

1.2.1. Reduktive und generative Grammatik

Eine Grammatik über einer endl. Zeichenmenge  $M$  ist ein Tripel  $(M, \rightarrow, z)$  wobei  $\rightarrow \subseteq M^* \times M^*$ , endlich  
 $z \in M$  heißt Axiom (ohne Wurzel)  
 $G = (M, \rightarrow, z)$  heißt auch Semi-Thue-Grammatik.

Grammatiken beschreiben formale Sprachen.

Generativ: Ein Wort  $w \in M^*$  ist in der durch  $G$  beschriebenen formalen Sprache, wenn gilt  $\langle z \rangle \xrightarrow{*} w$

Generativer Sprachschatz  $Lg(G) = \{w \in M^* : \langle z \rangle \xrightarrow{*} w\}$

Reduktiv: Ein Wort  $w \in M^*$  ist in der durch  $G$  beschriebenen formalen Sprache, wenn gilt:  $w \xrightarrow{*} \langle z \rangle$

Reduktiver Sprachschatz:  $Lr(G) = \{w \in M^* : w \xrightarrow{*} \langle z \rangle\}$

Sprechweise  $w \in Lg(G)$ :  $w$  wird durch  $G$  generiert (erzeugt)

$w \in Lr(G)$ :  $w$  wird durch  $G$  erkannt

Eine generative Grammatik heißt  $\epsilon$ -frei oder  $\epsilon$ -produktionsfrei, wenn die rechte Seite jeder Regel  $\neq \epsilon$ .

Eine reduktive Grammatik heißt  $\epsilon$ -frei oder  $\epsilon$ -produktionsfrei, wenn die linke Seite jeder Regel  $\neq \epsilon$ .

Semi-Thue-Grammatiken (S-T-G) haben enge Grenzen in der Ausdrucksmächtigkeit:

Beispiel:  $M = \{L\}$

Sprache  $S \subseteq M, S = \{L^{2^k} : k \in \mathbb{N}\}$

Behauptung: Diese Sprache ist durch S-T-G nicht beschreibbar

Beweis:

jede Regel hat die Form  $L^i \rightarrow L^j$

wir betrachten den Reduktiven Fall.

Wir benötigen mindestens eine Regel der Form  $L^i \rightarrow L$  mit  $i > 1$

$L$  muss Axiom sein

$L^i \xrightarrow{L^i \rightarrow L} L^{i-1} \rightarrow \dots \rightarrow L$

$\frac{L^i \rightarrow L}{S}$

$L^{i-1} \in S$  nicht in der Menge aber auf  $L$  ableitbar  $\rightarrow$  Ableitung

Chomsky-Grammatiken strukturieren den Zeichensatz  $M = T \cup N$  mit

$T \cap N = \emptyset$ .

$T$  Terminalzeichen - Zeichen in Wörtern d. formalen Sprache

$N$  Nichtterminalzeichen - Hilfszeichen für die Ableitung

Chomsky-Grammatik  $G = (T, N, \rightarrow, z)$

mit  $(T \cup N, \rightarrow, z)$  Semi-Thue-Grammatik

$z \in N$

$T \cap N = \emptyset$

$Lg(G) = Lg(T \cup N, \rightarrow, z) \cap T^*$

$Lr(G) = Lr(T \cup N, \rightarrow, z) \cap T^*$

Beispiel: Reduktive Chomsky-Grammatik zur Beschreibung der Sprache

$S = \{L^{2^k} : k \in \mathbb{N}\}$

$T = \{L\}$   $N = \{z, A, B, C\}$   $z$  Axiom

Regeln:  $L \rightarrow A$

$\epsilon \rightarrow B$

$AAB \rightarrow CB$

$AAC \rightarrow CA$

$BC \rightarrow BA$

$BAB \rightarrow z$

Wir betrachten den Sprachschatz der STG, die wir erhalten, indem wir die ersten zwei Regeln weglassen.

$\langle BAB \rangle$  ist im Sprachschatz



- (1) jedes Wort im Sprachschatz hat die Form  $\langle B \rangle^0 x \langle B \rangle$ ,  $x \in \{AC\}^*$   
 (2) In rückwärtsableitung entstehen  $C$  immer am linken Rand, um die  $C$  wieder zu entfernen, müssen sie nach rechts geschoben werden und sie verdoppeln dabei die Anzahl der  $A$ .

Ans (1) und (2) folgt:  
 d.h. die Wörter aus  $\{B, A\}^*$  im Sprachschatz haben die Form  $BA^{2^e}B$ .



Zwei reduktive (bzw. generative) Grammatiken heißen äquivalent, wenn sie die gleiche formale Sprache beschreiben.

Eine reduktive Grammatik heißt wortlängenmonoton, wenn für jede ihrer Regeln gilt:  $\underbrace{w \rightarrow w'}_{\text{ist Regel}} \Rightarrow |w| = |w'|$  Wörter nach Aufkantung d. Regel gleich lang

Gilt  $|w| > |w'|$  so heißt die Grammatik strikt wortlängenmonoton.  
 konsequent für Abb.:  $w_0 \rightarrow w_1 \rightarrow w_2 \rightarrow \dots \rightarrow w_n$

$$|w_0| \geq |w_1| \geq |w_2| \geq \dots \geq |w_n|$$

endlich!  $\rightarrow$  man kann einen Algorithmus finden

Eine Ersetzungsregel  $w \rightarrow w'$  heißt separiert, wenn  $w' \in N^+$

### 1.2.2. Die Sprachhierarchie nach Chomsky

Ch-G lassen sich nach der äußeren Form ihrer Regel klassifizieren.

Chomsky-0-Sprachen sind formale Sprachen die durch Chomsky Grammatiken beschrieben werden können.

Wir beschränken uns im Folgenden auf reduktive Grammatik (generativ analog):

Eine Ersetzungsregel  $u \alpha v \rightarrow u \alpha' v$   
 mit  $u, v \in (T \cup N)^*$ ,  $\alpha \in (T \cup N)^+$ ,  $\alpha' \in N$   
 heißt kontextsensitiv.

$a \rightarrow cb$  gilt aber nur im Kontext!

d.h. die Regel  $a \rightarrow cb$  darf nur im Kontext  $u \alpha v$  angewandt werden.

Eine Ch-G heißt kontextsensitiv oder Chomsky-1-Grammatik, wenn alle ihre Regeln ( $\epsilon$ -produktionsfrei) kontextsensitiv sind.

Trivialerweise gilt für Ch-1-G: Alle Regeln sind wortlängenmonoton.

Eine formale Sprache heißt kontextsensitiv oder Ch-1-Sprache, wenn es eine Ch-1-G gibt, die die Sprache beschreibt.

Ist eine Sprache  $S$  kontextsensitiv, dann wird  $S \cup \{c\}$  auch kontextsensitiv genannt.

Bsp.  $(a \neq a) \rightarrow (a)$  kontextsensitiv

Eine Regel  $a \rightarrow cb$  mit  $a \in (N \cup T)^*$ ,  $b \in N$  heißt Kontextfrei.

Sind alle Regeln einer Grammatik Kontextfrei, so heißt die Grammatik Chomsky-2-Grammatik oder Kontextfrei.  
 Jede Sprache, die durch eine Ch-2-G beschreibbar ist, heißt Kontextfrei.  
 Hinweis: BNF entspricht Ch-2.

$$\langle exp \rangle ::= 0 \mid 1 \mid \langle exp \rangle + \langle exp \rangle$$

$T = \{1, 0, +\}$        $0 \rightarrow E$   
 $N = \{E\}$              $1 \rightarrow E$   
 $E$  Wurzel             $E \rightarrow E$

Eine Kontextfreie Regel der Form  $m \langle a \rangle^n \rightarrow cb$  mit  $a, b \in N$ ;  $m, n \in T$  heißt beidseitig linear.

Gilt jedoch  $n = \epsilon$ , so heißt die Regel linkslinear. } einseitig linear.  
 Gilt  $m = \epsilon$ , so heißt die Regel rechtslinear.

Eine Regel der Form  $w \rightarrow cb$  mit  $w \in T^+$ ,  $b \in N$  heißt terminal.

Eine Ch-2-G, deren sämtliche Regeln terminal, deren sämtliche Regeln linkslinear oder rechtslinear (bzw. terminal oder linkslinear) sind, heißt Ch-3-Grammatik oder regulär.

Eine Sprache heißt regulär, wenn sie durch eine reguläre Grammatik beschreibbar ist (oder Ch-3-Sprache).

COL	Menge der Chomsky-0-Sprachen	
CFL	_____	1
CSL	_____	2
REG	_____	3

Satz:  $COL \supsetneq CSL \supsetneq CFL \supsetneq REG$   
Altkontextfrei      Kontextfrei      Kontextfrei      reguläre

### 1.2.3. Strukturelle Äquivalenz von Ableitungen

Ableitung über einer Grammatik  $w_0 \Rightarrow w_1 \Rightarrow \dots \Rightarrow w_n$

Fragen von Interesse:

- (1) Was ist ableitbar?  $\xrightarrow{*}$
- (2) welche Struktur hat eine Ableitung?

$\langle BAAAAB \rangle \rightarrow \langle BAACB \rangle \rightarrow \langle BCAB \rangle \rightarrow \langle BAAB \rangle \rightarrow \langle BCB \rangle \rightarrow \langle BAB \rangle$

Bei längerem Wort  $\langle B \dots \textcircled{C} \dots AAB \rangle$  zwei Regeln anwendbar zum gl. Zeitpunkt  
 → strukturelle Äquivalenz