NTIN071 A&G: Tutorial 7 – Pumping lemma for context-free languages, Chomsky normal form

Teaching goals: The student is able to

- give a formal statement of the Pumping lemma for context-free languages
- apply the Pumping lemma to prove that a given language is not context-free
- give the formal definition of Chomsky Normal Form and related notions
- explain the proof of the Pumping lemma for context-free languages
- convert a given context-free grammar to ChNF
- explain the CYK algorithm, apply to a given word and context-free grammar

In-class problems

Problem 1 (Pumping lemma: statement and proof)**.** (a) Formulate the Pumping Lemma for context-free languages (without consulting your notes).

(b) Compare the statement to the version for regular languages.

(c) Explain the idea behind its proof.

(d) Demonstrate pumping on the language $L = \{ww^R \mid w \in \{a,b\}^*\}$.

Problem 2 (Pumping lemma: application)**.** Decide if the following languages are context-free. Prove that your answer is correct.

(a) $L = \{0^i 1^i \mid i \geq 0\}$

(b) $L = \{0^i 1^j 0^i \mid 0 \leq j \leq i\}$

(c) $L = \{0^i 1^i 2^i \mid i \geq 0\}$

(d) $L = \{0^{2i} 1^{3i} 0^i \mid i \geq 0\}$

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

(f) $L = \{1^{n^2} \mid n \geq 0\}$

(g) $L = \{1^p \mid p \text{ is a prime}\}$

(h) $L = \{0^i 1^j \mid 0 \leq i \leq j^2\}$

Problem 3 (Convert to ChNF)**.** Convert the following context-free grammars to Chomsky normal form:

(a) $G_1 = (\{S, A, B\}, \{0,1\}, S, \mathcal{P})$, where

$$\begin{aligned}\mathcal{P} = \{S &\to 0AB,\\ A &\to 0A0 \mid 11,\\ B &\to 0\}\end{aligned}$$

(b) $G_2 = (\{S, A, B\}, \{0,1\}, S, \mathcal{P})$, where

$$\begin{aligned}\mathcal{P} = \{S &\to 0A10B10,\\ A &\to 1A0 \mid \epsilon,\\ B &\to 1B00 \mid \epsilon\}\end{aligned}$$

Problem 4 (The CYK algorithm)**.** Using the CYK algorithm determine if $w \in L(G)$.

(a) $w = 001100$, $G = G_1$ is the grammar from Problem 3(a)

(b) $w = 110011$, $G = G_1$ is the grammar from Problem 3(a)

Extra Practice and Thinking

Problem 5 (Pumping lemma: application)**.** Decide if the following languages are context-free. Prove that your answer is correct.

(a) $L = \{0^i 1^j 0^i \mid i, j \geq 0\}$
(b) $L = \{0^i 1^j 0^i \mid 0 \leq i \leq j\}$
(c) $L = \{0^i 1^j 2^k \mid 0 \leq i \leq j \leq k\}$
(d) $L = \{ww \mid w \in \{0,1\}^*\}$
(e) $L = \{ww^R \mid w \in \{0,1\}^*, |w|_0 = |w|_1\}$
(f) $L = \{1^{n^2+n+1} \mid n \geq 0\}$

Problem 6 (Pumping and right-linear grammars)**.** Give an alternative proof of the Pumping lemma for regular languages that is based on derivations from a right-linear grammar.

Problem 7 (Pumping linear languages)**.** Recall that a grammar is *linear*, if it only contains production rules of the form $A \to uBw$ and $A \to w$, where $A, B \in V$ and $u, w \in T^*$.
(a) Formulate a Pumping lemma for linear languages.
(b) Proof the statement using derivations from a (reduced) linear grammar.
(c) How does the pumping constant n from the lemma relate to a linear grammar for L?
(d) Show that the language $L = \{w \in \{0,1\}^* \mid |w|_0 = |w|_1\}$ is not linear.
(e) Where does L lie within the Chomsky hierarchy?

Problem 8 (About the conversion to ChNF)**.** Recall the process of converting a context-free grammar to Chomsky Normal Form. Then answer the following questions. Justify.
(a) Find an example of a grammar in which there is a generating variable only reachable via nongenerating variables.
(b) When reducing a grammar, which variables do we need to remove first: nongenerating or unreachable?
(c) Is it possible for a reachable generating variable to become nongenerating after the removal of unreachable variables?
(d) When we want to break up a production rule with long body, what is the minimal number of Chomsky Normal Form rules we need to create?

Problem 9 (Convert to ChNF)**.** Convert the following to Chomsky normal form:

(a) $G = (\{S, A, B\}, \{0, 1\}, S, \mathcal{P})$

$$\mathcal{P} = \{S \to A \mid 0SA \mid \epsilon, \quad A \to 1A \mid 1 \mid B1, \quad B \to 0B \mid 0 \mid \epsilon\}$$

(b) $G = (\{S, E, F\}, \{(,), *, +, , 1\}, S, \mathcal{P})$

$$\mathcal{P} = \{S \to (E), \quad E \to F + F \mid F * F, \quad F \to S \mid 1\}$$

Problem 10 (The CYK algorithm)**.** Using the CYK algorithm determine if $w \in L(G)$.
(a) $w = 01010010$, $G = G_2$ is the grammar from Problem 3(b)
(b) $w = 01010011$, $G = G_2$ is the grammar from Problem 3(b)