Einführung in die Computerlinguistik
Hausaufgabe (CFG 2), Abgabe 04.06.2019

Laura Kallmeyer

Sommer 2019, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1

1. *Betrachten Sie folgende CFG:*
   $G = \langle\{S, A, B, C, D, E\}, \{a, b, c\}, P, S\rangle$ *mit*
   $P = \{S \to ABC \mid AB, A \to a \mid acC, B \to bb \mid CBB, C \to D, E \to B\}$
   Geben Sie eine äquivalente Grammatik ohne nutzlose Symbole an.
2. $G = \langle\{S, T, U\}, \{a, b, c\}, \{S \to aS \mid abTb \mid cU, T \to c \mid TTc \mid \varepsilon, U \to ab \mid T\}, S\rangle$
   Geben Sie eine äquivalente Grammatik ohne ε*-Produktionen an.*

Lösung:

1. Symbole, aus denen sich terminale Ketten ableiten lassen: $\{A, B, E, S\}$
   neue Produktionsmenge ohne C, D: $S \to AB, A \to a, B \to bb, E \to B$
   Symbole, die vom Startsymbol erreichbar sind: $\{S, A, B, a, b\}$
   neue Produktionsmenge $S \to AB, A \to a, B \to bb$
   Grammatik: $\langle\{A, B, S\}, \{a, b\}, \{S \to AB, A \to a, B \to bb\}, S\rangle$
2. $N_\varepsilon = \{T, U\}$.
   Neue Produktionen: $S \to aS \mid abTb \mid cU \mid abb \mid c, T \to c \mid Tc \mid TTc, U \to ab \mid T$

Aufgabe 2 *Geben Sie für die Sprache* $\{(ab)^n cd^n \mid n \geq 0\}$

1. *eine CFG in Chomsky Normalform und*
2. *eine CFG in Greibach Normalform an.*

Lösung: (hier nur Produktionen, es sollte aber eigentlich das ganze Tupel hingeschrieben werden)

1. $S \to UT \mid c, \quad T \to SD, \quad U \to AB, \quad A \to a, \quad B \to b, \quad D \to d$
2. $S \to aBSD \mid c, \quad B \to b, \quad D \to d$

Aufgabe 3 *Betrachten Sie folgende CFG:*
$G = \langle\{S, T\}, \{a, b\}, \{S \to abTTb \mid ab, T \to aa \mid bS\}, S\rangle$
Geben Sie eine äquivalente Grammatik in Chomsky Normalform an.

Lösung:

Einführen von Präterminalen:

$S \to C_a C_b T T C_b \mid C_a C_b, T \to C_a C_a \mid C_b S, C_a \to a, C_b \to b$

Binarisierung:

$S \to C_a X_1 \mid C_a C_b, X \to C_b X_2, X_2 \to T X_3, X_3 \to T C_b,$

$T \to C_a C_a \mid C_b S,$

$C_a \to a, C_b \to b$

(Es gibt auch andere mögliche Binarisierungsreihenfolgen, dies ist die Links-nach-Rechts Binarisierung von den Vorlesungsfolien.)

Aufgabe 4 *Zeigen Sie, dass die Sprache* $\{wc^n \mid w \in \{a,b\}^*, |w|_a = |w|_b = n, n \geq 1\}$ *nicht kontextfrei ist.*

Tipp: Sie dürfen als bewiesen voraussetzen, dass $\{a^n b^n c^n \mid n \geq 1\}$ *nicht kontextfrei ist. Nutzen Sie die Abgeschlossenheit von kontextfreien Sprachen unter Schnittbildung mit regulären Sprachen.*

Lösung:

Wir nehmen an, $L = \{wc^n \mid w \in \{a,b\}^*, |w|_a = |w|_b = n, n \geq 1\}$ ist kontextfrei. Dann muss der Schnitt von L mit der regulären Sprache $L(a^*b^*c^*)$ auch kontextfrei sein. Dieser Schnitt ist aber $\{a^n b^n c^n \mid n \geq 1\}$, wovon wir schon wissen, dass es nicht kontextfrei ist.

Widerspruch, also ist unsere Annahme falsch und L ist nicht kontextfrei.