Hochschule für Angewandte
Wissenschaften Hamburg
Hamburg University of Applied Sciences

Automatentheorie und Formale Sprachen

– Kontextfreie Grammatiken –

Prof. Dr. Michael Neitzke

KFG1: Definition und Verwendung von KFG

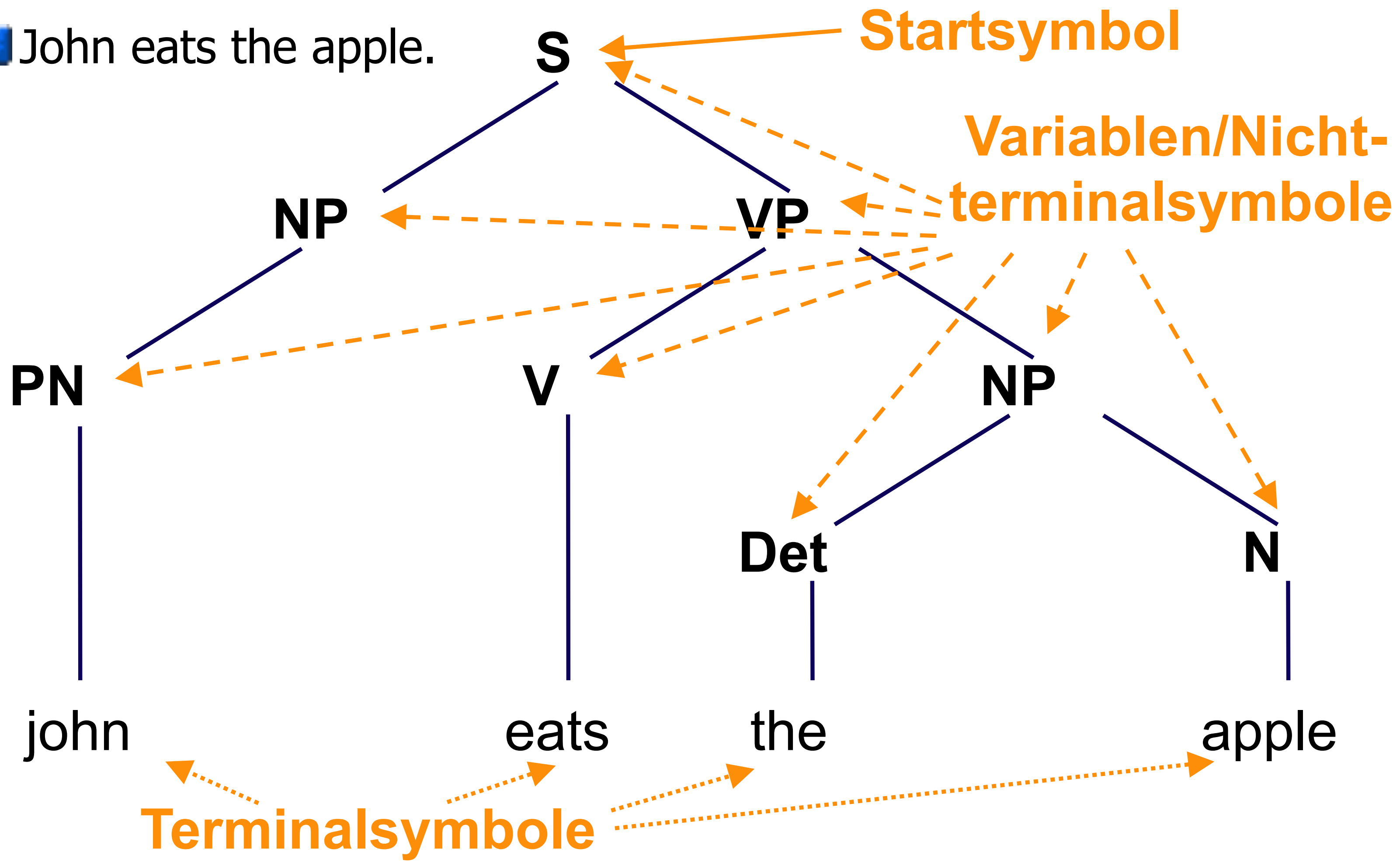
- Was sind Grammatiken
  - Beispiele
  - Formale Definition
- Mit Grammatiken arbeiten
- Begrifflichkeiten
  - Sprache einer Grammatik
  - Satzform

Kontextfreie Sprachen

- Reguläre Sprachen – Beschreibungsformen:
  - DEAs
  - NEAs
  - ε-NEAs
  - RA
- Nicht alle Sprachen sind regulär
  - Z. B. $\mathbf{0^n1^n}$
  - Palindrome, z. B.
    - **OTTO, MADAMIMADAM**
    - **0110, 101101101, 0**
- Kontextfreie Sprachen
  - Eine Beschreibungsform: Kontextfreie Grammatiken

Struktur natürlichsprachlicher Sätze - Beispiel

- John eats the apple.



Kontextfreie Grammatik zum Beispiel

- r1: S $\rightarrow$ NP VP
- r2: NP $\rightarrow$ PN
- r3: NP $\rightarrow$ Det N
- r4: VP $\rightarrow$ V
- r5: VP $\rightarrow$ V NP
- r6: PN $\rightarrow$ john | mary…
- r7: N $\rightarrow$ man | woman | dog | bird | apple…
- r8: V $\rightarrow$ sings | eats | bites | loves…
- r9: Det $\rightarrow$ the

Kontextfreie Grammatik zum Beispiel

- r1: Sentence → NominalPhrase VerbalPhrase
- r2: NominalPhrase → ProperName
- r3: NominalPhrase → Determiner Noun
- r4: VerbalPhrase → Verb
- r5: VerbalPhrase → Verb NominalPhrase
- r6: ProperName → john | mary...
- r7: Noun → man | woman | dog | bird | apple...
- r8: Verb → sings | eats | bites | loves...
- r9: Determiner → the

Ende V6
12.11.

Übung

Konstruieren Sie eine Grammatik für Palindrome über dem Alphabet {0,1}.

[BD: ... für $L = \{w \in \{0,1\}^* \mid w = 0^{2n}1^{2n}; n \in \mathbb{N}\}$

Kontextfreie Grammatik für Palindrome

1.	$P \to \varepsilon$
2.	$P \to 0$
3.	$P \to 1$
4.	$P \to 0P0$
5.	$P \to 1P1$

$L_1 = \{0^{2n} 1^{2n} : (n \geq 1)\}$

$S \to 00A11$

$A \to \varepsilon \mid 00A11$

einfachere Grammatik $S \to 00S11 \mid \varepsilon$

$\hookrightarrow L_2 = L_1 \cup \{\varepsilon\}$

Übung

- Aus welchen Komponenten besteht eine Grammatik?

S	$\rightarrow$	NP VP
NP	$\rightarrow$	PN
NP	$\rightarrow$	Det N
VP	$\rightarrow$	V
VP	$\rightarrow$	V NP
PN	$\rightarrow$	john | mary…
N	$\rightarrow$	man | woman | dog | bird | apple…
V	$\rightarrow$	sings | eats | bites | loves…
Det	$\rightarrow$	the

1.	$P \rightarrow \varepsilon$
2.	$P \rightarrow 0$
3.	$P \rightarrow 1$
4.	$P \rightarrow 0P0$
5.	$P \rightarrow 1P1$

Definition: Kontextfreie Grammatik

- Eine kfG umfasst 4 Komponenten:
  - **Variablen**, Menge von **nichtterminalen Symbolen**
  - **Alphabet**, Menge von **terminalen Symbolen**
  - Menge von **Produktionen**, **Regeln** bestehend aus
    - **Kopf**: Variable / nichtterminales Symbol
    - **Produktionssymbol**: $\rightarrow$
    - **Rumpf**: Zeichenreihe aus terminalen und nichtterminalen Symbolen (oder ε)
  - **Startsymbol**, besonderes nichtterminales Symbol
- Formal: **G = (V, T, P, S)**

Übungen

entspricht:
$A \to 0A$
$A \to \varepsilon$

- Gegeben sei folgende Grammatik:

$$S \to A1B$$
$$A \to 0A \mid \varepsilon$$
$$B \to 0B \mid 1B \mid \varepsilon$$

a) Beschreiben Sie die Grammatik formal als 4-Tupel. (V, T, P, S)

b) Welche Sprache wird erzeugt? Geben Sie einen RA an.

Übungen

$P = \{ \ldots, A \to \varepsilon, \ldots \}$

- Gegeben sei folgende Grammatik:

$$\left\{ \begin{array}{l} S \to A1B, \\ A \to 0A \mid \varepsilon, \\ B \to 0B \mid 1B \mid \varepsilon \end{array} \right\} = P$$

a) $G = (V, T, P, S)$ $\quad V = \{S, A, B\}$ $\quad T = \{0, 1\}$

b)

Übungen

$P = \{ \ldots, A \to \varepsilon, \ldots \}$

- Gegeben sei folgende Grammatik:

$$P = \left\{ \begin{array}{l} S \to A1B, \\ A \to 0A \mid \varepsilon, \\ B \to 0B \mid 1B \mid \varepsilon \end{array} \right\}$$

a) $G = (V, T, P, S)$ $V = \{S, A, B\}$ $T = \{0, 1\}$

b) $L(G) = L(0^* 1 (0+1)^*)$

Beispiel: kfG für arithmetische Ausdrücke

$T = \{ a, b, 0, 1, +, *, (,) \}$

1.	$E \to I$
2.	$E \to E + E$
3.	$E \to E * E$
4.	$E \to (E)$
5.	$I \to a$
6.	$I \to b$
7.	$I \to Ia$
8.	$I \to Ib$
9.	$I \to I0$
10.	$I \to I1$

- E: Expression
- I: Identifier
- Sprache für Identifier ist regulär!
  - $(a + b)(a + b + 0 + 1)^*$

- **Wie kann man überprüfen, ob**
  a*(a+b00)
  in der Sprache enthalten ist?

Möglichkeiten, mit einer Grammatik zu arbeiten

- **Bottom-up: Reduktion**
  - Ausgehend vom Wort werden die Regeln der Grammatik rückwärts angewandt, bis das Startsymbol errreicht ist.
  - Ist eine Zeichenreihe (ein Wort) in der Sprache einer Variablen (insbesondere des Startsymbols) enthalten?
- **Top-down: Ableitung**
  - Ausgehend vom Startsymbol werden die Regeln der Grammatik angewandt.
  - Lässt sich eine Zeichenreihe aus dem Startsymbol erzeugen?

Ableitung

- Symbol $\Rightarrow$ drückt Ableitungsschritt aus:
  - $\alpha A\beta \Rightarrow \alpha\gamma\beta$, wenn $(A \rightarrow \gamma) \in P$
- Beispiel: $E \Rightarrow E * E \Rightarrow I * E \Rightarrow a * E \Rightarrow a * (E) \Rightarrow$
  $a * (E + E) \Rightarrow a * (I + E) \Rightarrow a * (a + E) \Rightarrow$
  $a * (a + I) \Rightarrow a * (a + I0) \Rightarrow a * (a + I00) \Rightarrow$
  $a * (a + b00)$
- Symbol $\overset{*}{\Rightarrow}$ drückt Folge von Ableitungsschritten aus:
  - Z. B. $E \overset{*}{\Rightarrow} a * (a + b00)$
  - Aber auch: $E \overset{*}{\Rightarrow} E$

1.	$E \rightarrow I$
2.	$E \rightarrow E + E$
3.	$E \rightarrow E * E$
4.	$E \rightarrow (E)$
5.	$I \rightarrow a$
6.	$I \rightarrow b$
7.	$I \rightarrow Ia$
8.	$I \rightarrow Ib$
9.	$I \rightarrow I0$
10.	$I \rightarrow I1$

Mögliche Reduktionsfolgen

a ∗ (a + b00)	a ∗ (a + b00)
I ∗ (a + b00)	a ∗ (a + I00)
E ∗ (a + b00)	a ∗ (a + I0)
E ∗ (I + b00)	a ∗ (a + I)
E ∗ (E + b00)	a ∗ (a + E)
E ∗ (E + I00)	a ∗ (I + E)
E ∗ (E + I0)	a ∗ (E + E)
E ∗ (E + I)	a ∗ (E)
E ∗ (E + E)	a ∗ E
E ∗ (E)	I ∗ E
E ∗ E	E ∗ E
E	E

1.	$E \to I$
2.	$E \to E + E$
3.	$E \to E * E$
4.	$E \to (E)$
5.	$I \to a$
6.	$I \to b$
7.	$I \to Ia$
8.	$I \to Ib$
9.	$I \to I0$
10.	$I \to I1$

Was ist der Unterschied?

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Links- und rechtsseitige Ableitungen

- **Linksseitige Ableitung**:
  - Die am weitesten links stehende Variable wird ersetzt
  - Symbole: $(\overset{*}{\Rightarrow})$, $\overset{*}{\underset{lm}{\Rightarrow}}$

$\longrightarrow$ wenn nicht erwähnt wird, dass es eine linksseitige Ableitung ist (leftmost)

- **Rechtsseitige Ableitung**:
  - Die am weitesten rechts stehende Variable wird ersetzt
  - Symbole: $(\overset{*}{\Rightarrow})$, $\overset{*}{\underset{rm}{\Rightarrow}}$

— " — (rightmost)

Übung

- Erzeugen Sie die rechtsseitige Ableitung für den Ausdruck
  $a * (a + b00)$

1.	$E \to I$
2.	$E \to E + E$
3.	$E \to E * E$
4.	$E \to (E)$
5.	$I \to a$
6.	$I \to b$
7.	$I \to Ia$
8.	$I \to Ib$
9.	$I \to I0$
10.	$I \to I1$

Rechtsseitige Produktion = linksseit. Reduktion

$S \rightarrow E$

$E \rightarrow E\,Q\,F \mid F$

$F \rightarrow a$

$Q \rightarrow + \mid -$

S	
E	
EQF	
EQ	a
E	+a
EQF	+a
EQ	a+a
E	-a+a
F	-a+a
	a-a+a

	a-a+a
F	-a+a
E	-a+a
EQ	a+a
EQF	+a
E	+a
EQ	a
EQF	
E	
S	

- Wenn Eingabe von links nach rechts gelesen wird, bietet sich linksseitige Reduktion an

Diskussion

Wie könnte die Sprache einer Grammatik formal definiert sein?

Sprache einer Grammatik, Satzformen

- Sei $G = (V, T, P, S)$ eine kfG

$$L(G) = \{ w \in T^* \mid S \overset{*}{\underset{G}{\Rightarrow}} w \}$$

- **Sprache einer Grammatik**:
  Menge aller Zeichenreichen aus terminalen Symbolen, die sich vom Startsymbol ableiten lassen

- Alle $\alpha \in (T \cup V)^*$ mit $S \overset{*}{\underset{G}{\Rightarrow}} \alpha$ heißen **Satzformen**
  - Entsprechend der Art der Ableitung auch **linksseitige** oder **rechtsseitige** Satzformen

(bis 24.4)

Übungen

- Entwerfen Sie kontextfreie Grammatiken für die folgenden Sprachen:
  - $\{0^n 1^n \mid n \geq 1\}$
  - BDiff $\{a^i b^j c^k \mid i \neq j \ \vee j \neq k, i \geq 0, j \geq 0, k \geq 0\}$

Übungen

- $\{0^n 1^n \mid n \geq 1\}$

$S \to 01$

$S \to 0S1$

Übungen

- $\{a^i b^j c^k \mid i \neq j \lor j \neq k, i \geq 0, j \geq 0, k \geq 0\}$

$S \to AB \mid CD$

$A \to aA \mid \varepsilon$

$B \to bBc \mid E \mid cD$

$C \to$?

$D \to Dc \mid \varepsilon$

$E \to bE \mid b$

KFG1: Rückblick auf die Ziele

- Am Ende dieses Kapitels

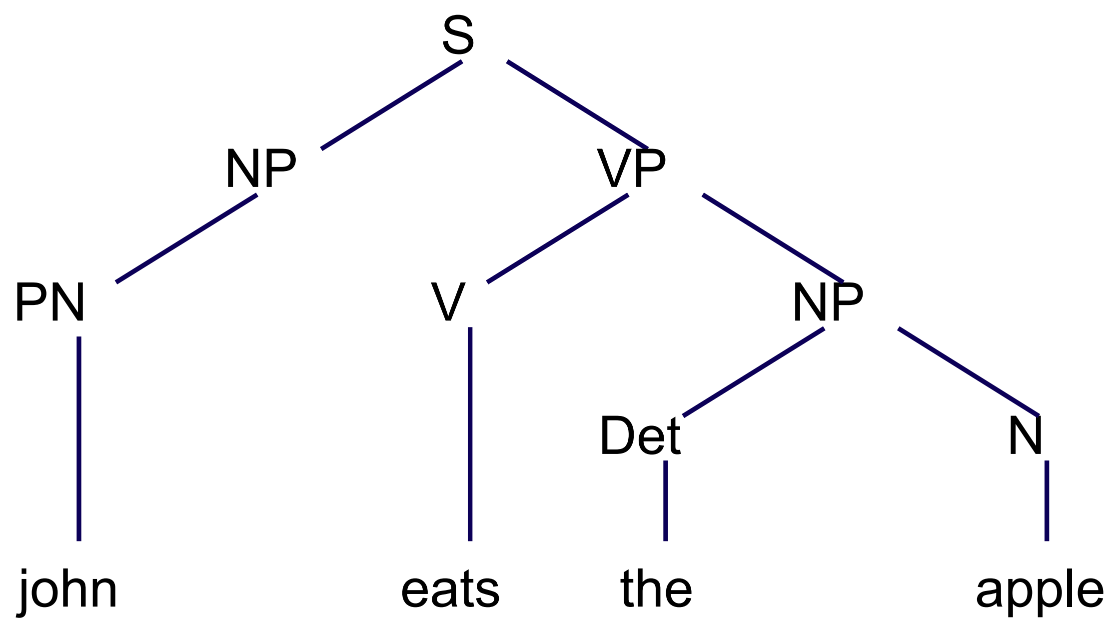
V7 19.11.

KFG2: Eigenschaften kontextfreier Grammatiken

- Parsebäume
- Mehrdeutige Grammatiken
- Reguläre Grammatiken

Parsebäume

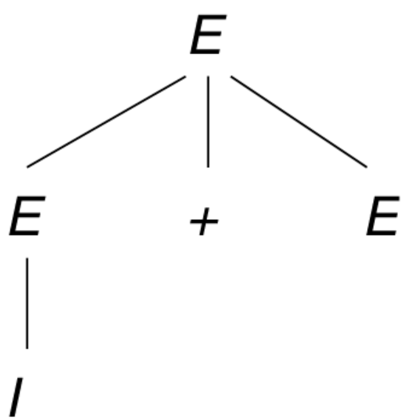
- Beispiel:



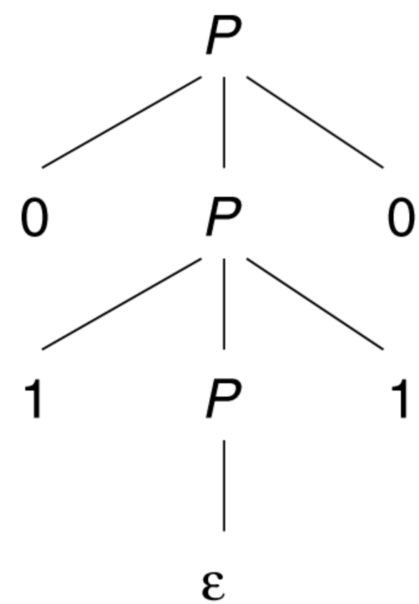
- Datenstruktur für das Ergebnis von Ableitungen oder Reduktionen
- Wird von Parsern erzeugt

Weitere Beispiele

- Parsebaum für $\mathbf{E} \overset{*}{\Rightarrow} \mathbf{I + E}$

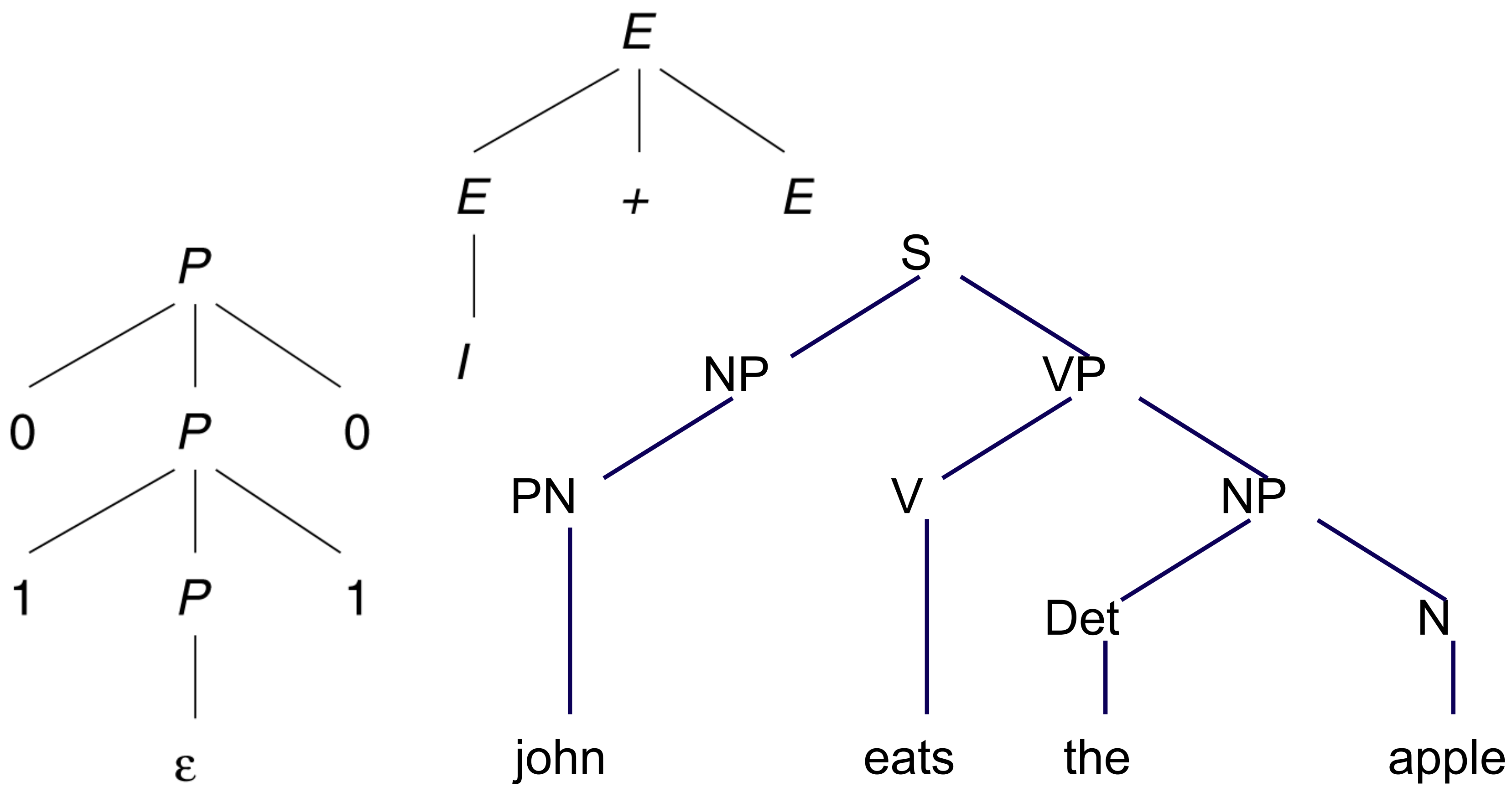


Parsebaum für $\mathbf{P} \overset{*}{\Rightarrow} \mathbf{0110}$



Übung

- Wie könnte ein Parsebaum definiert sein?



Definition Parsebaum

- **Parsebaum** für Grammatik $G = (V, T, P, S)$
- Beschriftungen der Knoten
  - Innere Knoten: mit Nichtterminalsymbol $A_i \in V$
  - Blattknoten: mit Nichtterminalsymbol, Terminalsymbol oder ε
  - Wenn Knoten mit A beschriftet und
    Nachfolger von links nach rechts mit X_1 , X_2 , … , X_k ,
    dann ist $A \rightarrow X_1 X_2 \ldots X_k$ in P enthalten
- Parsebäume für Zeichenreihen der durch G definierten Sprache:
  - Wurzel mit Startsymbol beschriftet
  - Blattknoten mit Terminalsymbolen oder ε beschriftet

Übungen

Gegeben sei folgende Grammatik:

$S \rightarrow A1B$

$A \rightarrow 0A \mid \varepsilon$

$B \rightarrow 0B \mid 1B \mid \varepsilon$

1. Geben Sie linksseitige und rechtsseitige Ableitungen für folgende Zeichenreihen an:
   a) **00101**
   b) **1001**
   c) **00011**
2. Zeichnen Sie die zugehörigen Parsebäume

Gegeben sei folgende Grammatik: $S \rightarrow A1B$

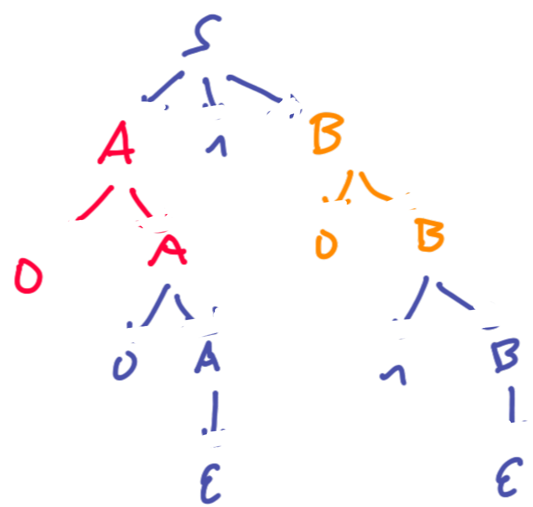
$A \rightarrow 0A \mid \varepsilon$

$B \rightarrow 0B \mid 1B \mid \varepsilon$

1. Geben Sie linksseitige und rechtsseitige Ableitungen für folgende Zeichenreihen an:
   - a) 00101
   - b) 1001
   - c) 00011
2. Zeichnen Sie die zugehörigen Parsebäume

LM $S \rightarrow A1B \rightarrow 0A1B \rightarrow 00A1B \rightarrow 001B \rightarrow 0010B$
$\rightarrow 00101B \rightarrow 00101$

RM $S \rightarrow A1B \rightarrow A10B \rightarrow A101B \rightarrow A101$
$\rightarrow 0A101 \rightarrow 00A101 \rightarrow 00101$



Baum "LM lesen"

Baum "RM lesen"

Gruppenarbeit

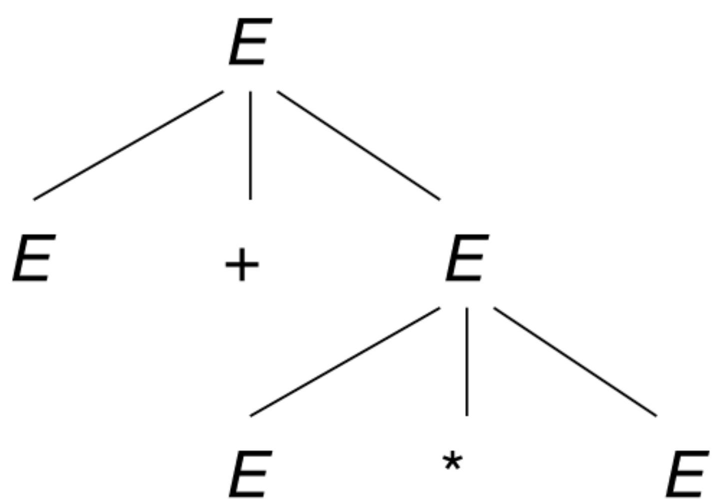
Wie viele (verschiedene linksseitige bzw. rechtsseitige Ableitungen und wie viele) verschiedene Parsebäume können Sie für die Worte **01110** bzw. **a + a * a** erzeugen?

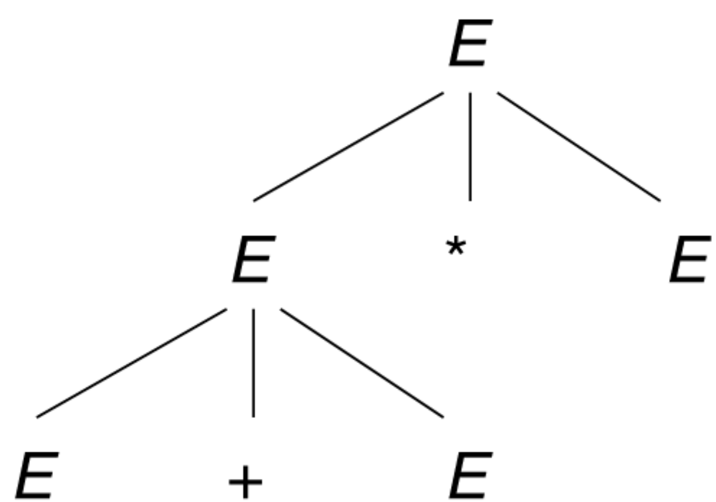
1.	$P \rightarrow \varepsilon$
2.	$P \rightarrow 0$
3.	$P \rightarrow 1$
4.	$P \rightarrow 0P0$
5.	$P \rightarrow 1P1$

1.	$E \rightarrow I$
2.	$E \rightarrow E + E$
3.	$E \rightarrow E * E$
4.	$E \rightarrow (E)$
5.	$I \rightarrow a$
6.	$I \rightarrow b$
7.	$I \rightarrow Ia$
8.	$I \rightarrow Ib$
9.	$I \rightarrow I0$
10.	$I \rightarrow I1$

Mehrdeutige Grammatiken (1)

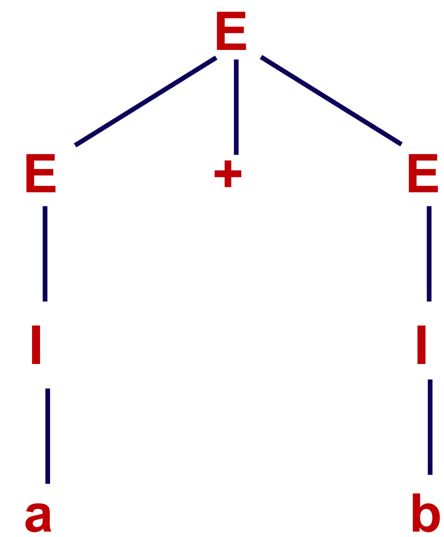
- Zwei Ableitungen bereits für $E + E * E$
  - $E \Rightarrow E + E \Rightarrow E + E * E$
  - $E \Rightarrow E * E \Rightarrow E + E * E$
- Zugehörige Parsebäume





Mehrdeutige Grammatiken (2)

- Meist viele Ableitungen, z. B. für **a + b**
  - $E \Rightarrow E+E \Rightarrow I+E \Rightarrow a+E \Rightarrow a+I \Rightarrow a+b$
  - $E \Rightarrow E+E \Rightarrow E+I \Rightarrow I+I \Rightarrow I+b \Rightarrow a+b$



- Kein Problem, wenn derselbe Parsebaum
- Unterschiedliche Parsebäume sind Merkmal mehrdeutiger Grammatiken

> Wenn eine Grammatik für <u>jede</u> Zeichenreihe nur einen Parsebaum erzeugt, ist sie **eindeutig**, sonst **mehrdeutig**.

- Und:

> Wenn eine Grammatik für <u>eine</u> Zeichenreihe über zwei unterschiedliche linksseitige (oder rechtsseitige) <u>Ableitungen</u> verfügt, ist sie **mehrdeutig**.

Diskussion

- Kann man eine gegebene Sprache durch unterschiedliche Grammatiken definieren?
- Könnte es sein, dass dieselbe Sprache durch eine eindeutige und durch eine mehrdeutige Grammatik beschrieben wird?

- Wann sollte man sagen, dass eine Sprache eindeutig ist?
- Wann sollte man sagen, dass eine Sprache mehrdeutig ist?

- Was ist schwieriger festzustellen?

Diskussion

- Kann man eine gegebene Sprache durch unterschiedliche Grammatiken definieren? JA!

$A \to Aa \mid \varepsilon$
$A \to aA \mid \varepsilon$

- Könnte es sein, dass dieselbe Sprache durch eine eindeutige und durch eine mehrdeutige Grammatik beschrieben wird?

JA (Folien 18 + 39)

- Wann sollte man sagen, dass eine Sprache eindeutig ist? $\exists$ eind. Grammatik
- Wann sollte man sagen, dass eine Sprache mehrdeutig ist? $\nexists$ – " –
- Was ist schwieriger festzustellen?

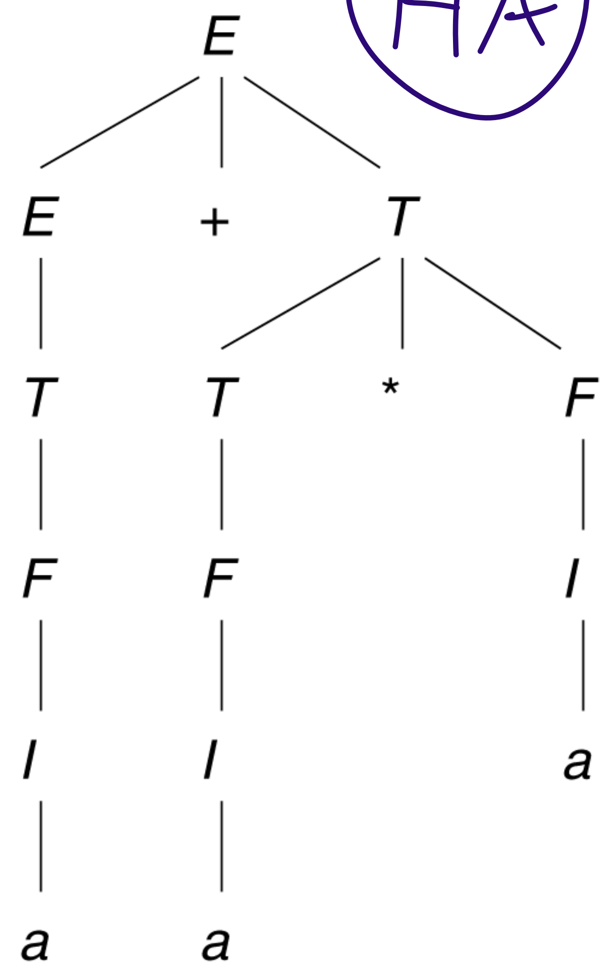
Mehrdeutigkeiten beseitigen

HA

- Es kann für ein und dieselbe Sprache unterschiedliche Grammatiken geben, eindeutige und mehrdeutige

- Beispiel:

  $$\mathbf{E} \rightarrow \mathbf{T} \mid \mathbf{E} + \mathbf{T}$$
  $$\mathbf{T} \rightarrow \mathbf{F} \mid \mathbf{T} * \mathbf{F}$$
  $$\mathbf{F} \rightarrow \mathbf{I} \mid (\mathbf{E})$$
  $$\mathbf{I} \rightarrow \mathbf{a} \mid \mathbf{b} \mid \mathbf{Ia} \mid \mathbf{Ib} \mid \mathbf{I0} \mid \mathbf{I1}$$



Diese Sprache ist also eindeutig!

Ein paar Fakten zu mehrdeutigen Grammatiken

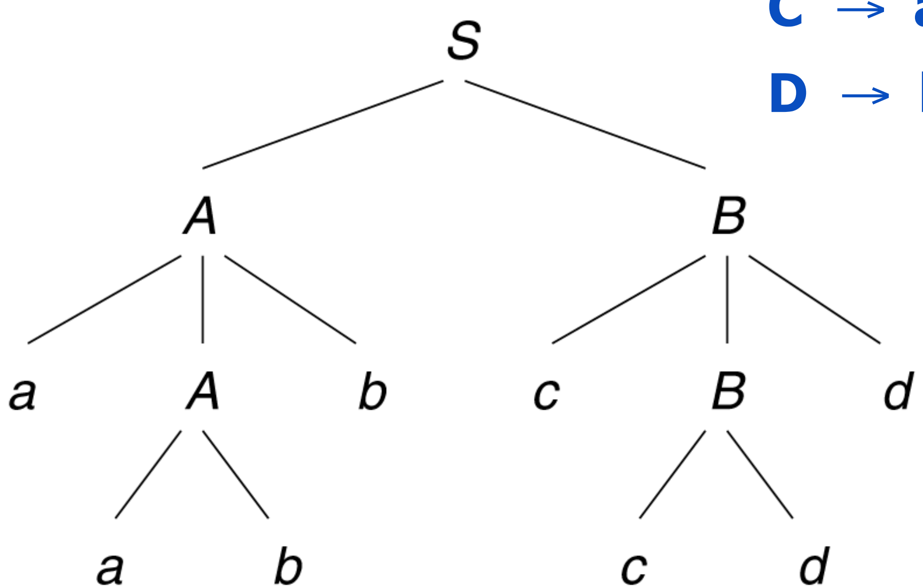
- Mehrdeutigkeiten können manchmal durch Veränderung der Grammatik beseitigt werden.
  - Ein und dieselbe Sprache kann durch unterschiedliche Grammatiken beschrieben werden
  - Eine **Sprache** ist **eindeutig**, wenn sie über mindestens eine eindeutige Grammatik verfügt
- Es gibt kontextfreie Sprachen, die ausschließlich mehrdeutige Grammatiken besitzen (inhärent mehrdeutig sind).
  - Also keine Möglichkeit, die Mehrdeutigkeiten zu umgehen
- Es gibt keinen Algorithmus, der feststellen kann, ob eine **kfG** mehrdeutig ist. d.h. Die Frage "Ist diese KFG mehrdeutig" ist nicht entscheidbar! !!!

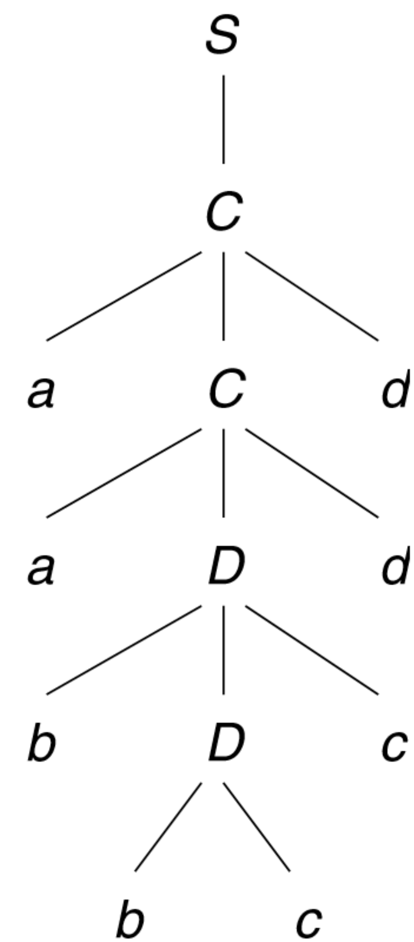
Beispiel für inhärent mehrdeutige Sprache

- $L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$
- Mögliche Grammatik:

$$
\begin{aligned}
S &\rightarrow AB \mid C \\
A &\rightarrow aAb \mid ab \\
B &\rightarrow cBd \mid cd \\
C &\rightarrow aCd \mid aDd \\
D &\rightarrow bDc \mid bc
\end{aligned}
$$

- Problematisch: $a^n b^n c^n d^n$





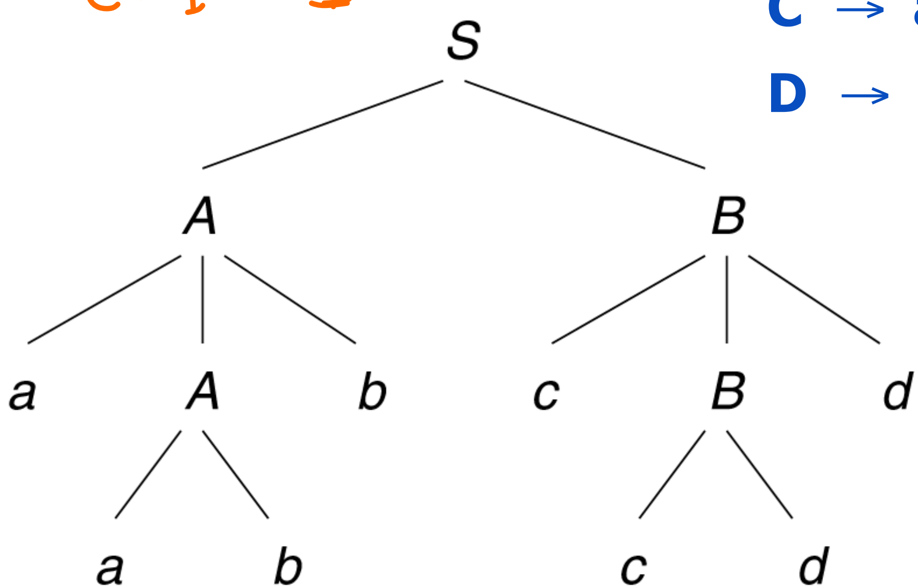
Beispiel für inhärent mehrdeutige Sprache

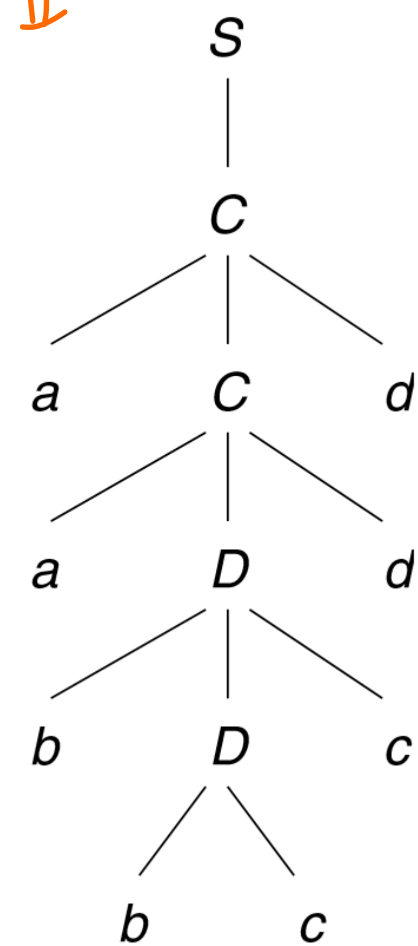
- $L = \underset{I}{\{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\}} \cup \underset{II}{\{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}}$
- Mögliche Grammatik:

$$
\begin{aligned}
S &\rightarrow \overset{I}{AB} \mid \overset{II}{C} \\
A &\rightarrow aAb \mid ab \\
B &\rightarrow cBd \mid cd \\
C &\rightarrow aCd \mid aDd \\
D &\rightarrow bDc \mid bc
\end{aligned}
$$

- Problematisch: $a^n b^n c^n d^n$

$\in M_I \cap M_{II}$





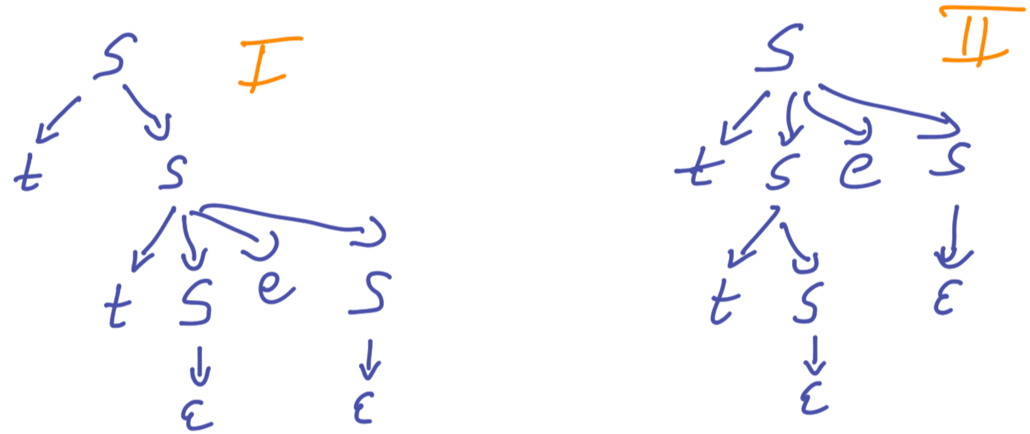
Übungen

1. Gegeben: $S \rightarrow tS \mid tSeS \mid \varepsilon$
   - Zeigen Sie, dass für **tte** zwei
     a) Parsebäume
     b) linksseitige Ableitungen
     c) rechtsseitige Ableitungen

     existieren.
2. BD Finden Sie für diese Sprache eine Grammatik, die eindeutig ist.

Übungen

a)



b)

lm:

$S \to tS \to ttSeS \to tteS \to tte$ I

$S \to tSeS \to ttSeS \to tteS \to tte$ II

c)

rm:

$S \to tS \to ttSeS \to ttSe \to tte$ I

$S \to tSeS \to tSe \to ttSe \to tte$ II

Grammatiken für reguläre Sprachen?

- Bisher:
  - Automaten, reguläre Ausdrücke für reguläre Sprachen
  - Grammatiken für kontextfreie Sprachen

- **Denken Sie daran, wie man mit regulären Ausdrücken Sprachen beschreibt und wie Automaten Wörter prüfen. Was ist die besondere Eigenschaft regulärer Sprachen?**
- **In welcher Hinsicht sind kontextfreie Sprachen mächtiger als reguläre?**
- **Wie schlägt sich dies in einer Grammatik nieder?**

Reguläre Grammatiken

- Einschränkung gegenüber kfG:
  - Maximal ein Nonterminalsymbol im rechten Teil einer Regel
  - Nonterminalsymbol nur am Ende
- Beispiel: L(01*2)

$$S \rightarrow 0A$$
$$A \rightarrow 1A$$
$$A \rightarrow 2$$

- "Rechtslineare" Grammatiken
- Äquivalent: Linkslineare Grammatiken

$$S \rightarrow A2$$
$$A \rightarrow A1$$
$$A \rightarrow 0$$

Reguläre - Übung

- Erstelle eine Rechtslineare Grammatik für $(0+1)^*+1.(01)^*$

Reguläre - Übung

- Erstelle eine Rechtslineare Grammatik für (0+1)*+1.(01)*

$$\underline{(0+1)^*} + \underline{1.(01)^*}$$

$$S \rightarrow \underline{A} \mid \underline{B}$$

$$A \rightarrow \varepsilon \mid 0A \mid 1A$$

$$B \rightarrow 1C$$

$$C \rightarrow \varepsilon \mid 01C$$

Reguläre Grammatiken: Was geht verloren?

- KfG beschreiben Verschachtelungen
  - Beispiel: Satzstruktur
  - Sequenz von Variablen
  - Hierarchien
  - Terminale, die um eine Variable gruppiert werden
- Reguläre Grammatiken:
  - Erzeugung von Terminalen nur von links nach rechts (oder umgekehrt)
  - Keine "globale" Strukturierung

KFG2: Rückblick auf die Ziele

	KFG1 bis KFG2
Kenntnis	- Definitionen der rot hervorgehobenen Begriffe kennen
Verständnis	- Eigenschaften regulärer und kontextfreier Grammatiken und Sprachen erläutern können - Die Beziehung zwischen Grammatik und Sprache erläutern können
Anwendung	- Ableitungen durchführen können - Parsebäume erstellen können
Analyse Synthese	- Kontextfreie Grammatiken auf Basis einer formalen Beschreibung oder einer Anwendungssituation entwickeln können - Grammatiken auf Mehrdeutigkeit prüfen können
Beurteilung	

KFG3: Normalformen

- Chomsky-Normalform
- Greibach-Normalform

Chomsky-Normalform

- Jede Kontextfreie Grammatik kann in die folgende Normalform gebracht werden:

  $A \rightarrow BC \mid a$

- Diese Form heißt **Chomsky-Normalform** (**CNF**)

Nur das Startsymbol darf zusätzlich zum leeren Wort abgeleitet werden:

$$S \rightarrow \varepsilon \mid BC \mid a$$

Umformung in die CNF: Beseitigung von ε-Regeln

1) Startsymbol: $\mathbf{S} \rightarrow \varepsilon$ wird ersetzt durch $\mathbf{S'} \rightarrow \mathbf{S} \mid \varepsilon$

- So wird gewährleistet, dass **S** unverändert in den Regelrümpfen stehen darf. Hier muss aber zusätzlich 2) beachtet werden.

2) Beliebige Variable: $\mathbf{A} \rightarrow \varepsilon$

- Regel wird gestrichen
- Wenn es keine weitere Regel für **A** gibt, wird **A** in allen Regelrümpfen gestrichen, also z.B.
  $\mathbf{X} \rightarrow \mathbf{aYAb}$ wird zu $\mathbf{X} \rightarrow \mathbf{aYb}$
- Wenn es weitere Regeln für **A** gibt, so wird **A** in den Regelrümpfen belassen, aber es wird eine Regel ohne **A** als weitere Alternative hinzugefügt, also z.B. $\mathbf{X} \rightarrow \mathbf{aYAb}$ wird zu $\mathbf{X} \rightarrow \mathbf{aYAb} \mid \mathbf{aYb}$

Beispiel: Beseitigung von ε-Regeln

$S \rightarrow \varepsilon \mid ABxC$

$A \rightarrow xAy \mid \varepsilon$

$B \rightarrow yBS \mid D$

$C \rightarrow \varepsilon$

$D \rightarrow B \mid z$

wird zu

$S' \rightarrow S \mid \varepsilon$

$S \rightarrow ABx \mid Bx$

$A \rightarrow xAy \mid xy$

$B \rightarrow yBS \mid yB \mid D$

$D \rightarrow B \mid z$

Eliminieren von Einheitsproduktionen

↳ nur ein Nonterminal rechts

3) Regeln **A → A** werden ersatzlos gestrichen.

4) Wenn eine Regel **A → B** existiert, so werden an Stelle von **B** die Regelrümpfe der Regeln für **B** eingesetzt (zirkuläre Abhängigkeiten erfordern Sonderbehandlung):

E → T | E + T
T → F | T * F
F → id | (E)

wird zu

E → id | (E) | T * F | E + T
T → id | (E) | T * F
F → id | (E)

Beispiel: Eliminieren von Einheitsproduktionen

$$S' \rightarrow S \mid \varepsilon$$
$$S \rightarrow ABx \mid Bx$$
$$A \rightarrow xAy \mid xy$$
$$B \rightarrow yBS \mid yB \mid D$$
$$D \rightarrow B \mid z$$

wird zu

$$S' \rightarrow ABx \mid Bx \mid \varepsilon$$
$$S \rightarrow ABx \mid Bx$$
$$A \rightarrow xAy \mid xy$$
$$B \rightarrow yBS \mid yB \mid z$$

Ersetzen von Terminalsymbolen

5) In allen Regeln mit mindestens zwei Symbolen im Regelrumpf werden die Terminalsymbole durch ihnen entsprechende Variablen ersetzt, die dann in eigenen Regeln auf das zugehörige Terminalsymbol abbilden.

$$X \rightarrow aYAb$$

wird zu

$$X \rightarrow N_aYAN_b$$
$$N_a \rightarrow a$$
$$N_b \rightarrow b$$

Beispiel: Ersetzen von Terminalsymbolen

$S' \rightarrow ABx \mid Bx \mid \varepsilon$

$S \rightarrow ABx \mid Bx$

$A \rightarrow xAy \mid xy$

$B \rightarrow yBS \mid yB \mid z$

wird zu

$S' \rightarrow ABX \mid BX \mid \varepsilon$

$S \rightarrow ABX \mid BX$

$A \rightarrow XAY \mid XY$

$B \rightarrow YBS \mid YB \mid z$

$X \rightarrow x$

$Y \rightarrow y$

Regelrümpfe auf zwei Variablen reduzieren

Jetzt kann es nur noch drei Arten von Regeln geben:

- $S' \to \varepsilon$
- Regeln, die ein einzelnes Terminalsymbol im Regelrumpf haben
- Regeln, deren Regelrumpf aus zwei oder mehr Variablen besteht

6) Regeln mit mehr als zwei Variablen werden durch Einführung zusätzlicher Variablen auf nur noch zwei Variablen im Rumpf reduziert:

$X \to ABCDE$ wird zu

$$X \to N_{AB}N_{CDE}$$
$$N_{AB} \to AB$$
$$N_{CDE} \to CN_{DE}$$
$$N_{DE} \to DE$$

Beispiel: Regelrümpfe auf zwei Variablen reduzieren

$S' \to ABX \mid BX \mid \varepsilon$		$S' \to AN \mid BX \mid \varepsilon$
$S \to ABX \mid BX$		$S \to AN \mid BX$
$A \to XAY \mid XY$	wird zu	$A \to XM \mid XY$
$B \to YBS \mid YB \mid z$		$B \to YO \mid YB \mid z$
$X \to x$		$N \to BX$
$Y \to y$		$M \to AY$
		$O \to BS$
		$X \to x$
		$Y \to y$

Greibach-Normalform

- Jede Kontextfreie Grammatik kann in die folgende Normalform gebracht werden: Nur das Startsymbol darf zusätzlich zum leeren Wort abgeleitet werden:

  $\mathbf{A \rightarrow aB_1B_2 .. B_k \qquad , k \geq 0}$

  $S \to \varepsilon \mid aB_1B_2B_3 \ldots$

- Diese Form heißt **Greibach-Normalform** (**GNF**)
- Umformung hat meist CNF als Basis
  - Sukzessive Ersetzung der ersten Variablen, bis Terminalsymbol erreicht
  - Kompliziert durch mögliche Zyklen
- Vorteil: Jeder Ableitungsschritt produziert exakt ein Terminalsymbol

Beispiel: CNF → GNF

$S' \rightarrow \underline{AN} \mid BX \mid \varepsilon$

$S \rightarrow AN \mid BX$

$A \rightarrow \underline{XM} \mid XY$

$B \rightarrow YO \mid YB \mid z$

$N \rightarrow BX$

$M \rightarrow AY$

$O \rightarrow BS$

$X \rightarrow x$

$Y \rightarrow y$

$S' \rightarrow AN \rightarrow XMN$
$\rightarrow xMN$

wird zu

$S' \rightarrow xMN \mid xYN \mid$
$yOX \mid yBX \mid zX \mid \varepsilon$

$S \rightarrow xMN \mid xYN \mid$
$yOX \mid yBX \mid zX$

$A \rightarrow xM \mid xY$

$B \rightarrow yO \mid yB \mid z$

$N \rightarrow yOX \mid yBX \mid zX$

$M \rightarrow xMY \mid xYY$

$O \rightarrow yOS \mid yBS \mid zS$

$X \rightarrow x$

$Y \rightarrow y$

Beispiel: CNF → GNF

$S' \to AN \mid BX \mid \varepsilon$
$S \to AN \mid BX$
$A \to XM \mid XY$
$B \to YO \mid YB \mid z$
$N \to BX$
$M \to AY$
$O \to BS$
$X \to x$
$Y \to y$

wird zu

$S' \to xMN \mid xYN \mid yOX \mid yBX \mid zX \mid \varepsilon$
$S \to xMN \mid xYN \mid yOX \mid yBX \mid zX$
$A \to xM \mid xY$
$B \to yO \mid yB \mid z$
$N \to yOX \mid yBX \mid zX$
$M \to xMY \mid xYY$
$O \to yOS \mid yBS \mid zS$
$X \to x$
$Y \to y$

$S' \to AN \to XMN \to xMN$ | $AN \to XYN \to xYN$

$S' \to$ | $BX \to YOX \to yOX$ | $BX \to YBX \to yBX$ | $BX \to zX$

$S \to AN$ (siehe S')

$S \to BX$ (siehe S')

$N \to BX$ (siehe S')

$M \to$ | $AY \to XMY \to xMY$ | $AY \to XYY \to xYY$

$O \to$ | $BS \to YOS \to yOS$ | $BS \to YBS \to yBS$ | $BS \to zS$

Ende V8

KFG3: Rückblick auf die Ziele

	KFG3
Kenntnis	- Definitionen der rot hervorgehobenen Begriffe kennen
Verständnis	
Anwendung	- Eine gegebene KFG in die CNF transformieren können - Von CNF nach GNF transformieren können
Analyse Synthese	
Beurteilung	