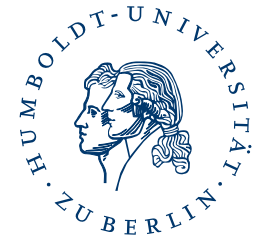
Einführung in die Theoretische Informatik

Johannes Köbler



Institut für Informatik
Humboldt-Universität zu Berlin

WS 2020/21

Die Chomsky-Hierarchie

Man unterscheidet vier Typen von Grammatiken $G = (V, \Sigma, P, S)$

Definition

1. G heißt vom Typ 3 oder regulär, falls für alle Regeln $u \to v$ gilt:
   $u \in V$ und $v \in \Sigma V \cup \Sigma \cup \{\varepsilon\}$
   (d.h. alle Regeln haben die Form $A \to aB$, $A \to a$ oder $A \to \varepsilon$)
2. G heißt vom Typ 2 oder kontextfrei, falls für alle Regeln $u \to v$ gilt:
   $u \in V$ (d.h. alle Regeln haben die Form $A \to v$)
3. G heißt vom Typ 1 oder kontextsensitiv, falls für alle Regeln $u \to v$ gilt:
   $|v| \geq |u|$ (mit Ausnahme der ε-Sonderregel, s. unten)
4. Jede Grammatik ist automatisch vom Typ 0

Die ε-Sonderregel

In einer kontextsensitiven Grammatik ist auch die Regel $S \to \varepsilon$ zulässig, falls das Startsymbol S nicht auf der rechten Seite einer Regel vorkommt

Kontextfreie Sprachen

Bemerkung

- Es ist klar, dass jede reguläre Grammatik auch kontextfrei ist
- Zudem ist die Sprache $L = \{a^n b^n \mid n \geq 0\}$ nicht regulär
- Es ist aber leicht, eine kontextfreie Grammatik für L anzugeben:

  $$G = (\{S\}, \{a, b\}, P, S) \text{ mit } P = \{S \to aSb, \varepsilon\}$$

- Also gilt REG $\subsetneq$ CFL
- Allerdings sind nicht alle kontextfreien Grammatiken kontextsensitiv
- Z.B. ist obige Grammatik G nicht kontextsensitiv, da sie die Regel $S \to \varepsilon$ enthält und S auf der rechten Seite der Regel $S \to aSb$ vorkommt
- Wir können G jedoch wie folgt in eine Grammatik G' umwandeln:
  - ersetze die Regel $S \to \varepsilon$ durch die Regel $S \to ab$ und
  - füge ein neues Startsymbol S' sowie die Regeln $S' \to S, \varepsilon$ hinzu
- Tatsächlich lässt sich jede kontextfreie Grammatik G in eine äquivalente kontextfreie Grammatik G' umwandeln, die auch kontextsensitiv ist

Chomsky-Normalform

Definition

Eine Grammatik $G = (V, \Sigma, P, S)$ ist in Chomsky-Normalform (CNF), falls $P \subseteq V \times (V^2 \cup \Sigma)$ ist, d.h. alle Regeln haben die Form $A \to BC$ oder $A \to a$

Satz

Zu jeder kontextfreien Grammatik G lässt sich eine CNF-Grammatik G' mit $L(G') = L(G) \setminus \{\varepsilon\}$ konstruieren

Anwendungen der Chomsky-Normalform

Korollar

$\mathrm{CFL} \subseteq \mathrm{CSL}$

Beweis

- Sei $L \in \mathrm{CFL}$ und sei $G = (V, \Sigma, P, S)$ eine CNF-Grammatik mit $L(G) = L \setminus \{\varepsilon\}$
- Im Fall $\varepsilon \notin L$ folgt sofort $L = L(G) \in \mathrm{CSL}$, da G kontextsensitiv ist
- Ist $\varepsilon \in L$, so erzeugt folgende kontextsensitive (und kontextfreie) Grammatik G' die Sprache $L = L(G) \cup \{\varepsilon\}$:

$$G' = (V \cup \{S_{neu}\}, \Sigma, P \cup \{S_{neu} \to S, \varepsilon\}, S_{neu})$$

□

Weitere Anwendungen der Chomsky-Normalform

- Der Beweis des Pumping-Lemmas für kontextfreie Sprachen basiert auf CNF-Grammatiken
- Zudem ermöglichen sie einen effizienten Algorithmus zur Lösung des Wortproblems für kontextfreie Sprachen

Das Pumping-Lemma für kontextfreie Sprachen

Zu jeder kontextfreien Sprache $L \in \mathrm{CFL}$ gibt es eine Zahl l, so dass sich alle Wörter $z \in L$ mit $|z| \geq l$ in $z = uvwxy$ zerlegen lassen mit

1. $vx \neq \varepsilon$,
2. $|vwx| \leq l$ und
3. $uv^i wx^i y \in L$ für alle $i \geq 0$

Das Wortproblem für kontextfreie Grammatiken

Gegeben: Eine kontextfreie Grammatik G und ein Wort x

Gefragt: Ist $x \in L(G)$?

Das Pumping-Lemma für kontextfreie Sprachen

Beispiel

- Betrachte die Sprache $L = \{a^n b^n | n \geq 0\}$
- Dann lässt sich jedes Wort $z = a^n b^n = a^{n-1} ab b^{n-1}$ in L mit $|z| \geq l = 2$ pumpen
- Zerlegen wir nämlich z in

  $$z = uvwxy \text{ mit } u = a^{n-1},\ v = a,\ w = \varepsilon,\ x = b \text{ und } y = b^{n-1},$$

  dann gilt

  1. $vx = ab \neq \varepsilon$
  2. $|vwx| = |ab| \leq 2$ und
  3. $uv^i wx^i y = a^{n-1} a^i b^i b^{n-1} \in L$ für alle $i \geq 0$

◁

Anwendung des Pumping-Lemmas

Beispiel

- Die Sprache $L = \{a^n b^n c^n \mid n \geq 0\}$ ist nicht kontextfrei
- Für eine vorgegebene Zahl $l \geq 0$ hat nämlich das Wort $z = a^l b^l c^l \in L$ die Länge $|z| = 3l \geq l$
- Dieses Wort lässt sich aber nicht pumpen:

  Für jede Zerlegung $z = uvwxy$ mit $vx \neq \varepsilon$ und $|vwx| \leq l$ gehört $z' = uv^0wx^0y$ nicht zu L:
  - Wegen $vx \neq \varepsilon$ ist $|z'| < |z|$
  - Wegen $|vwx| \leq l$ kommen in vx nicht alle drei Zeichen a, b, c vor
  - Kommt aber in vx beispielsweise kein a vor, so ist $\#_a(z) = \#_a(z')$ und somit gilt

    $$|z'| < |z| = 3\,\#_a(z) = 3\,\#_a(z')$$

  - Also gehört z' nicht zu L ◁

Abschlusseigenschaften von CFL

Satz

CFL ist abgeschlossen unter Vereinigung, Produkt und Sternhülle

Beweis

- Seien $G_1 = (V_1, \Sigma, P_1, S_1)$ und $G_2 = (V_2, \Sigma, P_2, S_2)$ kontextfreie Grammatiken mit $V_1 \cap V_2 = \varnothing$ und sei S eine neue Variable
- Dann gilt
  - $L(G_1) \cup L(G_2) = L(G_3)$ für die kontextfreie Grammatik
    $G_3 = (V_1 \cup V_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \to S_1, S_2\}, S)$
  - $L(G_1)L(G_2) = L(G_4)$ für die kontextfreie Grammatik
    $G_4 = (V_1 \cup V_2 \cup \{S\}, \Sigma, P_1 \cup P_2 \cup \{S \to S_1 S_2\}, S)$ und
  - $L(G_1)^* = L(G_5)$ für die kontextfreie Grammatik
    $G_5 = (V_1 \cup \{S\}, \Sigma, P_1 \cup \{S \to S_1 S, \varepsilon\}, S)$

□

Für $G_6 = (V_1, \Sigma, P_1 \cup \{S_1 \to S_1 S_1, \varepsilon\}, S_1)$ muss nicht $L(G_6) = L(G_1)^*$ gelten, da $L(G_6)$ im Fall $P_1 = \{S_1 \to aS_1 b, \varepsilon\}$ z.B. das Wort $aababb \notin L(G_1)^*$ enthält

Abschlusseigenschaften von CFL

Satz

CFL ist nicht abgeschlossen unter Schnitt und Komplement

Beweis von $L_1, L_2 \in \text{CFL} \not\Rightarrow L_1 \cap L_2 \in \text{CFL}$

- Folgende Sprachen sind kontextfrei (siehe Übungen):

$$L_1 = \{a^n b^m c^m \mid n, m \geq 0\} \quad \text{und} \quad L_2 = \{a^n b^n c^m \mid n, m \geq 0\}$$

- Nicht jedoch ihr Schnitt $L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ □

Beweis von $L \in \text{CFL} \not\Rightarrow \bar{L} \in \text{CFL}$

- Wäre CFL unter Komplement abgeschlossen, so wäre CFL wegen de Morgan auch unter Schnitt abgeschlossen
- Mit $A, B \in \text{CFL}$ wären dann nämlich auch $\overline{A}, \overline{B} \in \text{CFL}$, woraus wegen

$$\overline{A}, \overline{B} \in \text{CFL} \Rightarrow \overline{A} \cup \overline{B} = \overline{A \cap B} \in \text{CFL}$$

wiederum $A \cap B \in \text{CFL}$ folgen würde □

Umwandlung in Chomsky-Normalform

Satz

Zu jeder kontextfreien Grammatik G lässt sich eine CNF-Grammatik G' mit $L(G') = L(G) \setminus \{\varepsilon\}$ konstruieren

Beweis

Wir wandeln $G = (V, \Sigma, P, S)$ wie folgt in eine CNF-Grammatik G' um:

- Wir beseitigen zunächst alle Regeln der Form $A \to \varepsilon$ und danach alle Regeln der Form $A \to B$ (siehe folgende Folien)
- Dann fügen wir für jedes Terminal $a \in \Sigma$ eine neue Variable X_a und eine neue Regel $X_a \to a$ hinzu und ersetzen jedes Vorkommen von a, bei dem a nicht alleine auf der rechten Seite einer Regel steht, durch X_a
- Anschließend führen wir für jede Regel $A \to B_1 \dots B_k$, $k \geq 3$, neue Variablen $A_1, \dots, A_{k-2}$ ein und ersetzen sie durch die $k-1$ Regeln

$$A \to B_1 A_1,\ A_1 \to B_2 A_2,\ \dots\ A_{k-3} \to B_{k-2} A_{k-2},\ A_{k-2} \to B_{k-1} B_k$$

□

Falls G Regeln mit vielen Variablen auf der rechten Seite hat, empfiehlt es sich, Regeln der Form $A \to \varepsilon$ und $A \to B$ zuletzt zu beseitigen (s. Übungen)

Beseitigung von ε-Regeln

Satz

Zu jeder kontextfreien Grammatik $G = (V, \Sigma, P, S)$ gibt es eine kontextfreie Grammatik $G' = (V, \Sigma, P', S)$ ohne ε-Regeln mit $L(G') = L(G) \setminus \{\varepsilon\}$

Beweis

- Zuerst berechnen wir die Menge $E = \{A \in V \mid A \Rightarrow^* \varepsilon\}$ aller Variablen, die nach ε ableitbar sind:

```
1   E' := {A ∈ V | A → ε}
2   repeat
3     E := E'
4     E' := E ∪ {A ∈ V | ∃ B_1, ..., B_k ∈ E : A → B_1 ... B_k}
5   until E = E'
```

- Nun bilden wir P' wie folgt:

$$\left\{ A \to v' \;\middle|\; \begin{array}{l} \text{es ex. eine Regel } A \to_G v, \text{ so dass } v' \neq \varepsilon \text{ aus } v \text{ durch} \\ \text{Entfernen von beliebig vielen Variablen } A \in E \text{ entsteht} \end{array} \right\}$$

□

Beseitigung von ε-Regeln

Beispiel

Betrachte die Grammatik $G = (\{S, T, U, X, Y, Z\}, \{a, b, c\}, P, S)$ mit

$$\begin{array}{llll} P: & S \to aY, bX, Z & Y \to bS, aYY & T \to U \\ & X \to aS, bXX & Z \to \varepsilon, S, T, cZ & U \to abc \end{array}$$

- Berechnung von E:

E'	$\{Z\}$	$\{Z, S\}$
E	$\{Z, S\}$	$\{Z, S\}$

- Entferne $Z \to \varepsilon$ und füge die Regeln $Y \to b$ (wegen $Y \to bS$), $X \to a$ (wegen $X \to aS$) und $Z \to c$ (wegen $Z \to cZ$) hinzu:

$$\begin{array}{llll} P': & S \to aY, bX, Z & Y \to b, bS, aYY & T \to U \\ & X \to a, aS, bXX & Z \to c, S, T, cZ & U \to abc \end{array}$$

Beseitigung von Variablenumbenennungen

Satz

Zu jeder kontextfreien Grammatik $G = (V, \Sigma, P, S)$ gibt es eine kontextfreie Grammatik $G' = (V, \Sigma, P', S)$ ohne Regeln der Form $A \to B$ mit $L(G') = L(G)$

Beweis

- Zuerst entfernen wir sukzessive alle Zyklen $A_1 \to A_2 \to \cdots \to A_k \to A_1$
- Hierzu entfernen wir diese Regeln aus P und ersetzen alle Vorkommen der Variablen $A_2, \ldots, A_k$ in den übrigen Regeln durch A_1
- Befindet sich die Startvariable unter $A_1, \ldots, A_k$, so sei dies o.B.d.A. A_1
- Nun eliminieren wir sukzessive die restlichen Variablenumbenennungen, indem wir
  - eine Regel $A \to B$ wählen, so dass in P keine Variablenumbenennung $B \to C$ mit B auf der linken Seite existiert,
  - diese Regel $A \to B$ aus P entfernen und
  - für jede Regel $B \to v$ in P die Regel $A \to v$ zu P hinzunehmen □

Beseitigung von Variablenumbenennungen

Beispiel (Fortsetzung)

P: $S \to aY, bX, Z$ $\quad Y \to b, bS, aYY$ $\quad T \to U$

$X \to a, aS, bXX$ $\quad Z \to c, S, T, cZ$ $\quad U \to abc$

- Entferne den Zyklus $S \to Z \to S$ und ersetze Z durch S:

  $S \to aY, bX, c, T, cS$ $\quad Y \to b, bS, aYY$ $\quad T \to U$

  $X \to a, aS, bXX$ $\quad U \to abc$

- Ersetze die Regel $T \to U$ durch $T \to abc$ (wegen $U \to abc$):

  $S \to aY, bX, c, T, cS$ $\quad Y \to b, bS, aYY$ $\quad T \to abc$

  $X \to a, aS, bXX$ $\quad U \to abc$

- Ersetze dann auch die Regel $S \to T$ durch $S \to abc$ (wegen $T \to abc$):

  $S \to abc, aY, bX, c, cS$ $\quad Y \to b, bS, aYY$ $\quad T \to abc$

  $X \to a, aS, bXX$ $\quad U \to abc$

- Da T und U nirgends mehr auf der rechten Seite vorkommen, können wir die Regeln $T \to abc$ und $U \to abc$ weglassen:

  $S \to abc, aY, bX, c, cS$ $\quad Y \to b, bS, aYY$ $\quad X \to a, aS, bXX$

Bringe alle Regeln in die Form $A \to a$ und $A \to BC$

Beispiel (Schluss)

Betrachte die Grammatik $G = (\{S, X, Y, Z\}, \{a, b, c\}, P, S)$ mit

$$P : S \to abc, aY, bX, c, cS \quad Y \to b, bS, aYY \quad X \to a, aS, bXX$$

- Ersetze a, b und c durch A, B und C (außer wenn sie alleine auf der rechten Seite einer Regel stehen) und füge die Regeln $A \to a$, $B \to b$, $C \to c$ hinzu:

  $$S \to ABC, AY, BX, c, CS \quad Y \to b, BS, AYY \quad X \to a, AS, BXX$$
  $$A \to a \quad B \to b \quad C \to c$$

- Ersetze die Regeln $S \to ABC$, $Y \to AYY$ und $X \to BXX$ durch die Regeln $S \to AS'$, $S' \to BC$, $Y \to AY'$, $Y' \to YY$ und $X \to BX'$, $X' \to XX$:

  $$S \to AS', AY, BX, c, CS \quad S' \to BC \quad Y \to b, BS, AY' \quad Y' \to YY$$
  $$X \to a, AS, BX' \quad X' \to XX \quad A \to a \quad B \to b \quad C \to c$$

◁

Links- und Rechtsableitungen

Definition

Sei $G = (V, \Sigma, P, S)$ eine kontextfreie Grammatik

- Eine Ableitung

$$\underline{S} \Rightarrow l_1 \underline{A_1} r_1 \Rightarrow \cdots \Rightarrow l_{m-1} \underline{A_{m-1}} r_{m-1} \Rightarrow \alpha_m$$

  heißt Linksableitung von α_m (kurz $S \Rightarrow_L^* \alpha_m$), falls in jedem Ableitungsschritt die am weitesten links stehende Variable ersetzt wird, d.h. es gilt $l_i \in \Sigma^*$ für $i = 1, \ldots, m-1$

- Rechtsableitungen $S_0 \Rightarrow_R^* \alpha_m$ sind analog definiert
- G heißt mehrdeutig, wenn es ein Wort $x \in L(G)$ gibt, das mindestens zwei verschiedene Linksableitungen hat
- Andernfalls heißt G eindeutig

Für alle $x \in \Sigma^*$ gilt: $x \in L(G) \Leftrightarrow S \Rightarrow^* x \Leftrightarrow S \Rightarrow_L^* x \Leftrightarrow S \Rightarrow_R^* x$

Ein- und mehrdeutige Grammatiken

Beispiel

- In $G = (\{S\}, \{a, b\}, \{S \to aSbS, \varepsilon\}, S)$ gibt es 8 Ableitungen für $aabb$:

$$\underline{S} \Rightarrow_L a\underline{S}bS \Rightarrow_L aa\underline{S}bSbS \Rightarrow_L aabS\underline{b}S \Rightarrow_L aabb\underline{S} \Rightarrow_L aabb$$

$$\begin{aligned}
&\underline{S} \Rightarrow a\underline{S}bS \Rightarrow aa\underline{S}bSbS \Rightarrow aabSb\underline{S} \Rightarrow aab\underline{S}b \Rightarrow aabb\\
&\underline{S} \Rightarrow a\underline{S}bS \Rightarrow aa\underline{S}bSbS \Rightarrow aa\underline{S}bbS \Rightarrow aabb\underline{S} \Rightarrow aabb\\
&\underline{S} \Rightarrow a\underline{S}bS \Rightarrow aa\underline{S}bSbS \Rightarrow aaSbb\underline{S} \Rightarrow aa\underline{S}bb \Rightarrow aabb\\
&\underline{S} \Rightarrow a\underline{S}bS \Rightarrow aaSbS\underline{b}S \Rightarrow aa\underline{S}bS\underline{b} \Rightarrow aab\underline{S}b \Rightarrow aabb\\
&\underline{S} \Rightarrow a\underline{S}bS \Rightarrow aaSbS\underline{b}S \Rightarrow aaSb\underline{S}b \Rightarrow aa\underline{S}bb \Rightarrow aabb\\
&\underline{S} \Rightarrow aSb\underline{S} \Rightarrow a\underline{S}b \Rightarrow aa\underline{S}bSb \Rightarrow aab\underline{S}b \Rightarrow aabb
\end{aligned}$$

$$\underline{S} \Rightarrow_R aSb\underline{S} \Rightarrow_R a\underline{S}b \Rightarrow_R aaSb\underline{S}b \Rightarrow_R aa\underline{S}bb \Rightarrow_R aabb$$

- Darunter sind genau eine Links- und genau eine Rechtsableitung
- In $G' = (\{S\}, \{a, b\}, \{S \to aSbS, ab, \varepsilon\}, S)$ gibt es 3 Ableitungen für ab:

$$\underline{S} \Rightarrow ab \qquad \underline{S} \Rightarrow a\underline{S}bS \Rightarrow ab\underline{S} \Rightarrow ab \qquad \underline{S} \Rightarrow aSb\underline{S} \Rightarrow a\underline{S}b \Rightarrow ab$$

- Darunter sind zwei Links- und zwei Rechtsableitungen ◁