3 Kontextfreie Sprachen

Idee/Ziel: Beschreibung typischer Programmiersprachen und
des Verhaltens von rekursiven Programmen

Beispiel 3.1.

Mithilfe der folgenden Regeln werden Ausdrücke mit $+$, $-$, $\cdot$, $/$ und Konstanten 0 und 1 beschrieben.

$$
\begin{array}{lll}
\langle\mathsf{expr}\rangle \to \langle\mathsf{expr}\rangle + \langle\mathsf{expr}\rangle & \langle\mathsf{term}\rangle \to \langle\mathsf{term}\rangle \cdot \langle\mathsf{term}\rangle & \langle\mathsf{factor}\rangle \to 0 \\
\langle\mathsf{expr}\rangle \to \langle\mathsf{expr}\rangle - \langle\mathsf{expr}\rangle & \langle\mathsf{term}\rangle \to \langle\mathsf{term}\rangle / \langle\mathsf{term}\rangle & \langle\mathsf{factor}\rangle \to 1 \\
\langle\mathsf{expr}\rangle \to \langle\mathsf{term}\rangle & \langle\mathsf{term}\rangle \to \langle\mathsf{factor}\rangle & \langle\mathsf{factor}\rangle \to (\langle\mathsf{expr}\rangle)
\end{array}
$$

Dies soll jetzt verallgemeinert werden.

3.1 Kontextfreie Grammatiken

Definition. Eine *kontextfreie Grammatik* ist ein Tupel $G = (N, \Sigma, S, P)$ mit folgenden Komponenten:

- N ist eine endliche Menge von *Nichtterminalen*,
- Σ ist ein Alphabet von *Terminalen* mit $N \cap \Sigma = \emptyset$,
- $S \in N$ ist das *Startsymbol* und
- $P \subseteq N \times (N \cup \Sigma)^*$ ist eine endliche Menge von *Regeln*.

Eine *rechtslineare Grammatik* ist eine kontextfreie Grammatik (N, Σ, S, P) mit $P \subseteq N \times \Sigma N \cup \{\varepsilon\}$.

Beispiel 3.1 (Fortsetzung)
$N = \{\langle\mathsf{expr}\rangle, \langle\mathsf{term}\rangle, \langle\mathsf{factor}\rangle\}$
$\Sigma = \{+, -, \cdot, /, 0, 1, (,)\}$
$S = \langle\mathsf{expr}\rangle$
P ist die Menge der angegebenen Regeln (wobei $A \to v$ für (A, v) steht).
Dies ist keine rechtslineare Grammatik (z.B. wegen der Regel $\langle\mathsf{term}\rangle \to \langle\mathsf{term}\rangle \cdot \langle\mathsf{term}\rangle$).

Konvention Elemente von N: $\langle\mathsf{expr}\rangle$, A, B, ...
Elemente von Σ: a, b, ...

Definition. Sei $G = (N, \Sigma, S, P)$ eine kontextfreie Grammatik und $u, v \in (N \cup \Sigma)^*$.

- Das Wort v ist *in einem Schritt* aus dem Wort u *ableitbar*, wenn es Regel $A \to y$ in P und Wörter $x, z \in (N \cup \Sigma)^*$ gibt mit $u = xAz$ und $v = xyz$. Hierfür schreiben wir $u \Rightarrow_G v$.

- Das Wort v ist *(in beliebig vielen Schritten)* aus dem Wort u *ableitbar*, wenn es $n \in \mathbb{N}$ und Wörter $u_0, u_1, \ldots, u_n$ gibt mit $u = u_0 \Rightarrow_G u_1 \Rightarrow_G u_2 \cdots \Rightarrow_G u_n = v$. Hierfür schreiben wir $u \Rightarrow_G^* v$ (insbes. gilt mit $n = 0$ auch $u \Rightarrow_G^* u$).

- Die von G *erzeugte Sprache* ist

$$L(G) = \{u \in \Sigma^* \mid S \Rightarrow_G^* u\}.$$

- Eine Sprache L heißt *kontextfrei*, wenn es eine kontextfreie Grammatik G gibt mit $L = L(G)$.

- Eine Sprache L heißt *rechtslinear*, wenn es eine rechtslineare Grammatik G gibt mit $L = L(G)$.

Beispiel 3.1 (Fortsetzung)
Wir betrachten wieder die Grammatik mit den folgenden Regeln:

$$\begin{aligned}
\langle\mathsf{expr}\rangle &\to \langle\mathsf{expr}\rangle + \langle\mathsf{expr}\rangle\\
\langle\mathsf{expr}\rangle &\to \langle\mathsf{expr}\rangle - \langle\mathsf{expr}\rangle\\
\langle\mathsf{expr}\rangle &\to \langle\mathsf{term}\rangle\\
\langle\mathsf{term}\rangle &\to \langle\mathsf{term}\rangle \cdot \langle\mathsf{term}\rangle\\
\langle\mathsf{term}\rangle &\to \langle\mathsf{term}\rangle / \langle\mathsf{term}\rangle\\
\langle\mathsf{term}\rangle &\to \langle\mathsf{factor}\rangle\\
\langle\mathsf{factor}\rangle &\to 0\\
\langle\mathsf{factor}\rangle &\to 1\\
\langle\mathsf{factor}\rangle &\to (\langle\mathsf{expr}\rangle)
\end{aligned}$$

Dann gibt es z.B. die folgende Ableitung:

$$\begin{aligned}
\langle\mathsf{expr}\rangle &\Rightarrow_G \langle\mathsf{term}\rangle\\
&\Rightarrow_G \langle\mathsf{term}\rangle/\langle\mathsf{term}\rangle\\
&\Rightarrow_G \langle\mathsf{term}\rangle/\langle\mathsf{factor}\rangle\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{expr}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{expr}\rangle - \langle\mathsf{expr}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{expr}\rangle - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{term}\rangle - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(\langle\mathsf{factor}\rangle - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle \cdot \langle\mathsf{term}\rangle/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G \langle\mathsf{term}\rangle \cdot 0/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G 1 \cdot 0/(1 - \langle\mathsf{term}\rangle)\\
&\Rightarrow_G 1 \cdot 0/(1 - 0)
\end{aligned}$$

Beispiel. Sei $G = (N, \Sigma, S, P)$ die kontextfreie Grammatik mit $N = \{S\}$, $\Sigma = \{a, b\}$ und $P = \{S \to \varepsilon, S \to aSb\}$. Wegen $S \Rightarrow_G \varepsilon$ gilt $\varepsilon \in L(G)$. Außerdem $S \Rightarrow_G aSb \Rightarrow_G aaSbb \cdots \Rightarrow_G a^n S b^n \Rightarrow_G a^n \varepsilon b^n$ für alle $n \geq 1$. Da es keine weiteren Ableitungen gibt, gilt sogar

$$L(G) = \{a^n b^n \mid n \in \mathbb{N}\},$$

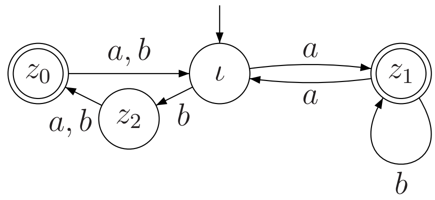
diese Sprache ist also kontextfrei. Im Kapitel 2 haben wir gesehen, daß diese Sprache nicht regulär ist. Es gibt also kontextfreie Sprachen, die nicht regulär sind.

Beispiel. Sei $G = (N, \Sigma, S, P)$ die kontextfreie Grammatik mit $N = \{S\}$, $\Sigma = \{a, b\}$ und $P = \{S \to \varepsilon, S \to aSa, S \to bSb\}$. Dann gilt $L(G) = \{ww^R \mid w \in \Sigma^*\}$. Diese Sprache ist also kontextfrei.

Satz. *Eine Sprache $L \subseteq \Sigma^*$ ist genau dann regulär, wenn sie rechtslinear ist.*

Beweis durch Beispiel:

Sei zunächst L regulär. Dann existiert ein DFA M mit $L(M) = L$, z.B. sei M der rechts stehende DFA:



Konstruiere eine rechtslineare Grammatik $G = (N, \Sigma, S, P)$ mit $L(G) = L(M) = L$:

- $N = \{z_0, \iota, z_1, z_2\}$,
- $\Sigma = \{a, b\}$,
- $S = \iota$ und

P umfaßt die folgenden Regeln:

$$
\begin{array}{lll}
\iota \to a z_1 & \iota \to b z_2 & \\
z_0 \to a\,\iota & z_0 \to b\,\iota & z_0 \to \varepsilon \\
z_1 \to a\,\iota & z_1 \to b z_1 & z_1 \to \varepsilon \\
z_2 \to a z_0 & z_2 \to b z_0 &
\end{array}
$$

Damit ist die Implikation „$\Longrightarrow$“ am Beispiel gezeigt.

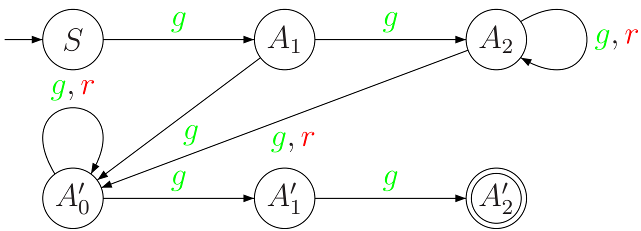
Sei umgekehrt $G = (N, \Sigma, S, P)$ eine rechtslineare Grammatik mit $L(G) = L$, z.B.

- $N = \{S, A_1, A_2, A_0', A_1', A_2'\}$,
- $\Sigma = \{g, r\}$ und

P umfaßt die folgenden Regeln:

$$
\begin{array}{rlrl}
S & \to gA_1 & A_0' & \to gA_0' \mid rA_0' \mid gA_1' \\
A_1 & \to gA_2 \mid gA_0' & A_1' & \to gA_2' \\
A_2 & \to gA_2 \mid rA_2 \mid gA_0' \mid rA_0' & A_2' & \to \varepsilon
\end{array}
$$

Hieraus konstruieren wir den folgenden NFA M mit $L(M) = L(G) = L$:



Also ist L nach Satz 2.10 regulär □

Bemerkung. Wir erhalten:
L regulär
$\Longrightarrow$ L ist Sprache einer rechtslinearen Grammatik
$\Longrightarrow$ L ist Sprache einer kontextfreien Grammatik
$\Longrightarrow$ L ist kontextfrei,
d.h. jede reguläre Sprache ist kontextfrei.

Aber nicht jede kontextfreie Sprache ist regulär (Bsp.: $\{a^n b^n \mid n \in \mathbb{N}\}$).

3.2 Ableitungsbäume

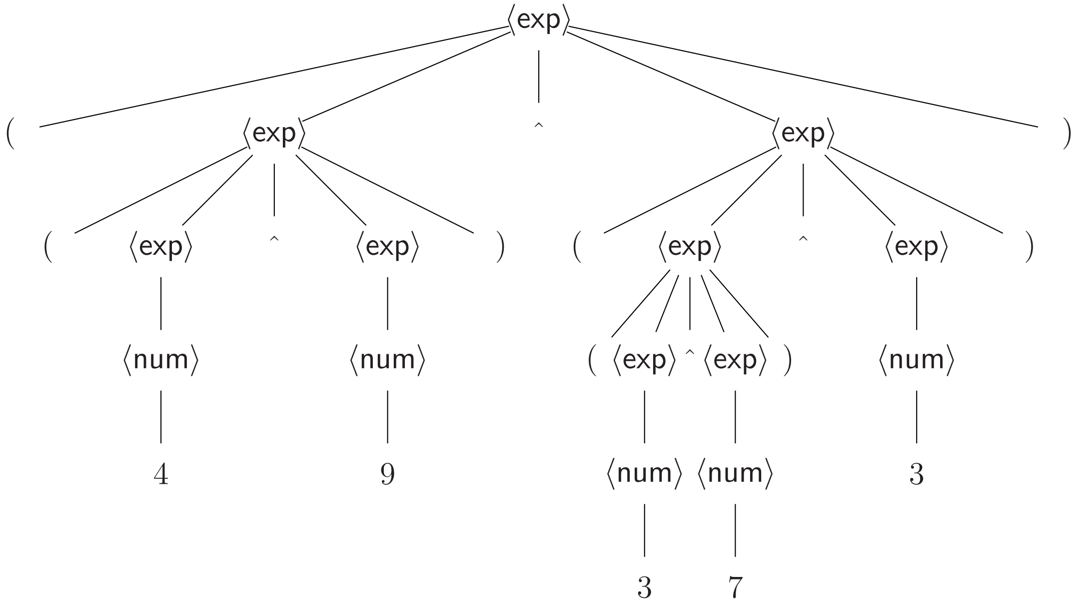
Wir betrachten als Beispiel die kontextfreie Grammatik G mit Startsymbol $\langle\mathsf{exp}\rangle$ und den Regeln

$$\langle\mathsf{num}\rangle \to 0 \mid 1 \mid 2 \cdots \mid 9 \qquad\qquad \langle\mathsf{exp}\rangle \to \langle\mathsf{num}\rangle \mid (\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle)$$

Sie hat z.B. die Ableitung

$$\begin{array}{llll}
\langle\mathsf{exp}\rangle & \Longrightarrow_G (\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle) & & \Longrightarrow_G (\langle\mathsf{exp}\rangle\hat{}(\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle)) \\
& \Longrightarrow_G (\langle\mathsf{exp}\rangle\hat{}(\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{num}\rangle)) & & \Longrightarrow_G (\langle\mathsf{exp}\rangle\hat{}(\langle\mathsf{exp}\rangle\hat{}3)) \\
& \Longrightarrow_G ((\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}(\langle\mathsf{exp}\rangle\hat{}3)) & & \Longrightarrow_G ((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}(\langle\mathsf{exp}\rangle\hat{}3)) \\
& \Longrightarrow_G ((4\hat{}\langle\mathsf{exp}\rangle)\hat{}(\langle\mathsf{exp}\rangle\hat{}3)) & & \Longrightarrow_G ((4\hat{}\langle\mathsf{exp}\rangle)\hat{}((\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}3)) \\
& \Longrightarrow_G ((4\hat{}\langle\mathsf{exp}\rangle)\hat{}((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}3)) & & \Longrightarrow_G ((4\hat{}\langle\mathsf{num}\rangle)\hat{}((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}3)) \\
& \Longrightarrow_G ((4\hat{}\langle\mathsf{num}\rangle)\hat{}((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{num}\rangle)\hat{}3)) & & \Longrightarrow_G ((4\hat{}\langle\mathsf{num}\rangle)\hat{}((\langle\mathsf{num}\rangle\hat{}7)\hat{}3)) \\
& \Longrightarrow_G ((4\hat{}9)\hat{}((\langle\mathsf{num}\rangle\hat{}7)\hat{}3)) & & \Longrightarrow_G ((4\hat{}9)\hat{}((3\hat{}7)\hat{}3))
\end{array}$$

Das Wort $((4\hat{}9)\hat{}((3\hat{}7)\hat{}3))$ gehört also zur von G erzeugten Sprache. Die Klammern in diesem Wort geben ihm eine „Struktur“, die wir in einem Baum veranschaulichen können. Dieser *Ableitungsbaum* entsteht aus der obigen Ableitung.


⟨exp⟩
(⟨exp⟩ ^ ⟨exp⟩)
(⟨exp⟩ ^ ⟨exp⟩) (⟨exp⟩ ^ ⟨exp⟩)
⟨num⟩ ⟨num⟩ (⟨exp⟩ ^ ⟨exp⟩) ⟨num⟩
4 9 ⟨num⟩ ⟨num⟩ 3
3 7

Um ihn zu konstruieren, geht man wie folgt vor:

- Initialisierung: erzeuge einen Knoten mit dem Startsymbol $S = \langle\mathsf{exp}\rangle$
- Für jeden Ableitungsschritt: Werde die Regel $A \to B_1 B_2 \cdots B_k$ angewandt.

  Der Knoten, der dem zu ersetzenden Nichtterminal A entspricht, bekommt k neue Kinder. Diese werden mit den Buchstaben B_1, B_2, ... B_k beschriftet.

Am Ende steht an den Blättern (von links nach rechts gelesen) das Wort ((4ˆ9)ˆ((3ˆ7)ˆ3)). Diese Wort heißt „Blattwort" des Ableitungsbaums.

Der angegebene Ableitungsbaum entsteht auch aus der Ableitung

$$
\begin{aligned}
\langle\mathsf{exp}\rangle &\Longrightarrow_G ((\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle)) && \Longrightarrow_G ((\langle\mathsf{exp}\rangle\hat{}((\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle))) \\
&\Longrightarrow_G ((\langle\mathsf{exp}\rangle\hat{}((\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{num}\rangle))) && \Longrightarrow_G ((\langle\mathsf{exp}\rangle\hat{}((\langle\mathsf{exp}\rangle\hat{}3)) \\
&\Longrightarrow_G (((\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}((\langle\mathsf{exp}\rangle\hat{}3)) && \Longrightarrow_G (((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}((\langle\mathsf{exp}\rangle\hat{}3)) \\
&\Longrightarrow_G ((4\hat{}\langle\mathsf{exp}\rangle)\hat{}((\langle\mathsf{exp}\rangle\hat{}3)) && \Longrightarrow_G ((4\hat{}\langle\mathsf{exp}\rangle)\hat{}(((\langle\mathsf{exp}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}3)) \\
&\Longrightarrow_G ((4\hat{}\langle\mathsf{exp}\rangle)\hat{}(((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}3)) && \Longrightarrow_G ((4\hat{}\langle\mathsf{num}\rangle)\hat{}(((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{exp}\rangle)\hat{}3)) \\
&\Longrightarrow_G ((4\hat{}\langle\mathsf{num}\rangle)\hat{}(((\langle\mathsf{num}\rangle\hat{}\langle\mathsf{num}\rangle)\hat{}3)) && \Longrightarrow_G ((4\hat{}\langle\mathsf{num}\rangle)\hat{}(((\langle\mathsf{num}\rangle\hat{}7)\hat{}3)) \\
&\Longrightarrow_G ((4\hat{}\langle\mathsf{num}\rangle)\hat{}((3\hat{}7)\hat{}3)) && \Longrightarrow_G ((4\hat{}9)\hat{}((3\hat{}7)\hat{}3))
\end{aligned}
$$

Aus jeder Ableitung kann also ein Ableitungsbaum gewonnen werden. Umgekehrt kann aus jedem Ableitungsbaum auch eine Ableitung gewonnen werden. Daher erhält man

Satz. *Sei* $G = (N, \Sigma, S, P)$ *eine kontextfreie Grammatik und* $w \in \Sigma^*$. *Dann sind äquivalent:*

(1) $w \in L(G)$

(2) Es gibt einen Ableitungsbaum mit Blattwort w.

Ableitungsbäume können kompliziert werden, wenn

- es Regeln $A \to w$ mit „langem" Wort w gibt,
- es Regeln $A_1 \to A_2$, $A_2 \to A_3$, ..., $A_{n-1} \to A_n$ mit „großem" n gibt, oder
- es Regeln $A \to \varepsilon$ gibt

Definition. Eine kontextfreie Grammatik $G = (N, \Sigma, S, P)$ ist *in Chomsky-Normalform*, wenn gelten:

- wenn es die Regeln $S \to \varepsilon$ gibt, so kommt S auf keiner rechten Seite vor
- darüber hinaus gibt es nur Regeln der Form $A \to a$ und $A \to BC$ mit $A, B, C \in N$ und $a \in \Sigma$.

Bemerkung. Sei G in Chomsky-Normalform, T ein Ableitungsbaum mit Blattwort $\neq \varepsilon$ und v ein Knoten, der mit einem Nichtterminal beschriftet ist. Dann gilt

- v hat genau zwei Kinder; diese sind mit Nichtterminalen beschriftet
  oder
- v hat genau ein Kind; dieses ist mit einem Terminal beschriftet.

Insbesondere ist kein Knoten mit ε beschriftet.

Satz 3.2. *Aus einer kontextfreien Grammatik G kann eine äquivalente kontextfreie Grammatik in Chomsky-Normalform berechnet werden.*

Beweis durch Beispiel:
Betrachte die kontextfreie Grammatik $G = (\{S\}, \{a, b\}, S, P)$ mit den Regeln

$$S \to \varepsilon \text{ und } S \to aSb\,.$$

<u>Schritt 0:</u> neues Startsymbol S':
$G_0 = (\{S, S'\}, \{a, b\}, S', P_0)$ mit den Regeln

$$S' \to S,\ S \to \varepsilon \text{ und } S \to aSb\,.$$

<u>Schritt 1:</u> Vereinzeln der Terminale:
$G_1 = (\{S, S', A, B\}, \{a, b\}, S', P_1)$ mit den Regeln

$$S' \to S,\ S \to \varepsilon \mid ASB,\ A \to a \text{ und } B \to b\,.$$

Schritt 2: Kürzen der rechten Seiten (durch Einführung neuer Nichtterminale):
$G_2 = (\{S, S', A, B, X\}, \{a, b\}, S', P_2)$ mit den Regeln

$$S' \to S \qquad S \to \varepsilon \mid AX \qquad A \to a \qquad B \to b$$
$$X \to SB$$

Schritt 3: Elimination von leeren rechten Seiten (durch Kontraktion von Ableitungen):
$G_3 = (\{S, S', A, B, X\}, \{a, b\}, S', P_3)$ mit den Regeln

$$S' \to S \mid \varepsilon \qquad S \to AX \qquad A \to a \qquad B \to b$$
$$X \to SB \mid B$$

Schritt 4: Elimination von Kettenregeln $A \to B$ (durch Kontraktion von Ableitungen):
$G_4 = (\{S, S', A, B, X\}, \{a, b\}, S', P_4)$ mit den Regeln

$$S' \to AX \mid \varepsilon \qquad S \to AX \qquad A \to a \qquad B \to b$$
$$X \to SB \mid b$$

$\square$

3.3 Nicht-kontextfreie Sprachen

Fragen:

(1) Existiert eine Sprache, die nicht kontextfrei ist?

(2) Wie kann man zeigen, daß eine Sprache nicht kontextfrei ist?

Satz (Pumpinglemma für kontextfreie Sprachen)**.** *Sei $L \subseteq \Sigma^*$ kontextfrei. Dann existiert $n > 0$, so daß für alle $z \in L$ mit $|z| \geq n$ gilt: es existieren Wörter $u, v, w, x, y \in \Sigma^*$ mit*

(i) $z = uvwxy$,

(ii) $|vwx| \leq n$,

(iii) $|vx| > 0$ *und*

(iv) für alle $m \in \mathbb{N}$ *gilt* $uv^m wx^m y \in L$.

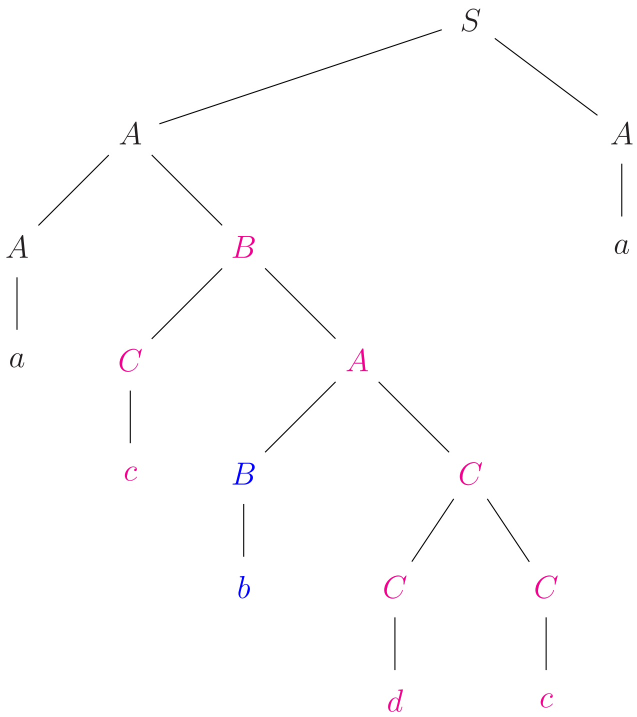
Beweis durch Beispiel:
Da L kontextfrei ist, existiert eine kontextfreie Grammatik G in *Chomsky-Normalform* mit $L = L(G)$, sei z.B. G die kontextfreie Grammatik mit den Regeln

$$S \to AA\,,\; A \to AB \mid BC \mid a\,,\; B \to CB \mid b\,,\; C \to CC \mid c \mid d$$

Das Wort $z = acbdca \in L(G)$ hat die Ableitung

$$\begin{aligned} \underline{S} &\Longrightarrow \underline{A}AS \Longrightarrow a\underline{B}A \Longrightarrow a\underline{C}AA \Longrightarrow ac\underline{A}A \Longrightarrow ac\underline{B}CA \Longrightarrow acb\underline{C}A \\ &\Longrightarrow acb\underline{C}CA \Longrightarrow acbd\underline{C}A \Longrightarrow acbdc\underline{A} \Longrightarrow acbdca \end{aligned}$$

und damit den folgenden Ableitungsbaum:



Auf dem Ast zum Terminal b kommt das Nichtterminal B mehrfach vor.
Die beiden Vorkommen definieren zwei „Teilbäume“ (den violetten und den blauen).
Wenn wir den violetten Teilbaum streichen, erhalten wir einen Ableitungsbaum für aba,
wenn wir ihn verdoppeln einen für $accbdcdca$.

Allgemeiner können wir Ableitungsbäume für alle Wörter der Form

$$a\,\underbrace{c\dots c}_{m \text{ mal } c}\,b\,\underbrace{dc\dots dc}_{m \text{ mal } dc}\,a$$

erhalten. Mit $u = a$, $v = c$, $w = b$, $x = dc$ und $y = a$ haben wir also $uv^m wx^m y \in L(G)$ für alle $m \geq 0$ gezeigt.

Wichtig war hierbei:

- Keine rechte Seite einer Regel ist leer oder besteht aus einzelnem Nichtterminal (was durch die Chomsky-Normalform sichergestellt ist).
- Es gibt einen Ast, auf dem sich ein Nichtterminal wiederholt.

  Hierfür reicht es, einen Ast mit $> |N|$ vielen Kanten zu haben.

  Hierfür wiederum reicht es, daß $|z| > 2^{|N|}$ gilt. Wir setzen also $n = 2^{|N|} + 1$.

$\square$

Beispiel 3.3. Sei $\Sigma = \{a, b, c\}$. Die Sprache $L = \{a^k b^k c^k \mid k \geq 0\}$ ist nicht kontextfrei.

Beweis:
Angenommen, L wäre kontextfrei. Nach dem Pumpinglemma für kontextfreie Sprachen existiert dann $n \geq 1$, so daß für alle $z \in L$ mit $|z| \geq n$ Wörter $u, v, w, x, y \in \Sigma^*$ existieren mit den Eigenschaften (i)-(iv) aus dem Pumpinglemma.
Betrachte das Wort $z = a^n b^n c^n \in L$. Wegen $|z| = 3n \geq n$ existieren $u, v, w, x, y \in \Sigma^*$ mit (i)-(iv)
Dann haben wir $uvwxy \stackrel{(i)}{=} z = a^n b^n c^n$ und $|vwx| \leq n$ (nach (ii)), also $|vwx|_a = 0$ oder $|vwx|_c = 0$. Damit folgt insbes. $|vx|_a = 0$ oder $|vx|_c = 0$.
In beiden Fällen zeigen wir mit $m = 0$, daß $|uv^m wx^m y| = 3n$ gilt:

- angenommen $|vx|_c = 0$. Dann gilt $|uwy|_c = |uvwxy|_c = |z|_c = n$. Nach (iv) haben wir $uwy = uv^0 wx^0 y \in L$, also $|uwy|_a = |uwy|_b = |uwy|_c = n$, also $|uwy| = 3n$.

- Im Fall $|vx|_a = 0$ kann analog argumentiert werden.

Aus $|uwy| = 3n = |z| = |uvwxy|$ folgt $|vx| = 0$, im Widerspruch zu (iii).
Also ist L nicht kontextfrei. $\square$

In Übung gezeigt: Klasse der kontextfreien Sprachen ist abgeschlossen unter Vereinigung, Verkettung und Iteration. Dies gilt auch für Klasse der regulären Sprachen (die auch unter Komplementierung und Schnitt abgeschlossen ist).

Folgerung. *Die Klasse der kontextfreien Sprachen ist nicht abgeschlossen unter Schnitt.*

Beweis:
Die kontextfreie Grammatik mit den Regeln

$$S \to AB \quad A \to \varepsilon \mid aA \quad B \to bBc \mid \varepsilon$$

erzeugt die kontextfreie Sprache $L_1 = \{a^k b^\ell c^\ell \mid k, \ell \geq 0\}$; auch $L_2 = \{a^k b^k c^\ell \mid k, \ell \geq 0\}$ ist kontextfrei.
Es gilt aber $L_1 \cap L_2 = \{a^k b^k c^k \mid k \geq 0\}$. Diese Sprache ist nach Beispiel 3.3 nicht kontextfrei, d.h. die Klasse der kontextfreien Sprachen ist nicht abgeschlossen unter Schnitten.

□

Folgerung. *Die Klasse der kontextfreien Sprachen ist nicht abgeschlossen unter Komplementierung.*

Beweis:
Betrachte die folgenden kontextfreien Sprachen:

- $L_1 = \{a, b, c\}^* \setminus \{a\}^*\{b\}^*\{c\}^*$: erzeugt durch die kf. Grammatik mit den Regeln

$$S_1 \to A_1 baA_1 \mid A_1 cbA_1 \mid A_1 caA_1 \qquad A_1 \to \varepsilon \mid aA_1 \mid bA_1 \mid cA_1 \,.$$

- $L_2 = \{a^k b^\ell c^m \mid k \neq \ell, m \geq 0\}$: erzeugt durch die kf. Grammatik mit den Regeln

$$\begin{array}{lll} S_2 \to A_2 G_2 C_2 \mid G_2 B_2 C_2 & G_2 \to aG_2 b \mid \varepsilon & \\ A_2 \to aA_2 \mid a & B_2 \to B_2 b \mid b & C_2 \to \varepsilon \mid cC_2 \,. \end{array}$$

- $L_3 = \{a^k b^\ell c^m \mid k \geq 0, \ell \neq m\}$: erzeugt durch die analoge kf. Grammatik mit den Regeln

$$\begin{array}{lll} S_3 \to A_3 G_3 C_3 \mid A_3 B_3 G_3 & G_3 \to bG_3 c \mid \varepsilon & \\ A_3 \to aA_3 \mid \varepsilon & B_3 \to B_3 b \mid b & C_3 \to c \mid cC_3 \,. \end{array}$$

Dann ist $L_1 \cup L_2 \cup L_3$ kontextfrei, da diese Sprache durch die Vereinigung der drei kontextfreie Grammatiken mit den neuen Regeln

$$S \to S_1 \mid S_2 \mid S_3$$

erzeugt wird.

Es gilt aber

$$\{a, b, c\}^* \setminus (L_1 \cup L_2 \cup L_3) = \{a^k b^k c^k \mid k \geq 0\},$$

d.h. das Komplement der kontextfreien Sprache $L_1 \cup L_2 \cup L_3$ ist die Sprache aus Beispiel 3.3, also nicht kontextfrei.

Also ist die Klasse der kontextfreien Sprachen nicht abgeschlossen unter Komplementierungen. $\square$

Beispiel. Mit Hilfe des Pumpinglemmas für kontextfreie Sprachen kann auch gezeigt werden, daß die folgenden Sprachen nicht kontextfrei sind (mit $\Sigma = \{a, b\}$):

- $\{a^i b a^j b a^{i \cdot j} \mid i, j \in \mathbb{N}\}$
- $\{uu \mid u \in \Sigma^*\}$
- $\{u_1 \$ u_2 \$ \cdots u_k \# v \mid k \in \mathbb{N}, u_1, \ldots, u_k, v \in \Sigma^*, \exists i\colon u_i = v\}$