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 3.4 (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* $uv^m wx^m y \in L$.

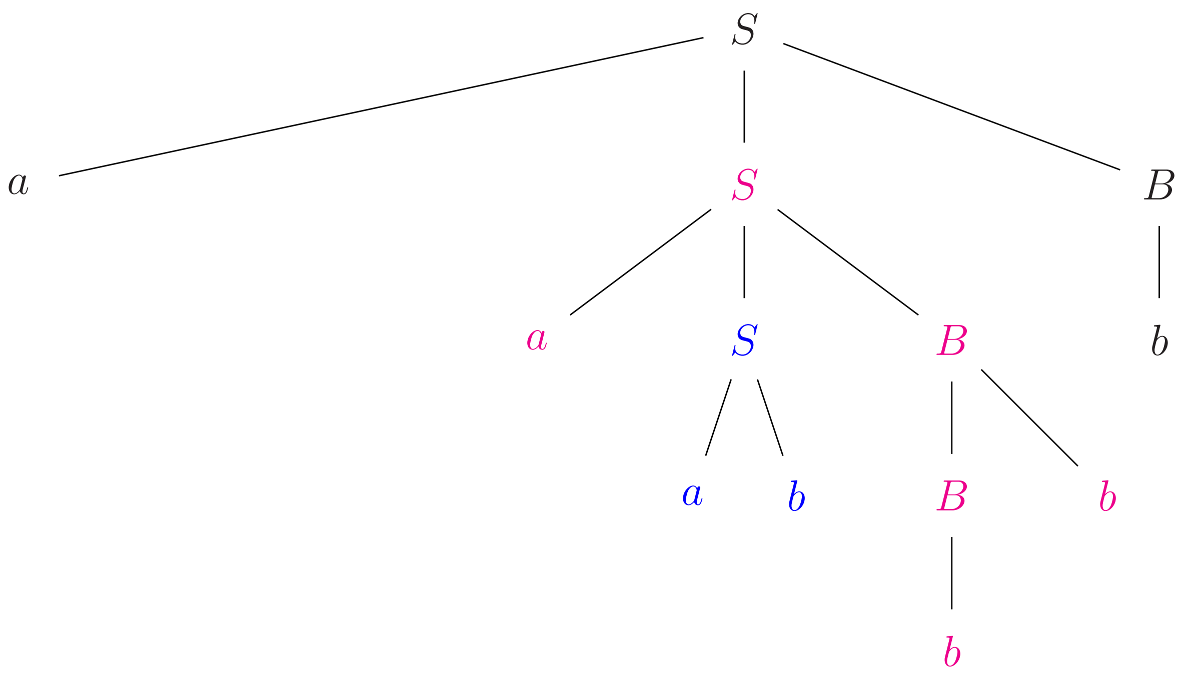
Beweis durch Beispiel:
Da L kontextfrei ist, existiert eine kontextfreie Grammatik G mit $L = L(G)$, sei z.B. G die kontextfreie Grammatik mit den Regeln

$$S \to aSB \mid ab \text{ und } B \to b \mid Bb\,.$$

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

$$\underline{S} \Rightarrow aS\underline{B} \Rightarrow a\underline{S}b \Rightarrow aaS\underline{B}b \Rightarrow aaS\underline{B}bb \Rightarrow a\underline{S}bbb \Rightarrow aaabbbb$$

und damit den folgenden Ableitungsbaum:



Auf dem Ast zum dritten a kommt das Nichtterminal S mehrfach vor. Die beiden unteren Vorkommen definieren zwei „Teilbäume“ (den violetten und den blauen). Wenn wir den violetten Teilbaum streichen, erhalten wir einen Ableitungsbaum für $aabb$, wenn wir ihn verdoppeln einen für $aa\,aabbb\,bbb$.

Allgemeiner können wir Ableitungsbäume für alle Wörter der Form

$$a\,\underbrace{a\ldots a}_{m \text{ mal } a}\,ab\,\underbrace{bb\ldots bb}_{m \text{ mal } bb}\,b$$

erhalten. Mit $u = a$, $v = a$, $w = ab$, $x = bb$ und $y = b$ 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 immer sichergestellt werden kann).

- 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 $|z| >$ maximale Länge einer rechten Seite hoch Anzahl der Nichtterminale ist. Wir setzen also $n = \max\{|w| \mid (A \to w) \in P\}^{|N|} + 1$.

$\square$

Beispiel 3.5. 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 $uvwxy \overset{(i)}{=} z = a^n b^n c^n$ 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. Diese sind auch abgeschlossen unter Komplementierung und Schnitt.

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 3.5 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
$$S_1 \to A_1baA_1 \mid A_1cbA_1 \mid A_1caA_1 \qquad A_1 \to \varepsilon \mid aA_1 \mid bA_1 \mid cA_1\,.$$
- $L_2 = \{a^kb^\ell c^m \mid k \neq \ell, m \geq 0\}$: erzeugt durch die kf. Grammatik mit den Regeln
$$\begin{array}{lll} S_2 \to A_2G_2C_2 \mid G_2B_2C_2 & G_2 \to aG_2b \mid \varepsilon & \\ A_2 \to aA_2 \mid a & B_2 \to B_2b \mid b & C_2 \to \varepsilon \mid cC_2\,. \end{array}$$
- $L_3 = \{a^kb^\ell c^m \mid k \geq 0, \ell \neq m\}$: erzeugt durch die analoge kf. Grammatik mit den Regeln
$$\begin{array}{lll} S_3 \to A_3G_3C_3 \mid A_3B_3G_3 & G_3 \to bG_3c \mid \varepsilon & \\ A_3 \to aA_3 \mid \varepsilon & B_3 \to B_3b \mid b & C_3 \to c \mid cC_3\,. \end{array}$$

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 3.5, 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 $\Sigma = \{a, b\}$):

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