Lösungen zu den Übungsaufgaben im Buch *Automaten und Sprachen: Theoretische Informatik für die Praxis* von Andreas Müller, ISBN 978-3-662-70145-4 (Softcover), ISBN 978-3-662-70146-1 (eBook), https://link.springer.com/book/10.1007/978-3-662-70146-1. Website zum Buch: https://autospr.ch

# Kapitel 8: Nicht kontextfreie Sprachen

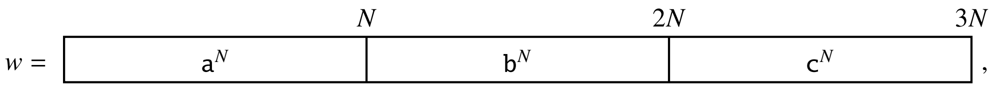
**8.1.** In Abschnitt 8.2 wurde gezeigt, dass die Sprache $\{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \ge 0\}$ über dem Alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ nicht kontextfrei ist. Wenn man die Bedingungen etwas lockert, und nur noch verlangt, dass die Anzahl der verschiedenen Zeichen übereinstimmt, die Reihenfolge aber beliebig sein darf, erhält man die Sprache

$$L = \{w \in \Sigma^* \mid |w|_\mathtt{a} = |w|_\mathtt{b} = |w|_\mathtt{c}\}.$$

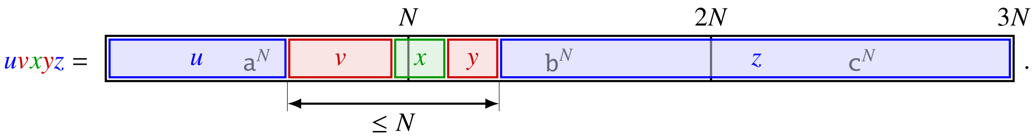
Ist $L$ kontextfrei?

*Lösung.* Nein, auch $L$ ist nicht kontextfrei. Man kann dies mit dem Pumping-Lemma für kontextfreie Sprachen nachweisen, wobei man sogar das gleiche Beispielwort verwenden kann wie im Falle der Sprache $\{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \ge 0\}$ .

❶ Man nimmt dazu an, dass $L$ kontextfrei sei. ❷ Gemäß Pumping-Lemma gibt es daher die Pumping-Length $N$. ❸ Wir konstruieren jetzt das Wort




welches die Voraussetzungen des Pumping-Lemma sicher erfüllt. ❹ Es muss also eine Unterteilung $w = uvxyz$ geben, so dass $|vxy| \le N$ ist.




Die Blöcke $vxy$ können auch anderso liegen, dürfen aber zusammen nicht länger als $N$ sein. ❺ Diese letzte Bedingung hat zur Folge, dass $v$ und $y$ höchstens in zwei der drei Blöcke $\mathtt{a}^N$, $\mathtt{b}^N$ oder $\mathtt{c}^N$ liegen können. Insbesondere ändert sich beim Pumpen nur die Anzahl von zwei der drei Buchstaben, nach auf- oder abpumpen erhält man also ein Wort, bei dem die Anzahl jedes der drei Buchstaben nicht mehr gleich ist, also kein Wort mehr aus $L$, im Widerspruch zur Aussage des Pumping-Lemma. ❻ Daher kann $L$ nicht kontextfrei sein. ○

*Diskussion.* Es genügt nicht, darauf zu verweisen, dass $L$ die Sprache $L' = \{\mathtt{a}^n\mathtt{b}^n\mathtt{c}^n \mid n \ge 0\}$ enthalte, und dass daher auch $L'$ nicht kontextfrei sein. Dann könnte man nämlich auch argumentieren, dass $L$ die kontextfreie Sprache $\emptyset$ enthalte, und daher kontextfreie sei!

Wenn man sich den Pumping-Lemma-Beweis genauer anschaut, kann man auch verstehen, warum es nicht reicht. Der Beweis beruht ja darauf, dass aufgepumpte Wörter nicht mehr in der Sprache sind. Wenn die Sprache aber mehr Wörter umfasst, dann ist es auch schwieriger "aus der Sprache herauszufallen", es ist ja jetzt leichter, die Bedingung der Sprache zu erfüllen. Damit ist es möglich, dass ein Wort, welches in $L'$ nicht aufgepumpt werden kann, in $L$ aufpumpbar wird.

**8.2.** Will man eine Matrix wie

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

in einem Computerprogramm verwenden, braucht man irgendeine Syntax, mit der man Matrizen eingeben kann. Octave und Matlab verwenden zum Beispiel:

```
A = [
    1, 2, 3;
    4, 5, 6;
    7, 8, 9;
]
```

Zeilen werden durch Strichpunkt getrennt, Element in einer Zeile durch Komma. Jede Zeile muss gleich viele Elemente enthalten. Können Sie eine Grammatik formulieren, welche genau korrekte Matrix-Definitionen produziert?

*Lösung.* Nein, die Sprache der korrekten Matrix-Definitionen ist nicht kontextfrei, wie man mit dem Pumping-Lemma zeigen kann.

Wäre das möglich, könnte man auch eine Grammatik aufstellen, die alle Ziffern und Whitespace weglässt, und die daher genau die Strings der Form

$$\texttt{[,}^n\texttt{;,}^n\texttt{;,}^n\texttt{;]} \tag{1}$$

produziert, wobei auch mehr als drei Zeilen möglich sein sollen. Aber auch dieses Sprache ist nicht kontextfrei.

Wäre diese Sprache kontextfrei, müsste es nach dem Pumping Lemma für kontextfreie Sprachen eine Zahl $N$ geben, so dass Wörter $w$ mit $|w| \ge N$ aufgepumpt werden können.

Wir wählen das Wort

$$w = \texttt{[,}^{2N}\texttt{;,}^{2N}\texttt{;,}^{2N}\texttt{;]}$$

Es ist sicher lang genug.

Gemäß dem Pumping Lemma gibt es eine Unterteilung $w = uvxyz$ derart, dass $|vxy| \le N$, $|vy| > 0$ und $uv^kxy^kz$ ist für jedes $k$ ein Ausdruck der Form (1). Wegen der Bedingung $|vxy| \le N$ kann sich $vxy$ über höchstens zwei der Gruppen von Kommas erstrecken, alle Zeilen müssen also immer $2N$ Kommas enthalten. Das Wort $uxz$ (der Fall $k = 0$) enthält zwei oder drei Zeilen. Wären es drei Zeilen, müssten beide $2N$ lang sein, d. h. es wäre gar nichts entfernt worden, im Widerspruch zur Bedingung $|vy| > 0$. Zwei Zeilen sind aber auch nicht möglich, denn dazu hätten $2N$ Kommas entfernt werden müssen, im Widerspruch zu $2N \le |yv| < |yzv| \le N$.

Somit kann $uxz$ nicht ein Wort der Form (1) sein, im Widerspruch zur Aussage des Pumping Lemmas. Dieser Widerspruch zeigt, dass die Sprache nicht kontextfrei sein kann. ○

**8.3.** In einer HTML-Tabelle sollte man in jedem `<tr>`-Element gleich viele `<td>`-Elemente angeben (sofern man nicht `<td>`-Elemente verwendet, die sich über mehrere Spalten oder Zeilen erstrecken). Es stellt sich die Frage, ob man dies durch eine kontextfreie Grammatik sicherstellen könnte. Betrachten Sie dazu die folgende vereinfachte Sprache $L$ über dem Alphabet $\Sigma = \{\texttt{r}, \texttt{d}\}$. Die Buchstaben sollen dabei für `<tr>`-Elemente und `<td>`-Elemente stehen. Die Wörter von $L$ sollen die Form

$$(\texttt{rd}^n)^k$$

haben. Damit ist gemeint, dass jedes Wort aus $k$ Teilstücken besteht, die alle identisch sind und jeweils aus einem `r` und $n$ Zeichen `d` bestehen. Die Wörter

$$\texttt{rrr},\quad \texttt{rddrddrdd},\quad \texttt{rdddddddrddddddd},\quad \texttt{rdddddddddddddd}$$

gehören also zu $L$, die Wörter

$$\texttt{rdr},\quad \texttt{rdrddrddd},\quad \texttt{rdddddddr},\quad \texttt{ddddddddddddddr}$$

aber nicht. Gibt es für diese Sprache eine kontextfreie Grammatik?

*Lösung.* Nein, wie man mit dem Pumping Lemma für kontextfreie Sprachen zeigen kann.

Angenommen $L$ sei kontextfrei, dann gibt es eine Pumping Length $N$, so dass Wörter mit mindestens Länge $N$ aufgepumpt werden können. Wir wählen als Wort

$$w = \texttt{r}\underbrace{\texttt{d}\dots\texttt{d}}_{N}\texttt{r}\underbrace{\texttt{d}\dots\texttt{d}}_{N}\texttt{r}\underbrace{\texttt{d}\dots\texttt{d}}_{N}$$

Es hat Länge $3N + 3$, es muss also eine Zerlegung $w = uvxyz$ geben so, dass $|vxy| \le N$ ist. Wie auch immer diese Zerlegung aussieht, $v$ oder $y$, vielleicht sogar beide, werden ausschließlich aus `d` bestehen. Außerdem können $v$ und $y$ höchstens zwei der Blöcke aus Zeichen $d$ im Wort $w$ überlappen. Der erste oder der letzte Block ist vollständig in $u$ bzw. $z$ enthalten. Insbesondere müssen in einem aufgepumpten Wort immer noch alle Blöcke aus Zeichen `d` gleich lang sein, damit es immer noch in $l$ ist.

Nehmen wir an, $v$ bestehe aus lauter `d`. Dann wird sich beim Aufpumpen "sein" Block verlängern, im Widerspruch zum eben gesagten. Ähnlich ergibt sich auch ein Widerspruch, wenn $y$ aus lauter `d` besteht.

Diese Widersprüche zeigen, dass die Sprache $L$ nicht kontextfrei sein kann. ○

**8.4.** Der Sprachforscher Stuart M. Shieber hat aus Beispielen wie dem Satz

De Jan säit das mer d'Chind em Hans es Hus händ wele laa hälfe aastriiche.

abgeleitet, dass das Schweizerdeutsch grammatische Konstruktionen zulässt, die auf Wörter der Form

$$wa^m b^n xc^m d^n y \tag{2}$$

hinaus laufen[1]. Zeigen Sie, dass die Sprache aus Wörtern der Form (2) nicht kontextfrei ist.

*Lösung.* Wir wenden das Pumping Lemma für kontextfreie Sprachen auf die Sprache

$$L = \{wa^m b^n xc^m d^n y \mid m, n \ge 0\}$$

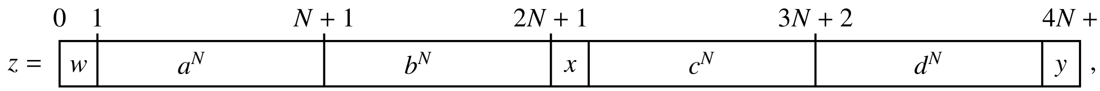
an.

1. Annahme: $L$ ist kontextfrei.

2. Nach dem Pumping-Lemma gibt es die Pumping-Length $N$.

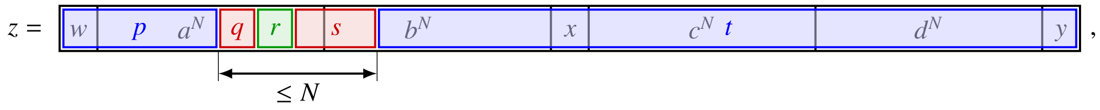
[1] Stuart M. Shieber, *Evidence against the context-freeness of natural language*, Linguistics and Philosophy, **8** (1985) 333–343

3. Wir konstruieren das Beispielwort




es ist offensichtlich in der Sprache $L$ und es ist ausreichend lang, dass die Schlussfolgerungen des Pumping-Lemma darauf anwendbar sind.

4. Gemäß dem Pumping Lemma gibt es eine Unterteilung




wobei $|qrs| \leq N$ gelten muss. Es folgt, dass $q$ und $s$ nur jeweils zwei benachbarte der vier "langen" Blöcke $a^N$, $b^N$, $c^N$ oder $d^N$ berühren können.

5. Beim Aufpumpen werden zwei benachbarte langen Blöcke verändert. Die Bedingung, dass alternierende Blöcke, also $a^N$ und $c^N$ bzw. $b^N$ und $d^N$ gleich lang sein müssen, wird nach dem Pumpen daher nicht mehr erfüllt sein. Ein aufgepumptes Wort wird daher nicht mehr in $L$ sein.

6. Dieser Widerspruch zeigt, dass die Annahme, $L$ sei kontextfrei, nicht haltbar ist. ○

*Diskussion.* Dies bedeutet natürlich nicht, dass Schweizerdeutsch eine nicht kontextfreie Sprache ist. In der Wirklichkeit sind die Sätze nämlich immer von beschränkter Länge, die Exponenten $n$ und $m$ können daher nicht beliebig groß sein. Das müssen Sie aber, denn der Pumping-Lemma-Beweis verlangt, dass man $n = m = N$ setze können muss, wobei die Pumping-Length $N$ eben sehr groß sein kann.

**8.5.** In Abschnitt 5.1.5 wurde gezeigt, dass die Sprache

$$L = \{1^n\texttt{+}1^m\texttt{=}1^{m+n} \mid n, m \geq 0\}$$

kontextfrei ist. In einer Bemerkung wurde darauf hingewiesen, dass die Sprache

$$L = \{1^n\texttt{*}1^m\texttt{=}1^{m\cdot n} \mid n, m \geq 0\}$$

nicht kontextfrei sei. Beweisen Sie dies.

*Lösung.* ○