

Remark

Note that we allow for  $\epsilon$ -productions in context-free grammars, i.e. productions are of the form  $A \rightarrow B$  with  $A \in V_N$  and  $B \in V^*$ .

Exercise 1

Define context-free grammars for the following languages:

- 1)  $\{a^n b^n \mid n \geq 0\}$
- 2)  $\{0^i 1^j \mid i \leq j \leq 2i \text{ \& } i \geq 0\}$
- 3)  $\{0^n 1^m 0^{m+n} \mid m+n \geq 1\}$

Solution:

$$1) G_1: S \rightarrow aSb \mid \epsilon$$

$$2) G_2: S \rightarrow 0S11 \mid 0S1 \mid \epsilon$$

$$3) G_3: S \rightarrow A1B \\ A \rightarrow 0A0 \mid B \mid 00 \\ B \rightarrow 1B0 \mid 10$$

Exercise 2

Eliminate the  $\varepsilon$ -productions from the grammars  $G_1$  and  $G_2$  of the solution to exercise 1

Remember that the grammars  $G'_1$  and  $G'_2$  obtained by eliminating the  $\varepsilon$ -productions from  $G_1$  and  $G_2$  are such that  $L(G'_1) = L(G_1) - \{\varepsilon\}$  and  $L(G'_2) = L(G_2) - \{\varepsilon\}$

Solution:

•  $G_1: S \rightarrow aSb \mid \varepsilon$

The set of nullable symbols is simply  $\{S\}$

Hence  $G'_1: S \rightarrow aSb \mid ab$  and  $L(G'_1) = \{a^n b^n \mid n \geq 1\}$

•  $G_2: S \rightarrow 0S11 \mid 0S1 \mid \varepsilon$

The set of nullable symbols is again  $\{S\}$

Hence  $G'_2: S \rightarrow 0S11 \mid 0S1 \mid 011 \mid 01$  and  $L(G'_2) = \{0^i 1^j \mid i \leq j \leq 2i \text{ and } i \geq 1\}$

Exercise 3

Eliminate the unit productions from

$$\begin{aligned} S &\rightarrow A \mid B \\ A &\rightarrow 0S1 \mid B \\ B &\rightarrow 1 \mid S1 \end{aligned}$$

Solution:

From reachability we get:  $S \Rightarrow^* A$ ,  $S \Rightarrow^* B$ , and  $A \Rightarrow^* B$

Hence we get that  $S \rightarrow 0S1 \mid 1 \mid S1$

$$A \rightarrow 0S1 \mid 1 \mid S1$$

$$B \rightarrow 1 \mid S1$$

Since both  $A$  and  $B$  are not reachable from  $S$  we can simplify the grammar to  $S \rightarrow 0S1 \mid 1 \mid S1$

Exercise 4

Consider the following grammar:

$$S \rightarrow OS | BD | EO$$

$$B \rightarrow C | \epsilon | 1$$

$$D \rightarrow O | AO | S | \epsilon$$

$$E \rightarrow BE | SE$$

$$C \rightarrow O$$

$$F \rightarrow C$$

a) Simplify the grammar

b) Transform the simplified grammar in Chomsky normal form

Solution:

a) Elimination of  $\epsilon$ -productions

Nullable symbols:  $\{B, D, S\}$

$$\text{We get: } S \rightarrow OS | O | BD | B | D | EO$$

$$B \rightarrow C | 1$$

$$D \rightarrow O | AO | S$$

$$E \rightarrow BE | SE$$

$$C \rightarrow O$$

$$F \rightarrow C$$

Elimination of unit productions

From reachability:  $S \Rightarrow^* \{B, C, D\}$ ,  $B \Rightarrow^* C$ ,  $D \Rightarrow^* \{S, B, C\}$ , and  $F \Rightarrow^* C$

$$\text{We get: } S \rightarrow OS | O | BD | 1 | AO | EO$$

$$B \rightarrow O | 1$$

$$D \rightarrow O | AO | OS | BD | 1 | EO$$

$$E \rightarrow BE | SE$$

$$C \rightarrow O$$

$$F \rightarrow O$$

$S \Rightarrow^* \{B, C, D\}$  means that  $S \Rightarrow^* B$ ,  $S \Rightarrow^* C$ ,  $S \Rightarrow^* D$  and similarly for  $D \Rightarrow^* \{S, B, C\}$

## Elimination of useless symbols

E8.4

→ Non-generating symbols

Generating symbols:  $\{S, B, D, C, F\}$

We get:  $S \rightarrow OS|O|BD|1$

$B \rightarrow 0|1$

$D \rightarrow 0|OS|BD|1$

$C \rightarrow 0$

$F \rightarrow 0$

→ Non-reachable symbols

Reachable symbols:  $\{S, B, D\}$

We get:  $S \rightarrow OS|O|BD|1$

$B \rightarrow 0|1$

$D \rightarrow 0|OS|BD|1$

b) Since we do not have long productions we only need to remove "mixed bodies" (such as OS) by introducing some new productions

We get:  $S \rightarrow GS|O|BD|1$

$B \rightarrow 0|1$

$D \rightarrow 0|GS|BD|1$

$G \rightarrow 0$

## Exercise 5

E8.5

Transform the grammar  $G_3$  of the solution to exercise 1 in Chomsky normal form

Solution:

$$G_3: S \rightarrow A|B$$

$$A \rightarrow 0A0|B|00$$

$$B \rightarrow 1B0|10$$

Elimination of unit productions

From reachability:  $S \Rightarrow^* A, S \Rightarrow^* B, A \Rightarrow^* B$

We get:  $S \rightarrow 0A0|1B0|10|00$

$$A \rightarrow 0A0|1B0|10|00$$

$$B \rightarrow 1B0|10$$

Remove "mixed bodies" ( $C \rightarrow 0, D \rightarrow 1$ )

We get:  $S \rightarrow CAC|DBC|DC|CC$

$$A \rightarrow CAC|DBC|DC|CC$$

$$B \rightarrow DBC|DC$$

$$C \rightarrow 0$$

$$D \rightarrow 1$$

"Factor" long productions ( $E \rightarrow AC, F \rightarrow BC$ )

We get:  $S \rightarrow CE|DF|DC|CC$

$$A \rightarrow CE|DF|DC|CC$$

$$B \rightarrow DF|DC$$

$$C \rightarrow 0$$

$$D \rightarrow 1$$

$$E \rightarrow AC$$

$$F \rightarrow BC$$

HW: Verify that for

$$G'_3: S \rightarrow A|B$$

$$A \rightarrow 0S0|00$$

$$B \rightarrow 1B0|10$$

we get the following grammar in Chomsky NF:

$$S \rightarrow CE|CC|DF|DC$$

$$B \rightarrow DF|DC$$

$$C \rightarrow 0$$

$$D \rightarrow 1$$

$$E \rightarrow SC$$

$$F \rightarrow BC$$

Note: we have eliminated the useless symbol A