Einführung in die Computerlinguistik
Hausaufgabe (CFG 3), Abgabe 21.06.2022, 8.30 Uhr

Laura Kallmeyer

Sommer 2022, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1

1. *Geben Sie für die Sprache* $\{wa^n a^n w^R \mid n > 0, w \in \{a,b\}^*\}$ *eine CFG in Chomsky Normalform an.*
2. *Betrachten Sie folgende CFG:*
   $G = \langle \{S, A, B\}, \{a, b\}, \{S \rightarrow aSB \mid ASB \mid ab, A \rightarrow aa, B \rightarrow b\}, S\rangle$
   Geben Sie eine äquivalente Grammatik in Chomsky Normalform an.
3. *Betrachten Sie folgende CFG:*
   $G = \langle \{S, A, B, C, D\}, \{a, b\}, \{S \rightarrow aSB \mid ASB \mid ab, A \rightarrow AA, B \rightarrow b \mid bC, D \rightarrow b \mid aS\}, S\rangle$
   Geben Sie eine äquivalente Grammatik ohne nutzlose Symbole an.
4. *Betrachten Sie folgende CFG:*
   $G = \langle \{S, A, B\}, \{a, b\}, \{S \rightarrow ASB \mid a, A \rightarrow BB, B \rightarrow b \mid \varepsilon\}, S\rangle$
   Geben Sie eine äquivalente Grammatik ohne ε*-Produktionen an.*

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

1. $S \rightarrow AT \mid BU \mid AX, \quad T \rightarrow SA, \quad U \rightarrow SB, \quad X \rightarrow a \mid X \rightarrow VA, \quad V \rightarrow AX, \quad A \rightarrow a, \quad B \rightarrow b$
2. $S \rightarrow C_a X \mid AX \mid C_a C_b, X \rightarrow SB, A \rightarrow C_a C_a, B \rightarrow b, C_a \rightarrow a, C_b \rightarrow b$
3. $S \rightarrow aSB \mid ab, B \rightarrow b$
4. $N_\varepsilon = \{A, B\}$.
   Nach Entfernen der ε-Produktion:
   $S \rightarrow ASB \mid SB \mid AS \mid S \mid a, A \rightarrow BB \mid B, B \rightarrow b$

Aufgabe 2

Aus der Vorlesung wissen Sie, dass $L_1 = \{a^n b^n c^n \mid n \geq 1\}$ *nicht kontextfrei ist (s. Folie 18).*

1. *Zeigen Sie unter Bezugnahme auf die Abgeschlossenheit von CFLs unter Homomorphismen, dass auch* $L_2 = \{a^n b^n d^n e^n f^m g^m \mid n \geq 1, m > 2\}$ *nicht kontextfrei ist.*
2. *Zeigen Sie, unter Bezugnahme auf die Abgeschlossenheit von CFLs unter Schnittbildung mit regulären Sprachen, dass auch* $L_3 = \{a^m b^m c^m a^k b^k c^k \mid m + k \geq 1, m \geq 0, k \geq 0\}$ *nicht kontextfrei ist.*

Tipp: Die Beweisführung in beiden Fällen ist ein Widerspruchsbeweis: Man nimmt das Gegenteil von dem an, was man beweisen möchte, und zeigt, dass dies zu einem Widerspruch führt.

Konkret: Man nimmt an, dass die jeweilige Sprache L_i *kontextfrei ist. Dann muss auch ihr Bild unter einem geeigneten Homomorphismus (bzw. ihr Schnitt mit einer geeigneten regulären Sprache) kontextfrei sein. Hier muss man Homomorphismus bzw. reguläre Sprache so wählen, dass man bei einer Srache landet, von der man schon gezeigt hat, dass sie nicht kontextfrei ist. Dies ist ein Widerspruch zu dem, was aus der Annahme der Kontextfreiheit von* L_i *folgt, und somit ist diese Annahme falsch.*

Lösung:

1. Annahme: L_2 ist kontextfrei. Dann ist auch $h(L_2)$ kontextfrei, wobei h ein Homomorphismus ist mit $h(a) = a$, $h(b) = b$, $h(d) = c$, $h(e) = h(f) = h(g) = \varepsilon$. $h(L_2) = L_1$. Wir wissen aber schon, dass L_1 nicht kontextfrei ist, es ergibt sich also ein Widerspruch. Damit kann L_2 auch nicht kontextfrei sein.

2. Annahme: L_3 ist kontextfrei. Dann ist auch der Schnitt von L_3 mit $L(a^+b^+c^+)$, also mit der von $a^+b^+c^+$ denotierten regulären Sprache, kontextfrei. $L_3 \cap L(a^+b^+c^+) = L_1$. Wir wissen aber schon, dass L_1 nicht kontextfrei ist, es ergibt sich also ein Widerspruch. Damit kann L_3 auch nicht kontextfrei sein.