

Chomsky Normalform

Wir illustrieren den Algorithmus, der eine kontextfreie Grammatik in Chomsky Normalform überführt. Ausgangspunkt ist die Grammatik mit $V = \{S, A, B, C, D\}$, $\Sigma = \{a, b\}$ und den folgenden Produktionen:

$$S \rightarrow aAa \mid bBb$$

$$A \rightarrow C \mid a$$

$$B \rightarrow C \mid b$$

$$C \rightarrow CD \mid \varepsilon$$

$$D \rightarrow A \mid B \mid ab$$

Schritt 1: Trennen der Terminalsymbole. Für jedes Terminalsymbol $\sigma \in \Sigma$ führen wir eine neue Variable V_σ ein und ersetzen jedes Vorkommen von σ durch V_σ . Danach fügen wir für jedes $\sigma \in \Sigma$ die neue Produktion $V_\sigma \rightarrow \sigma$ hinzu.

$$S \rightarrow V_aAV_a \mid V_bBV_b$$

$$A \rightarrow C \mid V_a$$

$$B \rightarrow C \mid V_b$$

$$C \rightarrow CD \mid \varepsilon$$

$$D \rightarrow A \mid B \mid V_aV_b$$

$$V_a \rightarrow a \quad V_b \rightarrow b$$

Schritt 2: Auflösen von langen Produktionen. Für jede Produktion der Form $X \rightarrow Y_1Y_2 \dots Y_k$ mit $k \geq 3$ führen wir neue Variablen Z_2, Z_3, \dots, Z_{k-1} ein und ersetzen $X \rightarrow Y_1Y_2 \dots Y_k$ durch $X \rightarrow Y_1Z_2, Z_2 \rightarrow Y_2Z_3, \dots, Z_{k-1} \rightarrow Y_{k-1}Y_k$.

$$S \rightarrow V_aE \mid V_bF$$

$$A \rightarrow C \mid V_a$$

$$B \rightarrow C \mid V_b$$

$$C \rightarrow CD \mid \varepsilon$$

$$D \rightarrow A \mid B \mid V_aV_b$$

$$E \rightarrow AV_a$$

$$F \rightarrow BV_b$$

$$V_a \rightarrow a \quad V_b \rightarrow b$$

Schritt 3: Beseitigen der ε -Produktionen. Zunächst bestimmen wir alle Variablen, aus denen sich ε ableiten lässt. Dazu markieren wir zunächst alle Variablen X , für die es eine Regel $X \rightarrow \varepsilon$ gibt. In unserem Beispiel wird C markiert.

Dann wiederholen wir die folgende Regel, solange es geht: wenn eine unmarkierte Variable X existiert, so dass auf der rechten Seite einer Produktion $X \rightarrow \alpha$ nur markierte Variablen vorkommen, markiere X .

In unserem Beispiel markieren wir A und B (wegen $A \rightarrow C$ und $B \rightarrow C$) und dann D (wegen $D \rightarrow A$). Zum Schluss sind A , B , C und D markiert, und das sind genau die Variablen, aus denen man ε ableiten kann.

Schließlich betrachten wir jede Produktion der Form $X \rightarrow YZ$. Wenn Y markiert ist, fügen wir die Produktion $X \rightarrow Z$ hinzu, wenn Z markiert ist, fügen wir die Produktion $X \rightarrow Y$ hinzu. Dann löschen wir die Produktionen $X \rightarrow \varepsilon$.

$$\begin{aligned}
 S &\rightarrow V_a E \mid V_b F \\
 A &\rightarrow C \mid V_a \\
 B &\rightarrow C \mid V_b \\
 C &\rightarrow CD \mid C \mid D \\
 D &\rightarrow A \mid B \mid V_a V_b \\
 E &\rightarrow AV_a \mid V_a \\
 F &\rightarrow BV_b \mid V_b \\
 V_a &\rightarrow a \quad V_b \rightarrow b
 \end{aligned}$$

Schritt 4: Beseitigen der Kettenregeln. Zunächst müssen wir Kreise der Form $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \dots \rightarrow X_1$ beseitigen. Die aktuelle Grammatik enthält den Kreis $A \rightarrow C \rightarrow D \rightarrow B \rightarrow C \rightarrow D \rightarrow A$. Wir führen eine neue Variable G ein und ersetzen alle Vorkommen von A , B , C , D durch G .

$$\begin{aligned}
 S &\rightarrow V_a E \mid V_b F \\
 G &\rightarrow G \mid V_a \\
 G &\rightarrow G \mid V_b \\
 G &\rightarrow GG \mid G \mid G \\
 G &\rightarrow G \mid G \mid V_a V_b \\
 E &\rightarrow GV_a \mid V_a \\
 F &\rightarrow GV_b \mid V_b \\
 V_a &\rightarrow a \quad V_b \rightarrow b
 \end{aligned}$$

Dann beseitigen wir alle Regeln der Form $G \rightarrow G$.

$$\begin{aligned}
 S &\rightarrow V_a E \mid V_b F \\
 G &\rightarrow GG \mid V_a \mid V_b \mid V_a V_b \\
 E &\rightarrow GV_a \mid V_a \\
 F &\rightarrow GV_b \mid V_b \\
 V_a &\rightarrow a \quad V_b \rightarrow b
 \end{aligned}$$

Jetzt enthält die Grammatik keine Kreise mehr. Wir ordnen die Variablen so, dass bei jeder Kettenregel $X \rightarrow Y$ die Variable X vor der Variablen Y kommt. Eine solche Ordnung ist zum Beispiel S, G, E, F, V_a, V_b . (Eine andere gültige Ordnung wäre S, F, E, G, V_a, V_b , aber nicht S, E, G, V_a, V_b, F , da es die Regel $F \rightarrow V_b$ gibt).

Dann gehen wir die Variablen in umgekehrter Reihenfolge durch. Für jede Variable X ersetzen wir jede Kettenregel $X \rightarrow Y$ durch neue Produktionen der Form $X \rightarrow \alpha$, für alle Produktionen $Y \rightarrow \alpha$.

In unserem Beispiel tun wir also folgendes: Wir betrachten zunächst V_b . Es ist nichts zu tun, weil es keine Produktion der Form $V_b \rightarrow X$, X Variable, gibt.

Dann betrachten wir V_a , wo es genauso aussieht.

Dann betrachten wir F . Es gibt die Produktion $F \rightarrow V_b$. Wir ersetzen sie durch $F \rightarrow b$, da es die Produktion $V_b \rightarrow b$ gibt.

Dann kommt E , wo wir $E \rightarrow V_a$ durch $E \rightarrow a$ ersetzen.

Danach ist G an der Reihe. Wir ersetzen $G \rightarrow V_a$ durch $G \rightarrow a$ und $G \rightarrow V_b$ durch $G \rightarrow b$.

Zum Schluss kommt S , wo nichts zu tun ist.

Die resultierende Grammatik ist in Chomsky Normalform:

$$\begin{aligned}
 S &\rightarrow V_a E \mid V_b F \\
 G &\rightarrow G G \mid a \mid b \mid V_a V_b \\
 E &\rightarrow G V_a \mid a \\
 F &\rightarrow G V_b \mid b \\
 V_a &\rightarrow a \quad V_b \rightarrow b
 \end{aligned}$$