Das Pumping-Lemma für kontextfreie Sprachen

Prof. Dr. Jasmin Blanchette

Lehr- und Forschungseinheit für
Theoretische Informatik

Stand: 23. Mai 2023
Folien ursprünglich von PD Dr. David Sabel

Widerlegen der Kontextfreiheit

Wir lernen eine Methode kennen zum Widerlegen der Kontextfreiheit:

- Pumping-Lemma für kontextfreie Sprachen

Es gibt weitere (allgemeinere) Formulierungen, z.B.

- Ogdens Lemma (benannt nach William F. Ogden)
  (ist im Skript, aber kein Prüfungsstoff)
- Interchange-Lemma

Einschub: Binärbäume (1)

Binärbaum: Baum, wobei jeder Knoten 0 oder 2 Kinder hat

Lemma

Sei B ein Binärbaum mit $\geq 2^k$ Blättern. Dann hat B einen Pfad der Länge $\geq k$.

Einschub: Binärbäume (1)

Binärbaum: Baum, wobei jeder Knoten 0 oder 2 Kinder hat

Lemma

Sei B ein Binärbaum mit $\geq 2^k$ Blättern. Dann hat B einen Pfad der Länge $\geq k$.

Beweis durch Induktion über k:

$k = 0$:

- Ein Baum mit $2^k = 2^0 = 1$ Blatt besteht genau aus diesem Blatt und hat einen Pfad der Länge $0 \geq 0$.

Einschub: Binärbäume (1)

Binärbaum: Baum, wobei jeder Knoten 0 oder 2 Kinder hat

Lemma

Sei B ein Binärbaum mit $\geq 2^k$ Blättern. Dann hat B einen Pfad der Länge $\geq k$.

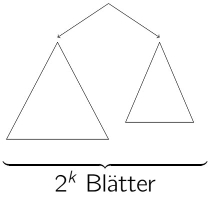
Beweis durch Induktion über k:

$k = 0$:

- Ein Baum mit $2^k = 2^0 = 1$ Blatt besteht genau aus diesem Blatt und hat einen Pfad der Länge $0 \geq 0$.

$k > 0$:

- Einer der beiden Teilbäume unter der Wurzel hat $\geq 2^{k-1}$ Blätter.
- Per Induktionsannahme hat dieser einen Pfad der Länge $\geq k - 1$.
- Daher hat der gesamte Baum einen Pfad der Länge $\geq k$. $\square$



Einschub: Binärbäume (2)

Lemma

Sei B ein Binärbaum von dem alle Pfade eine Länge $\leq k$ haben. Dann hat $B \leq 2^k$ Blätter.

Einschub: Binärbäume (2)

Lemma

Sei B ein Binärbaum von dem alle Pfade eine Länge $\leq k$ haben. Dann hat $B \leq 2^k$ Blätter.

Beweis durch Induktion über k:

$k = 0$:

- B besteht nur aus $2^0 = 1$ Blatt.

$k > 0$:

- B hat zwei Teilbäume unter der Wurzel, von denen alle Pfade eine Länge $\leq k - 1$ haben.
- Durch die Induktionshypothese haben die beiden Teilbäume jeweils $\leq 2^{k-1}$ Blätter.
- B als Ganzes hat dann $\leq 2^{k-1} + 2^{k-1} = 2^k$ Blätter. □

Pumping-Lemma für CFLs: Bemerkungen

- Erinnerung: Pumping-Lemma für reguläre Sprachen:
  Jede reguläre Sprache erfüllt die Pumping-Eigenschaft für reguläre Sprachen.

Pumping-Lemma für CFLs: Bemerkungen

- Erinnerung: Pumping-Lemma für reguläre Sprachen:
  Jede reguläre Sprache erfüllt die Pumping-Eigenschaft für reguläre Sprachen.
- Analog: Pumping-Lemma für kontextfreie Sprachen:
  Jede kontextfreie Sprache erfüllt die Pumping-Eigenschaft für kontextfreie Sprachen.

Pumping-Lemma für CFLs: Bemerkungen

- Erinnerung: Pumping-Lemma für reguläre Sprachen:
  Jede reguläre Sprache erfüllt die Pumping-Eigenschaft für reguläre Sprachen.
- Analog: Pumping-Lemma für kontextfreie Sprachen:
  Jede kontextfreie Sprache erfüllt die Pumping-Eigenschaft für kontextfreie Sprachen.
- Kann vor allem zum Widerlegen benutzt werden:
  Sprache **verletzt** die Pumping-Eigenschaft für CFLs.
  $\implies$ Sprache ist **nicht kontextfrei**.

Pumping-Lemma für CFLs: Bemerkungen

- Erinnerung: Pumping-Lemma für reguläre Sprachen:
  Jede reguläre Sprache erfüllt die Pumping-Eigenschaft für reguläre Sprachen.
- Analog: Pumping-Lemma für kontextfreie Sprachen:
  Jede kontextfreie Sprache erfüllt die Pumping-Eigenschaft für kontextfreie Sprachen.
- Kann vor allem zum Widerlegen benutzt werden:
  Sprache **verletzt** die Pumping-Eigenschaft für CFLs.
  $\Longrightarrow$ Sprache ist **nicht kontextfrei**.
- Pumping-Eigenschaft bei regulären Sprachen, informell:
  Man kann Wörter an einer Stelle aufpumpen und in der Sprache verbleiben ($uv^iw \in L$ für alle $i \in \mathbb{N}$).

Pumping-Lemma für CFLs: Bemerkungen

- Erinnerung: Pumping-Lemma für reguläre Sprachen:
  Jede reguläre Sprache erfüllt die Pumping-Eigenschaft für reguläre Sprachen.
- Analog: Pumping-Lemma für kontextfreie Sprachen:
  Jede kontextfreie Sprache erfüllt die Pumping-Eigenschaft für kontextfreie Sprachen.
- Kann vor allem zum Widerlegen benutzt werden:
  Sprache **verletzt** die Pumping-Eigenschaft für CFLs.
  $\Longrightarrow$ Sprache ist **nicht kontextfrei**.
- Pumping-Eigenschaft bei regulären Sprachen, informell:
  Man kann Wörter an einer Stelle aufpumpen und in der Sprache verbleiben ($uv^iw \in L$ für alle $i \in \mathbb{N}$).
- Pumping-Eigenschaft bei kontextfreien Sprachen, informell:
  Man kann Wörter an ***zwei*** *Stellen gleichzeitig aufpumpen und in der Sprache verbleiben ($uv^iwx^iy \in L$ für alle $i \in \mathbb{N}$).*

Das Pumping-Lemma für CFLs

Lemma (Pumping-Lemma für CFLs)

Sei L eine kontextfreie Sprache. Dann gibt es eine Zahl $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$, das Mindestlänge n hat (d.h. $|z| \geq n$), als $z = uvwxy$ geschrieben werden kann, sodass gilt:

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$.

Beweis des Pumping-Lemmas (1)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- Sei $G = (V, \Sigma, P, S)$ eine CFG in Chomsky-Normalform mit $L(G) = L \setminus \{\varepsilon\}$.

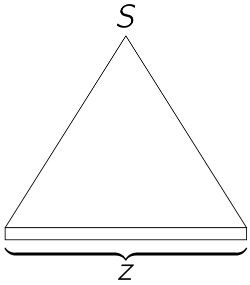
Beweis des Pumping-Lemmas (1)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- Sei $G = (V, \Sigma, P, S)$ eine CFG in Chomsky-Normalform mit $L(G) = L \setminus \{\varepsilon\}$.
- Sei $n = 2^{|V|}$.



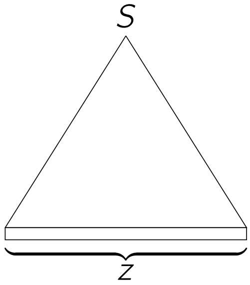
Beweis des Pumping-Lemmas (1)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- Sei $G = (V, \Sigma, P, S)$ eine CFG in Chomsky-Normalform mit $L(G) = L \setminus \{\varepsilon\}$.
- Sei $n = 2^{|V|}$.
- Betrachte Syntaxbaum eines Wortes z mit $|z| \geq 2^{|V|} = n$.



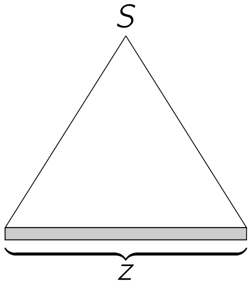
Beweis des Pumping-Lemmas (1)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- Sei $G = (V, \Sigma, P, S)$ eine CFG in Chomsky-Normalform mit $L(G) = L \setminus \{\varepsilon\}$.
- Sei $n = 2^{|V|}$.
- Betrachte Syntaxbaum eines Wortes z mit $|z| \geq 2^{|V|} = n$.
- Da G in Chomsky-Normalform, ist der Syntaxbaum binär, bis auf die letzte Schicht, die Produktionen $A \to a$ anwendet.



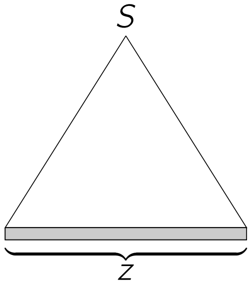
Beweis des Pumping-Lemmas (1)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- Sei $G = (V, \Sigma, P, S)$ eine CFG in Chomsky-Normalform mit $L(G) = L \setminus \{\varepsilon\}$.
- Sei $n = 2^{|V|}$.
- Betrachte Syntaxbaum eines Wortes z mit $|z| \geq 2^{|V|} = n$.
- Da G in Chomsky-Normalform, ist der Syntaxbaum binär, bis auf die letzte Schicht, die Produktionen $A \to a$ anwendet.
- Baum ohne letzte Schicht hat $|z| \geq 2^{|V|}$ Blätter.



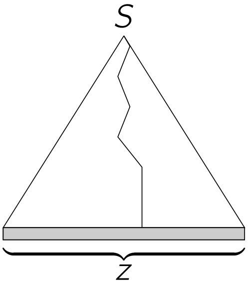
Beweis des Pumping-Lemmas (1)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- Sei $G = (V, \Sigma, P, S)$ eine CFG in Chomsky-Normalform mit $L(G) = L \setminus \{\varepsilon\}$.
- Sei $n = 2^{|V|}$.
- Betrachte Syntaxbaum eines Wortes z mit $|z| \geq 2^{|V|} = n$.
- Da G in Chomsky-Normalform, ist der Syntaxbaum binär, bis auf die letzte Schicht, die Produktionen $A \to a$ anwendet.
- Baum ohne letzte Schicht hat $|z| \geq 2^{|V|}$ Blätter.
- Daher hat der längste Pfad von der Wurzel zum Blatt eine Länge $\geq |V|$. Dieser besteht aus $\geq |V| + 1$ Knoten und jeder Knoten ist mit einer Variablen markiert.



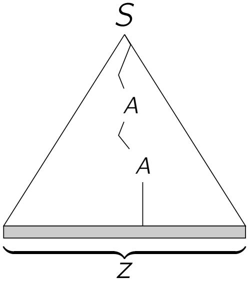
Beweis des Pumping-Lemmas (2)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- ...
- Da es nur $|V|$ Variablen gibt, kommt mindestens eine Variable mehrfach auf diesem Pfad vor.
- Wähle die Vorkommen der Variablen so, dass das obere Vorkommen am tiefsten ist. Sei A die Variable.



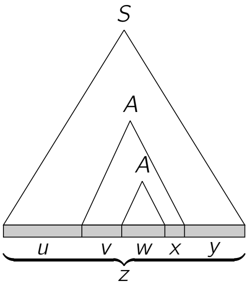
Beweis des Pumping-Lemmas (3)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- ...
- Betrachte die Teilbäume, die jeweils A als Wurzel haben.
- Sie entsprechen Ableitungen von Teilworten von z.
- Der Teilbaum mit dem unteren A als Wurzel erzeugt ein Teilwort des Teilbaums mit dem oberen A als Wurzel. D.h. $z = uvwxy$, wobei vwx vom oberen A und w vom unteren A erzeugt wird.
- Es gilt $|w| \geq 1$, da Variablen einer Grammatik in Chomsky-Normalform nur Wörter mit Länge ≥ 1 ableiten.
- vwx muss echt länger sein als w, da das obere A über dem unteren A steht. Daher folgt $|v| \geq 1$ oder $|x| \geq 1$ (d.h. $|vx| \geq 1$).



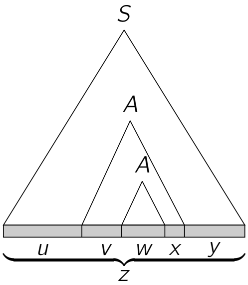
Beweis des Pumping-Lemmas (4)

Behauptung: Für jede CFL L gibt es $n \in \mathbb{N}_{>0}$, sodass jedes Wort $z \in L$ mit $|z| \geq n$ als $z = uvwxy$ geschrieben werden kann mit

- $|vx| \geq 1$
- $|vwx| \leq n$
- für alle $i \geq 0$: $uv^iwx^iy \in L$

Beweis:

- ...
- Da wir das tiefste Vorkommen der wiederholten Variable gewählt haben, kann der Pfad vom oberen A bis zur Blattebene nur aus $\leq |V| + 1$ Knoten bestehen und Länge $\leq |V|$ haben.
- Da der Pfad von der Wurzel maximaler Länge ist, müssen andere Pfade vom oberen A bis zur Blattebene kürzer oder gleich lang sein.
- Daraus folgt: $|vwx| \leq 2^{|V|} = n$.
- Aus dem Baum folgt: $A \Rightarrow^* w$ und $A \Rightarrow^* vAx$ und daher kann man auch $A \Rightarrow^* v^iwx^i$ für alle $i \in \mathbb{N}$ ableiten.
- Schließlich folgt daraus $S \Rightarrow^* uv^iwx^iy$ für alle $i \in \mathbb{N}$. □



Pumping-Lemma: Illustrationen

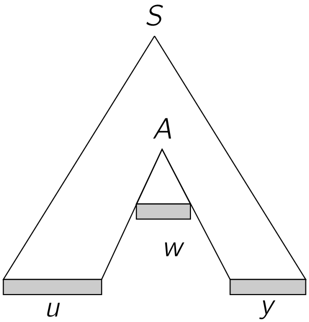


Illustration für uv^0wx^0y

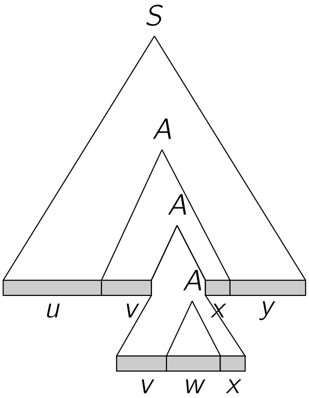


Illustration für uv^2wx^2y

Verwendung des Pumping-Lemma

- Die Pumping-Eigenschaft ist eine **notwendige** aber **keine hinreichende** Bedingung für CFLs.
- Daher kann das Pumping-Lemma **nicht** verwendet werden, um Kontextfreiheit zu zeigen.
- Aber es kann verwendet werden, um **Kontextfreiheit zu widerlegen**.

Pumping-Lemma zum Widerlegen der Kontextfreiheit

Formulierung des Pumping-Lemmas für CFGs zum Widerlegen der Kontextfreiheit

Sei L eine formale Sprache für die gilt:

Für jede Zahl $n \in \mathbb{N}_{>0}$ gibt es ein Wort $z \in L$, das Mindestlänge n hat (d.h. $|z| \geq n$), und für jede Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$, gibt es ein $i \geq 0$, sodass $uv^iwx^iy \notin L$.

Dann ist L nicht kontextfrei.

Beweis:

Umformung der negierten prädikatenlogischen Formel (siehe Skript), die sich aus dem Pumping-Lemma ergibt.

Pumping-Lemma als Spiel

Sei L die formale Sprache.

1. Der **Gegner** wählt die Zahl $n \in \mathbb{N}_{>0}$.
2. **Wir** wählen das Wort $z \in L$ mit $|z| \geq n$.
3. Der **Gegner** wählt die Zerlegung
   $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.
4. **Wir** gewinnen das Spiel, wenn wir ein $i \geq 0$ angeben können, sodass
   $uv^i wx^i y \notin L$.

Wenn wir **für jede Wahl des Gegners** das Spiel gewinnen können,
dann haben wir gezeigt, dass L nicht kontextfrei ist.

Beispiele

Satz

Die Sprache $L = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ ist nicht kontextfrei.

Beweis:

- Gegner wählt $n \in \mathbb{N}_{>0}$.

Beispiele

Satz

Die Sprache $L = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ ist nicht kontextfrei.

Beweis:

- Gegner wählt $n \in \mathbb{N}_{>0}$.
- Wir wählen $z = a^n b^n c^n$.

Beispiele

Satz

Die Sprache $L = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ ist nicht kontextfrei.

Beweis:

- Gegner wählt $n \in \mathbb{N}_{>0}$.
- Wir wählen $z = a^n b^n c^n$.
- Gegner wählt Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.

Beispiele

Satz

Die Sprache $L = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ ist nicht kontextfrei.

Beweis:

- Gegner wählt $n \in \mathbb{N}_{>0}$.
- Wir wählen $z = a^n b^n c^n$.
- Gegner wählt Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.
- Fall 1: vwx ist von der Form $a^i b^j$, $i + j \leq n$.
  Da $|vx| \geq 1$, gilt $\#_a(vx) \geq 1$ oder $\#_b(vx) \geq 1$, aber $\#_c(vx) = 0$
  Damit folgt $uv^0wx^0y \notin L$.

Beispiele

Satz

Die Sprache $L = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ ist nicht kontextfrei.

Beweis:

- Gegner wählt $n \in \mathbb{N}_{>0}$.
- Wir wählen $z = a^n b^n c^n$.
- Gegner wählt Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.
- Fall 1: vwx ist von der Form $a^i b^j$, $i + j \leq n$.
  Da $|vx| \geq 1$, gilt $\#_a(vx) \geq 1$ oder $\#_b(vx) \geq 1$, aber $\#_c(vx) = 0$
  Damit folgt $uv^0wx^0y \notin L$.
- Fall 2: vwx ist von der Form $b^i c^j$, $i + j \leq n$
  Da $|vx| \geq 1$, gilt $\#_b(vx) \geq 1$ oder $\#_c(vx) \geq 1$, aber $\#_a(vx) = 0$
  Damit folgt $uv^0wx^0y \notin L$.

Beispiele

Satz

Die Sprache $L = \{a^i b^i c^i \mid i \in \mathbb{N}\}$ ist nicht kontextfrei.

Beweis:

- Gegner wählt $n \in \mathbb{N}_{>0}$.
- Wir wählen $z = a^n b^n c^n$.
- Gegner wählt Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.
- Fall 1: vwx ist von der Form $a^i b^j$, $i + j \leq n$.
  Da $|vx| \geq 1$, gilt $\#_a(vx) \geq 1$ oder $\#_b(vx) \geq 1$, aber $\#_c(vx) = 0$
  Damit folgt $uv^0wx^0y \notin L$.
- Fall 2: vwx ist von der Form $b^i c^j$, $i + j \leq n$
  Da $|vx| \geq 1$, gilt $\#_b(vx) \geq 1$ oder $\#_c(vx) \geq 1$, aber $\#_a(vx) = 0$
  Damit folgt $uv^0wx^0y \notin L$.
- Andere Fälle sind nicht möglich. $\square$

Beispiele (2)

Satz

Die Sprache $L = \{a^i b^j c^i d^j \mid i, j \in \mathbb{N}_{>0}\}$ ist nicht kontextfrei.

Beweis:

- Gegner wählt $n \in \mathbb{N}_{>0}$.
- Wir wählen $z = a^n b^n c^n d^n$.
- Gegner wählt Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$.
- 1. Fall: $vwx = a^i b^j$ mit $i + j \leq n$. Da $|vx| \geq 1$, gilt $\#_a(vx) + \#_b(vx) \geq 1$ und $uv^0wx^0y = uwy = a^{i'} b^{j'} c^n d^n$ und $i' < n$ oder $j' < n$, d.h. $uwy \notin L$.
- 2. Fall: $vwx = b^i c^j$ mit $i + j \leq n$. Da $|vx| \geq 1$, gilt $\#_b(vx) + \#_c(vx) \geq 1$ und $uv^0wx^0y = uwy = a^n b^{i'} c^{j'} d^n$ und $i' < n$ oder $j' < n$, d.h. $uwy \notin L$.
- 3. Fall: $vwx = c^i d^j$ mit $i + j \leq n$. Da $|vx| \geq 1$, gilt $\#_c(vx) + \#_d(vx) \geq 1$ und $uv^0wx^0y = uwy = a^n b^n c^{i'} d^{j'}$ und $i' < n$ oder $j' < n$, d.h. $uwy \notin L$. $\square$

Unäres Alphabet

Satz

Sei L eine formale Sprache über einem unären Alphabet (d.h. $|\Sigma| = 1$). Dann ist L genau dann regulär, wenn L kontextfrei ist.

Beweis:

- Wenn L regulär ist, dann ist L auch kontextfrei.
- Rückrichtung: ist im Skript, aber kein Prüfungsstoff.
  (Der Beweis verwendet die Pumping-Eigenschaft für CFLs.)

Beispiele

Satz

Die Sprachen

$$\begin{aligned} L_1 &= \{a^p \mid p \text{ ist eine Primzahl}\} \\ L_2 &= \{a^n \mid n \text{ ist keine Primzahl}\} \\ L_3 &= \{a^n \mid n \text{ ist eine Quadratzahl}\} \\ L_4 &= \{a^{2^n} \mid n \in \mathbb{N}\} \end{aligned}$$

sind allesamt nicht kontextfrei.

Beweis: Wir haben für alle vier Sprachen gezeigt, dass sie nicht regulär sind. Da sie alle über einem unären Alphabet definiert sind, sind sie auch nicht kontextfrei. $\square$