PART II AUTOMATA AND FORMAL LANGUAGES
MICHAELMAS 2015-16
EXAMPLE SHEET 4

$*$ denotes a harder problem; $**$ denotes an even harder problem.

(1) Let G be the CFG given by

$$S \to ABS \mid AB, \quad A \to aA \mid a, \quad B \to bA$$

For each of the words $aabaab$, $aaaaba$, $aabbaa$, $abaaba$, determine whether or not they lie in $\mathcal{L}(G)$. If so, give a derivation and a parse tree; if not, explain why not.

(2) Convert the following CFG to CNF:

$$S \to aSbb \mid T, \quad T \to bTaa \mid S \mid \epsilon$$

(3) Give a CFG for each of the following CFL's, and then transform each such CFG into CNF:

(a) $\{a^n b^{2n} c^k \mid k, n \geq 1\}$

(b) $\{a^n b^k a^n \mid k, n \geq 1\}$

(c) $\{a^k b^m c^n \mid k, m, n \geq 1,\ 2k \geq n\}$

(4) For each of the following languages, either show that it is a CFL by constructing a CFG for it, or use the pumping lemma to show that it is not a CFL:

(a) $\{a^n b^m \mid n \neq m\}$

(b) $\{a^m b^n c^m d^n \mid m, n \geq 1\}$

(c) $\{a^n b^m c^k d^l \mid 2n = 3m \text{ and } 5k = 7l\}$

(d) $\{a^n b^m c^k d^l \mid 2n = 3k \text{ and } 5m = 7l\}$

(e) $\{a^n b^m c^k d^l \mid 2n = 3k \text{ or } 5m = 7l\}$

(f) $\{1^p \mid p \text{ is prime}\}$

(g) $\{a^{2^n} \mid n \geq 1\}$

(h) $\{ww \mid w \in \{a, b\}^*\}$

(i*) $\{a, b\}^* \setminus \{ww \mid w \in \{a, b\}^*\}$

(5) Construct NPDA's which accept, either by final state or by empty stack, each of the following languages:

(a) $\{w \in \{a,b\}^* \mid w \text{ contains the same number of } a\text{'s and } b\text{'s}\}$

(b) $\{a^i b^j c^k \mid i = j \text{ or } i = k\}$

(c) $\{ww^R \mid w \in \{0,1\}^*\}$

(6) Let $G = (N, \Sigma, P, S)$ be a CFG in CNF. Suppose we form a new CFG G' from G by adding the production $B \to \epsilon$ to P for every $B \in N$. Describe the new language $\mathcal{L}(G')$ in terms of the original language $\mathcal{L}(G)$.

(7) Give a CFG which generates the set of syntactically correct (though perhaps mathematically false) arithmetic equations over $\mathbb{N}$ with addition and subtraction. For example, $4 + 9 = 11 - 20$.
Take as the set of terminals $\Sigma = \{+, -, =, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. When designing your CFG, you may use words delimited by $<>$ as nonterminals. For example, you could use $<$ *start* $>$, $<$ *number* $>$, etc.

(8) Give a CFG which generates the set of regular expressions over the alphabet $\{0, 1\}$. Take as the set of terminals $\Sigma = \{\mathbf{0}, \mathbf{1}, (,), +, ^*, \emptyset, \epsilon\}$. Show that this language is not a regular language.

(9) Let L, M be CFL's, and let a be any symbol. Show that the following are all CFL's: $\emptyset$, $\{\epsilon\}$, $\{a\}$, $L \cup M$, LM, L^*, and L^R (the reverse of L). Conclude that every regular language is a CFL.

(10) (a) Show that the following two languages are both CFL's: $L_1 := \{a^n b^n c^i \mid n, i \geq 1\}$, and $L_2 := \{a^i b^n c^n \mid n, i \geq 1\}$.

(b) Show that the language $L := \{a^n b^n c^n \mid n \geq 1\}$ is not a CFL.

(c) Show that $L_1 \cap L_2 = L$, and hence that the intersection of two CFL's is not necessarily a CFL.

(d) Conclude that the complement of a CFL need not be a CFL.

(11) (a) Let G be a CFG in CNF, and $w \in \mathcal{L}(G)$ a word of length $n \geq 1$. Show that *any* derivation of w in G uses precisely $2n - 1$ steps.

(b) Let G be a CFG in CNF with m nonterminals. Show that, if $\mathcal{L}(G) \neq \emptyset$, then $\mathcal{L}(G)$ contains at least one word of length $< 2^{m+1}$.

(c*) Give an algorithm that, on input of a CFG G and a word w on the terminal symbols of G, decides if $w \in \mathcal{L}(G)$ or not.

(d*) Give an algorithm that, on input of a CFG G, decides if $\mathcal{L}(G) = \emptyset$ or not.

(12**) Prove that the intersection of a regular language with a CFL is a CFL.