LHS. (That finishes $S$, but we still need to maintain the stronger properties in the other rules.)
- $A \to aS$: Suppose $A \Longrightarrow^* x$ utrf. Then $x = ay$ where $S \Longrightarrow^* y$. By IH $P_S$ on RHS, $y$ has no $bb$ so neither does $x$, *and* the leading $a$ makes $ay = x$ immediately uphold the added clause of $P_A$. So $P_A$ is upheld on LHS.
- $A \to AA$: Suppose $A \Longrightarrow^* x$ utrf. Then $x =: yz$ where $A \Longrightarrow^* y$ and $A \Longrightarrow^* z$. By IH $P_A$ on RHS (twice), neither $y$ nor $z$ has an internal $bb$, and $y$ begins with $a$ which implies the same for $x$. Is that all we need to say, i.e., is it immaterial that $z$ begins with $a$? No, that is needed too, to say that no $bb$ occurs at the boundary between $y$ and $z$. So $P_A$ holds on the LHS.
- $B \to a$: Both clauses of $P_B$ are immediately upheld.
- $B \to bAaB$: Suppose $B \Longrightarrow^* x$ utrf. Then $x =: byaz$ where $A \Longrightarrow^* y$ and $B \Longrightarrow^* z$. By IH $P_A, P_B$ on RHS, those strings have no internal $bb$, and they don't touch and in fact surround an $a$, so the only $bb$ danger can come from the leading $b$. That is averted, however, because $P_A$ on RHS implies that $y$ begins with $a$. We're not done—we need to uphold "ends with $a$" as well to get $P_B$ on LHS, but this follows by "self-induction" since $z$ ends in $a$ by IH $P_B$ on RHS.

Thus the properties are upheld by all the rules, so $L(G) \subseteq E$ by structural induction.

**(4) (18 + 6 = 24 pts.)**

With reference to $G$ and $E$ in problem (3), prove that $E \subseteq L(G)$ (18 pts.) and also find two rules not used in your proof that can be deleted (6 pts.) while keeping $L(G) = E$. *Hints:* Note that if $x \in E$ and $x$ begins with $b$ and $|x| \geq 2$ then the next char must be $a$. Do you *need* to add induction hypotheses for the other variables besides $S$?

*Answer:* Prove $(\forall n \geq 0)P(n)$, where $P(n) \equiv$ for each $x \in \{a, b\}^n$, if $x$ has no substring $bb$ then $S \Longrightarrow^* x$. Let's see if we can do this without having to use a property $P'(n)$ that is "augmented" with clauses for the other variables.

*Basis* ($n = 0$): $\epsilon \in E$ and $S \Longrightarrow \epsilon$. Check.

*Another Basis* ($n = 1$): Both $a$ and $b$ belong to $E$. We have $S \Longrightarrow BS \Longrightarrow aS \Longrightarrow a$ and $S \Longrightarrow b$. (For reference at the end, note also the alternative $S \Longrightarrow SA \Longrightarrow A \Longrightarrow aS \Longrightarrow a$, which uses the rule $S \to SA$ instead of $S \to BS$ for the basis—the same idea also yields more parse trees for *babab* in 3(a).)

*Induction* ($n \geq 2$): We may assume (IH) the statements $P(m)$ for all $m$, $0 \leq m < n$. We will in fact only *use* $P(n-1)$ and $P(n-2)$, but it doesn't hurt to say the full slate is at your disposal. Let any $x \in \Sigma^n$ such that $x \in E$ be given. We try two mutually exhaustive cases: (i) $x$ begins with $a$ and (ii) $x$ begins with $ba$. Regarding the latter, note we don't have $x = b$ since $n \geq 2$ (here is where we'd be reminded to make a separate base case for it if we hadn't already), so there is a second char, and since the second char being $b$ would make $x$ begin with $bb$ and put $x \notin E$, case (ii) must hold as stated if case (i) doesn't.

*Case (i):* $x = ay$ where $|y| = n-1$. Since $x \in E$ we have $y \in E$, so IH $P(n-1)$ kicks in to imply $S \Longrightarrow^* y$. Then we build $S \Longrightarrow BS \Longrightarrow aS \Longrightarrow^* ay = x$. So $P(n)$ holds in this case.

*Case (ii):* $x = baz$ where $|z| = n-2$. Then $z$ too must be in $E$, so by IH $P(n-2)$, $S \Longrightarrow^* z$. This gives us $S \Longrightarrow SA \Longrightarrow bA \Longrightarrow baS \Longrightarrow^* baz = x$.

Thus $P(n)$ holds in each of two exhaustive cases, so it follows in general, which completes the induction showing $(\forall n)P(n)$, which says $E \subseteq L(G)$. (Putting problems (3) and (4) together, we get a full proof of $L(G) = E$.)

We nowhere used the rules $A \to AA$ and $B \to bAaB$ in the proof, so while sound they are superfluous. *Actually*, we could have avoided $S \to BS$ in case (i) as well as the $n = 1$ basis, by doing $S \Longrightarrow SA \Longrightarrow A \Longrightarrow aS \Longrightarrow^* ay = x$ instead. So $S \to BS$ is a third rule that can be removed—but any two were enough for full credit.

**(5) (9 pts.)**

The language $L = \{a^i b^j c^k : i < j \;\vee\; j < k\}$ **is** context-free.[1] But suppose you were being questioned on Capitol Hill by a prosecutor trying to prove otherwise. Say you give $p = 5$ as the pumping length and the prosecutor tells you to break down

$$s = a^p b^{p+1} c^{p+2} = aaaaabbbbbbccccccc.$$

Give a breakdown $s =: uvxyz$ into five substrings that "survives"—i.e., is such that for all $i \geq 0$, $uv^i xy^i z$ **does** belong to $L$. Briefly explain why. (There are multiple correct answers.)

*Answer:* One of many ways is to take

$$u = a^{p-1} \quad v = a \quad x = \epsilon \quad y = b \quad z = b^p c^{p+2}$$

(either abstractly or concretely with $p = 5$ was fine). Ditto taking $vxy = bc$ on the other side instead, and actually the OR allows you to give practically any breakdown with $|vxy| \leq 5$. The reason in the above particular case is that the property of having one less $a$ than $b$ doesn't change in $uv^i xy^i z$ for any $i$.

[1] A grammar starts with $S \to S_1 C \mid AS_2$, makes $A$ derive $a^*$ and $C$ derive $c^*$, and makes $S_1$ derive $\{a^i b^j : j > i\}$ by ways we have seen in lectures and notes and answer keys, likewise $S_2$ deriving $\{b^j c^k : k > j\}$. You can answer the question without caring about these details. END OF EXAM