Theoretische Informatik I
Vorlesung vom 07.07.00

<u>Fortsetzung des Beweises von Satz Z:</u>
"$\Leftarrow$": Sei $G=(N,T,P,\sigma)$ eine rechtslineare Grammatik. Wir konstruieren einen nichtdeterministischen endlicher Akzeptor A mit $L(G)=L(A)$.

Sei $A=(X, S, \delta, S_0, F)$ definiert durch
$S:= N\cup\{\alpha\}$, $\alpha\notin N$
$S_0:=\{\sigma\}$

$$F = \begin{cases} \{\sigma,\alpha\}, \text{falls} \sigma \to \varepsilon \in P \\ \{\alpha\}, \text{falls} \sigma \to \varepsilon \notin P \end{cases}$$

$B\in\delta(A,x)$, falls $A\to xB \in P$
$\alpha\in\delta(A,x)$, falls $A\to x \in P$

Nun gilt für $n\geq1$:
$w=x_1...x_n\in L(G)$ mit $x_i\in X \Leftrightarrow$ Es gibt eine Folge von Nonterminals $A_1, ... , A_{n-1}$ mit:
$\sigma\to x_1A_1\to x_1x_2A_2\to...\to x_1...x_{n-1}A_{n-1}\to x_1...x_n=w$
$\Leftrightarrow$ Es gibt eine Folge von Zuständen $A_1, ..., A_{n-1}$ mit:
$A_1\in\delta(\sigma,x_1)$, $A_2\in\delta(A_1,x_2)$, ..., $\alpha\in\delta(A_{n-1},x_n) \Leftrightarrow w\in L(A)$.

<u>5. Formale Sprachen und Grammatiken</u>

- Typisierung von Grammatiken und Sprachen
- Allgemeine Aussagen über Grammatiken und Darstellungen

<u>Def A:</u> (Typisierung von Grammatiken)
Sei $G=(N,T,P,\sigma)$ eine Grammatik

a) G heißt vom <u>Typ 3</u> oder <u>rechtslinear</u> oder <u>regulär,</u> wenn für alle $(\alpha,\beta)\in P$ gilt: $\alpha\in N$, $\beta\in T^*\bullet(\{\varepsilon\}\cup N)$.
b) G heißt vom <u>Typ 2</u> oder <u>kontextfrei</u>, wenn für alle $(\alpha,\beta)\in P$ gilt: $\alpha\in N$
c) G heißt vom <u>Typ 1</u> oder <u>kontextsensitiv</u>, wenn für alle $(\alpha,\beta)\in P$ gilt: $\alpha=\alpha_1B\alpha_2$, $\beta=\alpha_1\overline{\beta}\alpha_2$ für ein $B\in N$ und $\alpha_1, \alpha_2, \overline{\beta} \in (N\cup T)^*$. [Nonterminals dürfen nur ersetzt werden, wenn der Kontext (α_1,α_2) stimmt. Bei Typ 2 und Typ 3 ist der Kontext egal.]
d) G heißt vom <u>Typ 0</u>, wenn G Grammatik ist.
e) G heißt vom <u>Erweiterungstyp</u>, wenn für alle $(\alpha,\beta)\in P$ gilt: $|\alpha| \leq |\beta|$
f) G heißt <u>lineare</u> Grammatik, wenn G kontextfrei ist und für alle $(\alpha,\beta)\in P$ gilt: $\beta\in T^*\bullet N\bullet T^* \cup T^*$

<u>Def B:</u> a) eine Sprache $L\subseteq T^*$ heißt <u>Sprache vom Typ i</u>, i=0,1,2,3, wenn es eine Grammatik G gibt vom Typ i mit $L=L(G)$.
b) Klassifikation nach Noam Chomsky (*1928)
$CF:=\{L\subseteq T^* | L \text{ ist vom Typ } 2\}$
$CS:=\{L\subseteq T^* | L \text{ ist vom Typ } 1\}$
$Ch\text{-}0:=\{L\subseteq T^* | L \text{ ist vom Typ } 0\}$
$Lin:=\{L\subseteq T^* | \text{zu } L \text{ gibt es eine lineare Grammatik mit } L=L(G)\}$

<u>Bem:</u> Offenbar gilt: $\begin{matrix} Reg \subseteq \\ Lin \subseteq \end{matrix} CF \subseteq CS \subseteq Ch\text{-}0$ [Chomsky Hierarchie]

Ziele: Zeige jeweils $\subsetneqq$ und charakterisiere die jeweiligen Sprachklassen

Zwei Grammatiken G, G' heißen Äquivalent, wenn L(G)=L(G').

<u>5.1 Kontextfreie Sprachen</u>

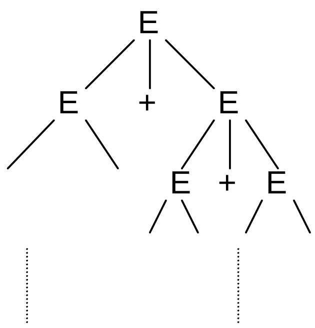
<u>Beispiel:</u> 1) Sei $G=(N,T,P,\sigma)$ mit $T=\{a,b\}$, $N=\{\sigma\}$, $P=\{\sigma\to a\sigma b, \sigma\to\varepsilon\}$.
G ist kontextfrei (sogar linear).
$L(G)=\{a^nb^n | n\geq0\}$. $L(G)\notin Reg \Rightarrow Reg\neq CF$

also gilt: Reg $\subsetneqq$ Lin $\subsetneqq$ CF

2) Sei $G'=(N',T',P',\sigma')$ mit $T'=\{(,),v,c,+,\cdot\}$, $N'=(E,T,F)$
E:expression
T:term
F:factor
$\sigma'=E$
$P'=\{E\rightarrow E+E,\ E\rightarrow T,\ T\rightarrow T\cdot T,\ T\rightarrow F,\ F\rightarrow (E),\ F\rightarrow v,\ F\rightarrow c\}$
L(G')={arithmetische Ausdrücke mit korrekter Klammerung mit Operatoren +,· und Operanden v,c}

Beispiel: $E\rightarrow E+E\rightarrow E+E+E\rightarrow E+T\cdot T+E\rightarrow E+T\cdot T\cdot T+E\rightarrow {}^{*}E+F\cdot F\cdot F+E\rightarrow {}^{*}v+v\cdot c\cdot v+v$

Ableitungsbäume: Graphische Darstellung von Ableitungen bei kontextfreien Grammatiken durch Bäume:



T∍Blätter: v + v · c · v + v

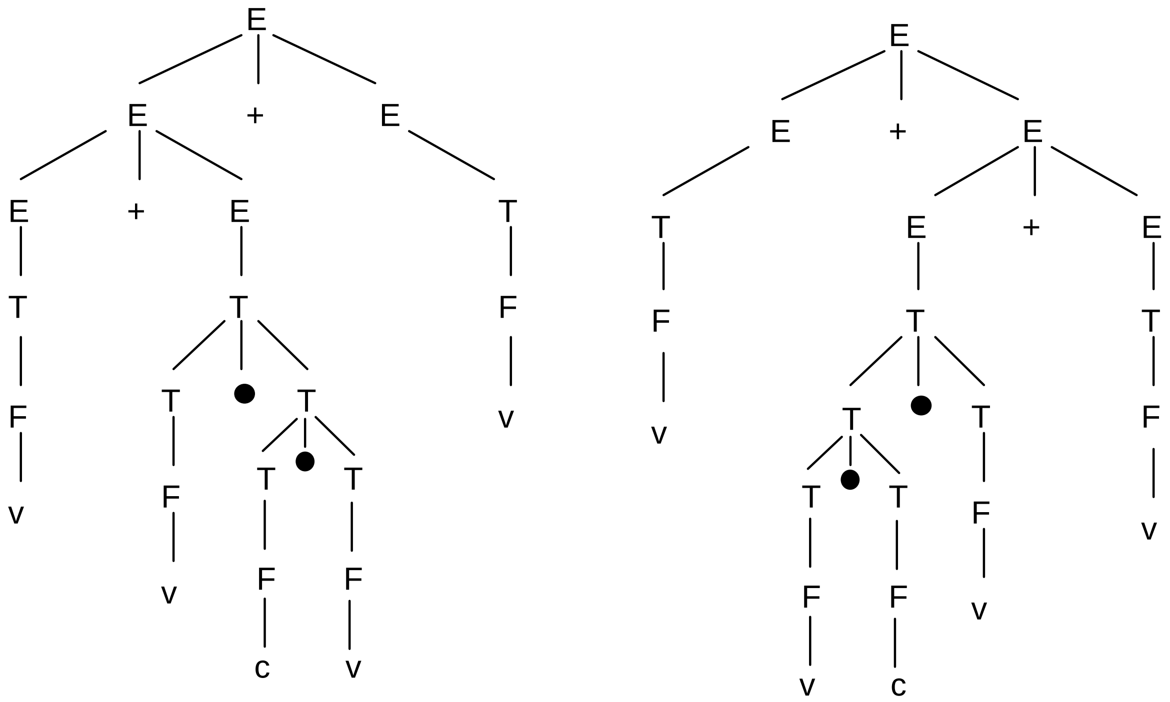
Def C: Sei $G=(N,T,P,\sigma)$ ein kontextfreie Grammatik. Ein <u>Ableitungsbaum</u> (Syntaxbaum) zu G ist ein geordneter, markierter Baum $B=(V,E,v_0)$ mit:

- Jeder Knoten $v\in V$ ist mit einem Symbol aus $N\cup T\cup\{\varepsilon\}$ markiert.
- Die Wurzel v_0 ist mit σ markiert.
- Jeder innere Knoten ist mit einem Nichtterminalsymbol markiert.
- Jedes Blatt ist mit einem Symbol aus $T\cup\{\varepsilon\}$
- Ist $v\in V$ ein innerer Knoten mit den Söhnen $v_1,\ldots,v_k\in V$ in dieser Reihenfolge, ist A die Markierung von v und A_i die Markierung von v_i, so ist $A\rightarrow A_1\ldots A_k\in P$
- Ein mit ε markiertes Blatt hat keinen Bruder

Satz D: Sei $G=(N,T,P,\sigma)$ eine kontextdreie Grammatik. Dann gilt für jedes $w\in T^*$, $w=x_1\ldots x_n$, $\sigma \xrightarrow[G]{}{}^{*} w \Leftrightarrow$
Es gibt einen Ableitungsbaum zu G, dessen Blätter von links nach rechts gelesen mit $x_1,\ldots,x_n$ markiert sind.
<u>Beweis:</u> „⇒“ durch Induktion über die Länge der Ableitung
„⇐“ durch Induktion über die Höhe des Baumes

Verschiedene Ableitungen können durch den denselben Ableitungsbaum dargestellt werden.
(1) $E\rightarrow E+E\rightarrow E+T\rightarrow T+T\rightarrow\ldots$
(2) $E\rightarrow E+E\rightarrow T+E\rightarrow F+E\rightarrow v+E\rightarrow v+T\rightarrow v+F\rightarrow v+v$
Linksableitung: jeweils linkestes Nichtterminal wird ersetzt.

Fazit: Jedem Ableitungsbaum läßt sich eine Linksableitung zuordnen: D.h. für jede kontextfreie Grammatik G gilt: $w\in L(G)\Leftrightarrow$Es gibt eine Linksableitung von w bezüglich G.
Andererseits: Es kann zu einer Grammatik G und einem Wort $w\in L(G)$ mehrere strukturell unterschiedliche Ableitungsbäume geben. Beispiel: $v+v\cdot c\cdot v+v$ → 2 Ableitungsbäume



Mehrdeutige Grammatik: Es gibt ein Wort w mit strukturell unterschiedlichen Ableitungsbäumen von w bezüglich G. (Sonst ist die Grammatik eindeutig)

Eindeutige Sprache: Es gibt eine eindeutige Grammatik, die die Sprache erzeugt. (Sonst ist die Sprache inhärent mehrdeutig)

Beispiel: 1) Eindeutige Grammatik für arithmetische Ausdrücke: N,T,σ wie oben,
$P'=\{E\rightarrow E+T,\ E\rightarrow T,\ T\rightarrow T\cdot F,\ T\rightarrow F,\ F\rightarrow (E),\ F\rightarrow v,\ F\rightarrow c\}$

2) Inhärent mehrdeutige Sprache: $L=\{a^n b^n c^m \mid n,m\geq 0\}\cup\{a^m b^n c^n \mid m,n\geq 0\}$
Genau die Wörter $a^i b^i c^i$ haben strukturell verschiedene Ableitungsbäume
Beweis: Aufwendig. [Wegener: Theoretische Informatik]

5.1.1 Backus-Naur-Form

Notation zur kompakten Beschreibung kontextfreier Grammatiken [J. Backus, P.Naur (→Algol60)]
Kürzungsregeln:

$A\rightarrow\beta_1$ ... $A\rightarrow\beta_n$	$\xrightarrow{\text{BNF}}$ $A\rightarrow\mid\beta_1\mid\ldots\mid\beta_n\mid$	| ist Metasymbol
$A\rightarrow\alpha\chi$ $A\rightarrow\alpha\beta\chi$	$\xrightarrow{\text{BNF}}$ $A\rightarrow\alpha[\beta]\chi$	"[...] β kann auch weggelassen werden."
$A\rightarrow\alpha\chi$ $A\rightarrow\alpha B\chi$ $B\rightarrow\beta$ $B\rightarrow\beta B$	$\xrightarrow{\text{BNF}}$ $A\rightarrow\alpha\{\beta\}\chi$	"{...} β kann beliebig oft, auch nullmal zwischen α und χ geschaltet werden."

Aussage: Durch Grammatiken in BNF können genau die kontextfreien Sprachen beschrieben werden.