

# FORMALE SYSTEME

## 14. Vorlesung: Abschlusseigenschaften kontextfreier Sprachen

Hannes Straß

Folien: © Markus Krötzsch, <https://iccl.inf.tu-dresden.de/web/FS2020>, CC BY 3.0 DE

TU Dresden, 25. November 2021

### Pumpen für kontextfreie Sprachen

**Satz (Pumping-Lemma):**

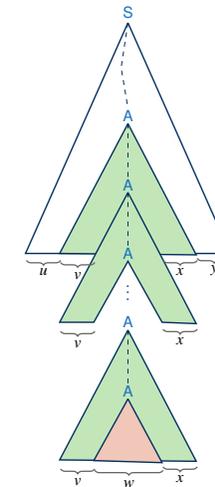
Für jede kontextfreie Sprache  $L$  gibt es eine Zahl  $n \geq 0$ , so dass gilt:  
 für jedes Wort  $z \in L$  mit  $|z| \geq n$   
 gibt es eine Zerlegung  $z = uvwxy$  mit  $|vx| \geq 1$  und  $|vwx| \leq n$ , so dass:  
 für jede Zahl  $k \geq 0$  gilt:  $uv^kwx^ky \in L$ .

Beispiel: Für die Sprache  $\{a^i b^i \mid i \geq 0\}$  gilt der Satz. Wir wählen  $n = 2$ . Sei  $z = a^i b^i$  mit  $i \geq 1$  ein beliebiges Wort mit  $|z| \geq 2$ . Wir wählen die Zerlegung  $u = a^{i-1}$ ,  $v = a$ ,  $w = \epsilon$ ,  $x = b$  und  $y = b^{i-1}$ .  
 Dann ist  $uv^kwx^ky = a^{i-1} a^k b^k b^{i-1} = a^{i+k-1} b^{i+k-1} \in L$  für alle  $k \geq 0$ .

## Rückblick: Das Pumping-Lemma

### Ableitungsbäume aufpumpen

Die Idee des Lemmas lässt sich gut am Ableitungsbaum darstellen:



Abgeleitetes Wort:  
 $uv^kwx^ky$

## Beispiel

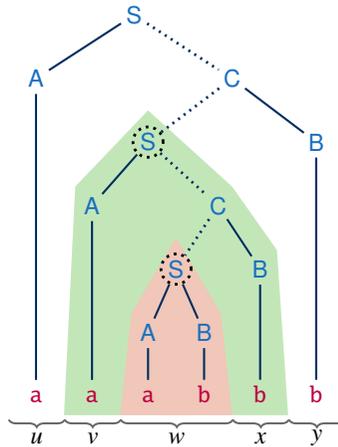
CNF-Grammatik für  $\{a^i b^i \mid i \geq 1\}$ :

$S \rightarrow AB \mid AC$   
 $A \rightarrow a$   
 $B \rightarrow b$   
 $C \rightarrow SB$

Konstante aus dem Beweis des Pumping-Lemmas:  $n = 2^4 = 16$

Aber man kann auch schon ab  $n = 4$  pumpen.

Beispiel: Ableitung für **aaabbb**



## Beispiel (2)

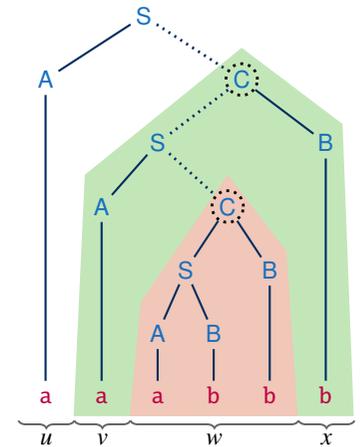
CNF-Grammatik für  $\{a^i b^i \mid i \geq 1\}$ :

$S \rightarrow AB \mid AC$   
 $A \rightarrow a$   
 $B \rightarrow b$   
 $C \rightarrow SB$

Konstante aus dem Beweis des Pumping-Lemmas:  $n = 2^4 = 16$

Aber man kann auch schon ab  $n = 4$  pumpen.

Beispiel: Ableitung für **aaabbb**



## Beispiel (3)

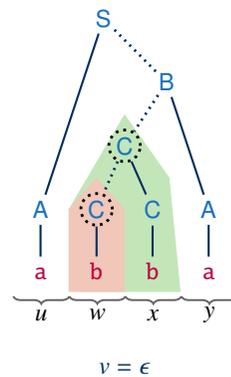
CNF-Grammatik für  $ab^+a$ :

$S \rightarrow AB$   
 $A \rightarrow a$   
 $B \rightarrow CA$   
 $C \rightarrow CC \mid b$

Konstante aus dem Beweis des Pumping-Lemmas:  $n = 2^4 = 16$

Aber man kann auch schon ab  $n = 4$  pumpen.

Beispiel: Ableitung für **abba**



## Abschlusseigenschaften kontextfreier Sprachen

## Rückblick: Reguläre Sprachen

**Satz:** Wenn  $L$ ,  $L_1$  und  $L_2$  reguläre Sprachen sind, dann beschreiben auch die folgenden Ausdrücke reguläre Sprachen:

- (1)  $L_1 \cup L_2$  (Abschluss unter Vereinigung)
- (2)  $L_1 \cap L_2$  (Abschluss unter Schnitt)
- (3)  $\bar{L}$  (Abschluss unter Komplement)
- (4)  $L_1 \circ L_2$  (Abschluss unter Konkatenation)
- (5)  $L^*$  (Abschluss unter Kleene-Stern)

Wie sieht es bei den kontextfreien Sprachen aus?

## Abschluss für kontextfreie Sprachen?

Bei kontextfreien Sprachen ergibt sich ein anderes Bild:

**Satz:** Wenn  $L$ ,  $L_1$  und  $L_2$  kontextfreie Sprachen sind, dann sind auch die folgenden Sprachen kontextfrei:

- (1)  $L_1 \cup L_2$  (Abschluss unter Vereinigung)
- (2)  $L_1 \circ L_2$  (Abschluss unter Konkatenation)
- (3)  $L^*$  (Abschluss unter Kleene-Stern)

Aber:

**Satz:** Es gibt kontextfreie Sprachen  $L$ ,  $L_1$  und  $L_2$ , so dass die folgenden Sprachen nicht kontextfrei sind:

- (1)  $L_1 \cap L_2$  (Nichtabschluss unter Schnitt)
- (2)  $\bar{L}$  (Nichtabschluss unter Komplement)

## Abschlusseigenschaften von Typ-2-Sprachen

**Satz:** Wenn  $L$ ,  $L_1$  und  $L_2$  kontextfreie Sprachen sind, dann sind auch die folgenden Sprachen kontextfrei:

- (1)  $L_1 \cup L_2$  **Abschluss unter Vereinigung**
- (2)  $L_1 \circ L_2$  (Abschluss unter Konkatenation)
- (3)  $L^*$  (Abschluss unter Kleene-Stern)

## Abschluss unter Vereinigung

**Beweisansatz:** Wir konstruieren eine Vereinigungsgrammatik.

Gegeben seien zwei formale Grammatiken  $G_1 = \langle V_1, \Sigma, P_1, S_1 \rangle$  und  $G_2 = \langle V_2, \Sigma, P_2, S_2 \rangle$  mit  $V_1 \cap V_2 = \emptyset$  (o.B.d.A.).

Die **Vereinigungsgrammatik**  $G_1 \uplus G_2$  ist gegeben durch

$$G_1 \uplus G_2 = \langle V_1 \cup V_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \rightarrow S_1 \mid S_2\}, S \rangle,$$

wobei  $S$  ein neues Startsymbol ist, das nicht in  $V_1 \cup V_2$  vorkommt.

In Worten: Die neuen Ableitungsregeln  $S \rightarrow S_1 \mid S_2$  ermöglichen es, dass  $G_1 \uplus G_2$  entweder Wörter aus  $G_1$  oder aus  $G_2$  generiert.

Es ist daher leicht zu sehen:

**Satz:**  $L(G_1 \uplus G_2) = L(G_1) \cup L(G_2)$ .

## Vereinigungen in Typ 2?

Bisher haben wir nur erkannt:

**Satz:**  $L(G_1 \uplus G_2) = L(G_1) \cup L(G_2)$ .

Für die Abschlusseigenschaft sollte noch mehr gelten:

**Satz:** Wenn  $G_1$  und  $G_2$  kontextfrei sind, dann ist auch  $G_1 \uplus G_2$  kontextfrei.

Das ist leicht zu sehen, da die zusätzliche Regel kontextfrei ist.

Daher gilt sogar noch ein stärkerer Satz:

**Satz:** Wenn  $G_1$  und  $G_2$  von Typ  $i \in \{2, 1, 0\}$  sind, dann ist auch  $G_1 \uplus G_2$  von Typ  $i$ .

→ Typ-0-Sprachen, kontextsensitive Sprachen und kontextfreie Sprachen sind unter Vereinigung abgeschlossen.

(Reguläre Sprachen auch, aber das haben wir anders gezeigt)

## Quiz: Vereinigungsgrammatik

Gegeben seien zwei Grammatiken  $G_1 = \langle V_1, \Sigma, P_1, S_1 \rangle$  und  $G_2 = \langle V_2, \Sigma, P_2, S_2 \rangle$  mit  $V_1 \cap V_2 = \emptyset$  (o.B.d.A.). Die Vereinigungsgrammatik  $G_1 \uplus G_2$  ist gegeben durch  $G_1 \uplus G_2 = \langle V_1 \cup V_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \rightarrow S_1 \mid S_2\}, S \rangle$  wobei  $S \notin V_1 \cup V_2$  gilt.

**Quiz:** Wir betrachten die Grammatiken  $G_1 = \langle V_1, \Sigma, P_1, S_1 \rangle$  und  $G_2 = \langle V_2, \Sigma, P_2, S_2 \rangle$  mit:...

## Abschlusseigenschaften von Typ-2-Sprachen

**Satz:** Wenn  $L$ ,  $L_1$  und  $L_2$  kontextfreie Sprachen sind, dann sind auch die folgenden Sprachen kontextfrei:

- (1)  $L_1 \cup L_2$  (Abschluss unter Vereinigung)
- (2)  $L_1 \circ L_2$  Abschluss unter Konkatination
- (3)  $L^*$  (Abschluss unter Kleene-Stern)

## Konkatination von Grammatiken

Wir erinnern uns:  $L_1 \circ L_2 = \{w_1 w_2 \mid w_1 \in L_1 \text{ und } w_2 \in L_2\}$ .

Es ist nicht schwer, eine passende Grammatik zu finden:

Gegeben seien zwei formale Grammatiken  $G_1 = \langle V_1, \Sigma, P_1, S_1 \rangle$  und  $G_2 = \langle V_2, \Sigma, P_2, S_2 \rangle$  mit  $V_1 \cap V_2 = \emptyset$  (o.B.d.A.).

Die Grammatik  $G_1 \circ G_2$  ist gegeben durch

$$G_1 \circ G_2 = \langle V_1 \cup V_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \rightarrow S_1 S_2\}, S \rangle,$$

wobei  $S$  ein neues Startsymbol ist, das nicht in  $V_1 \cup V_2$  vorkommt.

In Worten: Die neue Ableitungsregel  $S \rightarrow S_1 S_2$  ermöglicht es, dass  $G_1 \circ G_2$  Wörter aus  $G_1$  gefolgt von Wörtern aus  $G_2$  generiert.

## Beispiel Konkatenation

Wir betrachten die Grammatiken

$$G_1 : S_1 \rightarrow aS_1b \mid \epsilon$$

$$G_2 : S_2 \rightarrow cS_2 \mid \epsilon$$

Damit ist  $L(G_1) = \{a^i b^i \mid i \geq 0\}$  und  $L(G_2) = \{c^k \mid k \geq 0\}$ .

Die Konkatenation der Grammatiken enthält die folgenden Regeln:

$$G_1 \circ G_2 : \quad S \rightarrow S_1 S_2 \quad S_1 \rightarrow aS_1b \mid \epsilon \quad S_2 \rightarrow cS_2 \mid \epsilon$$

Damit ergibt sich  $L(G_1 \circ G_2) = \{a^i b^i c^k \mid i \geq 0, k \geq 0\}$ .

## Abschlusseigenschaften von Typ-2-Sprachen

**Satz:** Wenn  $L$ ,  $L_1$  und  $L_2$  kontextfreie Sprachen sind, dann sind auch die folgenden Sprachen kontextfrei:

- (1)  $L_1 \cup L_2$  (Abschluss unter Vereinigung)
- (2)  $L_1 \circ L_2$  (Abschluss unter Konkatenation)
- (3)  $L^*$  **Abschluss unter Kleene-Stern**

## Korrektheit

**Hypothese:**  $L(G_1 \circ G_2) = L(G_1) \circ L(G_2)$

Leider stimmt das nicht (für alle Typen von Grammatiken):

Gegenbeispiel:

$G_1$  sei die Grammatik mit der einen Regel  $S_1 \rightarrow a$  und  $G_2$  die Grammatik mit den Regeln  $S_2 \rightarrow b$  und  $aS_2 \rightarrow c$ .

Dann ist  $L(G_1) = \{a\}$  und  $L(G_2) = \{b\}$ .

Trotzdem erlaubt  $G_1 \circ G_2$  die Ableitung  $S \Rightarrow S_1 S_2 \Rightarrow aS_2 \Rightarrow c$ .

Korrekt ist dagegen:

**Satz:** Wenn  $G_1$  und  $G_2$  kontextfrei sind, dann ist  $L(G_1 \circ G_2) = L(G_1) \circ L(G_2)$ .

Zudem ist  $G_1 \circ G_2$  in diesem Fall kontextfrei.

**Beweis:** (Skizze) Aus Ableitungsbäumen von  $G_1$  und  $G_2$  lässt sich mittels der Regel  $S \rightarrow S_1 S_2$  ein Ableitungsbaum für  $G_1 \circ G_2$  konstruieren; da jede Ableitung von  $G_1 \circ G_2$  diese Regel verwendet, lässt sich diese Konstruktion auch immer umkehren.  $\square$

## Kleene-Stern für Grammatiken

Wir erinnern uns:  $L^* = \{w_1 w_2 \cdots w_i \mid i \geq 0, w_1, \dots, w_i \in L\} = \bigcup_{i \geq 0} L^i$

Auch hier kann man leicht eine passende Grammatik finden:

Gegeben sei eine formale Grammatik  $G = \langle V, \Sigma, P, S \rangle$ .

Die Grammatik  $G^*$  ist gegeben durch

$$G^* = \langle V \cup \{S'\}, \Sigma, P \cup \{S' \rightarrow \epsilon \mid SS'\}, S' \rangle,$$

wobei  $S'$  ein neues Startsymbol ist, das nicht in  $V$  vorkommt.

In Worten: Die neue Ableitungsregel  $S' \rightarrow \epsilon \mid SS'$  ermöglicht es, dass  $G^*$  beliebig lange Ketten von Wörtern aus  $G$  generiert.

## Quiz: Kleene-Stern für Grammatiken

Gegeben sei eine Grammatik  $G = \langle V, \Sigma, P, S \rangle$ . Die Grammatik  $G^*$  ist gegeben durch  $G^* = \langle V \cup \{S'\}, \Sigma, P \cup \{S' \rightarrow \epsilon \mid SS'\}, S' \rangle$  für ein neues Startsymbol  $S' \notin V$ .

**Quiz:** Wir betrachten die Grammatik  $G = \langle V, \Sigma, P, S \rangle$  für  $\Sigma = \{0, 1\}$  mit  $P$  wie folgt:...

## Korrektheit beim Kleene-Abschluss

Wie schon beim Konkatenieren funktioniert diese Operation auf kontextfreien Grammatiken wie erwünscht:

**Satz:** Wenn  $G$  kontextfrei ist, dann ist  $L(G^*) = L(G)^*$ .  
Zudem ist  $G^*$  in diesem Fall kontextfrei.

**Beweis:** (Skizze) Ableitungen in  $G^*$  verwenden Ableitungen von  $G$  und die zusätzliche Regel; dies kann wieder in beide Richtungen verwendet werden, um Sprachgleichheit zu zeigen.  $\square$

Für nicht-kontextfreie Grammatiken gilt dieser Satz im Allgemeinen nicht:

Beispiel: Die (kontextsensitive) Grammatik  $G$  mit der einen Regel  $SS \rightarrow 0$  kann nichts hervorbringen.

$$G^* : \quad S' \rightarrow \epsilon \mid SS', \quad SS \rightarrow 0$$

Trotzdem erlaubt die Grammatik  $G^*$  die Ableitung eines Wortes:

$$S' \Rightarrow SS' \Rightarrow SSS' \Rightarrow SS \Rightarrow 0$$

## Beispiel Kleene-Stern

Wir betrachten die Grammatik

$$G : \quad S \rightarrow SaSb \mid SbSa \mid \epsilon$$

Übung: Welche Sprache erzeugt diese Grammatik?

Der Kleene-Abschluss dieser Grammatik ist

$$G^* : \quad S' \rightarrow \epsilon \mid SS' \quad S \rightarrow SaSb \mid SbSa \mid \epsilon$$

Übung: Welche Sprache erzeugt diese Grammatik?

## Nichtabschlusseigenschaften

**Satz:** Es gibt kontextfreie Sprachen  $L$ ,  $L_1$  und  $L_2$ , so dass die folgenden Sprachen nicht kontextfrei sind:

- (1)  $L_1 \cap L_2$       Nichtabschluss unter Schnitt
- (2)  $\bar{L}$               (Nichtabschluss unter Komplement)

## Nichtabschluss unter $\cap$

**Beweisansatz:** Wir müssen kontextfreie Sprachen  $L_1$  und  $L_2$  finden, für die  $L_1 \cap L_2$  nicht kontextfrei ist.

Welche nichtkontextfreien Sprachen kennen wir?

$$\{a^i b^i c^i \mid i \geq 0\}$$

Welche ähnlichen kontextfreien Sprachen kennen wir?

$$\{a^i b^i c^k \mid i \geq 0, k \geq 0\}$$

## Nichtabschlusseigenschaften

**Satz:** Es gibt kontextfreie Sprachen  $L$ ,  $L_1$  und  $L_2$ , so dass die folgenden Sprachen nicht kontextfrei sind:

- (1)  $L_1 \cap L_2$  (Nichtabschluss unter Schnitt)
- (2)  $\bar{L}$  Nichtabschluss unter Komplement

## Nichtabschluss unter $\cap$ (2)

**Satz:** Es gibt kontextfreie Sprachen  $L$ ,  $L_1$  und  $L_2$ , so dass  $L_1 \cap L_2$  keine kontextfreie Sprache ist.

**Beweis:** Die folgenden Sprachen sind kontextfrei:

$$\begin{aligned} \{a^i b^i \mid i \geq 0\} & \quad \text{(zuvor gezeigt)} \\ \{b^i c^i \mid i \geq 0\} & \quad \text{(analog)} \\ \{a^i \mid i \geq 0\} & \quad \text{(regulär)} \\ \{c^i \mid i \geq 0\} & \quad \text{(regulär, zuvor gezeigt)} \end{aligned}$$

Dank Abschluss unter Konkatenation sind also auch diese Sprachen kontextfrei:

$$\begin{aligned} L_1 &= \{a^i b^i \mid i \geq 0\} \circ \{c^i \mid i \geq 0\} = \{a^i b^i c^k \mid i \geq 0, k \geq 0\} \\ L_2 &= \{a^i \mid i \geq 0\} \circ \{b^i c^i \mid i \geq 0\} = \{a^i b^k c^k \mid i \geq 0, k \geq 0\} \end{aligned}$$

Der Schnitt  $L_1 \cap L_2$  ist aber  $\{a^i b^i c^i \mid i \geq 0\}$  und daher nicht kontextfrei (zuvor gezeigt).  $\square$

## Nichtabschluss unter Komplement

**Satz:** Es gibt eine kontextfreie Sprache  $L$ , so dass  $\bar{L}$  keine kontextfreie Sprache ist.

**Beweis:** Die Behauptung folgt aus den bisherigen Resultaten:

- Angenommen, Typ-2-Sprachen wären unter Komplement abgeschlossen.
- Seien  $L_1$  und  $L_2$  kontextfreie Sprachen.
- Dann sind auch  $\bar{L}_1$  und  $\bar{L}_2$  kontextfreie Sprachen.
- Wir wissen bereits, dass Typ-2-Sprachen unter Vereinigung abgeschlossen sind.
- Demnach ist auch  $\bar{L}_1 \cup \bar{L}_2$  eine kontextfreie Sprache.
- Wegen Abschlusses unter Komplement ist also auch  $\overline{\bar{L}_1 \cup \bar{L}_2}$  kontextfrei.
- Laut den Gesetzen der Mengenlehre gilt:

$$L_1 \cap L_2 = \overline{\bar{L}_1 \cup \bar{L}_2} \quad \text{(de Morgan)}$$

- Da  $L_1$  und  $L_2$  beliebig waren, folgt, dass Typ-2-Sprachen unter Schnitt abgeschlossen sind.
- Widerspruch.  $\square$

## Eine Warnung zum Nichtabschluss

Nichtabschluss unter Schnitt und Komplement bedeutet nicht, dass Schnitte bzw. Komplemente kontextfreier Sprachen grundsätzlich nicht kontextfrei sein können.

Beispiel: Alle regulären Sprachen sind kontextfrei, und ihre Schnitte und Komplemente sind weiterhin regulär, also auch kontextfrei.

Beispiel: Das Komplement der nichtregulären Sprache  $L = \{a^i b^i \mid i \geq 0\}$  über dem Alphabet  $\{a, b\}$  ist die Vereinigung der folgenden Sprachen:

$$L_1 = \{a, b\}^* \{ba\} \{a, b\}^* \quad L_2 = \{a\}^+ L \quad L_3 = L \{b\}^+$$

$L_1$  ist regulär, also kontextfrei.  $L_2$  und  $L_3$  sind kontextfrei, da sie als Konkatination zweier kontextfreier Sprachen entstehen. Die Vereinigung  $L_1 \cup L_2 \cup L_3$  ist also auch kontextfrei.

## Ein nichtkontextfreies Komplement

Unser Widerspruchsbeweis zum Abschluss unter Komplement liefert uns noch kein konkretes Beispiel eines nichtkontextfreien Komplements.

Wir können es aber aus den Beweisen ableiten:

- Seien  $L_1 = \{a^i b^k c^k \mid i \geq 0, k \geq 0\}$  und  $L_2 = \{a^i b^k c^k \mid i \geq 0, k \geq 0\}$
- $L_1 \cap L_2 = \overline{L_1} \cup \overline{L_2} = \{a^i b^i c^i \mid i \geq 0\} = L$  ist nicht kontextfrei
- $\overline{L_1} \cup \overline{L_2}$  ist die Vereinigung der folgenden Typ-2-Sprachen:
  - $\{a, b, c\}^* \circ \{ba, ca, cb\} \circ \{a, b, c\}^*$  (falsche Reihenfolge)
  - $\{a\}^+ \circ \{a^i b^i \mid i \geq 0\} \circ \{c\}^*$  (zu viele a für  $L_1$ )
  - $\{a^i b^i \mid i \geq 0\} \circ \{b\}^+ \circ \{c\}^*$  (zu viele b für  $L_1$ )
  - $\{a\}^* \circ \{b\}^+ \circ \{b^i c^i \mid i \geq 0\}$  (zu viele b für  $L_2$ )
  - $\{a\}^* \circ \{b^i c^i \mid i \geq 0\} \circ \{c\}^+$  (zu viele c für  $L_2$ )
- $\overline{L_1} \cup \overline{L_2}$  ist daher kontextfrei, aber ihr Komplement nicht.

## Zusammenfassung und Ausblick

Kontextfreie Sprachen sind abgeschlossen unter Vereinigung, Konkatination und Kleene-Stern.

Kontextfreie Sprachen sind nicht abgeschlossen unter Komplement und Schnitt.

Abschlüsse beruhen auf Grammatikoperationen, die man auf beliebige Grammatiken anwenden könnte, aber nur der Abschluss unter  $\cup$  für Typ 0 und Typ 1 kann so gezeigt werden.

Offene Fragen:

- Haben kontextfreie Sprachen ein Berechnungsmodell?
- Welche Probleme auf kontextfreien Grammatiken kann man lösen?
- Was gibt es zu Typ 1 und Typ 0 zu sagen?