Aufgaben zum Tut am 7.11.2006

Thomas Pajor

7. November 2006

Aufgabe 1.

Sei $G = (V, \Sigma, P, S)$ eine Grammatik mit $\Sigma := \{a, b, c\}$, $V := \{S, A\}$ und

$$P := \{S \to AA, \quad A \to AAA \mid abA \mid aAb \mid bAa \mid baA \mid c\}$$

(a) Geben Sie den (maximalen) Chomsky–Typ der Grammatik G an.

(b) Bestimmen Sie alle Wörter von $L(G)$, die sich mit vier oder weniger Ableitungen erzeugen lassen.

(c) Ist das Wort $w = acbbca$ in $L(G)$ enthalten? Falls ja, bestimmen Sie einen Syntaxbaum. Ist er eindeutig?

Lösung.

(a) G ist offensichtlich nicht vom Typ CH-3, denn die Regel $S \to AA \in P$ entspricht nicht den Einschränkungen für eine Typ 3 Grammatik. Allerdings gilt für alle Produktionen $l \to r \in P$, dass $l \in V$ und $r \in (\Sigma \cup V)^*$. Die Grammatik entspricht also den Einschränkungen des Typs 2.

(b) Wir betrachten sukzessive die Mengen erzeugbarer Wörter mit $i = 1, \ldots, 4$ Ableitungsschritten. Sie seien mit M_i bezeichnet.

- Mit nur einer Ableitung lässt sich kein Wort erzeugen, denn $S \to AA$ ist die einzige Produktion, die aus dem Startsymbol hervorgeht. Es gilt also $M_0 = \emptyset$.
- Mit zwei Ableitungen lässt sich ebenfalls nichts erzeugen. Somit ist $M_2 = \emptyset$.
- Die einzigen Ableitungsketten der Länge drei, die Wörter aus Σ^* erzeugen sind $S \Rightarrow AA \Rightarrow cA \Rightarrow cc$ bzw. $S \Rightarrow AA \Rightarrow Ac \Rightarrow cc$. Also ist $M_3 = \{cc\}$.

- Wir betrachten nun die Wörter, die durch Anwedung von vier Ableitungen erzeugt werden können.

  Es sind mindestens zwei Ableitungen nötig, um das Symbol A in ein Wort aus $M_p := \{abc, cab, bca, bac\}$ zu überführen. Da S nur in AA überführt werden kann, ist die einzige Möglichkeit mit vier Ableitungen ein Wort zu erzeugen, das Überführen des einen As in ein Wort aus M_p und das Überführen des anderen As in ein c. Die Ableitungsregel $A \to AAA$ brauchen wir garnicht zu betrachten, da hier mindestens 6 Schritte nötig wären um ein Terminal–Wort zu erzeugen. Wir können also festhalten, dass $M_4 = M_p \cdot \{c\} \cup \{c\} \cdot M_p$.

Es folgt, dass die Menge M der Wörter, die sich durch mindestens vier Ableitungen erzeugen lassen die folgende ist

$$\begin{aligned} M &= \bigcup_{i=1}^{4} M_i \\ &= \emptyset \cup \emptyset \cup \{cc\} \cup \{abcc, cabc, bcac, bacc, cabc, ccab, cbca, cbac\} \\ &= \{cc, abcc, cabc, bcac, bacc, cabc, ccab, cbca, cbac\} \end{aligned}$$

(c) Das Wort $w = acbbca$ ist in $L(G)$ enthalten. Der Syntaxbaum aus Abbildung 1 illustriert, wie sich das Wort aus S ableiten lässt.

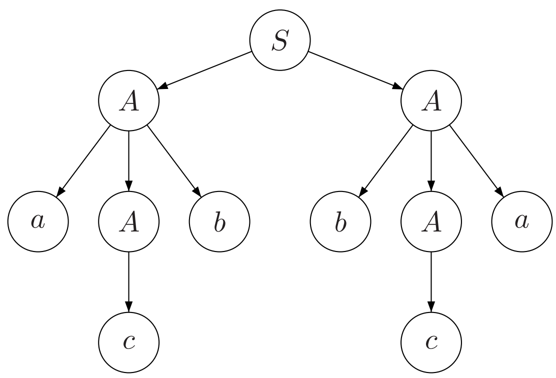

Abbildung 1: Syntaxbaum für das Wort $acbbca$

Der Baum ist außerdem eindeutig, denn die einzige Möglichkeit c zu produzieren, ist durch direkte Ableitung von einem A. Um die Zeichenkette acb bzw. bca zu produzieren, muss diese zwangsläufig aus aAb bzw. bAa abgeleitet worden sein. Die einzige Möglichkeit diese wiederrum zu produzieren ist aus dem Wort AA, was aus S abgeleitet wird. Der Baum muss also eindeutig sein.

Aufgabe 2.

Gegeben sei folgende Grammatik $G := (\{S\}, \{a, b\}, P, S)$ mit

$$
\begin{aligned}
P := \{S &\to aSa, &(1)\\
S &\to bSb, &(2)\\
S &\to aa, &(3)\\
S &\to bb\} &(4)
\end{aligned}
$$

(a) Bestimmen Sie den (maximalen) Chomsky–Typ der Grammatik G.

(b) Bestimmen Sie die durch G erzeugte Sprache $L(G)$ und beweisen Sie ihre Behauptung.

Lösung.

(a) Gleich wegen der ersten Ableitungsregel kann die Grammatik nicht vom Typ CH–3 sein. Da es aber keine Ableitungsregel gibt, die im Widerpsruch zu CH–2 steht, ist G vom Typ CH–2.

(b) Behauptung: Die durch G erzeugte Sprache ist $L = \{ww^R \mid w \in \Sigma^+\}$, also die Sprache aller Palindrome gerader Länge über Σ. Wir beweisen $L = L(G)$ in zwei Schritten:

1. $L \subseteq L(G)$:
   Wir führen Induktion über die Wortlänge der Wörter in L. Das kürzeste Wort in L hat offensichtlich Länge 2, also
   - IA: Sei $w \in L$ mit $|w| = 2$. Dann ist $w = aa$ oder $w = bb$. Diese werden durch G erzeugt in dem man beginnend vom Startsymbol S die dritte oder vierte Ableitungsregel anwendet.
   - IV: Die Behauptung gelte für ein beliebiges n (n gerade).
   - IS: $n \rightsquigarrow n + 2$:
     Sei $w \in L$ mit $|w| = n + 2$. Dann lässt sich w schreiben als $w = xvx$ mit $v \in L$ und $x \in \Sigma$.
     Da $|v| = n$ ist nach IV $v \in L(G)$, es existiert also eine Folge von Produktionen mit $S \overset{*}{\Rightarrow} v$. Je nach x fügen wir nun die erste bzw. zweite Produktion an den Anfang der Ableitungskette an, und erhalten somit $S \Rightarrow xSx \overset{*}{\Rightarrow} xvx = w$, woraus $w \in L(G)$ folgt.
2. $L(G) \subseteq L$:
   Wir führen Induktion über die Anzahl n der Ableitungsregeln, die Wörter w mit $w \in \Sigma^*$ erzeugen[1].

[1] „Unfertige“ Ableitungsfolgen, bei denen das Wort noch Variablen enthält, wollen wir außer acht lassen.

IA: Sei $n = 1$, es kommen also nur die Ableitungsfolgen $S \to aa$ bzw. $S \to bb$ in Frage. Sowohl aa alsauch bb sind in L enthalten.

IV: Die Behauptung gelte für ein beliebiges n.

IS: $n \rightsquigarrow n + 1$:
Betrachte die Folge von Ableitungsregeln $S \Rightarrow xSx \overset{*}{\Rightarrow} xvx =: w$. Nach Induktionsvoraussetzung ist $v \in L$, da v nach n Ableitungen produziert wird. Dann ist aber auch $xvx \in L$.

$\Rightarrow L = L(G)$. $\square$