Lösungen zu Übungsblatt 6

Aufgabe 1

Um nachzuweisen, dass eine Sprache L *nicht* kontextfrei ist, genügt es nach dem (starken) Pumping Lemma für kontextfreie Sprachen zu zeigen:

Für jedes $n \geq 0$ existiert ein $z \in L$ mit $|z| \geq n$ so, dass für jede Zerlegung $z = uvwxy$ mit $vx \neq \varepsilon$ (und $|vwx| \leq n$) gilt: Es existiert ein Pumpfaktor $i \in \mathbb{N}$ mit $uv^iwx^iy \notin L$.

Man beachte, dass man für *jedes* $n \in \mathbb{N}$ ein geeignetes Wort $z \in L$ angeben muss und anschließend für *jede* Zerlegung $z = uvwxy$ mit den genannten Eigenschaften einen geeigneten Pumpfaktor $i \in \mathbb{N}$. Man darf also nur z und i selbst auswählen, nicht die Zahl n oder die Zerlegung des Wortes z.

Für die Auswahl eines geeigneten Wortes z (für jedes $n \in \mathbb{N}$) lässt sich kein allgemeines Rezept angeben, aber es gibt einige Strategien, die (allein oder in Kombination miteinander) zum Erfolg führen können, z.B.:

- Man wählt ein Wort, das 'gerade noch' in der Sprache L liegt. Dann hat man eine gute Chance, dass man schon durch Pumpen mit dem Faktor 0 oder 2 die Sprache L verlässt. Diese Strategie wird in Teilaufgabe **b.** verfolgt.

- Man wählt ein Wort, das eine 'einfache' oder 'regelmäßige' Struktur hat. Dann hat man eine bessere Kontrolle darüber, wie die Teilwörter v und x aussehen können, die beim Pumpen wiederholt werden. Diese Strategie wird in **c.** und **d.** verfolgt.

a. $L_1 = \{a^n b^{n^2} \mid n \in \mathbb{N}\}$.

Sei $n \geq 1$. Man wähle $z = a^n b^{n^2}$. Dann ist $z \in L_1$ mit $|z| = n + n^2 \geq n$. Sei nun $z = uvwxy$ eine Zerlegung von z mit $vx \neq \varepsilon$ und $|vwx| \leq n$.

1. Fall: vx enthält nur as, also $vx = a^k$ mit $k \geq 1$. Dann ist $uv^2wx^2y = a^{n+k}b^{n^2} \notin L_2$, da $n^2 \neq (n+k)^2$.

2. Fall: vx enthält mindestens ein b. Sei $m = \#_b(vx)$. Dann gilt $0 < m \leq |vx| \leq |vwx| \leq n$ und $\#_b(uv^2wx^2y) = \#_b(z) + \#_b(vx) = n^2 + m$. Wegen $n^2 < n^2 + m \leq n^2 + n < (n+1)^2$ ist $n^2 + m$ keine Quadratzahl und damit $uv^2wx^2y \notin L_1$.

Also kann man in beiden Fällen den Pumpfaktor $i = 2$ wählen.

b. $L_2 = \{a^i b^j c^k \mid i < j < k\}$.

Sei $n \geq 1$. Man wähle $z = a^n b^{n+1} c^{n+2}$. Dann ist $z \in L_2$ mit $|z| = 3n + 3 \geq n$. Sei nun $z = uvwxy$ mit $vx \neq \varepsilon$ und $|vwx| \leq n$. Wegen $|vwx| \leq n$ enthält vwx und damit auch vx höchstens zwei der drei Zeichen a, b, c. Also existieren zwei *benachbarte* Zeichen (a, b oder b, c), von denen eines in vx vorkommt und das andere nicht. Damit bleiben die folgenden vier Fälle zu betrachten.

1. Fall: vx enthält mindestens ein a, aber kein b. Mit Pumpfaktor 2 erhält man $\#_a(uv^2wx^2y) > \#_a(z) = n$ und $\#_b(uv^2wx^2y) = \#_b(z) = n+1$, also $uv^2wx^2y \notin L_2$.

2. Fall: vx enthält mindestens ein b, aber kein c. Analog zum 1. Fall.

3. Fall: vx enthält mindestens ein b, aber kein a. Mit Pumpfaktor 0 erhält man $\#_b(uv^0wx^0y) < \#_b(z) = n+1$ und $\#_a(uv^0wx^0y) = \#_a(z) = n$, also $uv^0wx^0y \notin L_2$.

4. Fall: vx enthält mindestens ein c, aber kein b. Analog zum 3. Fall.

Damit ist in allen vier Fällen ein Pumpfaktor i gefunden mit $uv^iwx^iz \notin L_2$, also ist L_2 nicht kontextfrei.

c. $L_3 = \{w \in \{a,b\}^* \mid \#_a(w) = \#_b(w) = \#_c(w)\}$.

Sei $n \geq 1$. Man wähle $z = a^nb^nc^n$. Dann ist $z \in L_3$ mit $|z| = 3n \geq n$. Sei nun $z = uvwxy$ mit $vx \neq \varepsilon$ und $|vwx| \leq n$. Wegen $|vwx| \leq n$ enthält vwx und damit auch vx höchstens zwei der drei Zeichen a, b, c, also erhöht sich beim Pumpen mit $i = 2$ die Anzahl von (mindestens) *einem* der drei Zeichen und die Anzahl von (mindestens) einem der übrigen Zeichen bleibt gleich. Also ist $uv^2wx^2y \notin L_3$, d.h. L_3 ist nicht kontextfrei.

d. $L_4 = \{www \mid w \in \{a,b\}^*\}$.

Sei $n \geq 1$. Man wähle $z = a^nba^nba^nb$. Dann ist $z \in L_4$ mit $|z| = 3n+3 \geq n$. Sei nun $z = uvwxy$ mit $vx \neq \varepsilon$ und $|vwx| \leq n$. Wegen $|vwx| \leq n$ enthält vwx und damit auch vx höchstens ein b.

1. Fall: vx enthält (genau) ein b. Dann gilt $\#_b(uv^0wx^0y) = 2$, also $uv^0wx^0y \notin L_4$.

2. Fall: vx enthält nur as. Dann stammen diese as aus höchstens zwei der drei Teilwörter a^n. Also gilt $uv^0wx^0y = a^{n-i}ba^{n-j}ba^{n-k}b$, wobei (mindestens) eine der drei Zahlen i, j, k größer als 0 ist und (mindestens) eine der anderen gleich 0. Da ein Wort der Form $a^{n-i}ba^{n-j}ba^{n-k}b$ nur drei bs enthält, kann es aber nur dann von der Form www sein, wenn $w = a^{n-i}b = a^{n-j}b = a^{n-k}b$, d.h. wenn $i = j = k$ gilt. Damit ist gezeigt, dass $uv^0wx^0y \notin L_4$.

Also kann man in beiden Fällen den Pumpfaktor $i = 0$ wählen.

Aufgabe 2

Per Definition ist eine Sprache L deterministisch kontextfrei, wenn es einen deterministischen Kellerautomaten M gibt, der die Sprache $L\$$ akzeptiert.

Es genügt also, einen deterministischen Kellerautomaten M anzugeben, und zu beweisen, dass $L(M) = L\$$ gilt. Man beachte, dass dazu—wie bei jeder Mengengleichheit—zwei Teilmengenbeziehungen nachzuweisen sind:

(a) $L\$ \subseteq L(M)$, d.h. jedes Wort $w \in L\$$ wird von M akzeptiert,

(b) $L(M) \subseteq L\$$, d.h. jedes Wort $w \notin L\$$ wird von M abgelehnt.

Ein exakter Beweis beider Richtungen kann sehr aufwändig sein und wird deshalb hier nur für den einfachen Kellerautomaten in **a.** angegeben. Bei den übrigen Teilaufgaben wird—etwas informal—in Worten begründet, warum der Kellerautomat richtig arbeitet. Man beachte, dass auch bei solch informalen Begründungen *beide* Teilmengenbeziehungen erwähnt werden sollten.

a. $L_1 = \{wcw^R \mid w \in \{a,b\}^*\}$

Sei $M_1 = (\Sigma, \Gamma, Q, s, F, \Delta)$ mit $\Sigma = \{a, b, c, \$\}, \Gamma = \{a, b\}, Q = \{s, q, f\}$, $F = \{f\}$ und

$$\begin{aligned}
\Delta = \{\ & ((s,a,\varepsilon),(s,a)), && (1)\\
& ((s,b,\varepsilon),(s,b)), && (2)\\
& ((s,c,\varepsilon),(q,\varepsilon)), && (3)\\
& ((q,a,a),(q,\varepsilon)), && (4)\\
& ((q,b,b),(q,\varepsilon)), && (5)\\
& ((q,\$,\varepsilon),(f,\varepsilon))\ \} && (6)
\end{aligned}$$

Es gilt $L_1\$ = L(M_1)$, denn:

$\subseteq$: Sei $w \in L_1\$$, d.h. $w = ucu^R\$$ mit $u \in \{a,b\}^*$. Dann gilt

$$\begin{aligned}
(s,w,\varepsilon) = (s, ucu^R\$, \varepsilon) &\vdash^*_{M_1} (s, cu^R\$, u^R) && \text{mit (1) und (2)}\\
&\vdash_{M_1} (q, u^R\$, u^R) && \text{mit (3)}\\
&\vdash^*_{M_1} (q, \$, \varepsilon) && \text{mit (4) und (5)}\\
&\vdash_{M_1} (f, \varepsilon, \varepsilon) && \text{mit (6)}
\end{aligned}$$

also $w \in L(M_1)$.

$\supseteq$: Sei $w \in L(M_1)$, d.h. $(s, w, \varepsilon) \vdash^*_{M_1} (f, \varepsilon, \varepsilon)$. Da man nur mit (3) durch Lesen eines c von s nach q und nur mit (6) durch Lesen des Endmarkers von q nach f gelangen kann, muss diese Folge von Übergangsschritten folgende Form haben:

$$\begin{aligned}
(s,w,\varepsilon) = (s, ucv\$) &\vdash^*_{M_1} (s, cv\$, u^R) && \text{mit (1) und (2)}\\
&\vdash_{M_1} (q, v\$, u^R) && \text{mit (3)}\\
&\vdash^*_{M_1} (q, \$, \varepsilon) && \text{mit (4) und (5)}\\
&\vdash_{M_1} (f, \varepsilon, \varepsilon) && \text{mit (6)}
\end{aligned}$$

Da sich mit (1) und (2) nur as und bs verschieben lassen, gilt $u \in \{a,b\}^*$, und weil sich mit (4) und (5) nur zwei gleiche Zeichen gegeneinander aufheben, gilt $v = u^R$, also $w = ucu^R\$ \in L_1\$$.

b. $L_2 = \{a^m b^n \mid m \neq n\}$

Man wähle $M_2 = (\Sigma, \Gamma, Q, s, F, \Delta)$ mit $\Sigma = \{a, b, \$\}$, $\Gamma = \{a, b, \#\}$, $Q = \{s, q_1, q_2, f_1, f_2, f\}$, $F = \{f\}$ und

$$
\begin{aligned}
\Delta = \{\ & ((s,\varepsilon,\varepsilon),(q_1,\#)), && (1)\\
& ((q_1,a,\varepsilon),(q_1,a)), && (2)\\
& ((q_1,b,a),(q_2,\varepsilon)), && (3)\\
& ((q_2,b,a),(q_2,\varepsilon)), && (4)\\
& ((q_1,b,\#),(f_1,\varepsilon)), && (5)\\
& ((q_2,b,\#),(f_1,\varepsilon)), && (6)\\
& ((q_1,\$,a),(f_2,\varepsilon)), && (7)\\
& ((q_2,\$,a),(f_2,\varepsilon)), && (8)\\
& ((f_1,b,\varepsilon),(f_1,\varepsilon)), && (9)\\
& ((f_1,\$,\varepsilon),(f,\varepsilon)), && (10)\\
& ((f_2,\varepsilon,a),(f_2,\varepsilon)), && (11)\\
& ((f_2,\varepsilon,\#),(f,\varepsilon))\ \} && (12)
\end{aligned}
$$

M_2 arbeitet wie folgt:

Mit (1) wird # in den Keller gelegt. Dann werden mit (2) as in den Keller verschoben und mit (3) und (4) werden bs in der Eingabe mit as im Keller abgeglichen.

Sind mehr bs als as vorhanden, so stößt man im Keller auf # bevor die bs in der Eingabe aufgebraucht sind. In diesem Fall gelangt man mit (5) oder (6) in den Zustand f_1, entfernt mit (9) die restlichen bs aus der Eingabe und erreicht schließlich mit (10) die Endkonfiguration $(f, \varepsilon, \varepsilon)$.

Sind mehr as als bs vorhanden, so stößt man in der Eingabe auf \$ bevor die as im Keller aufgebraucht sind. In diesem Fall gelangt man mit (7) oder (8) in den Zustand f_2, entfernt mit (11) die restlichen as aus dem Keller und erreicht mit (12) die Endkonfiguration $(f, \varepsilon, \varepsilon)$.

Sind dagegen gleich viele as und bs vorhanden, dann gerät man in die Konfiguration $(q_1, \$, \#)$ oder $(q_2, \$, \#)$, aus der man nicht weiter kommt. Und wenn die as und bs nicht in der richtigen Reihenfolge stehen, d.h. wenn nach dem ersten b noch ein a auftaucht, dann lässt sich dieses a nicht mehr aus der Eingabe entfernen, weil man sich nicht mehr im Zustand q_1 befindet.

Damit ist klar, dass sich M_2 in allen Fällen richtig verhält, d.h. dass er alle Wörter aus $L_2\$$ akzeptiert und die übrigen ablehnt.

Man beachte, dass das Zeichen $\#$ tatsächlich benötigt wird, denn nur mit einer solchen 'Kellerbodenmarkierung' kann ein deterministischer Kellerautomat das Ende des Kellers erkennen (im Gegensatz zu einem nichtdeterministischen, der 'erraten' kann, dass der Keller leer ist).

c. $L_3 = \{ca^m b^n \mid m \neq n\} \cup \{da^m b^{2m} \mid m \geq 0\}$

Ein deterministischer Kellerautomat M_3, der die Sprache $L_3\$$ erkennt, sollte wie folgt arbeiten:

Wenn das Eingabewort mit c beginnt, dann arbeitet er weiter wie der Automat M_2 aus Teilaufgabe **b.**, d.h. er überprüft ob das Restwort von der Form $a^m b^n\$$ mit $m \neq n$ ist. Wenn das Eingabewort mit d beginnt, dann muss er überprüfen ob das Restwort von der Form $a^m b^{2m}\$$ ist.

Man wähle $M_3 = (\Sigma, \Gamma, Q, s, F, \Delta)$ mit $\Sigma = \{a, b, c, d, \$\}$, $\Gamma = \{a, b, \#\}$, $Q = \{s, p_1, p_2, q_1, q_2, f_1, f_2, f\}$, $F = \{f\}$ und

$$
\begin{array}{rll}
\Delta = \{ & ((s, c, \varepsilon), (q_1, \#)), & (1) \\
& \vdots & (2) \text{ bis } (12) \text{ wie in } M_2 \\
& ((s, d, \varepsilon), (p_1, \#)), & (13) \\
& ((p_1, a, \varepsilon), (p_1, a)), & (14) \\
& ((p_1, bb, a), (p_2, \varepsilon)), & (15) \\
& ((p_2, bb, a), (p_2, \varepsilon)), & (16) \\
& ((p_1, \$, \#), (f, \varepsilon)), & (17) \\
& ((p_2, \$, \#), (f, \varepsilon)) \ \} & (18)
\end{array}
$$

Nach den vorhergehenden Überlegungen ist es schon klar, dass ein mit c beginnendes Wort genau dann von M_3 akzeptiert wird, wenn es von der Form $ca^m b^n\$$ mit $m \neq n$ ist. Es bleibt also nur noch der Fall zu betrachten, dass das Eingabewort mit d beginnt.

In diesem Fall wird mit (13) das Zeichen $\#$ in den Keller gelegt, mit (14) werden zunächst as in den Keller verschoben und anschließend mit (15) und (16) je zwei bs in der Eingabe gegen ein a im Keller aufgehoben. Wenn genau doppelt so viele bs wie as vorhanden sind, dann stößt man im Zustand p_1 oder p_2 gleichzeitig auf $\$$ in der Eingabe und auf $\#$ im Keller und gelangt mit (17) oder (18) in die Endkonfiguration $(f, \varepsilon, \varepsilon)$. In allen anderen Fällen stößt man *nicht* gleichzeitig auf $\$$ und $\#$ und erreicht deshalb auch nicht die Endkonfiguration $(f, \varepsilon, \varepsilon)$. Damit ist klar, dass ein mit d beginnendes Wort genau dann akzeptiert wird, wenn es von der Form $da^m b^{2m}\$$ ist.

d. $L_4 = \{a^m cb^n \mid m \neq n\} \cup \{a^m db^{2m} \mid m \geq 0\}$

Einen deterministischen Kellerautomaten M_4, der die Sprache $L_4\$$ erkennt, kann man erhalten, indem man den Automaten M_3 aus Teilaufgabe **c.** etwas abändert.

M_3 entscheidet schon am ersten Zeichen des Eingabewortes, ob er in den Zustand q_1 oder in den Zustand p_1 wechselt, d.h. ob er ein Wort der Form $a^n b^m\$$ mit $m \neq n$ oder ein Wort der Form $a^m b^{2m}\$$ erwartet. Aber in der Anfangsphase benötigt M_3 diese Entscheidung noch gar nicht: Sowohl im Zustand q_1 als auch im Zustand p_1 verschiebt er so lange as in den Keller bis entweder das erste b oder der Endmarker $\$$ auftaucht. Erst von diesem Zeitpunkt an spielt es eine Rolle, ob er zu Beginn ein c oder ein d gelesen hat.

Diese Überlegung legt nahe, wie man einen deterministischen Kellerautomaten M_4 für die Sprache $L_4\$$ erhält, in der das 'entscheidende' Zeichen c oder d erst hinter den as steht: Anstelle der beiden Zustände p_1 und q_1 von M_3 benutzt M_4 nur einen einzigen Zustand p, um as in den Keller zu verschieben. Von p aus wechselt er in den Zustand q_2 sobald ein c auftaucht und in den Zustand p_2 sobald ein d auftaucht und arbeitet dann jeweils weiter wie M_3.

Man wähle $M_4 = (\Sigma, \Gamma, Q, s, F, \Delta)$ mit $\Sigma = \{a, b, c, d, \$\}$, $\Gamma = \{a, b, \#\}$, $Q = \{s, p, p_2, q_2, f_1, f_2, f\}$, $F = \{f\}$ und

$$\begin{array}{rcl} \Delta & = & \{\ ((s,\varepsilon,\varepsilon),(p,\#)), ((p,a,\varepsilon),(p,a)), \\ & & \quad ((p,c,\varepsilon),(q_2,\varepsilon)), ((p,d,\varepsilon),(p_2,\varepsilon))\ \} \\ & \cup & \{(4),\ (6),\ (8) \text{ bis } (12),\ (16) \text{ und } (18) \text{ aus } M_3\} \end{array}$$

Dass tatsächlich $L(M_4) = L_4\$$ gilt, sieht man wie folgt:

Zunächst ist klar, dass ein Wort höchstens dann von M_4 akzeptiert wird, wenn es von der Form $a^m cu$ oder $a^m du$ (mit $m \geq 0$) ist, denn nur mit Wörtern dieser Form kann M_4 vom Startzustand s über p in einen der anderen Zustände gelangen. Für ein solches Wort gilt dann

- $(s, a^m cu, \varepsilon) \vdash_{M_4} (p, a^m cu, \#) \vdash^*_{M_4} (p, cu, a^m\#) \vdash_{M_4} (q_2, u, a^m\#)$
- $(s, a^m du, \varepsilon) \vdash_{M_4} (p, a^m du, \#) \vdash^*_{M_4} (p, du, a^m\#) \vdash_{M_4} (p_2, u, a^m\#)$

und im Zustand q_2 oder p_2 arbeitet M_4 weiter wie M_3, d.h.

- $(q_2, u, a^m\#) \vdash^*_{M_4} (f, \varepsilon, \varepsilon) \;\Leftrightarrow\; u = b^n\$$ mit $m \neq n$
- $(p_2, u, a^m\#) \vdash^*_{M_4} (f, \varepsilon, \varepsilon) \;\Leftrightarrow\; u = b^{2m}\$$

Also ist

$$\begin{array}{rcl} w \in L(M_4) & \Leftrightarrow & (s, w, \varepsilon) \vdash^*_{M_4} (f, \varepsilon, \varepsilon) \\ & \Leftrightarrow & w = a^m cb^n\$ \text{ mit } m \neq n \text{ oder } w = a^m db^{2m}\$ \\ & \Leftrightarrow & w \in L_4\$ \end{array}$$

Aufgabe 3

Wegen $\varepsilon \notin L(G)$ dürfen wir—nach den Sätzen 3.27 und 3.28 der Vorlesung—annehmen, dass G weder ε-Produktionen noch Einheitsproduktionen enthält.

a. Für jedes $a \in \Sigma$ sei B_a ein neues Nichtterminalzeichen, d.h. ein Nichtterminalzeichen, das nicht in N liegt, und für jedes $\gamma \in (N \cup \Sigma)^+$ mit $|\gamma| > 1$ sei γ' das Wort, das aus γ entsteht, indem man jedes $a \in \Sigma$ durch B_a ersetzt. Dann wählen wir $G' = (\Sigma, N', S, P')$ mit $N' = N \cup \{B_a \mid a \in \Sigma\}$ und

$$\begin{aligned} P' &= \{A \to a \mid (A \to a) \in P \wedge a \in \Sigma\} \\ &\cup \{A \to \gamma' \mid (A \to \gamma) \in P \wedge \gamma \notin \Sigma\} \\ &\cup \{B_a \to a \mid a \in \Sigma\} \end{aligned}$$

Dann gilt offensichtlich $L(G) = L(G')$ und jede rechte Seite einer Produktion in G' ist entweder ein Terminalzeichen $a \in \Sigma$ oder ein Wort $\gamma' \in (N')^+$ mit $|\gamma'| \geq 2$.

b. Jede Produktion der Form $A \to B_1 \ldots B_k$ mit $k > 2$ kann man ersetzen durch 2 Produktionen $A \to B_1 A'$ und $A' \to B_2 \ldots B_k$, wobei A' ein neues Nichtterminalzeichen ist. Diesen Umformungsschritt führt man so lange mit der Grammatik G' aus Teilaufgabe **a.** durch, bis keine Produktionen $A \to \gamma$ mit $|\gamma| > 2$ mehr existieren. Dann sind alle rechten Seiten von der gewünschten Form.

c. Nach dem Verfahren aus **a.** erhält man $G' = (\Sigma, N', S, P')$ mit $N' = \{S, A, B, B_a, B_b\}$ und

$$\begin{aligned} P' &= \{S \to B_a A,\ S \to B_b B,\ A \to B B_a B_a,\ A \to B_b B_a\} \\ &\cup \{B \to B_b A A,\ B \to B_a B_b,\ B_a \to a,\ B_b \to b\} \end{aligned}$$

Nach dem Verfahren aus **b.** erhält man daraus eine Grammatik in Chomsky-Normalform: $G'' = (\Sigma, N'', S, P'')$ mit $N'' = N' \cup \{A', B'\}$ und

$$\begin{aligned} P'' &= \{S \to B_a A,\ S \to B_b B,\ A \to B A', A' \to B_a B_a,\ A \to B_b B_a\} \\ &\cup \{B \to B_b B',\ B' \to A A,\ B \to B_a B_b,\ B_a \to a,\ B_b \to b\} \end{aligned}$$