

FORMALE SYSTEME

14. Vorlesung: Abschlusseigenschaften kontextfreier Sprachen

Markus Krötzsch

Professur für Wissensbasierte Systeme

TU Dresden, 27. November 2023

Rückblick: Das Pumping Lemma

Pumpen für kontextfreie Sprachen

Satz (Pumping Lemma): Für jede kontextfreie Sprache \mathbf{L}

gibt es eine Zahl $n \geq 0$, so dass gilt:

für jedes Wort $z \in \mathbf{L}$ mit $|z| \geq n$

gibt es eine Zerlegung $z = uvwxy$ mit $|vx| \geq 1$ und $|vwx| \leq n$, s.d.:

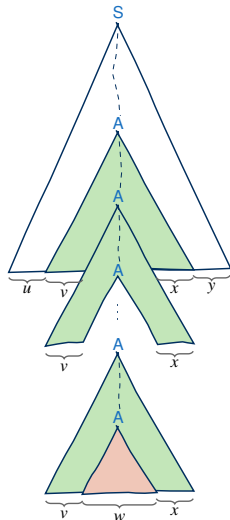
für jede Zahl $k \geq 0$ gilt: $uv^kwx^ky \in \mathbf{L}$

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 $uv^kwx^ky = a^{i-1} a^k b^k b^{i-1} = a^{i+k-1} b^{i+k-1} \in \mathbf{L}$ für alle $k \geq 0$.

Ableitungsbäume aufpumpen

Die Idee des Lemmas lässt sich gut am Ableitungsbaum darstellen:



Abgeleitetes Wort:
 uv^kwx^ky

Beispiel

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

$$S \rightarrow AB \mid AC$$

$$A \rightarrow a$$

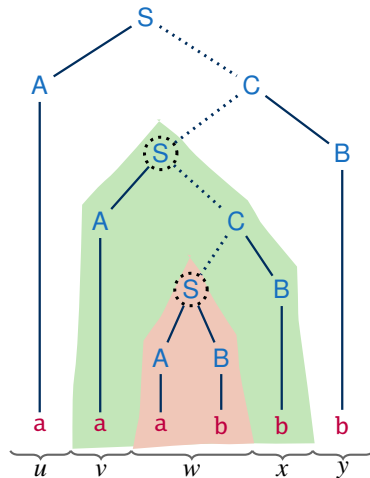
$$B \rightarrow b$$

$$C \rightarrow SB$$

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

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

Beispiel: Ableitung für **aaabbb**



Beispiel (3)

CNF-Grammatik für ab^+a :

$S \rightarrow AB$

$A \rightarrow a$

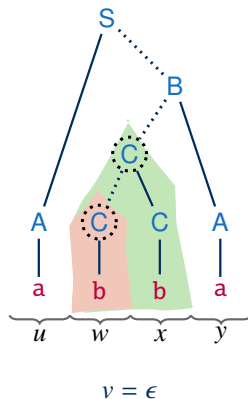
$B \rightarrow CA$

$C \rightarrow CC \mid b$

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

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 beschreiben auch die folgenden Ausdrücke kontextfreie Sprachen:

- (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 Ausdrücke keine kontextfreien Sprachen 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 beschreiben auch die folgenden Ausdrücke kontextfreie Sprachen:

- (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

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: $\mathbf{L}(G_1 \uplus G_2) = \mathbf{L}(G_1) \cup \mathbf{L}(G_2)$.

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)

Abschlusseigenschaften von Typ-2-Sprachen

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

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

Konkatenation von Grammatiken

Wir erinnern uns: $\mathbf{L}_1 \circ \mathbf{L}_2 = \{w_1w_2 \mid w_1 \in \mathbf{L}_1 \text{ und } w_2 \in \mathbf{L}_2\}$

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

$$G_1 \circ G_2 = \langle V_1 \cup V_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \rightarrow S_1S_2\}, S \rangle,$$

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

In Worten: Die neue Ableitungsregel $S \rightarrow S_1S_2$ 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 $\mathbf{L}(G_1) = \{a^i b^i \mid i \geq 0\}$ und $\mathbf{L}(G_2) = \{c^k \mid k \geq 0\}$.

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

Korrektheit

Hypothese: $\mathbf{L}(G_1 \circ G_2) = \mathbf{L}(G_1) \circ \mathbf{L}(G_2)$

Leider stimmt das nicht:

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

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

Korrekt ist dagegen:

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

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

Beweis: einfach (man betrachte den Ableitungsbaum).

Damit erhalten wir den gewünschten Abschluss.

Abschlusseigenschaften von Typ-2-Sprachen

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

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

Kleene-Stern für Grammatiken

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

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 neuen Ableitungsregeln $S' \rightarrow \epsilon \mid SS'$ ermöglichen es, dass G^* beliebig lange Ketten aus Wörtern aus G generiert.

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

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

Übung: Welche Sprache erzeugt diese Grammatik?

Korrektheit beim Kleene-Abschluss

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

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

Beweis: einfach (man betrachte den Ableitungsbaum).

Für nicht kontextfreie Grammatiken gilt dieser Satz im Allgemeinen nicht:

Beispiel: Die (kontextsensitive) Grammatik G mit der einen Regel $SS \rightarrow \emptyset$ kann nichts hervorbringen.
Trotzdem erlaubt die Grammatik G^* die Ableitung eines Wortes:

$$S' \Rightarrow SS \Rightarrow \emptyset$$

Nichtabschlusseigenschaften

Satz: Es gibt kontextfreie Sprachen L , L_1 und L_2 , so dass die folgenden Ausdrücke keine kontextfreien Sprachen 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?

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

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:

$\{a^i b^i \mid i \geq 0\}$ (zuvor gezeigt)

$\{b^i c^i \mid i \geq 0\}$ (analog)

$\{a^i \mid i \geq 0\}$ (regulär)

$\{c^i \mid i \geq 0\}$ (regulär, zuvor gezeigt)

Dank Abschluss unter Konkatenation sind also auch kontextfrei:

$$L_1 = \{a^i b^i \mid i \geq 0\} \circ \{c^i \mid i \geq 0\} = \{a^i b^i c^k \mid i \geq 0, k \geq 0\}$$

$$L_2 = \{a^i \mid i \geq 0\} \circ \{b^i c^i \mid i \geq 0\} = \{a^i b^k c^k \mid i \geq 0, k \geq 0\}.$$

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

Nichtabschlusseigenschaften

Satz: Es gibt kontextfreie Sprachen L , L_1 und L_2 , so dass die folgenden Ausdrücke keine kontextfreien Sprachen sind:

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

Nichtabschluss unter Komplement

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

Beweis: Die Behauptung folgt unmittelbar aus dem bereits gezeigten:

- Angenommen, Typ-2-Sprachen wären unter Komplement abgeschlossen
- Wir wissen bereits, dass Typ-2-Sprachen unter Vereinigung abgeschlossen sind
- Laut den Gesetzen der Mengenlehre gilt:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}} \quad (\text{De Morgan})$$

- Daraus folgt, dass Typ-2-Sprachen unter Schnitt abgeschlossen sind –
Widerspruch □

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, aber 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:

$$L_1 = \{a, b\}^* \{ba\} \{a, b\}^*$$

$$L_2 = \{a\}^+ L$$

$$L_3 = L \{b\}^+$$

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 $L_1 = \{a^i b^i c^k \mid i \geq 0, k \geq 0\}$ und $L_2 = \{a^i b^k c^k \mid i \geq 0, k \geq 0\}$
- $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}} = \{a^i b^i c^i \mid i \geq 0\} = L$ 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, Konkatenation 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?