Allgemeinere Grammatiken. Die Chomsky-Hierarchie

Regeln der Form $\alpha \to \beta$

mit $\alpha \in (\Sigma \cup V)^* - \Sigma^*$, $\beta \in (\Sigma \cup V)^*$

z.B. $01AB100 \to 11C$

Typ 0: keine weiteren Einschränkungen

Typ 1: $|\alpha| \le |\beta|$* <u>kontextsensitive</u> Grammatik

*Ausnahme: Die Regel $S \to \varepsilon$ ist erlaubt, aber dann darf S nie in β vorkommen.

Typ 2: $\alpha \in V$ <u>kontextfreie</u> Grammatik

Typ 3: $\alpha \in V$, $\beta \in \Sigma V \cup \Sigma \cup \{\varepsilon\}$

<u>rechtslineare</u> Grammatik

Typ-0/1/2/3-<u>Sprachen</u>: Sprachen, die durch eine Typ 0/1/2/3-Grammatik erzeugt werden.

Die Typ-3-Sprachen sind die regulären Sprachen. (Übung)

Die Typ-0-Sprachen sind die rekursiv aufzählbaren Sprachen.

Noam Chomsky, geb. 1928, amerikanischer Linguist und politischer Aktivist

<u>Chomsky-Hierarchie</u>: Typ-3-Spr. $\subset$ Typ-2-Spr $\subset$ Typ-1-Spr $\subset$ Typ-0-Spr.

kontextsensitive Grammatiken

Beispiel $\{0^n 1^n 0^n \mid n \geq 0\}$ $\qquad \Sigma = \{0, 1\}$

$$S \to \varepsilon \mid A10$$
$$A \to 0$$
$$A1 \to 011B$$
$$B1 \to 1B$$
$$B0 \to C00$$
$$1C \to C1$$
$$0C \to 0A$$

$$S \Rightarrow A10 \Rightarrow 011B0 \Rightarrow 011C00 \overset{*}{\Rightarrow} \underbrace{0C}_{0A}1100$$
$$\overset{*}{\Rightarrow} 000A\,1111\,0000$$
$$\Rightarrow 0000\,11B\,111\,0000$$
$$\Rightarrow \ldots\ 1B11 \ldots$$
$$\Rightarrow \ldots\ 1B1 \ldots$$
$$\Rightarrow 0000\,1111\,1B\,0000$$
$$\Rightarrow \ldots\ldots\ 1C00000$$
$$C1$$
$$C$$
$$\overset{*}{\Rightarrow} 0000\,C11111\,00000$$
$$\Rightarrow 0000A\,11111\,00000$$

SATZ: Kontextsensitive Sprachen sind entscheidbar.

Eingabe: Typ-1-Grammatik G, $w \in \Sigma^*$

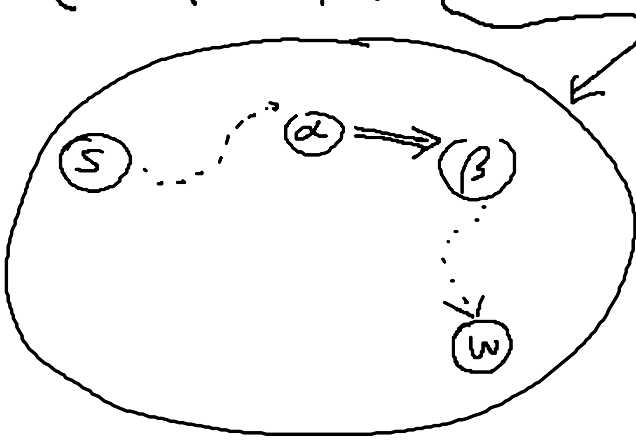
Frage: Ist $w \in L(G)$? (das Wortproblem)

Beweis: Wegen $|\alpha| \leq |\beta|$ können in einer Ableitung von w nur Satzformen der Länge $\leq n := |w|$ vorkommen.

(Ausnahme $w = \varepsilon$)

Das sind nur endlich viele Satzformen.

$$\left(s := |\Sigma \cup V| \ ; \ \underbrace{s + s^2 + s^3 + \ldots + s^n} \right)$$



Ist die Satzform w von S über eine Folge von Ableitungsschritten ($\Rightarrow$) erreichbar? $\square$

- Nicht jede entscheidbare Sprache ist kontextsensitiv.

SATZ:

Typ-0-Sprachen $\equiv$ rekursiv aufzählbare Sprachen

Beweis: „$\Rightarrow$": Typ-0-Grammatik G gegeben.

Algorithmus für $L(G)$:

Eingabe $w \in \Sigma^*$

Für $n = 0, 1, 2, \ldots$

Probiere alle Ableitungen in G mit n Schritten

Wenn eine Ableitung zu w führt:

AKZEPTIERE w

„$\Leftarrow$": (Skizze)

Turingmaschine M gegeben

Annahme: M akzeptiert nur mit leerem Band

IDEE: Die Ableitung der Grammatik erzeugt die Konfigurationen einer akzeptierenden Berechnung in <u>umgekehrter Reihenfolge</u>

Zustände von M $\leftrightarrow$ Variablen von G

δ	1	B
q_0	$(q_1, 1, R)$	(q_1, B, R)
q_1	(q_0, B, L)	$(q_2, 1, R)$
q_2	(q_1, B, R)	$(q_3, 1, L)$
q_3	$(q_3, 1, N)$	(q_3, B, N)

$\Sigma = \{1\}$

$Q_0 1B \to 1 Q_1 1$

$Q_0 BB \to B Q_1 1$

$1 Q_2 \to Q_1 B$

$AB \to A$

$A \to AB$

$BZ \to Z$

$Z \to BZ$

$S \to A Q_3 Z$

$AQ_0 \to Y \qquad Y1 \to 1Y \qquad YZ \to \varepsilon$

$S \Rightarrow A Q_3 Z \stackrel{*}{\Rightarrow}$

$A 1 \underline{1 Q_2} 1 1 Z$

$\Rightarrow A 1 \underline{Q_1 B} 1 1 Z \quad [\vdash A 1 1 Q_2 1 1 Z]$

Randbegrenzungen

$\stackrel{*}{\Rightarrow} A Q_0 1111 Z \Rightarrow 1111 \in L(G)$