Die Chomsky-Normalform (CNF) für kontextfreie Grammatiken

Def: Eine kontextfreie Grammatik $G = (\Sigma, V, P, S)$ ist in Chomsky-Normalform, wenn alle Regeln folgende Gestalt haben:

$$A \to BC, \quad A, B, C \in V$$
$$A \to a \quad A \in V, a \in \Sigma$$

Ausnahme: Die Regel $S \to \varepsilon$ ist erlaubt, aber dann darf S auf keiner rechten Seite vorkommen.

Bsp. $S \to SS \mid ST \mid 0$ $\quad$ ~~$T \to S0T1$~~ $\quad$ $S \to$ ~~$TTTT$~~

$T \to TS \mid SS \mid 1$ $\quad$ $S \to$ ~~T~~ $\quad$ $T \to$ ~~ε~~

Σ und V gemischt $\quad$ zu kurz $\quad$ zu lang

SATZ: Zu jeder kontextfreien Grammatik G gibt es eine äquivalente Grammatik G' in CNF.

$L(G') = L(G)$

Folgerung: Jede kontextfreie Sprache ist auch eine kontextsensitive Sprache.

Beweis: Transformation von $G \rightsquigarrow G'$

Schritt 1: Trennung von Variablen und Terminalsymbolen

- Füge für jedes $a \in \Sigma$ eine neue Variable V_a hinzu

- Ersetze auf den rechten Seiten jedes $a \in \Sigma$ durch V_a
- Füge die Regeln $V_a \to a$ für alle $a \in \Sigma$ hinzu

Bsp: $T \to S0T1 \rightsquigarrow T \to SV_0TV_1$ $\quad V_1 \to 1$, $V_0 \to 0$

Schritt 2: Behandlung zu <u>langer</u> Regeln

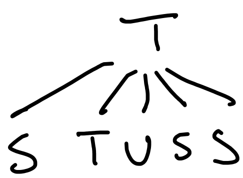
Bsp: $T \to STUSS \rightsquigarrow T \to SW_1$

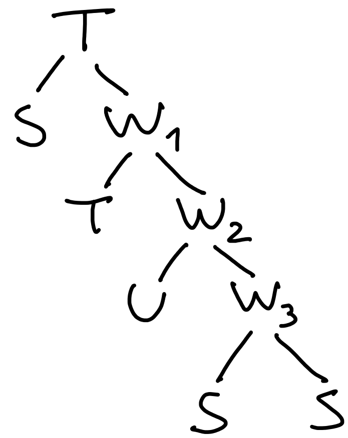
neue Variablen:

$W_1 \to TW_2$

$W_2 \to UW_3$

$W_3 \to SS$





Für jede rechte Seite mit Länge $k \geq 3$ führe $k-2$ neue Variablen und $k-1$ neue Regeln ein, die die ursprünglichen Regeln ersetzen.

Schritt 3: Elimination der ε-Regeln

3a. Bestimme die Menge $V_\varepsilon := \{A \in V \mid A \overset{*}{\Rightarrow} \varepsilon\}$

3b. Vereinfachung der Regeln mit Variablen $\in V_\varepsilon$ auf der rechten Seite

IDEE$_\varepsilon$: $S \to TU$ und $U \overset{*}{\Rightarrow} \varepsilon$:

füge die Regel $S \to T$ hinzu

(3a.) $V_0 := \{A \in V \mid A \to \varepsilon \in P\}$

Wenn $A \to BC$ und $B, C \in V_0 \Rightarrow$ Füge A zu V_0 hinzu

Wenn $A \to B$, und $B \in V_0 \Rightarrow$ Füge A zu V_0 hinzu

Wiederhole, bis sich V_0 nicht mehr ändert.

Am Ende ist $V_0 = V_\varepsilon$.

(3b) Wenn $A \to BC$ und $C \in V_\varepsilon \Rightarrow$ zusätzliche Regel $A \to B$

Wenn $A \to BC$ und $B \in V_\varepsilon \Rightarrow$ zusätzliche Regel $A \to C$

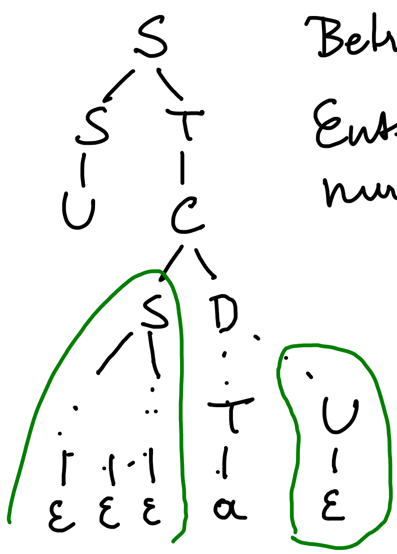
Entferne alle Regeln der Form $A \to \varepsilon$.

G vor Schritt 3b, G' nach Schritt 3b

Behauptung: $\forall A \in V \ \forall w \in \Sigma_i^*: \ A \overset{*}{\underset{G'}{\Rightarrow}} w \Leftrightarrow A \overset{*}{\underset{G}{\Rightarrow}} w$ und $w \neq \varepsilon$

Beweis: „$\Rightarrow$" ✓

„$\Leftarrow$"



Betrachte den Ableitungsbaum für w in G

Entferne alle Teilbäume, die nur ε-Blätter enthalten

Es entsteht ein Ableitungsbaum für w in G'

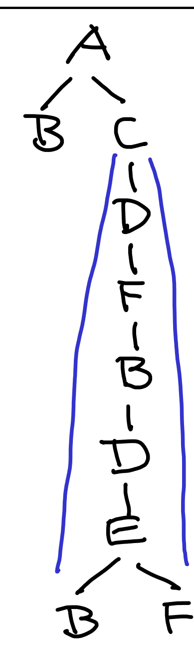
□

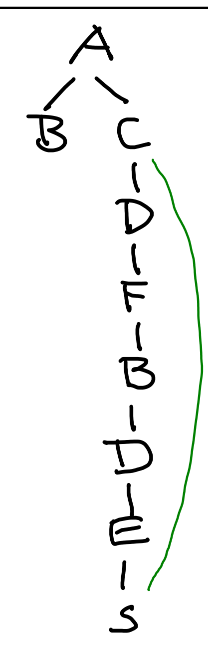
Wenn $\varepsilon \in L(G)$ ist:

Verwende ein neues Startsymbol $\bar{S}$

zusätzliche Regeln $\bar{S} \to \varepsilon \mid S$

- Terminalregeln $A \to a, a \in \Sigma_i$ ✓
- Paarregeln $A \to BC$ ✓
- Einheitsregeln $A \to B, B \in V$ ✗





Schritt (4). Elimination der Einheitsregeln

Wenn $A \overset{*}{\Rightarrow} B$ und $B \to CD$,

Füge Regel $A \to CD$ hinzu

Wenn $A \overset{*}{\Rightarrow} B$ und $B \to a$ $(a \in \Sigma_i)$,

Füge Regel $A \to a$ hinzu

Lass alle Einheitsregeln weg. → neue Grammatik G'

Behauptung $L(G) = L(G')$ „$\supseteq$" ✓

„$\subseteq$" Auf jede maximale Kette $A_1 \to A_2 \to \ldots \to A_k$ von Einheitsregeln im Ableitungsbaum muss eine Paarregel $A_k \to BC$ oder Terminalregel $A_k \to a$ folgen.

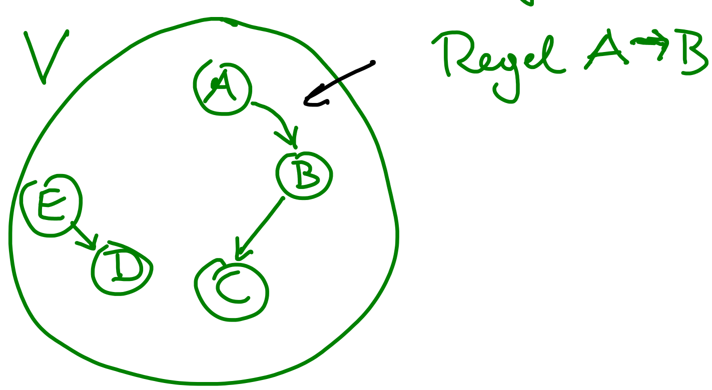
Die Anwendung der Einheitsregeln in der Kette wird durch die neue Regel $A_1 \to BC$ bzw. $A_1 \to a$ ersetzt.

→ Ableitungsbaum ohne Einheitsregeln

Nun ist die Grammatik in CNF. $\square$

* Finden aller Paare $A \overset{*}{\Rightarrow} B$:

Erreichbarkeitsproblem im „Graphen der Einheitsregeln"



Beispiel 1.

$$
\begin{aligned}
S &\to aAa \mid bBb \qquad \Sigma = \{a, b\} \\
A &\to C \mid a \\
B &\to C \mid b \\
C &\to CD \mid \varepsilon \\
D &\to A \mid B \mid ab
\end{aligned}
$$

Schritt (1): Trennen von Variablen und Terminalsymbolen

$$S \to V_a A V_a \mid V_b B V_b \qquad V_a \to a$$
$$A \to C \mid V_a \qquad V_b \to b$$
$$B \to C \mid V_b$$
$$C \to CD \mid \varepsilon$$
$$D \to A \mid B \mid V_a V_b$$

Schritt (2): Behandlung zu langer Regeln

$$S \to V_a E \mid V_b F \qquad V_a \to a$$
$$A \to C \mid V_a \qquad V_b \to b$$
$$B \to C \mid V_b \qquad E \to A V_a$$
$$C \to CD \mid \varepsilon \qquad F \to B V_b$$
$$D \to A \mid B \mid V_a V_b$$

Schritt ③: Elimination der ε-Regeln

$S \to V_a E \mid V_b F$ $\qquad$ $V_a \to a$

$A \to C \mid V_a$ $\qquad$ $V_b \to b$

$B \to C \mid V_b$ $\qquad$ $E \to A V_a \mid V_a$

$C \to \underline{CD} \mid \cancel{\varepsilon} \mid C \mid D$ $\qquad$ $F \to B V_b \mid V_b$

$D \to A \mid B \mid V_a V_b$

③a Bestimme die Menge $V_\varepsilon := \{A \in V \mid A \stackrel{*}{\Rightarrow} \varepsilon\}$

③b Zusätzliche Regeln

Schritt ④. Elimination der Einheitsregeln

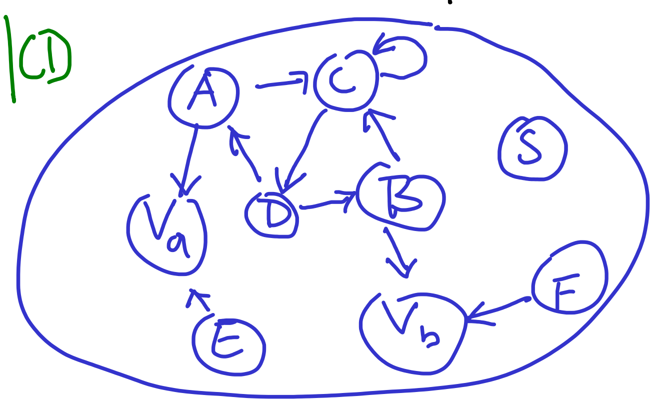
✓ $S \to V_a E \mid V_b F$ $\qquad$ ✓ $V_a \to a$

$A \to \cancel{C} \mid \cancel{V_a} \mid a \mid b \mid CD \mid V_a V_b$ $\qquad$ ✓ $V_b \to b$

$B \to \cancel{C} \mid \cancel{V_b} \mid a \mid b \mid CD \mid V_a V_b$ $\qquad$ ✓ $E \to A V_a \mid \cancel{V_a} \mid a$

✓ $C \to CD \mid \cancel{C} \mid \cancel{D} \mid a \mid b \mid V_a V_b$ $\qquad$ ✓ $F \to B V_b \mid \cancel{V_b} \mid b$

✓ $D \to \cancel{A} \mid \cancel{B} \mid V_a V_b \mid a \mid b \mid CD$



$\boxed{\text{Wenn } A \stackrel{*}{\Rightarrow} B \text{ und } B \to a, \text{ Füge Regel } A \to a \text{ hinzu}}$

$\boxed{\text{Wenn } A \stackrel{*}{\Rightarrow} B \text{ und } B \to CD, \text{ Füge Regel } A \to CD \text{ hinzu}}$

Beispiel 2: eindeutige Grammatik für arithmetische Ausdrücke

„Ausdruck" $A \to T \mid A + T$

„Term" $T \to F \mid T * F$

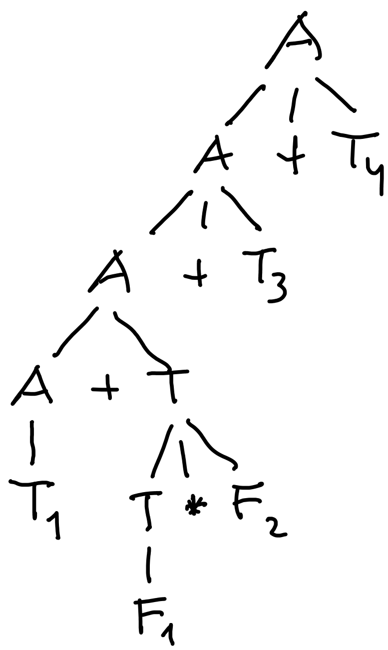
„Faktor" $F \to Z \mid (A)$

„Zahl" $Z \to 1 \mid 2 \mid 3$

jeder Ausdruck ist die Summe von 1 oder mehr Termen, durch „+" verbunden.

$$\underbrace{\underbrace{1}_{T_1}+\underbrace{(2+2)*3}_{T_2}+\underbrace{(1*1)}_{T_3}}_{A}+\underbrace{3}_{T_4}$$

jeder Term ist das Produkt von 1 oder mehr Faktoren, durch „*" verbunden.



$A \to \cancel{T} \mid AB \mid 1 \mid 2 \mid 3 \mid V_(D \mid TC$, $B \to V_+ T$

$T \to \cancel{F} \mid TC \mid 1 \mid 2 \mid 3 \mid V_(D$, $C \to V_* F$

$F \to \cancel{Z} \mid V_(D \mid 1 \mid 2 \mid 3$, $D \to A V_)$

$Z \to 1 \mid 2 \mid 3$

$V_+ \to +$ $V_* \to *$

$V_(\to ($ $V_) \to)$