

## 5. Modellierung von Strukturen

### 5.1 Kontextfreie Grammatiken

**Kontextfreie Grammatik (KFG):** formaler Kalkül, Ersetzungssystem; definiert

- **Sprache** als Menge von Sätzen; jeder **Satz** ist eine **Folge von Symbolen**
- **Menge von Bäumen**; jeder Baum repräsentiert die **Struktur eines Satzes** der Sprache

#### Anwendungen:

- Programme einer **Programmiersprache** und deren Struktur, z. B. Java, Pascal, C
- Sprachen als Schnittstellen zwischen Software-Werkzeugen, **Datenaustauschformate**, z. B. HTML, XML
- Bäume zur Repräsentation **strukturierter Daten**, z. B. in HTML
- Struktur von **Protokollen** beim Austausch von Nachrichten zwischen Geräten oder Prozessen

#### Beispiel zu HTML:

```
<table>
  <tr>
    <td>Mo</td>
    <td>11-13</td>
    <td>AM</td>
  </tr>
  <tr>
    <td>Fr</td>
    <td>9-11</td>
    <td>AM</td>
  </tr>
</table>
```

## Vorlesung Modellierung WS 2000/2001 / Folie 501

#### Ziele:

Einsatz von KFGn kennenlernen

#### in der Vorlesung:

Erläuterungen zu den Anwendungen

#### nachlesen:

G. Goos: Vorl. über Informatik Bd.1, Abschnitt 1.6.3

# Kontextfreie Grammatik

Eine kontextfreie Grammatik  $G = (T, N, P, S)$  besteht aus:

$T$	<b>Menge der Terminalsymbole</b> (kurz: <b>Terminale</b> )
$N$	<b>Menge der Nichtterminalsymbole</b> (kurz: <b>Nichtterminale</b> )
	$T$ und $N$ sind disjunkte Mengen
$S \in N$	<b>Startsymbol</b> (auch Zielsymbol)
$P \subseteq N \times V^*$	<b>Menge der Produktionen</b>
$V = T \cup N$	heißt auch <b>Vokabular</b> , seine Elemente heißen <b>Symbole</b>

Eine Produktion wird notiert  $A ::= x$  wobei  $x$  eine evtl. leere Folge von Symbolen aus  $V$  ist.  
 Man sagt „ $A$  steht auf der **linken Seite** und  $x$  auf der **rechten Seite** der Produktion.“  
 Man gibt Produktionen häufig **Namen**:  $p1: A ::= x$

**Beispiel:**

<b>Terminale</b>	$T = \{ (, ) \}$	<b>Name</b>	<b>Produktionsmenge P</b>
<b>Nichtterminale</b>	$N = \{ \text{Klammern}, \text{Liste} \}$	$p1:$	$\text{Klammern} ::= '()' \text{ Liste }$
<b>Startsymbol</b>	$S = \text{Klammern}$	$p2:$	$\text{Liste} ::= \text{Klammern} \text{ Liste}$
		$p3:$	$\text{Liste} ::=$
			}

## Vorlesung Modellierung WS 2000/2001 / Folie 502

### Ziele:

KFG Definition lernen

### in der Vorlesung:

- Erläuterung der Begriffe an dem Beispiel
- Erläuterung der Notation von Produktionen
- Kennzeichnung von Terminalen, um Verwechslungen zu vermeiden: ' (

### nachlesen:

G. Goos: Vorl. über Informatik Bd.1, Abschnitt 1.6.3

# Bedeutung der Produktionen

Eine Produktion  $A ::= x$  ist eine **Strukturregel**: A besteht aus x

## Beispiele:

DeutscherSatz	$::=$	Subjekt Prädikat Objekt
Ein DeutscherSatz	<i>besteht aus (der Folge)</i>	Subjekt Prädikat Objekt
Klammern	$::=$	'( ' Liste ')'
Zuweisung	$::=$	Variable ':=' Ausdruck
Variable	$::=$	Variable '[' Ausdruck ']'

Produktion als gewurzelter Baum mit geordneten Kanten und mit Symbolen als Knotenmarken:



Eine Produktion  $A ::= x$  ist auch eine **Ersetzungsregel**:

In einer Symbolfolge  $u A v$  kann A ersetzt werden durch x; Ergebnis:  $u x v$

Unter Anwendung der Produktion **Klammern**  $::= ' ( ' \text{Liste} ' ) '$  kann man z. B. in **Klammern Klammern Liste** das 2. Symbol ersetzen: **Klammern ( Liste ) Liste**

## Vorlesung Modellierung WS 2000/2001 / Folie 503

### Ziele:

Produktionen verstehen

### in der Vorlesung:

Erläuterungen der beiden Rollen von Produktionen:

- Definition von Struktur: "besteht aus"
- Definition von Ersetzungen

### nachlesen:

G. Goos: Vorl. über Informatik Bd.1, Abschnitt 1.6.3

## Ableitungen

Produktionen sind **Ersetzungsregeln**: Ein Nichtterminal A in einer Symbolfolge u A v kann durch die rechte Seite x einer Produktion  $A ::= x$  ersetzt werden.

Das ist ein **Ableitungsschritt**; er wird notiert als  $u A v \Rightarrow u x v$

z. B. **Klammern Klammern Liste**  $\Rightarrow$  **Klammern ( Liste ) Liste**  
mit Produktion p1

Beliebig viele Ableitungsschritte nacheinander angewandt heißen **Ableitung**; notiert als  $u \Rightarrow^* v$

Eine kontextfreie Grammatik **definiert eine Sprache**; das ist eine Menge von Sätzen.

Jeder Satz ist eine Folge von Terminalsymbolen, die aus dem Startsymbol ableitbar ist:

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

Die Grammatik auf Mod-5.2 definiert z. B. geschachtelte Folgen paariger Klammern als Sprachmenge:

$$\{ ( ) , ( ( ) ) , ( ( ( ) ) ) , ( ( ( ( ) ) ( ) ) ) , \dots \} \subseteq L(G)$$

## Vorlesung Modellierung WS 2000/2001 / Folie 504

### Ziele:

Ableitungsbegriff verstehen

### in der Vorlesung:

Erläuterungen dazu

- Beispiele für Ableitungen
- Beispiele für Sprachen

### nachlesen:

G. Goos: Vorl. über Informatik Bd.1, Abschnitt 1.6.3

# Ableitungsbäume

Jede Ableitung kann man als **gewurzelten Baum** darstellen:

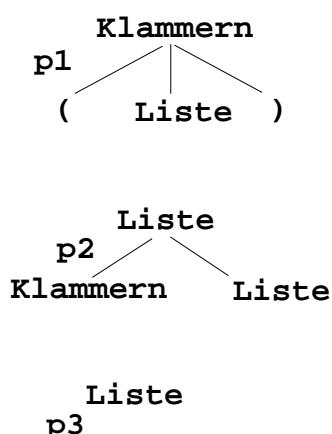
Die **Knoten** mit ihren Marken repräsentieren **Vorkommen von Symbolen**.

Ein Knoten mit seinen direkten Nachbarn repräsentiert die **Anwendung einer Produktion**.

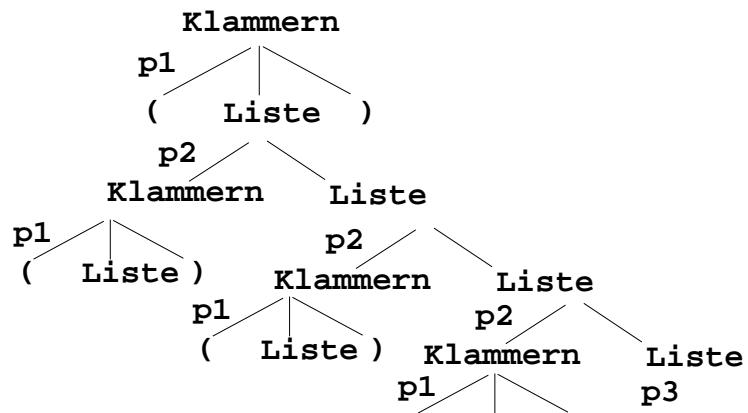
Die **Wurzel** ist mit dem **Startsymbol** markiert.

**Terminale** kommen nur an **Blättern** vor.

Produktionen:



ein Ableitungsbau:



## Beispiel: Ausdrucksgrammatik

p1: Ausdruck ::= Ausdruck BinOpr Ausdruck

p2: Ausdruck ::= Zahl

p3: Ausdruck ::= Bezeichner

p4: Ausdruck ::= '(' Ausdruck ')'

p5: BinOpr ::= '+'

p6: BinOpr ::= '-'

p7: BinOpr ::= '\*\*'

p8: BinOpr ::= '/'

Startsymbol: Ausdruck

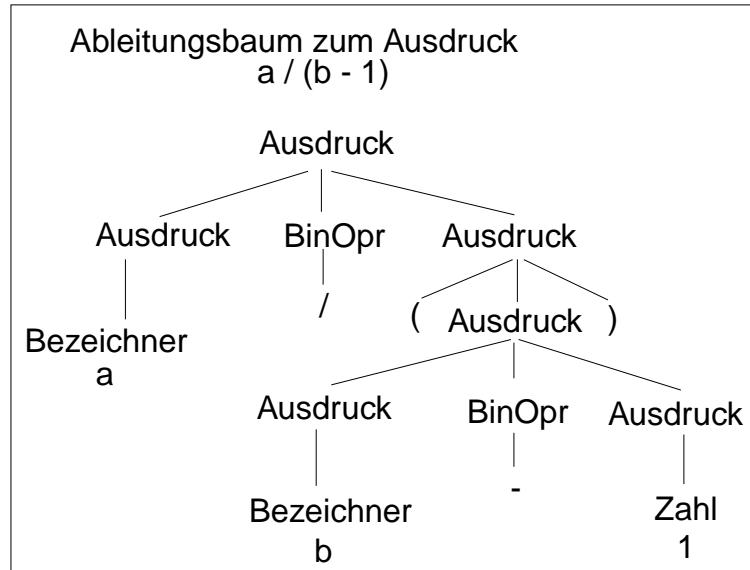
Terminale:

{ Zahl, Bezeichner, (, ), +, -, \*, / }

Schreibweise der Terminate

Zahl und Bezeichner wird

nicht in der KFG definiert.



## Vorlesung Modellierung WS 2000/2001 / Folie 506

### Ziele:

Vollständiges Beispiel sehen

### in der Vorlesung:

- Erläuterungen dazu.
- Vergleich mit Kantorowitsch-Bäumen.
- Es gibt Sätze zu der Grammatik, die mehrere Ableitungsbäume haben, z. B. a+b+c. Solche Grammatiken nennt man "mehrdeutig".

### nachlesen:

G. Goos: Vorl. über Informatik Bd.1, Abschnitt 1.6.3

## Beispiel: Tabellen in HTML

**HTML:** Hypertext Markup Language zur Darstellung von verzweigten Dokumenten, insbesondere im WWW verwendet.

**typisch:** geklammerte Strukturen mit Klammern der Form `<x>...</x>`.

hier: vereinfachter Ausschnitt aus der Sprache zur Darstellung von Tabellen.

### Produktionen der kontextfreien Grammatik:

```
Table    ::= '<table>' Rows '</table>'  
Rows    ::= Row *  
Row     ::= '<tr>' Cells '</tr>'  
Cells   ::= Cell *  
Cell    ::= '<td>' Text '</td>'  
Cell    ::= '<td>' Table '</td>'
```

### Erweiterung der Notation von KFGn:

**X \*** auf der rechten Seite einer Produktion steht für eine **beliebig lange Folge von X**

(gleiche Bedeutung wie bei Wertebereichen)

### Beispieltext in HTML:

```
<table>  
  <tr> <td>Tag</td>  
        <td>Zeit</td>  
        <td>Raum</td></tr>  
  <tr> <td>Mo</td>  
        <td>11:00-12.30</td>  
        <td>AM</td></tr>  
  <tr> <td>Fr</td>  
        <td>9:15-10:45</td>  
        <td>AM</td></tr>  
</table>
```

### Darstellung der Tabelle:

Tag	Zeit	Raum
Mo	11:00-12.30	AM
Fr	9:15-10:45	AM

## Vorlesung Modellierung WS 2000/2001 / Folie 507

### Ziele:

HTML-Ausschnitt verstehen

### in der Vorlesung:

Erläuterungen

- zum \*-Operator (siehe Mod-2.8a),
- zur Struktur von HTML,
- zum Beispiel,
- zur Baumdarstellung