Theoretische Informatik 1

WiSe 23/24

Prof. Dr. Sebastian Siebertz
AG Theoretische Informatik
MZH, Raum 3160
siebertz@uni-bremen.de

Universität
Bremen

Wiederholung Grammatiken

Definition Grammatik

Eine *Grammatik* ist von der Form $G = (N, \Sigma, P, S)$, wobei

- N und Σ endliche, disjunkte Alphabete von *Nichtterminalsymbolen* bzw. *Terminalsymbolen* sind,
- $S \in N$ das *Startsymbol* ist,
- $P \subseteq (N \cup \Sigma)^+ \times (N \cup \Sigma)^*$ eine endliche Menge von Ersetzungsregeln (*Produktionen*) ist.

- Wir schreiben Produktionen $(u, v) \in P$ gewöhnlich als $u \longrightarrow v$.
- Wir schreiben Elemente von N mit Großbuchstaben und Elemente von Σ mit Kleinbuchstaben.

Ableitbarkeit, erzeugte Sprache

Definition Ableitbarkeit, erzeugte Sprache

Sei $G = (N, \Sigma, P, S)$ eine Grammatik und x, y Wörter aus $(N \cup \Sigma)^*$.

1) *y aus x direkt ableitbar:*
   $x \vdash_G y \Leftrightarrow x = x_1 u x_2$ und $y = x_1 v x_2$ mit $u \longrightarrow v \in P$ und $x_1, x_2 \in (N \cup \Sigma)^*$
2) *y aus x in n Schritten ableitbar:*
   $x \vdash_G^n y \Leftrightarrow x \vdash_G x_1 \vdash_G \cdots \vdash_G x_{n-1} \vdash_G y$ für $x_1, \ldots, x_{n-1} \in (N \cup \Sigma)^*$
3) *y aus x ableitbar:*
   $x \vdash_G^* y \Leftrightarrow x \vdash_G^n y$ für ein $n \geq 0$
4) Die durch G *erzeugte Sprache* ist
   $L(G) := \{w \in \Sigma^* \mid S \vdash_G^* w\}$.

Die Chomsky-Hierarchy

Definition Chomsky-Hierarchy

Es sei $G = (N, \Sigma, P, S)$ eine Grammatik.

Typ 0		Jede Grammatik G
Typ 1	monoton	Falls alle Regeln in G *nicht verkürzend* sind, also die Form $w \longrightarrow u$ haben wobei $w, u \in (\Sigma \cup N)^+$ und $\lvert u \rvert \geq \lvert w \rvert$
Typ 2	kontextfrei	Falls alle Regeln in G die Form $A \longrightarrow w$ haben mit $A \in N, w \in (\Sigma \cup N)^*$
Typ 3	rechtslinear	Falls alle Regeln in G die Form $A \longrightarrow uB$ oder $A \longrightarrow u$ haben mit $A, B \in N$, $u \in \Sigma^*$.

Sprachklassen der Chomsky-Hierarchie

Definition Sprachklassen

Für $i = 0, 1, 2, 3$ ist die *Klasse der Typ-i-Sprachen definiert* als

$$\mathcal{L}_i := \{L(G) \mid G \text{ ist Grammatik vom Typ } i\}.$$

Beobachtung:

- Jede Typ 3 Grammatik ist auch eine Typ 2 Grammatik. ($\mathcal{L}_3 \subseteq \mathcal{L}_2$)
- Jede Typ 1 Grammatik ist auch eine Typ 0 Grammatik. ($\mathcal{L}_1 \subseteq \mathcal{L}_0$)
- Typ 2 und Typ 3 Grammatiken erlauben das Verkürzen des abgeleiteten Wortes, sie sind also nicht immer Typ 1 Grammatiken!
- Heute: Jede Typ-2-Grammatik kann in äquivalente Typ-1 Grammatik umgewandelt werden.

Sprachklassen der Chomsky-Hierarchie

Wir werden zeigen: Die Sprachtypen bilden eine *strikte Hierarchie*.

Satz: Chomsky-Hierarchie

Es gilt $\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0$.

- Nach Definition der Grammatiktypen gilt $\mathcal{L}_3 \subseteq \mathcal{L}_2$ und $\mathcal{L}_1 \subseteq \mathcal{L}_0$.
- Die Inklusion $\mathcal{L}_2 \subseteq \mathcal{L}_1$ werden wir später zeigen.
- Auch die Striktheit der Inklusionen werden wir teilweise heute beweisen.

Satz: Typ 3 = regulär

Die Typ-3-Sprachen sind genau die regulären Sprachen, d. h.:

$$\mathcal{L}_3 = \{L \mid L \text{ ist regulär}\}$$

Regulär vs. kontextfrei

Satz

$\mathcal{L}_3 \subset \mathcal{L}_2$.

- Nach Definition der Grammatiktypen gilt $\mathcal{L}_3 \subseteq \mathcal{L}_2$.
- Die Inklusion ist strikt, denn die kontextfreie Grammatik

  $S \longrightarrow aSb$

  $S \longrightarrow \varepsilon$

  erzeugt $L = \{a^n b^n : n \geq 0\}$.
- L ist nicht regulär, also nicht Typ 3.

Weiteres Beispiel kontextfrei und regulär

- $L = \{a^n b^m \mid n \neq m\}$.
- Kontextfreie Grammatik $G = (N = \{S, A, B\}, \Sigma = \{a, b\}, P, S)$:

$$
\begin{array}{rcl}
P & = & \{S \longrightarrow aA, \quad S \longrightarrow Bb, \\
 & & \;\; A \longrightarrow aAb, \quad B \longrightarrow aBb, \\
 & & \;\; A \longrightarrow aA, \quad B \longrightarrow Bb, \\
 & & \;\; A \longrightarrow \varepsilon, \quad B \longrightarrow \varepsilon\}
\end{array}
$$

- $A \vdash_G^* w \in \{a, b\}^* \quad \Rightarrow \quad w = a^n b^m$ mit $n \geq m$,
- $B \vdash_G^* w \in \{a, b\}^* \quad \Rightarrow \quad w = a^n b^m$ mit $n \leq m$,
- Also $S \vdash_G^* w \in \{a, b\}^* \quad \Rightarrow$
  - $w = aa^n b^m$ mit $n \geq m$ oder
  - $w = a^n b^m b$ mit $n \leq m$, d. h. $L(G) = \{a^n b^m \mid n \neq m\}$.

Vereinfachung kontextfreier Grammatiken

Terminierende und erreichbare Symbole; reduzierte Grammatik

Es sei $G = (N, \Sigma, P, S)$ eine kontextfreie Grammatik.

1) $A \in N$ heißt *terminierend*, falls es ein $w \in \Sigma^*$ gibt mit $A \vdash_G^* w$.
2) $A \in N$ heißt *erreichbar*, falls es $u, v \in (\Sigma \cup N)^*$ gibt mit $S \vdash_G^* uAv$.
3) G heißt *reduziert*, falls alle Elemente von N *erreichbar* und *terminierend* sind.

Zwei Grammatiken heißen *äquivalent*, falls sie dieselbe Sprache erzeugen.

Terminierende Symbole

Lemma

Für jede kontextfreie Grammatik $G = (N, \Sigma, P, S)$ ist die Menge der terminierenden Symbole (in polynomieller Zeit) berechenbar.

Beweis.

Erinnerung: $A \in N$ heißt *terminierend*, falls es ein $w \in \Sigma^*$ gibt mit $A \vdash_G^* w$.

- Wir definieren
  - $T_1 := \{A \in N \mid \exists w \in \Sigma^* : A \longrightarrow w \in P\}$
  - $T_{i+1} := T_i \cup \{A \in N \mid \exists w \in (\Sigma \cup T_i)^* : A \longrightarrow w \in P\}$
- Es gilt $T_1 \subseteq T_2 \subseteq \cdots \subseteq N$.
- Da N endlich ist, gibt es ein k mit $T_k = T_{k+1} = \bigcup_{i \geq 1} T_i$.
- Behauptung: $T_k = \{A \in N \mid A \text{ ist terminierend}\}$.
  - Beweis per Induktion (siehe Skript).

Beispiel

- Sei $G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$ die Grammatik mit

$$
\begin{aligned}
P \;=\; \{ \quad & S \longrightarrow A \mid aCbBa, \\
& A \longrightarrow aBAc, \\
& B \longrightarrow Cab, \\
& C \longrightarrow AB \mid aa \quad \}.
\end{aligned}
$$

- $T_1 = \{C\} \quad \subsetneq \quad T_2 = \{C, B\} \quad \subsetneq \quad T_3 = \{C, B, S\} \quad = \quad T_4.$
- Es ist also A das einzige nichtterminierende Symbol.

Leerheitsproblem für kontextfreie Grammatiken

Definition: Leerheitsproblem für kontextfreie Grammatiken

- Gegeben: kontextfreie Grammatik G.
- Frage: gilt $L(G) = \varnothing$?

Satz

Das Leerheitsproblem ist für kontextfreie Grammatiken in polynomieller Zeit entscheidbar.

Beweis.

- Es gilt $L(G) \neq \varnothing \quad \Leftrightarrow \quad \exists w \in \Sigma^* : S \vdash_G^* w \quad \Leftrightarrow \quad S$ ist terminierend.
- Unsere Konstruktion zum Berechnen der terminierenden Symbole kann in polynomieller Zeit ausgeführt werden.

Erreichbare Symbole

Lemma

Für jede kontextfreie Grammatik $G = (N, \Sigma, P, S)$ ist die Menge der erreichbaren Nichtterminalsymbole (polynomieller Zeit) berechenbar.

Beweis.

Erinnerung: $A \in N$ heißt *erreichbar*, falls es $u, v \in (\Sigma \cup N)^*$ gibt mit $S \vdash_G^* uAv$.

- Wir definieren
  - $E_0 := \{S\}$
  - $E_{i+1} := E_i \cup \{A \mid \exists B \in E_i \text{ mit Regel } B \longrightarrow u_1 A u_2 \in P\}$.
- Es gilt $E_0 \subseteq E_1 \subseteq E_2 \subseteq \cdots \subseteq N$.
- Da N endlich ist, gibt es ein k mit $E_k = E_{k+1}$ und damit $E_k = \bigcup_{i \geq 0} E_i$.
- Behauptung: $E_k = \{A \in N \mid A \text{ ist erreichbar}\}$.
  - Beweis per Induktion (siehe Skript).

Beispiel

- Sei $G = (N, \Sigma, P, S)$ die Grammatik mit

$$
\begin{aligned}
P \;=\; \{ \quad & S \longrightarrow aS \mid SB \mid SS \mid \varepsilon, \\
& A \longrightarrow ASB \mid C, \\
& B \longrightarrow Cb, \qquad\qquad\qquad \}
\end{aligned}
$$

- $E_0 = \{S\} \quad \subsetneq \quad E_1 = \{S, B\} \quad \subsetneq \quad E_2 = \{S, B, C\} \quad = \quad E_3$
- Es ist also A das einzige unerreichbare Symbol.

Reduzierte kontextfreie Grammatik

Korollar

Zu jeder kontextfreien Grammatik G mit $L(G) \neq \emptyset$ kann in (polynomieller Zeit) eine äquivalente reduzierte kontextfreie Grammatik konstruiert werden.

- Eine Grammatik G mit $L(G) = \emptyset$ kann *niemals reduziert* sein, da
  - jede Grammatik ein Startsymbol S haben muss und
  - S in G nicht terminierend sein kann.

Beweis des Korollars:

- Eliminiere *erst* die nichtterminierenden, *dann* die unerreichbaren Symbole.
  - Das Eliminieren der nichtterminierenden Symbole kann neue unerreichbare Symbole erzeugen,
  - aber nicht umgekehrt.

Chomsky-Normalform

Definition

Eine kontextfreie Grammatik (N, Σ, P, S) ist in *Chomsky-Normalform*, wenn alle Ableitungsregeln die folgende Form haben:

- $A \longrightarrow a$ oder $A \longrightarrow BC$ mit $A, B, C \in N,\ a \in \Sigma$
- $S \longrightarrow \varepsilon$ ist erlaubt, wenn es keine Regeln $A \to BC$ mit $S \in \{B, C\}$ gibt.

Beobachtungen:

- Wir werden zeigen: jede kontextfreie Grammatik lässt sich in Chomsky-Normalform umwandeln.
- Es folgt: $\mathcal{L}_2 \subseteq \mathcal{L}_1$.
- Die Chomsky-Normalform ist Basis für effiziente Lösung des Wortproblems für kontextfreie Grammatiken.

Chomsky-Normalform

Definition

Eine kontextfreie Grammatik (N, Σ, P, S) ist in *Chomsky-Normalform*, wenn alle Ableitungsregeln die folgende Form haben:

- $A \longrightarrow a$ oder $A \longrightarrow BC$ mit $A, B, C \in N,\ a \in \Sigma$
- $S \longrightarrow \varepsilon$ ist erlaubt, wenn es keine Regeln $A \to BC$ mit $S \in \{B, C\}$ gibt.

Vorgehen zum erstellen der CNF:

1. Aufheben der Mischung von Terminalen und Nichtterminalen auf den rechten Seiten von Produktionen. [→ separierte Grammatiken]
2. Aufbrechen langer Wörter auf den rechten Seiten von Produktionen.
3. Eliminieren von Regeln der Form $A \longrightarrow \varepsilon$ (ε-*Regeln*).
4. Eliminieren von Regeln der Form $A \longrightarrow B$ (*Kettenregeln*).

Separierte Grammatik

Definition

Eine kontextfreie Grammatik (N, Σ, P, S) heißt *separiert*, wenn alle Ableitungsregeln die folgende Form haben:

- $A \longrightarrow \varepsilon$ für $A \in N$,
- $A \longrightarrow a$ für $A \in N$, $a \in \Sigma$
- $A \longrightarrow N^*$ für $A \in N$

Lemma

Jede kontextfreie Grammatik lässt sich umformen in eine äquivalente separierte kontextfreie Grammatik.

Separierte Grammatik

Beweis.

- Sei (N, Σ, P, S) eine kontextfreie Grammatik. Wir konstruieren eine äquivalente separierte Grammatik (N', Σ, P', S) wie folgt.
- Führe für jedes $a \in \Sigma$ ein neues Nichtterminalsymbol X_a und die Produktion $X_a \longrightarrow a$ ein.
- Ersetze in jeder Produktion $A \longrightarrow w$ mit $w \notin N^*$ alle Terminalsymbole a durch die zugehörigen X_a.
- Offensichtlich gilt $L(G) = L(G')$.

Beispiel

- Sei $G = (\{S, A, B, C\}, \{a, b\}, P, S)$ die kontextfreie Grammatik mit

$$\begin{aligned} P \;=\; \{ & S \longrightarrow aS \mid SS \mid bA \mid A, \\ & A \longrightarrow BB, \\ & B \longrightarrow CC \mid aAbC, \\ & C \longrightarrow \varepsilon \} \end{aligned}$$

- Äquivalente separierte Grammatik $G' = \{S, A, B, C, X_a, X_b\}, \{a, b\}, P', S')$ mit

$$\begin{aligned} P' = \{ \quad & S \longrightarrow X_aS \mid SS \mid X_bA \mid A, \\ & A \longrightarrow BB, \\ & B \longrightarrow CC \mid X_aAX_bC, \\ & C \longrightarrow \varepsilon, \\ & X_a \longrightarrow a, \\ & X_b \longrightarrow b\}. \end{aligned}$$

Aufbrechen langer Wörter

Lemma

Jede separierte kontextfreie Grammatik lässt sich umformen in eine äquivalente Grammatik, die nur Regeln der folgenden Form enthält:

- $A \longrightarrow \varepsilon$ *für* $A \in N$
- $A \longrightarrow a$ *für* $A \in N,\ a \in \Sigma$
- $A \longrightarrow B$ *für* $A, B \in N$
- $A \longrightarrow BC$ *für* $A, B, C \in N$

Beweis.

Wir ersetzen alle Produktionen $A \longrightarrow B_1 \cdots B_n$ für $n > 2$ durch

$$A \longrightarrow B_1 C_1,\ C_1 \longrightarrow B_2 C_2,\ \ldots,\ C_{n-2} \longrightarrow B_{n-1} B_n$$

wobei die C_i jeweils neue Symbole sind.

Beispiel fortgesetzt

- Wir ersetzen die Regeln

$$
\begin{aligned}
P' = \{ \quad S &\longrightarrow X_aS \mid SS \mid X_bA \mid A, \\
A &\longrightarrow BB, \\
B &\longrightarrow CC \mid X_aAX_bC, \\
C &\longrightarrow \varepsilon, \\
X_a &\longrightarrow a, \\
X_b &\longrightarrow b\} \text{ durch}
\end{aligned}
$$

- $$
\begin{aligned}
P'' = \{ \quad S &\longrightarrow X_aS \mid SS \mid X_bA \mid A, \\
A &\longrightarrow BB, \\
B &\longrightarrow CC \mid X_aC_1, \\
C_1 &\longrightarrow AC_2, \\
C_2 &\longrightarrow X_bC, \\
C &\longrightarrow \varepsilon, \\
X_a &\longrightarrow a, \\
X_b &\longrightarrow b\}
\end{aligned}
$$

Eliminieren von ε-Regeln

Definition

Eine kontextfreie Grammatik (N, Σ, P, S) heißt ε-frei, falls gilt:

1) $A \longrightarrow \varepsilon \notin P$ für $A \in N \setminus \{S\}$.
2) Falls $S \longrightarrow \varepsilon \in P$, so kommt S nicht auf der rechten Seite einer Regel vor.

Lemma

Jede kontextfreie Grammatik lässt sich umformen (in polynomieller Zeit) in eine äquivalente ε-freie kontextfreie Grammatik.

Eliminieren von ε-Regeln

Beweis.

- Wir finden zunächst alle $A \in N$ mit $A \vdash_G^* \varepsilon$:
  - $N_1 := \{A \in N \mid A \longrightarrow \varepsilon \in P\}$
  - $N_{i+1} := N_i \cup \{A \in N \mid A \longrightarrow B_1 \cdots B_n \in P \text{ mit } B_1, \ldots, B_n \in N_i\}$
- Es gibt ein k mit $N_k = N_{k+1} = \bigcup_{i \geq 1} N_i$.
- Behauptung: $N_k = \{A \mid A \vdash_G^* \varepsilon\}$.
  - Beweis per Induktion (siehe Skript).

Fortsetzung Eliminieren von ε-Regeln

- Entferne alle Regeln $A \longrightarrow \varepsilon$.
- Füge für alle Regeln

$$A \longrightarrow u_1 B_1 \cdots u_n B_n u_{n+1}$$

  mit $B_1, \ldots, B_n \in N_k$ und $u_1, \ldots, u_{n+1} \in (\Sigma \cup N \setminus N_k)^*$die Regeln

$$A \longrightarrow u_1 \beta_1 u_2 \cdots u_n \beta_n u_{n+1}$$

  hinzu für alle $\beta_1 \in \{B_1, \varepsilon\}, \ldots, \beta_n \in \{B_n, \varepsilon\}$ mit $u_1 \beta_1 u_2 \cdots u_n \beta_n u_{n+1} \neq \varepsilon$.
- Falls $\varepsilon \notin L(G) \rightarrow$ fertig.
- Sonst füge neues Startsymbol S_0 und die Produktionen
  - $S_0 \longrightarrow S$ und
  - $S_0 \longrightarrow \varepsilon$ ein.

Beispiel fortgesetzt

- Für die Regeln

$$
\begin{aligned}
P'' = \{ \quad & S \longrightarrow X_a S \mid SS \mid X_b A \mid A, \\
& A \longrightarrow BB, \\
& B \longrightarrow CC \mid X_a C_1, \\
& C_1 \longrightarrow AC_2, \\
& C_2 \longrightarrow X_b C, \\
& C \longrightarrow \varepsilon, \\
& X_a \longrightarrow a, \\
& X_b \longrightarrow b\}
\end{aligned}
$$

  finde zunächst alle $A \in N$ mit $A \vdash_G^* \varepsilon$:

- $N_1 = \{C\} \subsetneq N_2 = \{B, C\} \subsetneq N_3 = \{A, B, C\} \subsetneq N_4 = \{A, B, C, S\} = N_5$.

Beispiel fortgesetzt

- Entferne alle Regeln $A \longrightarrow \varepsilon$ $(A \in N)$.

$$
\begin{aligned}
P''' = \{ \quad & S \longrightarrow X_a S \mid SS \mid X_b A \mid A, \\
& A \longrightarrow BB, \\
& B \longrightarrow CC \mid X_a C_1, \\
& C_1 \longrightarrow AC_2, \\
& C_2 \longrightarrow X_b C, \\
& C \longrightarrow \varepsilon, \\
& X_a \longrightarrow a, \\
& X_b \longrightarrow b\}
\end{aligned}
$$

Beispiel fortgesetzt

- Füge für alle Regeln $A \longrightarrow u_1 B_1 \cdots u_n B_n u_{n+1}$ mit $B_1, \ldots, B_n \in N_k$ und $u_1, \ldots, u_{n+1} \in (\Sigma \cup N \setminus N_k)^*$ die Regeln

$$A \longrightarrow u_1 \beta_1 u_2 \cdots u_n \beta_n u_{n+1}$$

hinzu für alle $\beta_1 \in \{B_1, \varepsilon\}, \ldots, \beta_n \in \{B_n, \varepsilon\}$ mit $u_1 \beta_1 u_2 \cdots u_n \beta_n u_{n+1} \neq \varepsilon$.

$$\begin{aligned} P'''' = \{ \;& S \longrightarrow X_a S \mid SS \mid X_b A \mid A \mid X_a \mid S \mid X_b, \\ & A \longrightarrow BB \mid B, \\ & B \longrightarrow CC \mid X_a C_1 \mid C, \\ & C_1 \longrightarrow AC_2 \mid C_2, \\ & C_2 \longrightarrow X_b C \mid X_b, \\ & X_a \longrightarrow a,\ X_b \longrightarrow b\} \end{aligned}$$

$N_k = \{S, A, B, C\}$.

Beispiel fortgesetzt

- Es gilt $\varepsilon \in L(G)$, also füge neues Startsymbol S_0 und die Produktionen
  - $S_0 \longrightarrow S$ und
  - $S_0 \longrightarrow \varepsilon$ ein.

$$
\begin{aligned}
P'''' = \{ \; & S_0 \longrightarrow S \mid \varepsilon \\
& S \longrightarrow X_a S \mid SS \mid X_b A \mid A \mid X_a \mid S \mid X_b, \\
& A \longrightarrow BB \mid B, \\
& B \longrightarrow CC \mid X_a C_1 \mid C, \\
& C_1 \longrightarrow AC_2 \mid C_2, \\
& C_2 \longrightarrow X_b C \mid X_b, \\
& X_a \longrightarrow a, \; X_b \longrightarrow b \}
\end{aligned}
$$

Korollar

Korollar

Es gilt $\mathcal{L}_2 \subseteq \mathcal{L}_1$.

Beweis.

- Jede ε-freie kontextfreie Grammatik ist eine Typ-1-Grammatik, da keine der verbleibenden Regeln verkürzend ist.
- Ausnahme: $S \longrightarrow \varepsilon$, wobei dann S aber wie auch bei Typ 1 gefordert auf keiner rechten Regelseite auftritt.

Elimination von Kettenregeln

Lemma

Jede kontextfreie Grammatik lässt sich umformen (in polynomieller Zeit) in eine äquivalente kontextfreie Grammatik, die keine Kettenregeln enthält.

Beweis.

- Bestimme die Relation $K := \{(A,B) \mid A \vdash_G^* B\}$:
  - $K_0 \ := \ \{(A,A) \ \mid \ A \in N\}$
  - $K_{i+1} \ := \ K_i \ \cup \ \{(A,B) \in N \times N \ \mid \ \exists (A,B') \in K_i : \ B' \longrightarrow B \in P\}.$
- Es gibt ein k mit $K_k(A) = K_{k+1}(A)$.
- Behauptung: $K_k = K$
  - Beweis per Induktion (siehe Skript).

Elimination von Kettenregeln fortgesetzt

- Entferne alle Kettenregeln.
- Füge für jede verbleibende Regel $B \longrightarrow w$ und jedes $(A, B) \in K$ auch $A \longrightarrow w$ hinzu:

$$P' = \{A \longrightarrow w \ : \ (A, B) \in K, \ B \longrightarrow w \in P \ \text{ und } w \notin N\}$$

Beispiel fortgesetzt

- Für die Regeln

$$
\begin{aligned}
P'''' = \{ \; & S_0 \longrightarrow S \mid \varepsilon \\
& S \longrightarrow X_a S \mid SS \mid X_b A \mid A \mid X_a \mid S \mid X_b, \\
& A \longrightarrow BB \mid B, \\
& B \longrightarrow CC \mid X_a C_1 \mid C, \\
& C_1 \longrightarrow A C_2 \mid C_2, \\
& C_2 \longrightarrow X_b C \mid X_b, \\
& X_a \longrightarrow a, \; X_b \longrightarrow b\}
\end{aligned}
$$

berechne K:

  - $K_1 \setminus K_0 = \{(S_0, S), (S, A), (S, X_a), (S, X_b), (A, B), (B, C), (C_1, C_2), (C_2, X_b)\}$.
  - $K_2 \setminus K_1 = \{(S_0, A), (S_0, X_a), (S_0, X_b), (S, B), (A, C), (C_1, X_b)\}$.
  - $K_3 \setminus K_2 = \{(S_0, B), (S, C)\}, K_4 \setminus K_3 = \{(S_0, C)\}, K_5 = K_4 = K$.

Beispiel fortgesetzt

- Entferne alle Kettenregeln.

$$
\begin{aligned}
P'''' = \{\ & S_0 \longrightarrow S \mid \varepsilon \\
& S \longrightarrow X_a S \mid SS \mid X_b A \mid A \mid X_a \mid S \mid X_b, \\
& A \longrightarrow BB \mid B, \\
& B \longrightarrow CC \mid X_a C_1 \mid C, \\
& C_1 \longrightarrow AC_2 \mid C_2, \\
& C_2 \longrightarrow X_b C \mid X_b, \\
& X_a \longrightarrow a,\ X_b \longrightarrow b\}
\end{aligned}
$$

Elimination von Kettenregeln fortgesetzt

- Füge für jede verbleibende Regel $B \longrightarrow w$ und jedes $(A, B) \in K$ auch $A \longrightarrow w$ hinzu

$$
\begin{aligned}
P'''' = \{\ & S_0 \longrightarrow \varepsilon \mid X_a S \mid SS \mid X_b A, \\
& S \longrightarrow X_a S \mid SS \mid X_b A, \\
& A \longrightarrow BB, \\
& B \longrightarrow CC \mid X_a C_1, \\
& C_1 \longrightarrow A C_2, \\
& C_2 \longrightarrow X_b C, \\
& X_a \longrightarrow a,\ X_b \longrightarrow b\}
\end{aligned}
$$

- $K \setminus K_0 = \{(S_0, S), (S, A), (S, X_a), (S, X_b), (A, B), (B, C), (C_1, C_2), (C_2, X_b), (S_0, A), (S_0, X_a), (S_0, X_b), (S, B), (A, C), (C_1, X_b), (S_0, B), (S, C), (S_0, C)\}$.

Elimination von Kettenregeln fortgesetzt

- Füge für jede verbleibende Regel $B \longrightarrow w$ und jedes $(A, B) \in K$ auch $A \longrightarrow w$ hinzu

$$
\begin{aligned}
P'''' = \{\ & S_0 \longrightarrow \varepsilon \mid X_aS \mid SS \mid X_bA \mid BB \mid CC \mid X_aC_1 \mid a \mid b,\\
& S \longrightarrow X_aS \mid SS \mid X_bA \mid BB \mid CC \mid X_aC_1 \mid a \mid b,\\
& A \longrightarrow BB \mid CC \mid X_aC_1,\\
& B \longrightarrow CC \mid X_aC_1,\\
& C_1 \longrightarrow AC_2 \mid X_bC \mid b,\\
& C_2 \longrightarrow X_bC \mid b,\\
& X_a \longrightarrow a,\ X_b \longrightarrow b\}
\end{aligned}
$$

- $K \setminus K_0 = \{(S_0, S), (S, A), (S, X_a), (S, X_b), (A, B), (B, C), (C_1, C_2), (C_2, X_b), (S_0, A), (S_0, X_a), (S_0, X_b), (S, B), (A, C), (C_1, X_b), (S_0, B), (S, C), (S_0, C)\}$.

Chomsky-Normalform

Definition

Eine kontextfreie Grammatik (N, Σ, P, S) ist in *Chomsky-Normalform*, wenn alle Ableitungsregeln die folgende Form haben:

- $A \longrightarrow a$ oder $A \longrightarrow BC$ mit $A, B, C \in N,\ a \in \Sigma$
- $S \longrightarrow \varepsilon$ ist erlaubt, wenn es keine Regeln $A \to BC$ mit $S \in \{B, C\}$ gibt.

Satz

Jede kontextfreie Grammatik lässt sich umformen in eine äquivalente Grammatik in Chomsky-Normalform.

- Das Herstellen der Chomsky-Normalform wie im Beweis kann dazu führen, dass Nichtterminale überflüssig werden.
- Dies verstößt nicht gegen die Chomsky-Normalform. Wir können die entstandene Grammatik nochmals reduzieren ohne die Chomsky-Normalform zu verletzen.