NAME: ______________________________

COMPSCI 501
Formal Language Theory
Midterm Spring 2023

D. A. M. Barrington 10 April 2023

DIRECTIONS:

- Answer the problems on the exam pages.
- There are eight problems on pages 2-8, some with multiple parts, for 100 total points plus 10 extra credit. Final scale will be determined after the exam.
- The supplemental page 9 has definitions for your use and should not be handed in.
- If you need extra space use the back of a page.
- No books, notes, calculators, or collaboration.

1	/20
2	/30
3	/10
4	/10
5	/10
6	/10
7	/10
8	/+10
Total	/100+10

Question 1 (20): These are ten true/false statements, with no justification needed or wanted (2 points each):

- (a, 2) The language $\{\langle M \rangle : L(M) \in CFL\}$, where M is a TM, is Turing decidable.

- (b, 2) The language $\{\langle M \rangle : L(M) \in CFL\}$, where M is a TM, is Turing recognizable.

- (c, 2) If $A \leq_m B$, and B is not co-TR, it is possible that A is Turing recognizable.

- (d, 2) If Σ is a finite alphabet, then every string in Σ^* is finite.

- (e, 2) If X and Y are each *nonempty* TD languages, then $X \leq_m Y$ must be true.

- (f, 2) If X is a language, and both X and its complement $\overline{X}$ are both context-free, it is possible X is not a regular language.

- (g, 2) If A is not a CFL, and $A \leq_m B$, then it is possible that B could be a CFL.

- (h, 2) Let G be a context-free grammar, let p be a pumping length for it in the CFLPL, and assume that $L(G)$ is an infinite language. Then it is possible that $L(G)$ contains no strings w such that $p \leq |w| \leq 2p$.

- (i, 2) Recall that $K(x)$ is the length of the minimal description of a binary string x. It is not the case that for every string, such that x is a member of any regular language, $K(x)$ is strictly less than $|x|$.

- (j, 2) There exists a Turing machine M that accepts a string if and only that string is same length as the description of M.

Question 2 (30): These are five true-false questions, with brief justification required. Three points for each correct boolean answer, and up to three points per question for the justification:

- (a, 6) The following language is undecidable: $Q = \{\langle G, x\rangle : \exists u : \exists y : uxy \in L(G)\}$. This language Y is the set of pairs consisting of a context-free grammar G and a string x such that G can generate *some* superstring of x.

- (b, 6) The following language is decidable: $R = \{\langle G, x\rangle : \forall u : \forall y : uxy \in L(G)\}$. This language Y is the set of pairs consisting of a context-free grammar G and a string x such that G can generate *every* superstring of x.

- (c, 6) Let $\Sigma = \{0, 1\}$. Let f be any function from Σ to Σ^* (from letters to strings). We extend f to a **homomorphism** f from Σ^* to Σ^* by defining $f(a_1 a_2 \dots a_n)$ as $f(a_1)f(a_2)\dots f(a_n)$. Then if S is any regular language, it is possible that the language $f(S) = \{f(w) : w \in S\}$ is not a regular language.

- (d, 6) Let f be a homomorphism of languages as defined in Question 2(c). Then if T is a regular language, then it is possible that the language $f^{-1}(T) = \{w : f(w) \in T\}$ is not a regular language.

- (e, 6) The following language is undecidable: U is the set of all pairs $\langle M, w \rangle$ such that M is a single-tape TM, w is a string, and M accepts w while never modifying the part of the tape containing the input.

Definitions: Some of Questions 3-8 deal with a new definition. If L_1 and L_2 are any two languages over the same alphabet, the **quotient language** L_1/L_2 is the set of all strings x such that there exists a string w such that $xw \in L_1$ and $w \in L_2$. An important special case is when $L_2 = \{w\}$, where w is a specific string, and we write this quotient language as L_1/w.

Question 3 (10): Prove that if V is a regular language, and w is any fixed string, then the language V/w is regular.

Question 4 (10): Prove that is L is a context-free language, and w is any fixed string, then the language L/w is context-free. (**Hint:** It may be easier to solve this by generalizing. If you assume that if L is a context-free language and that B is regular, and you prove L/B is context-free, this suffices – why?)

Question 5 (10): A Turing machine M, with one tape, is defined to **cycle** on any input w if there exists a configuration C such that if started in w, M reaches C *more than once*. Prove that if L is a TR language, then there exists a Turing machine M such that $L(M) = L$ and M does not cycle on any input w.

Question 6 (10): Consider the language $DOUBLE_{TM}$, the set of Turing machines $\langle M \rangle$ such that $L(M)$ contains a string of the form ww. Is $DOUBLE_{TM}$ a TR language? Is it a co-TR language? Prove your answers. Do not use the result of Question 7.

Question 7 (10): Prove that the language $DOUBLE_{CFG}$, the set of context-free grammars $\langle G \rangle$ such that $L(G)$ contains a string of the form ww, is undecidable. (**Hint:** Make a reduction $PCP \leq_m DOUBLE_{CFG}$ as follows. Let the set of dominoes for the PCP problem be $\{[\frac{x_1}{y_1}], \ldots, [\frac{x_k}{y_k}]\}$ and let a_i be a unique terminal for each domino. Build the grammar G with a rule $S \to A'B'$ and, for each i with $1 \leq i \leq k$, rules $A' \to x_i A a_i$, $A \to x_i A a_i$, $A \to \epsilon$, $B' \to y_i B a_i$, $B \to y_i B a_i$, and $B \to \epsilon$. Prove that the PCP instance has a match if and only if $L(G)$ contains a string of the form ww.

Question 8 (10 extra credit): In this problem, we will prove that the language REG_{CFG}, the set of context-free grammars G (with $\Sigma = \{0, 1\}$) such that $L(G)$ is regular, is undecidable. We will build a reduction $ALL_{CFG} \leq_m REG_{CFG}$, as follows. Throughout the problem, Σ denotes $\{0, 1\}$. Let Z be the language $\{0^n 1^n : n \geq 0\}$, which we know not to be regular. Given a grammar G, with alphabet Σ, we define a grammar G' (over the alphabet $\{0, 1, \#\}$) such that $L(G') = Z\#\Sigma^* \cup \Sigma^*\#L(G)$.

1. Describe how to build the grammar G' from the grammar G.

2. Prove that if $\langle G \rangle \in ALL_{CFG}$, then $L(G')$ is a regular language.

3. Suppose that some string $w \in \Sigma^*$ is not in $L(G)$. Prove that the quotient language $L(G')/\#w$ is not regular.

4. Using the result of Question 3, whether you solved it or not, explain how we may now conclude that REG_{CFG} is undecidable.

Supplemental Page for COMPSCI Midterm Spring 2023

Language A_{TM} (similarly A_{REG}, etc.) Set of pairs (M, w) such that M accepts w

Language ALL_{TM}**:** (similarly ALL_{REG}, etc.) Set of machines that accept all possible strings over their alphabet

Computable Function: Function f from strings to strings such that some Turing machine, on any input w, always halts with $f(w)$ on its tape

Context-Free Grammar (CFG): Grammar where rules allow single non-terminals to be replaced by strings

Context-Free Language (CFL): Definable by a context-free grammar or a PDA

co-TR: A language A is co-TR if and only if its complement $\overline{A}$ is TR.

Language E_{TM}**:** (similarly E_{REG}, etc.) Set of machines with empty languages

Mapping Reduction ($\leq_m$**):** $A \leq_m B$ means that there exists a computable function f such that $\forall w : w \in A \leftrightarrow f(w) \in B$

Pushdown Automaton (PDA): Nondeterministic finite-state machine with an added stack

Language REG_{TM}**:** (similarly REG_{REG}, etc.) Set of machines M such that $L(M)$ is a regular language

Regular Language: Can be defined by a DFA, NFA, or regular expression

Turing Decidable (TD): Is the language of a TM that always halts

Turing Recognizable (TR): Is the language of any TM