

Theoretische Grundlagen der Informatik

Vorlesung am 11. Januar 2018



Das Pumping-Lemma für kontextfreie Sprachen



Pumping-Lemma für kontextfreie Sprachen

Für jede kontextfreie Sprache L gibt es eine Konstante $n \in \mathbb{N}$, so dass sich jedes Wort $z \in L$ mit $|z| \ge n$

so als

z = uvwxy

schreiben lässt, dass

- $|vx| \geq 1$,
- $|vwx| \le n$ und
- für alle $i \ge 0$ das Wort $uv^i wx^i y \in L$ ist.

Ogden's Lemma für kontextfreie Sprachen



Ogden's Lemma für kontextfreie Sprachen

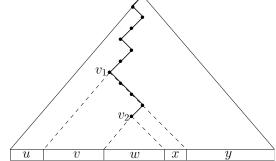
Für jede kontextfreie Sprache L gibt es eine Konstante $n \in \mathbb{N}$, so dass für jedes Wort $z \in L$ mit $|z| \ge n$ gilt:

Wenn wir in z mindestens n Buchstaben markieren, so lässt sich z so als z = uvwxy schreiben,

- dass von den mindestens n markierten Buchstaben
 - mindestens einer zu vx gehört und
 - höchstens n zu vwx gehören und
- für alle $i \ge 0$ das Wort $uv^i wx^i y \in L$ ist.

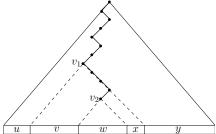


- Sei L kontextfreie Sprache
- Sei *G* Grammatik zu *L* mit Variablen *V* in Chomsky-Normalform, d.h. alle Regeln sind von der Form $A \rightarrow BC$ oder $A \rightarrow a$.
- Setze $n := 2^{|V|+1}$.
- Wähle beliebiges Wort $z \in L$ mit $|z| \ge n$
- Betrachte einen Syntaxbaum *T* zu *z*.



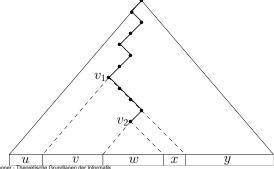


- T hat |z| Blätter und alle inneren Knoten außer den Vorgängern der Blätter haben Grad 2, ansonsten Grad 1.
- Seien mindestens n Blätter markiert.
- Durchlaufe einen Weg von der Wurzel zu einem Blatt.
 Wähle stets den Nachfolger, auf dessen Seite die größere Anzahl markierter Blätter liegt.
- Nenne Knoten auf dem Weg, für die rechter und linker Unterbaum markierte Blätter hat, Verzweigungsknoten.





- Wegen $n > 2^{|V|}$ liegen auf dem Weg mindestens |V| + 1Verzweigungsknoten
- Von den letzten |V| + 1 Verzweigungsknoten entsprechen mindestens zwei Knoten v_1, v_2 derselben Variablen A.
- Sei vwx Wort unter Teilbaum mit Wurzel v₁
- Sei w Wort unter Teilbaum mit Wurzel v₂.
- Damit sind *u* und *y* eindeutig bestimmt.





- Da v₁ Verzweigungsknoten ist, enthält vx mindestens einen markierten Buchstaben.
- Da der Unterbaum von v_1 inkl. v_1 nur |V| + 1 Verzweigungsknoten enthält, gibt es in vwx höchstens $2^{|V|+1} = n$ markierte Buchstaben.
- Zu G existieren die Ableitungen

$$S \stackrel{*}{\to} uAy$$
, $A \stackrel{*}{\to} vAx$, $A \stackrel{*}{\to} w$.

Daraus kann z abgeleitet werden durch

$$S \stackrel{*}{\rightarrow} uAy \stackrel{*}{\rightarrow} uvAxy \stackrel{*}{\rightarrow} uvwxy = z$$
,

aber auch uv^iwx^iy für jedes $i \ge 1$ durch

$$S \stackrel{*}{ o} uAy \stackrel{*}{ o} uvAxy \stackrel{*}{ o} uv^2Ax^2y \stackrel{*}{ o} \cdots o uv^iAx^iy o uv^iwx^iy.$$

Also ist auch $uv^iwx^iy \in L$ für $i \ge 0$.

Bemerkung



 Der Spezialfall von Odgen's Lemma, in dem alle Buchstaben von z markiert sind, ist gerade das Pumping-Lemma.



Die Chomsky-Hierarchie ist echt, d.h.

$$\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0$$
 ,

wobei \mathcal{L}_i , $0 \le i \le 3$, Klasse der durch Typ-i-Grammatiken erzeugten Sprachen.



Es gibt eine kontextfreie Sprache, die nicht regulär ist.

Die Sprache

$$L = \{a^i b^i | i \ge 1\}$$

ist kontextfrei und wird durch die Grammatik

$$V = \{S\}$$

 $\Sigma = \{0, 1\}$
 $R = \{S \rightarrow 01 \mid 0S1\}$.

erzeugt. Sie ist aber nicht regulär.

(Siehe auch Beispiele zum Pumping-Lemma für reguläre Sprachen)



Es gibt eine kontextsensitive Sprache, die nicht kontextfrei ist.

Die Sprache

$$L = \{a^i b^i c^i | i \ge 1\}$$

ist kontextsensitiv.

- L kontextsensitiv ⇔ es gibt NTM mit linearem Speicherbedarf für L
- Eingabe $w \in \{a, b, c\}^*$
- Überprüfe deterministisch, ob $w = a^i b^j c^k$
- Überprüfe deterministisch, ob j = i und k = i
- Speicherbedarf: i + j + k, also linear
- ⇒ L kontextsensitiv



Es gibt eine kontextsensitive Sprache, die nicht kontextfrei ist.

Die Sprache

$$L = \{a^i b^i c^i | i \ge 1\}$$

ist nicht kontextfrei.



Pumping-Lemma für kontextfreie Sprachen

Für jede kontextfreie Sprache L gibt es eine Konstante $n \in \mathbb{N}$, so dass sich jedes Wort $z \in L$ mit $|z| \ge n$

so als

z = uvwxy

schreiben lässt, dass

- $|vx| \geq 1,$
- $|vwx| \le n$ und
- für alle $i \ge 0$ das Wort $uv^i wx^i y \in L$ ist.



Die Sprache $L = \{a^i b^i c^i | i \ge 1\}$ ist nicht kontextfrei.

- Annahme: L sei kontextfrei. Sei dann n wie im PL gefordert.
- Wähle das Wort $z = a^n b^n c^n \in L$.
- Wir betrachten eine Zerlegung z = uvwxy wie im PL gefordert:
 - $|vx| \geq 1$,
 - $|vwx| \le n$ und
 - für alle $i \ge 0$ ist das Wort $uv^i wx^i y \in L$.



Die Sprache $L = \{a^i b^i c^i | i \ge 1\}$ ist nicht kontextfrei.

- Annahme: L sei kontextfrei. Sei dann n wie im PL gefordert.
- Wähle das Wort $z = a^n b^n c^n \in L$.
- Wir betrachten eine Zerlegung z = uvwxy wie im PL gefordert:
 - $|vx| \geq 1$,
 - $|vwx| \le n$ und
 - für alle $i \ge 0$ ist das Wort $uv^i wx^i y \in L$.
- Fallunterscheidung, Fall 1: vwx besteht nur aus a und b
 - Dann enthält vx mindestens ein a oder b.
 - Damit ist $uv^0wx^0y = a^ib^jc^n \notin L$ weil entweder i < n oder j < n.
 - Dies ist ein Widerspruch zum PL.



Die Sprache $L = \{a^i b^i c^i | i \ge 1\}$ ist nicht kontextfrei.

- Annahme: L sei kontextfrei. Sei dann n wie im PL gefordert.
- Wähle das Wort $z = a^n b^n c^n \in L$.
- Wir betrachten eine Zerlegung z = uvwxy wie im PL gefordert:
 - $|vx| \geq 1$,
 - $|vwx| \le n$ und
 - für alle $i \ge 0$ ist das Wort $uv^i wx^i y \in L$.
- Fallunterscheidung, Fall 2: vwx besteht nur aus b und c
 - Dann enthält vx mindestens ein b oder c.
 - Damit ist $uv^0wx^0y = a^nb^jc^j \notin L$ weil entweder i < n oder j < n.
 - Dies ist ein Widerspruch zum PL.



Die Sprache $L = \{a^i b^i c^i | i \ge 1\}$ ist nicht kontextfrei.

Ogden's Lemma für kontextfreie Sprachen

Für jede kontextfreie Sprache L gibt es eine Konstante $n \in \mathbb{N}$, so dass für jedes Wort $z \in L$ mit $|z| \ge n$ gilt:

Wenn wir in z mindestens n Buchstaben markieren, so lässt sich z so als z = uvwxy schreiben,

- dass von den mindestens n markierten Buchstaben
 - mindestens einer zu vx gehört und
 - höchstens *n* zu *vwx* gehören und
- für alle $i \ge 0$ das Wort $uv^i wx^i y \in L$ ist.



Die Sprache $L = \{a^i b^i c^i | i \ge 1\}$ ist nicht kontextfrei.

Alternativer Beweis mit Odgen's Lemma

- Annahme: L sei kontextfrei.
- Sei dann n wie in Odgen's Lemma gefordert.
- Wähle das Wort $z = a^{n+1}b^{n+1}c^{n+1} \in L$.
- Markiere alle b.
- Damit enthält vwx mindestens ein b aber kein a oder kein c.
- Es enthalte vwx kein c (anderer Fall analog)
- Damit ist $uv^0wx^0y = a^ib^jc^n \notin L$ weil entweder i < n oder j < n.
- Dies ist ein Widerspruch zu Odgen's Lemma.



Es gibt eine semi-entscheidbare Sprache, die nicht kontextsensitiv ist.

Es sei L_u die universelle Sprache.

Wiederholung

Die **universelle Sprache** L_u über $\{0, 1\}$ ist definiert durch

$$L_{u}:=\left\{ wv\mid v\in L(T_{w})\right\} .$$

 L_u ist also die Menge aller Wörter wv für die T_w bei der Eingabe v hält und v akzeptiert.



Es gibt eine semi-entscheidbare Sprache, die nicht kontextsensitiv ist.

Es sei L_u die universelle Sprache.

- lacksquare L_u ist semi-entscheidbar, aber nicht entscheidbar (Kapitel 3).
- Wegen der Semi-entscheidbarkeit gilt $L_u \in \mathcal{L}_0$.
- Annahme: $L_u \in \mathcal{L}_1$.
- Dann gibt es eine NTM, die Lu mit linearem Speicher erkennt.
- Mit linearem Speicher können nur exponentiell viele verschiedene Konfigurationen auftreten.
- Diese könnte durch eine DTM durch Ausprobieren aller möglichen Konfigurationen simuliert werden.
- Dies wäre ein Widerspruch zur Nichtentscheidbarkeit von L_u .

Nutzlose Variablen



Sei G eine kontextfreie Grammatik. Eine Variable A heißt **nutzlos**, falls es keine Ableitung $S \stackrel{*}{\to} w$ gibt, $w \in \Sigma^*$, in der A vorkommt.

Satz:

Für eine kontextfreie Grammatik kann die Menge der nutzlosen Variablen (in polynomialer Zeit) berechnet werden.

Beweis:

Wir benutzen ein zweistufiges Verfahren.

Schritt 1



Bestimme alle Variablen, die ein Wort erzeugen können

Formal: Berechne $V' = \{A \in V \mid \exists w \in \Sigma^* : A \stackrel{*}{\rightarrow} w\}$

- Initialisiere eine leere Queue Q.
- Füge alle $A \in V$ mit $A \to w$ für ein $w \in \Sigma^*$ in Q und V' ein.
- Entferne der Reihe nach jedes Element A aus Q
 - Ersetze jede Regel $B \to \alpha A \beta$ mit $\alpha, \beta \in (V \cup \Sigma)^*$ durch die Regeln $B \to \alpha w \beta$, wobei $w \in \Sigma^*$ und $A \to w$ Regel.
 - Wenn dabei eine Regel der Form $B \rightarrow w', \ w' \in \Sigma^*,$ entsteht und $B \notin V'$, füge B in Q und V' ein.
- Das Verfahren endet, wenn Q leer ist.

Schritt 1



Bemerkung 1

- Falls $S \notin V'$, breche das Verfahren ab.
- G erzeugt dann die leere Sprache und alle Variablen sind nutzlos.

Bemerkung 2

- Für jede Variable A mit $A \stackrel{*}{\rightarrow} w$ für ein $w \in \Sigma^*$ gilt:
- Per Induktion über die Länge der kürzesten Ableitungsregel der Form $A \stackrel{*}{\to} w$ kann für A gezeigt werden, dass $A \in V'$.



Grammatik $G = (\Sigma, V, S, R)$ mit Produktionen R gegeben durch

 $S \rightarrow Aa|B|Cab$

 $A \rightarrow bc|A$

 $B \rightarrow Bd|Cd$

 $C \rightarrow aBc$

 $D \rightarrow Ab$

 $E \rightarrow SD$



Füge alle $A \in V$ mit $A \to w$ für ein $w \in \Sigma^*$ in Q und V' ein.

$$egin{array}{lll} S &
ightarrow & Aa|B|Cab \ A &
ightarrow & bc|A \ B &
ightarrow & Bd|Cd \ C &
ightarrow & aBc \ D &
ightarrow & Ab \ E &
ightarrow & SD \ \end{array}$$

$$S
ightarrow Aa|B|Cab$$

 $A
ightarrow bc|A$
 $B
ightarrow Bd|Cd$
 $C
ightarrow aBc$
 $D
ightarrow Ab$
 $E
ightarrow SD$

 $V' = \{A\}$ $Q = \{A\}$

11.01.2018



- Ersetze jede Regel $B \to \alpha A\beta$ mit $\alpha, \beta \in (V \cup \Sigma)^*$ durch die Regeln $B \to \alpha w \beta$, wobei $w \in \Sigma^*$ und $A \to w$ Regel.
- Wenn dabei eine Regel der Form $B \to w'$, $w' \in \Sigma^*$ entsteht und $B \notin V'$, füge B in Q und V' ein.



- Ersetze jede Regel $B \to \alpha A \beta$ mit $\alpha, \beta \in (V \cup \Sigma)^*$ durch die Regeln $B \to \alpha w \beta$, wobei $w \in \Sigma^*$ und $A \to w$ Regel.
- Wenn dabei eine Regel der Form $B \to w'$, $w' \in \Sigma^*$ entsteht und $B \notin V'$, füge B in Q und V' ein.

$$S
ightharpoonup bca|B|Cab$$
 $S
ightharpoonup bca|B|Cab$ $A
ightharpoonup bc|A$ $A
ightharpoonup bc|A$ $B
ightharpoonup Bd|Cd$ $B
ightharpoonup Bd|Cd$ $C
ightharpoonup aBc$ $C
ightharpoonup aBc$ $D
ightharpoonup bcb$ $E
ightharpoonup bcb$ $E
ightharpoonup bca|Cd$ $C
ightharpoonup aBc$ $D
ightharpoonup bcb$ $D
ightharpoonup bca|Cd$ $D
ighth$



- Ersetze jede Regel $B \to \alpha A \beta$ mit $\alpha, \beta \in (V \cup \Sigma)^*$ durch die Regeln $B \to \alpha w \beta$, wobei $w \in \Sigma^*$ und $A \to w$ Regel.
- Wenn dabei eine Regel der Form $B \to w'$, $w' \in \Sigma^*$ entsteht und $B \notin V'$, füge B in Q und V' ein.

$$S
ightharpoonup bca|B|Cab$$
 $S
ightharpoonup bca|B|Cab$ $A
ightharpoonup bc|A$ $A
ightharpoonup bc|A$ $B
ightharpoonup Bd|Cd$ $B
ightharpoonup Bd|Cd$ $C
ightharpoonup aBc$ $C
ightharpoonup aBc$ $D
ightharpoonup bcb$ $E
ightharpoonup bcabcb$

$$V' = \{A, S, D\}$$
 $V' = \{A, S, D, E\}$ $Q = \{D\}$ $Q = \{E\}$



- Ersetze jede Regel $B \to \alpha A \beta$ mit $\alpha, \beta \in (V \cup \Sigma)^*$ durch die Regeln $B \to \alpha w \beta$, wobei $w \in \Sigma^*$ und $A \to w$ Regel.
- Wenn dabei eine Regel der Form $B \to w'$, $w' \in \Sigma^*$ entsteht und $B \notin V'$, füge B in Q und V' ein.

$$S
ightarrow bca|B|Cab$$
 $S
ightarrow bca|B|Cab$ $A
ightarrow bc|A$ $A
ightarrow bc|A$ $B
ightarrow Bd|Cd$ $B
ightarrow Bd|Cd$ $C
ightarrow aBc$ $C
ightarrow aBc$ $D
ightarrow bcb$ $D
ightarrow bcb$ $E
ightarrow bcabcb$

$$V' = \{A, S, D, E\}$$
 $V' = \{A, S, D, E\}$ $Q = \{E\}$ $Q = \{B\}$

Schritt 2



Bestimme alle Variablen in V', die vom Startsymbol aus "erreicht" werden können.

Formal: Berechne $\{A \in V' \mid S = A \text{ oder } \exists \alpha, \beta \in (V' \cup \Sigma)^* : S \stackrel{*}{\to} \alpha A \beta \}$

- Starte mit $V'' = \{S\}$
- Füge zu allen Regeln $A \to \alpha B\beta$ mit $\alpha, \beta \in (V' \cup \Sigma)^*, A \in V'', B \in V'$ die Variable B in V'' ein.
- Wiederhole den letzen Schritt, bis sich V" nicht mehr ändert.

Per Induktion über die Länge der kürzesten Ableitungsregel der Form $S \to \alpha A \beta$, $\alpha, \beta \in (V' \cup \Sigma)^*$, kann dann wieder die Korrektheit bewiesen werden.

Fazit: Nach Ende von Schritt 2 ist V'' die Menge aller nützlichen Variablen.

11.01.2018



Starte mit
$$V'' = \{S\}$$

$$S \rightarrow Aa|B|Cab$$

$$A \rightarrow bc|A$$

$$B \rightarrow Bd|Cd$$

$$C \rightarrow aBc$$

$$D \rightarrow Ab$$

$$E \rightarrow SD$$

$$V' = \{A, S, D, E\}$$

$$V'' = \{\}$$

$$S \rightarrow Aa|B|Cab$$

$$A \rightarrow bc|A$$

$$B \rightarrow Bd|Cd$$

$$D \rightarrow Ab$$

$$E \rightarrow SD$$

$$V' = \{A, S, D, E\}$$

$$V'' = {S \choose S}$$



Füge zu allen Regeln $A \to \alpha B\beta$ mit $\alpha, \beta \in (V' \cup \Sigma)^*, A \in V'', B \in V'$ die Variable B in V'' ein.

$$S
ightarrow Aa|B|Cab$$

 $A
ightarrow bc|A$
 $B
ightarrow Bd|Cd$
 $C
ightarrow aBc$

$$C \rightarrow aBC$$
 $D \rightarrow Ab$
 $E \rightarrow SD$

$$V' = \{A, S, D, E\}$$

 $V'' = \{S\}$

$$S \rightarrow Aa|B|Cab$$

$$A \rightarrow bc|A$$

$$B \rightarrow Bd|Cd$$

$$C \rightarrow aBc$$

$$D \rightarrow Ab$$

 $E \rightarrow SD$

$$V' = \{A, S, D, E\}$$

$$V'' = \{S, A\}$$

$$V'' = \{S, A\}$$



Wiederhole den letzen Schritt, bis sich V'' nicht mehr ändert.

$$S \rightarrow Aa|B|Cab$$

$$A \rightarrow bc|A$$

$$B \rightarrow Bd|Cd$$

$$C \rightarrow aBc$$

$$D \rightarrow Ab$$

$$E \rightarrow SD$$

$$V' = \{A, S, D, E\}$$

$$V'' = \{S, A\}$$

$$V'' = \{S, A\}$$

$$S \rightarrow Aa|B|Cab$$

$$A \rightarrow bc|A$$

$$B \rightarrow Bd|Cd$$

$$C \rightarrow aBc$$

 $D \rightarrow Ab$

$$E \rightarrow SD$$

$$V' = \{A, S, D, E\}$$

$$V'' = \{S, A\}$$

$$V'' = \{S, A\}$$

Korollar



Korollar

Für eine kontextfreie Grammatik G kann (in polynomialer Zeit) entschieden werden, ob $L(G)=\emptyset$ ist.

Beweis:

■ $L(G) = \emptyset$ genau dann, wenn S nutzlos.



Für eine kontextfreie Grammatik $G = (\Sigma, V, S, R)$ kann (in polynomialer Zeit) entschieden werden, ob L(G) endlich ist.

- Entferne alle nutzlosen Variablen
- Überführe G in eine äquivalente Grammatik in Chomsky-Normalform.
- Betrachte den gerichteten Graphen (V, E) mit
 - Knotenmenge V ist gleich der Variablenmenge von G
 - Kantenmenge $E = \{(A, B) \mid \exists C \in V : A \rightarrow BC \in R \lor A \rightarrow CB \in R\}$
- Mit Tiefensuche kann entschieden werden, ob dieser Graph einen Kreis enthält.
- Man kann sich leicht überlegen, dass L(G) genau dann endlich ist, wenn der entsprechende Graph keinen Kreis enthält.

Beispielgraph



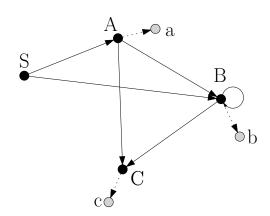
$$S \rightarrow AB$$

$$A \rightarrow BC$$

$$B \rightarrow BC$$

$$B \rightarrow b$$

$$C \rightarrow c$$





Die Klasse der kontextfreien Sprachen ist abgeschlossen bzgl. Vereinigung, Konkatenation und Kleenschem Abschluss.

Beweis:

- Seien L_1 kontextfreie Sprache mit Grammatik $G_1 = (\Sigma, V_1, S_1, R_1)$
- Seien L_2 kontextfreie Sprache mit Grammatik $G_2 = (\Sigma, V_2, S_2, R_2)$
- o.B.d.A. sei $V_1 \cap V_2 = \emptyset$.

Vereinigung: Die Grammatik

$$V = V_1 \cup V_2 \cup \{S\}$$

 S neues Startsymbol
 $R = R_1 \cup R_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}$

erzeugt $L_1 \cup L_2$.



Die Klasse der kontextfreien Sprachen ist abgeschlossen bzgl. Vereinigung, Konkatenation und Kleenschem Abschluss.

Beweis:

- Seien L_1 kontextfreie Sprache mit Grammatik $G_1 = (\Sigma, V_1, S_1, R_1)$
- Seien L_2 kontextfreie Sprache mit Grammatik $G_2 = (\Sigma, V_2, S_2, R_2)$
- o.B.d.A. sei $V_1 \cap V_2 = \emptyset$.

Konkatenation: Die Grammatik

$$V = V_1 \cup V_2 \cup \{S\}$$

$$R = R_1 \cup R_2 \cup \{S \rightarrow S_1 S_2\}$$

erzeugt $L_1 \cdot L_2$.



Die Klasse der kontextfreien Sprachen ist abgeschlossen bzgl. Vereinigung, Konkatenation und Kleenschem Abschluss.

Beweis:

- Seien L_1 kontextfreie Sprache mit Grammatik $G_1 = (\Sigma, V_1, S_1, R_1)$
- Seien L_2 kontextfreie Sprache mit Grammatik $G_2 = (\Sigma, V_2, S_2, R_2)$
- o.B.d.A. sei $V_1 \cap V_2 = \emptyset$.

Kleenscher Abschluss: Die Grammatik

$$V = V_1 \cup \{S\}$$

S neues Startsymbol

$$R = R_1 \cup \{S \rightarrow \varepsilon, S \rightarrow SS, S \rightarrow S_1\}$$

erzeugt L_1^* .



Die Klasse der kontextfreien Sprachen ist nicht abgeschlossen bzgl. Komplementbildung und Durchschnitt.

Beweis Schnitt: Betrachte die kontextfreien Sprachen

$$L_1 = \{a^n b^n | n \ge 1\}$$
 $L_2 = \{c\}^*$
 $L_3 = \{a\}^*$ $L_4 = \{b^n c^n | n \ge 1\}$

Nach dem letzen Satz sind dann auch $L_1 \cdot L_2$ und $L_3 \cdot L_4$ kontextfrei. Es ist dann

$$L := L_1L_2 \cap L_3L_4 = \{a^nb^nc^n|n \geq 1\}$$
.

Diese Sprache ist nicht kontextfrei.



Die Klasse der kontextfreien Sprachen ist nicht abgeschlossen bzgl. Komplementbildung und Durchschnitt.

Beweis Komplementbildung:

- Angenommen, die Klasse der kontextfreien Sprachen wäre bzgl. Komplementbildung abgeschlossen.
- Dann würde für beliebige kontextfreie Sprachen L_1 , L_2 gelten $(L_1^c \cup L_2^c)^c = L_1 \cap L_2$ ist wieder kontextfrei.
- Dies ist ein Widerspruch zur ersten Aussage des Satzes.