Sei $\Sigma = \{a, b, c, d, e\}$. Finden Sie Sprachen $\emptyset \subsetneq L_1 \subseteq L_2 \subseteq L_3 \subsetneq \Sigma^*$, so dass

(a) L_1 kontextfrei, aber nicht regulär,

(b) L_2 regulär und

(c) L_3 nicht kontextfrei ist.

LÖSUNG: (a) Sei $L_1 = \{a^n b^n \mid n \geq 0\}$. Diese Sprache ist kontextfrei, da Sie von folgender kontextfreier Grammatik erzeugt wird: $G = (\{S\}, \{a, b\}, \{S \to aSb, S \to \varepsilon\}, S)$. Aber L_1 nicht regulär. Sei $n \in \mathbb{N}$ beliebig, aber fest. Sei $x = a^n b^n$, $x \in L_1$ und $|x| \geq n$. Für alle möglichen Zerlegungen $x = uvw$ mit $|v| \geq 1$ und $|uv| \leq n$ gilt: $\exists k \geq 0, j \geq 1 : u = a^k, v = a^j, w = a^{n-i-j} b^n$ mit $j + k \leq n$. Nun ist $uv^2w = a^{n+j} b^n \notin L$ da $j \geq 1$ ist. Folglich kann nach Pumping-Lemma L_1 nicht regulär sein.

(b) Sei $L_2 = \{a^n b^m \mid n, m \geq 0\}$. Offensichtlich gilt $L_1 \subseteq L_2$. L_2 ist regulär, da folgender endlicher Automat L_2 erkennt: $A = (\{q, p\}, \{a, b\}, \delta, q, \{q, p\})$ mit $\delta(q, a) = \{q, p\}, \delta(p, a) = \delta(q, b) = \emptyset, \delta(p, b) = \{p\}$.

(c) Sei $L_3 = \{a^n b^m c^l d^l e^l \mid n, m, l \geq 0\}$. Offensichtlich gilt $L_2 \subseteq L_3$. L_3 ist nicht kontextfrei. Angenommen L_3 ist kontextfrei. Dann müsste die Konklusion des Pumping-Lemmas für kontextfreie Sprachen gelten. Sei $n \geq 0$ beliebig aber fest. Dann ist $w = c^n d^n e^n \in L_3$ und $|w| \geq n$. Nun gilt für alle Zerlegungen $w = uvxyz$ mit $|vxy| \leq n$ und $|vy| \geq 1$, dass in vxy nicht cs, ds und es sein können. Also kann uv^2xy^2z nicht aus L_3 sein, da es nicht mehr gleich viele cs wie ds wie es enthält.

Wikipedia: Chomsky-Normalform

Eine formale Grammatik $G=(V,\Sigma,P,S)$ ist in Chomsky-Normalform, wenn jede Produktion aus P eine der folgenden Formen hat:

- $A \to BC$
- $A \to a$
- $S \to \varepsilon$

wobei A, B und C Nichtterminalsymbole aus V sind und a ein Terminalsymbol aus Σ ist. S ist das Startsymbol und ε das leere Wort. Wenn die Produktion $S \to \varepsilon$ zur Grammatik gehört, dann darf S nicht auf der rechten Seite einer Produktion stehen.

Läßt man bei der ersten Produktion auf der rechten Seite beliebig viele anstatt zwei Nichtterminalsymbole zu, so spricht man von einer schwachen Chomsky-Normalform.

Konstruktion:

Liegt eine kontextfreie Grammatik $G=(V,\Sigma,P,S)$ vor, so läßt sich daraus schrittweise eine Grammatik G' in Chomsky-Normalform generieren, die dieselbe Sprache erzeugt.

) Ausnahme $S \to \varepsilon$ behandeln

Enthält die Grammatik G die Regel $S \to \varepsilon$, wird ein neues Startsymbol S' für G' eingeführt. Anschließend erhält die neue Grammatik die Regeln $S' \to \varepsilon$ und $S' \to S$. Damit ist sichergestellt, daß die Grammatik weiterhin das leere Wort ermöglicht und das ursprüngliche Startsymbol weiterhin auf der rechten Seite der Produktion verwendet werden kann.

) Eine schwache Chomsky-Normalform erzeugen.

Jedem Terminalsymbol a wird ein Nichtterminalsymbol X_a zugeordnet. Auf der rechten Seite jeder Produktion werden sämtliche Terminalsymbole a durch das entsprechende Nichtterminalsymbol X_a ersetzt. Abschließend werden alle Produktionen $X_a \to a$ der Grammatik hinzugefügt.

Rechte Seiten mit mehr als zwei Nichtterminalen ersetzen

Sind auf der rechten Seite einer Produktion mehr als zwei Nichtterminale, so werden zwei benachbarte Nichtterminale AB durch ein neues Nichtterminal Y_{AB} ersetzt. Die Produktion $Y_{AB} \to AB$ wird zur Grammatik hinzugefügt. Dies wiederholt man solange, bis keine Produktion mit mehr als zwei Nichtterminalen mehr vorkommt.

④ ε-Produktion entfernen

Streiche die Regeln $A \to \varepsilon$, außer $S' \to \varepsilon$ (falls vorhanden).

Gab es vorher genau eine Produktion mit A auf der linken Seite (also $A \to \varepsilon$), so streiche das A überall auf den rechten Seiten der Produktionen, denn es kann nicht zu einem Terminal abgeleitet werden.

Gab es vorher mehrere Produktionen mit A auf der linken Seite, so füge für jede Regel, die ein solches A auf der rechten Seite enthält, eine Regel hinzu, in der das A gestrichen wurde, denn es muß der Fall betrachtet werden, in dem das A als leeres Wort abgeleitet werden oder etwa nicht. Die Regel $C \to AB$ wird dann beispielsweise um die Regel $C \to B$ ergänzt.

Aus $C \to AB$ wird also:

$C \to B$

$C \to AB$

⑤ Kettenregeln (Produktionen der Form $A \to B$) entfernen

Wenn man eine Kettenregel, d.h. eine Produktion der Form $A \to B$, entfernt, fügt man für jede vorhandene Produktion der Form $B \to w$ eine neue Produktion $A \to w$ hinzu, falls diese keine bereits entfernte Kettenregel ergibt. w ist hierbei ein beliebiges Wort; die vorangegangenen Änderungen gewährleisten aber, daß W entweder genau ein Terminalsymbol oder ein Wort aus genau zwei Nichtterminalsymbolen ist.

Vgl.: Zu jeder kontextsensitiven Grammatik existiert eine Grammatik in Kuroda-Normalform mit Produktionsregeln der Form

- $A \to a$
- $A \to B$
- $A \to BC$
- $AB \to CD$

Eine Sprache heißt pumpbar gdw. es gibt ein $k \in \mathbb{N}$, sodaß für alle Wörter $w \in L$ mit $|w| \geq k$ es eine Zerlegung von $w = xyz$ gibt, für die gilt:

1. $|xy| \leq k$
2. $\forall i \geq 0 : xy^i z \in L$
3. $|y| > 0$

Pumpinglemma für reguläre Sprache:

L ist regulär $\Rightarrow$ L ist pumpbar

Beweis: Sei L regulär, dann existiert ein deterministischer endlicher Automat $M = (Q, \Sigma, \delta, q_0, F)$ der L erkennt. Wir setzen $k = |Q| + 1$.

Jeder Lauf für Wort w mit $|w| \geq k$ besucht mindestens $k = |Q|+1$ Zustände

$\rightarrow$ Zumindest 1 Zustand doppelt

Sei p der erste Zustand, der mehrfach besucht wurde

Lauf: $$\underbrace{q_0 \to \bigcirc \to \bigcirc \to p}_{\substack{\text{Teillauf zum 1. } p \\ x}} \underbrace{\to \bigcirc \to p}_{\substack{\text{vom 1. zum 2. } p \\ y}} \underbrace{\to \bigcirc \to p \to \bigcirc}_{\substack{\text{Teillauf ab 2. } p \\ z}}$$

- Eigenschaft 1: $|y| > 0$
- Eigenschaft 2: $|xy| \leq k$
- Eigenschaft 3: $\forall i \geq 0 : xy^i z \in L$

Wikipedia

Pumping-Lemma für kontextfreie Sprachen

Für jede kontextfreie Sprache L gibt es eine natürliche Zahl n, sodass gilt: Jedes Wort z in L mit Mindestlänge n hat eine Zerlegung $z = uvwxy$ mit den folgenden drei Eigenschaften:

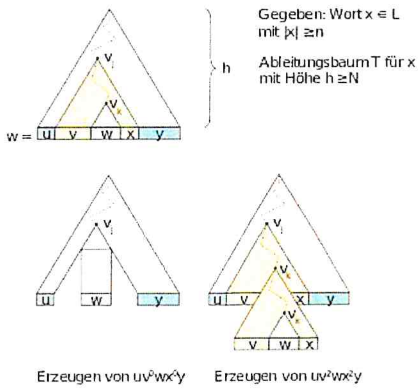
1. Die Wörter v, w und x haben zusammen höchstens die Länge n, d. h. $|vwx| \leq n$.
2. Eines der Wörter v, x ist nicht leer. Also $|vx| \geq 1$.
3. Für jede natürliche Zahl (mit 0) i ist das Wort $uv^i wx^i y$ in der Sprache L, d. h. die Wörter uwy, $uvwxy$, $uvvwxxy$ usw. liegen alle in L.

Neben den kontextfreien Sprachen gibt es auch nicht kontextfreie Sprachen, die dieses Pumping-Lemma erfüllen. Die Umkehrung des Lemmas gilt im Allgemeinen also nicht. Eine Verallgemeinerung des Pumping-Lemmas für kontextfreie Sprachen ist Ogdens Lemma.

Beweis

Gegeben sei eine kontextfreie Grammatik G in Chomsky-Normalform mit N Variablen, für die gilt, dass sie gerade die gewünschte Sprache beschreibt. Sei nun ein Wort x aus dieser Sprache gegeben, für das gilt: $|x| \geq 2^{|N|} =: n$.

Betrachten wir nun einen Ableitungsbaum T für x mit Höhe h. Da unsere Sprache in CNF angegeben wurde, hat T die Form eines Binärbaumes. Daraus folgt für die Höhe von T $h \geq \log_2(n) = |N|$. Es gibt also einen Pfad $v_0 \ldots v_h$ in T von der Wurzel zu einem Blatt, für den gilt, dass er Länge $h + 1 > |N|$ hat. Es existieren also zwei Knoten v_j, v_k auf diesem Pfad mit $0 \leq j < k \leq h$, welche die gleichen Variablen von G A_j, A_k repräsentieren.



Die Idee des Pumping-Lemmas für kontextfreie Sprachen ist, dass ein Wortteil durch mehrfaches Ableiten derselben Variablen beliebig oft wiederholt werden kann.

Betrachtet man den Teilbaum T_k, welcher von v_k aus aufgespannt wird, so bilden dessen Blätter den Teilstring w. Der Teilbaum T_j, welcher von v_j aufgespannt wird, besitzt als Teilbaum den Baum T_k. Man kann also die Blätter von T_j aufteilen in Blätter links von T_k und Blätter rechts von T_k und erhält somit eine Aufteilung der Blätter von T_j der Form vwx. Ebenso unterteilt der Teilbaum T_j den gesamten Ableitungsbaum in drei Teile u, vwx, y. Wir erhalten also als Aufteilung die Teilstrings u, y, welche im Ableitungsbaum links bzw. rechts von dem von v_j aufgespannten Teilbaum liegen, die Teilstrings v, x, welche in dem Teilbaum T_j liegen nicht jedoch in T_k, und zu guter Letzt den Teilstring w, welcher in T_k liegt. Da v_j und v_k die gleichen Variablen unserer Grammatik repräsentieren, folgt daraus, dass der Pfad von v_j nach v_k beliebig oft wiederholt werden kann. Durch eine Wiederholung des Pfades würden wir Worte der Form $uv^i wx^i y$ erzeugen, ohne unsere Sprache zu verlassen. Womit wir das Pumping-Lemma für kontextfreie Sprachen bewiesen hätten.

Beispiel

Ist die Sprache $L = \{a^m b^m c^m \mid m \geq 1\}$ kontextfrei?

Abschlußeigenschaften

	Regulär	Kontextfrei	Kontextsensitiv	Typ-0
Vereinigung	+	+	+	+
Konkatenation	+	+	+	+
Kleene Hülle	+	+	+	+
Schnitt	+	(−)	+	+
Komplement	+	(−)	+	(−)

Bemerkung 1. CF ist nicht abgeschlossen gegen Durchschnitt und Komplement.

Beweis: Seien $L = \{a^n b^n c^m \mid n,m \geq 0\}$ und $L' = \{a^m b^n c^n \mid n,m \geq 0\}$

Dann gilt $L \cap L' = \{a^n b^n c^n \mid n \geq 0\}$.

Mit Hilfe des Pumping-Lemmas kann man zeigen, daß $L \cap L'$ nicht kontexfrei ist (aber kontextsensitiv).

Nach den deMorganschen Regeln gilt $\overline{\bar{L_1} \cap \bar{L_2}} = L_1 \cup L_2$. Da CF gegen Vereinigung abgeschlossen ist,

$\Updownarrow$

$\bar{L_1} \cap \bar{L_2} = \overline{L_1 \cup L_2}$

würde aus dem Abschluß gegen Komplement der Abschluß gegen Durchschnitt folgen. $\square$.

Bemerkung 2. Die Klasse der Typ-0-Sprachen ist nicht unter Komplement abgeschlossen.

Beweis: Eine Sprache ist vom Typ-0 gdw. sie semi-entscheidbar ist.

Das Selbsthalteproblem H ist semi-entscheidbar, aber nicht entscheidbar.

Wäre H^c auch vom Typ-0 (semi-entscheidbar), dann wäre

H entscheidbar ↯

Bemerkung 3. Die regulären Sprachen sind nicht abgeschlossen unter unendlicher Vereinigung.

Beweis: Für alle $k \in \mathbb{N}$ ist die Sprache $L_k := \{a^k b^k\}$ ein Singleton. regulär

Jedoch ist die unendliche Vereinigung $L := \bigcup_{k \in \mathbb{N}} L_k = \{a^n b^n \mid n \in \mathbb{N}\}$ nach dem Pumping-Lemma nicht regulär.

$\square$