Erzeugte Sprachen, Mehrdeutige Grammatiken und Sprachen, Entfernen von ε-Produktionen

Prof. Dr. David Sabel

LFE Theoretische Informatik

Letzte Änderung der Folien: 26. April 2022

Wiederholung: Die Chomsky-Hierarchie

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

G ist vom Typ 0

G ist automatisch vom Typ 0.

G ist vom Typ 1 (kontextsensitive Grammatik), wenn ...

für alle $(\ell \to r) \in P$: $|\ell| \leq |r|$.

G ist vom Typ 2 (kontextfreie Grammatik), wenn ...

G ist vom Typ 1 und für alle $(\ell \to r) \in P$ gilt: $\ell = A \in V$

G ist vom Typ 3 (reguläre Grammatik), wenn ...

G ist vom Typ 2 und für alle $(A \to r) \in P$ gilt: $r = a$ oder $r = aA'$ für $a \in \Sigma, A' \in V$
(die rechten Seiten sind Worte aus $(\Sigma \cup (\Sigma V))$)

Beispiel (kontextsensitive Grammatik)

$G = (\{S, B, C\}, \{a, b, c\}, P, S)$ mit

$$P = \{S \to aSBC,\ S \to aBC,\ CB \to BC,\ aB \to ab,\ bB \to bb,\ bC \to bc,\ cC \to cc\}$$

Beispiel-Ableitung:

$$\begin{aligned} S &\Rightarrow aSBC \Rightarrow aaSBCBC \Rightarrow aaaSBCBCBC \Rightarrow aaaaBCBCBCBC \\ &\Rightarrow aaaabCBCBCBC \Rightarrow aaaabBCCBCBC \Rightarrow aaaabbCCBCBC \\ &\Rightarrow aaaabbCBCCBC \Rightarrow aaaabbBCCCBC \Rightarrow aaaabbBCCBCC \\ &\Rightarrow aaaabbBCBCCC \Rightarrow aaaabbBBCCCC \Rightarrow aaaabbbBCCCC \\ &\Rightarrow aaaabbbbCCCC \Rightarrow aaaabbbbcCCC \Rightarrow aaaabbbbccCC \\ &\Rightarrow aaaabbbbccCC \Rightarrow aaaabbbbcccc \end{aligned}$$

Steckengebliebene Folge von Ableitungsschritten:

$$S \Rightarrow aSBC \Rightarrow aaBCBC \Rightarrow aabCBC \Rightarrow aabcBC$$

Grammatik erzeugt $\{a^n b^n c^n \mid n \in \mathbb{N}_{>0}\}$

Satz

$L(G) = \{a^n b^n c^n \mid n \in \mathbb{N}_{>0}\}$ für $G = (\{S, B, C\}, \{a, b, c\}, P, S)$ mit
$P = \{S \to aSBC, S \to aBC, CB \to BC, aB \to ab, bB \to bb, bC \to bc, cC \to cc\}$

„$\supseteq$“: Zeige $a^n b^n c^n \in L(G)$ für alle $n \in \mathbb{N}_{>0}$

- Wende $n-1$ mal $S \to aSBC$ und dann einmal $S \to aBC$ an:
  $S \Rightarrow^* a^{n-1} S (BC)^{n-1} \Rightarrow a^n (BC)^n$
- Wende $CB \to BC$ solange an, bis es kein Teilwort CB mehr gibt.
  $a^n (BC)^n \Rightarrow^* a^n B^n C^n$
- Wende $aB \to ab$ und anschließend $n-1$ mal $bB \to bb$ an.
  $a^n B^n C^n \Rightarrow a^n b B^{n-1} C^n \Rightarrow^* a^n b^n C^n$
- Wende einmal $bC \to bc$ und anschließend $n-1$ mal $cC \to cc$ an
  $a^n b^n C^n \Rightarrow a^n b^n c C^{n-1} \Rightarrow^* a^n b^n c^n$

Zusammensetzen aller Ableitungsschritte zeigt $S \Rightarrow^* a^n b^n c^n$.

Grammatik erzeugt $\{a^n b^n c^n \mid n \in \mathbb{N}_{>0}\}$ (2)

Satz

$L(G) = \{a^n b^n c^n \mid n \in \mathbb{N}_{>0}\}$ für $G = (\{S, B, C\}, \{a, b, c\}, P, S)$ mit
$P = \{S \to aSBC, S \to aBC, CB \to BC, aB \to ab, bB \to bb, bC \to bc, cC \to cc\}$

„$\subseteq$“: Zeige, dass alle von G erzeugten Worte von der Form $a^n b^n c^n$ sind.

- Für $S \Rightarrow_G^* u$ mit u Satzform zeigen die Regeln:
  $\#_a(u) = \#_b(u) + \#_B(u) = \#_c(u) + \#_C(u)$
- Für $S \Rightarrow_G^* w$ mit $w \in \{a, b, c\}^*$ gilt: a's werden ganz links erzeugt, d.h. $w = a^n w'$ mit $w' \in \{b, c\}^*$ und $n = \#_b(w') = \#_c(w')$
- Es gilt $w' = bw_1$, da jedes auf a folgende Symbol durch $aB \to ab$ erzeugt wird und die Regeln keine Terminalsymbole vertauschen.

Grammatik, die $\{a^n b^n c^n \mid n \in \mathbb{N}_{>0}\}$ erzeugt (3)

Satz

$L(G) = \{a^n b^n c^n \mid n \in \mathbb{N}_{>0}\}$ für $G = (\{S, B, C\}, \{a, b, c\}, P, S)$ mit
$P = \{S \to aSBC, S \to aBC, CB \to BC, aB \to ab, bB \to bb, bC \to bc, cC \to cc\}$

„$\subseteq$“: Zeige, dass alle von G erzeugten Worte von der Form $a^n b^n c^n$ sind.

- ...
- Ebenso können die Terminalsymbole des Wortes $w' \in \{b, c\}^*$ nur durch $bB \to bb$, $bC \to bc$ und $cC \to cc$ erzeugt worden sein.
  Diese Produktionen erlauben nur einen Wechsel von b zu c und keine Wechsel von c zu b. Auch ein Umordnen der Terminalsymbole ist nicht möglich
  (da es keine Produktion dafür gibt).
- Daher gilt $w' = b^i c^j$ und mit $n = \#_b(w') = \#_c(w')$ sogar $w' = b^n c^n$. □

Beispiel einer Typ 0-Grammatik

Grammatik $G = (\{S, T, A, B, \$\}, \{a, b\}, P, S)$ mit

$$P = \{S \to \$T\$,\ T \to aAT,\ T \to bBT,\ T \to \varepsilon,\ \$a \to a\$,\ \$b \to b\$,\ Aa \to aA,\\ Ab \to bA,\ Ba \to aB,\ Bb \to bB,\ A\$ \to \$a,\ B\$ \to \$b,\ \$\$ \to \varepsilon\}$$

Eine Ableitung:

$$\begin{aligned} S &\Rightarrow \$T\$ \Rightarrow \$aAT\$ \Rightarrow \$aAaAT\$ \Rightarrow \$aAaAbBT\$\\ &\Rightarrow \$aAaAbB\$ \Rightarrow \$aaAAbB\$ \Rightarrow \$aaAbAB\$ \Rightarrow \$aabAAB\$\\ &\Rightarrow \$aabAA\$b \Rightarrow \$aabA\$ab \Rightarrow \$aab\$aab \Rightarrow a\$ab\$aab\\ &\Rightarrow aa\$b\$aab \Rightarrow aab\$\$aab \Rightarrow aabaab \end{aligned}$$

Beachte: $L(G) = \{ww \mid w \in \{a, b\}^*\}$ und $L(G)$ ist Typ 1-Sprache

Beispiel einer Typ 0-Grammatik (2)

Grammatik $G = (\{S, T, A, B, \$\}, \{a, b\}, P, S)$ mit

$$P = \{S \to \$T\$,\ T \to aAT,\ T \to bBT,\ T \to \varepsilon,\ \$a \to a\$,\ \$b \to b\$,\ Aa \to aA,\\ Ab \to bA,\ Ba \to aB,\ Bb \to bB,\ A\$ \to \$a,\ B\$ \to \$b,\ \$\$ \to \varepsilon\}$$

Begründung dafür, dass $L(G) = \{ww \mid w \in \{a, b\}^*\}$ gilt:

- Mit $S \to \$T\$$ wird zunächst eine Umrahmung mit \$\$ erzeugt
- Mit $T \to aAT$, $T \to bBT$, $T \to \varepsilon$ wird ein Wort aus 2er Blöcken aA und bB erzeugt
- Mit $Aa \to aA, Ab \to bA, Ba \to aB, Bb \to bB$ werden A's und B's bis vor \$ geschoben
- Mit $A\$ \to \$a$ und $B\$ \to \$b$ werden die A's und B's in a's und b's verwandelt, indem sie über das rechte \$ hüpfen.
- Mit $\$a \to a\$$, $\$b \to b\$$ wird das linke \$ zum rechten geschoben, mit $\$\$ \to \varepsilon$ werden sie dann eliminiert.
- Bei allen Schritten wird die relative Lage aller a zu b sowie aller A zu B nicht geändert.

Mehrdeutige Grammatiken

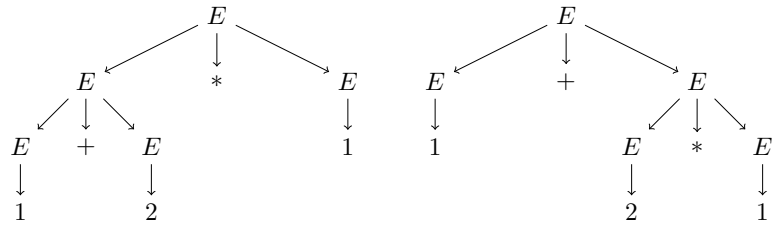
Beispiel:

$$(E, \{*, +, 1, 2\}, \{E \to E * E,\ \ E \to E + E,\ \ E \to 1,\ \ E \to 2\}, E)$$

Zwei Ableitungen für $1 + 2 * 1$:

- $E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow 1 + E * E \Rightarrow 1 + 2 * E \Rightarrow 1 + 2 * 1$
- $E \Rightarrow E + E \Rightarrow E + E * E \Rightarrow 1 + E * E \Rightarrow 1 + 2 * E \Rightarrow 1 + 2 * 1.$

Syntaxbäume dazu:



Mehrdeutige Grammatiken (2)

Mehrdeutige Grammatik

Eine Typ 2-Grammatik ist mehrdeutig, wenn es verschieden strukturierte Syntaxbäume für dasselbe Wort w gibt.

Inhärent mehrdeutige Sprache

Eine Typ 2-Sprache ist inhärent mehrdeutig, wenn es nur mehrdeutige Grammatiken gibt, die diese Sprache erzeugen.

Die Sprache

$$\{a^m b^m c^n d^n \mid m, n \in \mathbb{N}_{>0}\} \cup \{a^m b^n c^n d^m \mid m, n \in \mathbb{N}_{>0}\}$$

ist inhärent mehrdeutig (Beweis z.B. in Hopcroft, Motwani, Ullman, 2006)

ε-Regel für Typ 1,2,3-Grammatiken

- Das leere Wort ε kann bisher nicht für Typ 1,2,3 Grammatiken erzeugt werden:

  Produktion $S \to \varepsilon$ erfüllt die Typ 1-Bedingung $|S| \leq |\varepsilon|$ nicht. Daher Sonderregel:

ε-Regel für Typ 1-Grammatiken

Eine Grammatik $G = (V, \Sigma, P, S)$ vom Typ 1 darf eine Produktion $(S \to \varepsilon) \in P$ enthalten, vorausgesetzt, dass keine rechte Seite einer Produktion in P, die Variable S enthält.

Sonderregel erlaubt nicht:

$G = (\{S\}, \{a\}, \{S \to \varepsilon, S \to aSa\}, S)$

Sonderregel erlaubt:

$G = (\{S', S\}, \{a\}, \{S' \to \varepsilon, S' \to aSa, S' \to aa, S \to aSa, S \to aa\}, S')$

Leeres Wort hinzufügen geht mit Sonderregel immer

Satz

Sei $G = (V, \Sigma, P, S)$ vom Typ $i \in \{1, 2, 3\}$ mit $\varepsilon \notin L(G)$. Sei $S' \notin V$.
Dann erzeugt $G' = (V \cup \{S'\}, \Sigma, P \cup \{S' \to \varepsilon\} \cup \{S' \to r \mid S \to r \in P\}, S')$ die Sprache $L(G') = L(G) \cup \{\varepsilon\}$ und
G' erfüllt die ε-Regel für Typ 1,2,3-Grammatiken und G' ist vom Typ i.

Beweis:

- Da S' neu, kommt S' auf keiner rechten Seite vor.
- Da $S \to r \in P$ vom Typ i, sind auch $S' \to r$ vom Typ i
- Da $S' \Rightarrow \varepsilon$, gilt $\varepsilon \in L(G')$
- Für $w \neq \varepsilon$ gilt: $S \Rightarrow_G^* w$ g.d.w. $S' \Rightarrow_{G'}^* w$
  Der jeweils erste Ableitungsschritt muss ausgetauscht werden, d.h. $S \Rightarrow_G r$ vs. $S' \Rightarrow_{G'} r$

ε-Produktionen für Typ 2- und Typ 3-Grammatiken

Sonderregel für Typ 2- und Typ 3-Grammatiken:

ε-Produktionen in kontextfreien und regulären Grammatiken

In Grammatiken des Typs 2 und des Typs 3 erlauben wir Produktionen der Form $A \to \varepsilon$ (sogenannte ε-Produktionen).

Das ist keine echte Erweiterung, denn:

Satz (Entfernen von ε-Produktionen)

Sei $G = (V, \Sigma, P, S)$ eine kontextfreie (bzw. reguläre) Grammatik mit $\varepsilon \notin L(G)$. Dann gibt es eine kontextfreie (bzw. reguläre) Grammatik G' mit $L(G) = L(G')$ und G' enthält keine ε-Produktionen.

Beweis: Algorithmus auf der nächsten Folie.

Algorithmus 1: Entfernen von ε-Produktionen

Eingabe: Typ i-Grammatik $G = (V, \Sigma, P, S)$ mit $\varepsilon \notin L(G)$, $i \in \{2, 3\}$
Ausgabe: Typ i-Grammatik G' ohne ε-Produktionen, sodass $L(G) = L(G')$
Beginn
  finde die Menge $W \subseteq V$ aller Variablen A für die gilt $A \Rightarrow^* \varepsilon$:
  Beginn
    $W := \{A \mid (A \to \varepsilon) \in P\}$;
    wiederhole
      füge alle A zu W hinzu mit $A \to A_1 \dots A_n \in P$ und $A_1, \dots, A_n \in W$;
    bis *sich W nicht mehr ändert*;
  Ende
  $P' := P \setminus \{A \to \varepsilon \mid (A \to \varepsilon) \in P\}$; /* lösche Regeln $A \to \varepsilon$ */
  wiederhole
    für alle *Produktionen $A' \to uAv \in P'$ mit $|uv| > 0$ und $A \in W$* **tue**
      füge die Produktion $A' \to uv$ zu P' hinzu;
      /* für $A' \to u'Av'Aw'$ gibt es (mindestens) zwei Hinzufügungen: Für das Vorkommen von A nach u' als auch für das Vorkommen direkt vor w' */
    Ende
  bis *sich P' nicht mehr ändert*;
  gebe $G' = (V, \Sigma, P', S)$ als Ergebnisgrammatik aus;
Ende

Die neuen Produktionen nehmen den Ableitungsschritt $A \to \varepsilon$ vorweg.

Für reguläre Produktion $A' \to aA$ wird $A' \to a$ hinzugefügt (bleibt regulär!)

Beispiel: Entfernen von ε-Produktionen

$G = (\{A, B, C, D, S\}, \{0, 1\}, P, S)$ mit

$$P = \{S \to 1A,\ A \to AB,\ A \to DA,\ A \to \varepsilon,\ B \to 0,\\ B \to 1,\ C \to AAA,\ D \to 1AC\}.$$

- Menge W der Variablen, die ε herleiten:

  $W = \{A, C\}$ da $A \to \varepsilon$ und $C \to AAA$
- Starte mit

  $P' = \{S \to 1A,\ A \to AB,\ A \to DA,\ B \to 0,\\ B \to 1,\ C \to AAA,\ D \to 1AC\}.$
- Hinzufügen von Produktionen für Vorkommen von A und C

  $P' = \{S \to 1A,\ S \to 1,\ A \to AB,\ A \to B,\ A \to DA,\ A \to D,\\ B \to 0,\ B \to 1,\ C \to AAA,\ C \to AA,\ C \to A,\\ D \to 1AC,\ D \to 1A,\ D \to 1C,\ D \to 1\}.$