University of Waterloo
CS 462 — Formal Languages and Parsing
Winter 2020
Problem Set 1

Distributed Friday, January 10 2020.
Due Friday, January 17 2020 by 5 PM. Hand in to LEARN.

All answers should be accompanied by proofs. In all problems the underlying alphabet Σ is assumed to be finite.

For the definitions of terms like reversal, subword, palindrome, perfect shuffle, overlap, conjugate, etc., see the textbook.

Generally speaking, the level of detail in all problem sets should be *enough to convince a skeptical TA*.

Please use a document preparation system like LaTeX or Word for your solutions. Do not handwrite solutions! For diagrams only, feel free to draw them by hand if you like, and scan them.

1. [10 marks] Call a finite word w "reversal-closed" if it has the following property: x a subword of w implies that x^R is a subword of w.

   Prove that a finite word w is reversal-closed iff it is a palindrome.

2. [10 marks] Give a recursive definition of the perfect shuffle of two identical-length words, $x \amalg y$. Then use that definition to give a complete formal proof by induction of the identity $(x \amalg y)^R = y^R \amalg x^R$.

   Hint: use the recursive definition of reverse: $\epsilon^R = \epsilon$, and $(xa)^R = ax^R$ for $a \in \Sigma$, $x \in \Sigma^*$. (You also might have to prove some other properties of $\amalg$ first.)

3. [10 marks] An *overlap* is a word of the form $axaxa$, where a is a single letter, and x is a (possibly empty) word. For example, the word `entente` is an overlap, with $a =$ `e` and $x =$ `nt`.

   A word is said to *avoid circular overlaps* if no conjugate of a subword starts with an overlap. For example, `abcadeab` avoids circular overlaps. However, `eabcadeabcf` does not avoid circular overlaps: consider the subword `abcadeabc`; it has the conjugate `abcabcade` which starts with an overlap.

   Prove that there is a function $f(k)$ such that if the alphabet size is k, and x avoids circular overlaps, then $|x| \leq f(k)$. You should state what $f(k)$ is as explicitly as possible.

   The asymptotically best bound $f(k)$ found by a student gets some extra credit marks.