Normalformen

Eine kontextfreie Grammatik $G = (V, \Sigma, R, S)$ heißt in *Chomsky Normalform*, falls alle Produktionsregeln in R eine der beiden folgenden Formen haben:

$$A \to \sigma \quad \text{für ein } \sigma \text{ aus } \Sigma$$

$$A \to BC \quad \text{für } B, C \in V$$

Syntaxbäume sind „bis auf die letzten Schritte" binär.

Beispiel einer Transformation in Chomsky Normalform:

$$
\begin{array}{lcl}
S & \rightarrow & aACa \\
A & \rightarrow & B \mid a \\
B & \rightarrow & C \mid c \\
C & \rightarrow & cC \mid \varepsilon
\end{array}
$$

Elimination von ε-Regeln ($U \rightarrow \varepsilon$):

$$
\begin{array}{lcl}
S & \longrightarrow & aACa \\
A & \longrightarrow & B \mid a \\
B & \longrightarrow & C \mid c \\
C & \longrightarrow & cC \mid \varepsilon
\end{array}
$$

$$
\begin{array}{lcl}
S & \longrightarrow & aACa \mid aCa \mid aAa \mid aa \\
A & \longrightarrow & B \mid a \\
B & \longrightarrow & C \mid c \\
C & \longrightarrow & cC \mid c
\end{array}
$$

Elimination von Kettenregeln ($U \rightarrow W$):

$$
\begin{array}{lcl}
S & \rightarrow & aACa \mid aCa \mid aAa \mid aa \\
A & \rightarrow & B \mid a \\
B & \rightarrow & C \mid c \\
C & \rightarrow & cC \mid c
\end{array}
$$

$$
\begin{array}{lcl}
S & \rightarrow & aACa \mid aCa \mid aAa \mid aa \\
A & \rightarrow & cC \mid c \mid a \\
B & \rightarrow & cC \mid c \\
C & \rightarrow & cC \mid c
\end{array}
$$

Elimination nichtisolierter Terminalsymbole

$$\begin{array}{lcl} S & \to & aACa \mid aAa \mid aCa \mid aa \\ A & \to & a \mid c \mid cC \\ B & \to & c \mid cC \\ C & \to & c \mid cC \end{array}$$

$$\begin{array}{lcl} S & \to & T_aACT_a \mid T_aAT_a \mid T_aCT_a \mid T_aT_a \\ A & \to & a \mid c \mid T_cC \\ B & \to & c \mid T_cC \\ C & \to & c \mid T_cC \\ T_a & \to & a \\ T_c & \to & c \end{array}$$

und langer rechter Seiten:

$$
\begin{array}{lcl}
S & \longrightarrow & T_aACT_a \mid T_aAT_a \mid T_aCT_a \mid T_aT_a \\
 & \vdots & \\
T_c & \longrightarrow & c
\end{array}
$$

$$
\begin{array}{lcl}
S & \longrightarrow & T_aS_1 \mid T_aS_3 \mid T_aS_4 \mid T_aT_a \\
S_1 & \longrightarrow & AS_2 \\
S_2 & \longrightarrow & CT_a \\
S_3 & \longrightarrow & AT_a \\
S_4 & \longrightarrow & CT_a \\
A & \longrightarrow & a \mid c \mid T_cC \\
B & \longrightarrow & c \mid T_cC \\
C & \longrightarrow & c \mid T_cC \\
T_a & \longrightarrow & a \\
T_c & \longrightarrow & c
\end{array}
$$

Transformation allgemein:

Elimination von ε-Regeln:
Sei $G = (V, \Sigma, R, S)$ eine kontextfreie Grammatik mit $\varepsilon \notin L(G)$.

Sei $V_\varepsilon = \{A \in V \mid A \Rightarrow_G^* \varepsilon\}$. Wir streichen alle Regeln der Form $T \to \varepsilon$ und fügen für jede Regel der Form $U \to vEw$ mit $E \in V_\varepsilon$ und $vw \neq \varepsilon$ eine Regel $U \to vw$ ein.
Falls es auf einer rechten Seite mehrere Vorkommen von Variablen aus V_ε gibt, so müssen wir für alle möglichen Kombinationen dieser Variablen neue Regeln einführen.

Elimination von Kettenregelzyklen:
Sei $G = (V, \Sigma, R, S)$ eine kontextfreie Grammatik ohne ε-Regeln.

Falls es eine Menge von Variablen $T_1, \ldots, T_k$ mit $T_1 \rightarrow T_2$, ..., $T_{k-1} \rightarrow T_k$ und $T_k \rightarrow T_1$ gibt, ersetzen wir alle Vorkommen von $T_1, \ldots, T_k$ durch eine einzige neue Variable T.

Elimination von Kettenregeln:
Sei $G = (V, \Sigma, R, S)$ eine kontextfreie Grammatik ohne ε-Regeln und Kettenregelzyklen.

Da es keine Kettenregelzyklen gibt, können wir die Variablen so bezeichnen, dass $V = \{A_1, \ldots, A_n\}$ und aus $A_i \to A_j$ folgt, dass $i < j$.

Wir gehen die Regeln für $k = n-1, \ldots, 1$ durch:
Falls es eine Regel $A_k \to A_{k'}$ mit $k' > k$ gibt mit Regeln

$$A_{k'} \to \alpha_1 \mid \cdots \mid \alpha_m$$

streichen wir $A_k \to A_{k'}$ und fügen die Regeln

$$A_k \to \alpha_1 \mid \cdots \mid \alpha_m$$

hinzu.

Elimination von nichtisolierten Terminalsymbolen auf rechten Seiten:
Sei $G = (V, \Sigma, R, S)$ eine kontextfreie Grammatik ohne ε-Regeln und Kettenregeln.

Bei jeder Regel, die ein Terminalsymbol σ auf einer rechten Seite der Länge mindestens zwei enthält, ersetzen wir das Terminalsymbol durch eine neue Variable T_σ und fügen die Regel $T_\sigma \to \sigma$ hinzu.

Elimination von langen rechten Seiten:

Sei $G = (V, \Sigma, R, S)$ eine kontextfreie Grammatik, bei der alle Regel von der Form $A \to \sigma$ mit $\sigma \in \Sigma$ oder von der Form

$$A \to B_1 B_2 \cdots B_k$$

sind.

Falls bei einer Regel der zweiten Form $k \geq 3$ ist, führen wir neue Variablen $C_2, \ldots, C_{k-1}$ ein und ersetzen die Regel durch

$$\begin{array}{lcl} A & \to & B_1 C_2 \\ C_2 & \to & B_2 C_3 \\ & \vdots & \\ C_{k-1} & \to & B_{k-1} B_k \end{array}$$

Zwei Grammatiken G und G' heißen *äquivalent* genau dann wenn $L(G') = L(G)$.

Satz:

Sei G eine kontextfreie Grammatik mit $\varepsilon \notin L(G)$. Dann gibt es eine äquivalente kontextfreie Grammatik G' in Chomsky Normalform.

Beweisskizze: Transformiere G wie oben beschrieben in eine kontextfreie Grammatik G' in Chomsky Normalform. Dann gilt $L(G') = L(G)$.

∎

Eine kontextfreie Grammatik $G = (V, \Sigma, R, S)$ heißt in (strenger) *Greibach Normalform*, falls $R \subseteq V \times \Sigma V^*$, d.h., alle Produktionsregeln haben die Form

$$A \to \sigma B_1 B_2 \ldots B_k$$

für ein $\sigma \in \Sigma$, $k \geq 0$ und $B_1, B_2, \ldots, B_k \in V$.

Satz:

Sei G eine kontextfreie Grammatik mit $\varepsilon \notin L(G)$. Dann gibt es eine äquivalente kontextfreie Grammatik G' in Greibach Normalform.