

Kapitel IV Formale Sprachen und Grammatiken

1. Begriffe und Notationen

Sei Σ ein (endliches) Alphabet. Dann

Definition 42

- 1 ist Σ^* das **Monoid** über Σ , d.h. die Menge aller endlichen Wörter über Σ ;
- 2 ist Σ^+ die Menge aller nichtleeren endlichen Wörter über Σ ;
- 3 bezeichnet $|w|$ für $w \in \Sigma^*$ die Länge von w ;
- 4 ist Σ^n für $n \in \mathbb{N}_0$ die Menge aller Wörter der Länge n in Σ^* ;
- 5 eine Teilmenge $L \subseteq \Sigma^*$ eine **formale Sprache**.

2. Sprachkonzept

- Eine (formale) Sprache L besteht aus Wörtern w über einem Alphabet Σ
- Eine formale Sprache L ist also i.a. einfach eine Teilmenge von Σ^* :

$$L \subseteq \Sigma^*$$

- Wir betrachten insbesondere formale Sprachen, deren Wörter entlang gewisser Regeln erzeugt werden
- Diese Regeln werden als Grammatik G für die Sprache bezeichnet
- Die Menge der so bildbaren Wörter heißt Wortschatz oder Sprachschatz der Grammatik

- Eine Grammatik wird durch
 - ① Σ , eine endliche Menge von **Terminalzeichen** (Alphabet)
 - ② V , eine endliche Menge von **Nichtterminalzeichen** (grammatische Begriffe)
 - ③ P , eine endliche Menge von **Produktionen** (**Ableitungs-/Ersetzungsregeln**, grammatische Regeln)
 - ④ S , ein spezielles Element aus V , das **Axiom** oder **Startsymbol**
 vollständig beschrieben.
- Die Vereinigungsmenge $V \cup \Sigma$ der Nichtterminale und Terminale heißt das **Vokabular** der Grammatik.
- V und Σ werden o.B.d.A. als **disjunkt** vorausgesetzt, d.h.

$$V \cap \Sigma = \emptyset$$

- die Produktionen $p \in P$ haben ganz allgemein die Gestalt $A ::= B$, mit $A, B \in (V \cup \Sigma)^*$

Beispiel 43

Wir betrachten folgende Grammatik:

$\langle \text{Satz} \rangle$	$::=$	$\langle \text{Subjekt} \rangle \langle \text{Prädikat} \rangle \langle \text{Objekt} \rangle$
$\langle \text{Subjekt} \rangle$	$::=$	$\langle \text{Artikel} \rangle \langle \text{Attribut} \rangle \langle \text{Substantiv} \rangle$
$\langle \text{Artikel} \rangle$	$::=$	ϵ
$\langle \text{Artikel} \rangle$	$::=$	$\text{der} \text{die} \text{das} \text{ein} \dots$
$\langle \text{Attribut} \rangle$	$::=$	$\epsilon \langle \text{Adjektiv} \rangle \langle \text{Adjektiv} \rangle \langle \text{Attribut} \rangle$
$\langle \text{Adjektiv} \rangle$	$::=$	$\text{gross} \text{klein} \text{schön} \dots$

Die vorletzte Ersetzungsregel ist **rekursiv**, die durch diese **Grammatik** definierte Sprache deshalb unendlich.
Zur Darstellungsform dieser Grammatik **später** mehr!

Beispiel 44

- $L_1 = \{aa, aaaa, aaaaaa, \dots\} = \{(aa)^n, n \in \mathbb{N}\}$
($\Sigma_1 = \{a\}$)
- $L_2 = \{ab, abab, ababab, \dots\} = \{(ab)^n, n \in \mathbb{N}\}$
($\Sigma_2 = \{a, b\}$)
- $L_3 = \{ab, aabb, aaabbb, \dots\} = \{a^n b^n, n \in \mathbb{N}\}$
($\Sigma_3 = \{a, b\}$)
- $L_4 = \{a, b, aa, ab, bb, aaa, aab, abb, bbb, \dots\}$
 $= \{a^m b^n, m, n \in \mathbb{N}_0, m + n > 0\}$ ($\Sigma_4 = \{a, b\}$)

2.1 Erzeugung und Sprachschatz

- Sei G eine Grammatik mit Terminalalphabet Σ , Nichtterminalalphabet V , Produktionen P , $p = A ::= B \in P$ und seien $x, y \in (\Sigma \cup V)^*$
- y heißt **mittels G aus x in einem Schritt ableitbar**, falls es $v, w \in (\Sigma \cup V)^*$ gibt, so dass

$$x = vAw \text{ und } y = vBw$$

- y wird also aus x erzeugt, indem man in x ein Vorkommen der linken Seite von p durch deren rechte Seite ersetzt
- Wir schreiben dafür i.Z. auch

$$x \rightarrow_G y$$

- Überhaupt schreiben wir dementsprechend auch für Produktionen $A ::= B$ oft $A \rightarrow_G B$ (oder nur $A \rightarrow B$)

- Mit \rightarrow^* bezeichnen wir die **reflexive, transitive Hülle** von \rightarrow .
- Es gilt also $x \rightarrow^* y$ gdw es $x^{(0)}, x^{(1)}, \dots, x^{(k)} \in (\Sigma \cup V)^*$ gibt, so dass $k \geq 0$, $x = x^{(0)}$ und $y = x^{(k)}$ und

$$x^{(i)} \rightarrow_G x^{(i+1)} \text{ für alle } i = 0, \dots, k - 1$$

- Falls S das Axiom der Grammatik G ist und $S \rightarrow_G x$ gilt, so nennen wir x auch eine (mittels G erzeugbare) **Sprach-** oder **Satzform**

2.2 Darstellungsformen

2.2.1 Backus-Naur-Form

- Zur Beschreibung der Programmiersprache **ALGOL 60** erstmals benutzt
- Nichtterminale werden durch Winkelklammern kenntlich gemacht, z.B. $\langle \text{Ausdruck} \rangle$, diese können entfallen, wenn weiterhin die Eindeutigkeit der Regeln gegeben ist
- Produktionsregeln mit identischer linker Seite können mit dem „|“ Symbol zusammengefasst werden:
 $A ::= W_1$ und $A ::= W_2$ wird so zu $A ::= W_1|W_2$
- Im Rahmen der Vorlesung werden häufig Kleinbuchstaben für Terminale und Großbuchstaben für Nichtterminale gewählt

- Zur Vereinfachung von rekursiven Produktionen (das zu definierende Nichtterminal steht auch auf der rechten Seite), werden Metasymbole verwendet:
 - () Klammer für Zusammenfassung
 - [] Inhalt der Klammer optional (0-mal oder 1-mal)
 - { } Inhalt der Klammer beliebig oft mal (0,1,2,...)
- Diese erweiterte Notation wird als **ExtendedBNF** oder **EBNF** bezeichnet

Beispiel 45 (Grammatik für die Menge aller arithmetischen Ausdrücke in BNF)

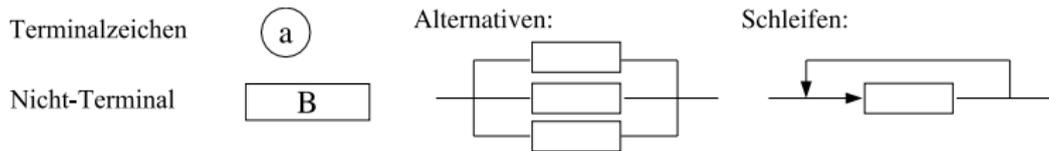
$\langle \text{Terminal} \rangle$	$::=$	$+ \mid - \mid * \mid / \mid \uparrow \mid (\mid) \mid v \mid z$
$\langle \text{Ausdruck} \rangle$	$::=$	$\langle \text{Term} \rangle \mid + \langle \text{Term} \rangle \mid - \langle \text{Term} \rangle \mid$ $\langle \text{Ausdruck} \rangle + \langle \text{Term} \rangle \mid$ $\langle \text{Ausdruck} \rangle - \langle \text{Term} \rangle$
$\langle \text{Term} \rangle$	$::=$	$\langle \text{Faktor} \rangle \mid \langle \text{Term} \rangle * \langle \text{Faktor} \rangle \mid$ $\langle \text{Term} \rangle / \langle \text{Faktor} \rangle$
$\langle \text{Faktor} \rangle$	$::=$	$\langle \text{Elementarausdruck} \rangle \mid$ $\langle \text{Faktor} \rangle \uparrow \langle \text{Elementarausdruck} \rangle$
$\langle \text{Elementarausdruck} \rangle$	$::=$	$z \mid v \mid (\langle \text{Ausdruck} \rangle)$

Einsatz der EBNF-Metasymbole vereinfacht die Produktionsregel für $\langle \text{Faktor} \rangle$ zu

$$\langle \text{Faktor} \rangle ::= \{ \langle \text{Faktor} \rangle \uparrow \} \langle \text{Elementarausdruck} \rangle$$

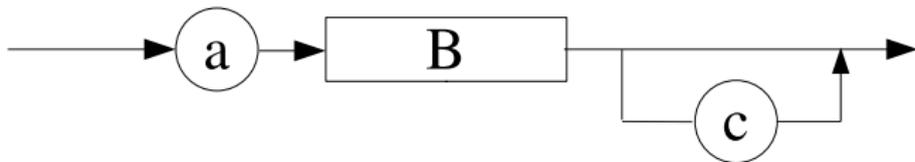
2.2.2 Syntaxdiagramme

- Produktionsregeln der BNF können direkt in ein **Syntaxdiagramm** überführt werden
- Terminalzeichen sind durch Kreise bzw. abgerundete Rechtecke, syntaktischen Variablen durch Rechtecke dargestellt
- Produktionsregeln mit Wiederholungen und Alternativen werden durch Verbindungen zwischen den Elementen dargestellt



Graphische Darstellung der BNF

Beispiel 46



Syntaxdiagramm der Produktion $S ::= aB[c]$

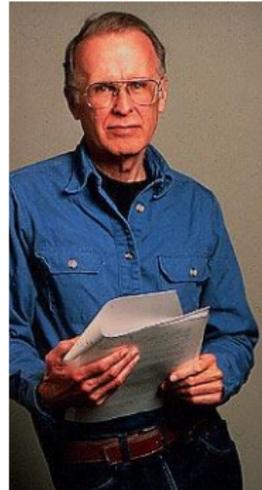
Peter Naur

- ✘ 25. Oktober 1928
- ✘ Dänischer Informatiker
- ✘ Niels Bohr Institute
Technical University of Denmark
Copenhagen University
- ✘ Entwicklung von ALGOL 60
- ✘ 2005 ACM Turing Award



John Backus

- ✘ 3. Dezember 1924 – 17. März 2007
- ✘ Amerikanischer Informatiker
- ✘ IBM-Mitarbeiter
- ✘ Entwicklung von FORTRAN
- ✘ 1977 ACM Turing Award



3. Die Chomsky-Hierarchie

Diese Sprachenhierarchie ist nach **Noam Chomsky** [MIT, 1976] benannt.

3.1 Phrasenstrukturgrammatik, Chomsky-Grammatik

Grammatiken bestehen aus

- 1 einem **Terminalalphabet** Σ (manchmal auch T), $|\Sigma| < \infty$
- 2 einem endlichen Vorrat von **Nichtterminalzeichen** (Variablen)
 $V, V \cap \Sigma = \emptyset$
- 3 einem **Startsymbol** (Axiom) $S \in V$
- 4 einer endliche Menge P von **Produktionen** (Ableitungsregeln)
der Form $l \rightarrow r$ (oder $l ::= r$), mit $l \in (V \cup \Sigma)^*$, $r \in (V \cup \Sigma)^*$

Eine **Phrasenstrukturgrammatik** (Grammatik) ist ein Quadrupel
 $G = (V, \Sigma, P, S)$.

Sei $G = (V, \Sigma, P, S)$ eine Phrasenstrukturgrammatik.

Definition 47

Wir schreiben

- 1 $z \rightarrow_G z'$ gdw
 $(\exists x, y \in (V \cup \Sigma)^*, l \rightarrow r \in P)[z = xly, z' = xry]$
- 2 $z \rightarrow_G^* z'$ gdw $z = z'$ oder
 $z \rightarrow_G z^{(1)} \rightarrow_G z^{(2)} \rightarrow_G \dots \rightarrow_G z^{(k)} = z'$. Eine solche Folge von Ableitungsschritten heißt eine **Ableitung für z' von z in G** (der Länge k).
- 3 Die von G **erzeugte Sprache** ist

$$L(G) := \{z \in \Sigma^*; S \rightarrow_G^* z\}$$

Zur Vereinfachung der Notation schreiben wir gewöhnlich \rightarrow und \rightarrow^* statt \rightarrow_G und \rightarrow_G^*

Vereinbarung:

Wir bezeichnen **Nichtterminale** mit großen und **Terminale** mit kleinen Buchstaben!

Beispiel 48

Wir erinnern uns:

- $L_2 = \{ab, abab, ababab, \dots\} = \{(ab)^n, n \in \mathbb{N}\}$
($\Sigma_2 = \{a, b\}$)
- Grammatik für L_2 mit folgenden Produktionen:

$$S \rightarrow ab, S \rightarrow abS$$

Beispiel 48 (Forts.)

- $L_4 = \{a, b, aa, ab, bb, aaa, aab, abb, bbb \dots\}$
 $= \{a^m b^n, m, n \in \mathbb{N}_0, m + n > 0\} \quad (\Sigma_4 = \{a, b\})$
- Grammatik für L_4 mit folgenden Produktionen:

$$\begin{aligned} S &\rightarrow A, S \rightarrow B, S \rightarrow AB, \\ A &\rightarrow a, A \rightarrow aA, \\ B &\rightarrow b, B \rightarrow bB \end{aligned}$$

3.2 Die Chomsky-Hierarchie

Sei $G = (V, \Sigma, P, S)$ eine Phrasenstrukturgrammatik.

- 1 Jede Phrasenstrukturgrammatik (Chomsky-Grammatik) ist (zunächst) automatisch vom **Typ 0**.
- 2 Eine Chomsky-Grammatik heißt (längen-)monoton, falls für alle Regeln

$$\alpha \rightarrow \beta \in P \text{ mit } \alpha \neq S$$

gilt:

$$|\alpha| \leq |\beta| ,$$

und, falls $S \rightarrow \epsilon \in P$, dann das Axiom S auf keiner rechten Seite vorkommt.

- ③ Eine Chomsky-Grammatik ist vom **Typ 1** (auch: **kontextsensitiv**), falls sie monoton ist und für alle Regeln $\alpha \rightarrow \beta$ in P mit $\alpha \neq S$ gilt:

$$\alpha = \alpha' A \alpha'' \text{ und } \beta = \alpha' \beta' \alpha''$$

für geeignete $A \in V$, $\alpha', \alpha'' \in (V \cup \Sigma)^*$ und $\beta' \in (V \cup \Sigma)^+$.

- ④ Eine Chomsky-Grammatik ist vom **Typ 2** (auch: **kontextfrei**), falls sie monoton ist und für alle Regeln $\alpha \rightarrow \beta \in P$ gilt:

$$\alpha \in V .$$

Bemerkung: Manchmal wird “kontextfrei” auch ohne die Monotonie-Bedingung definiert; **streng monoton** schließt dann die Monotonie mit ein, so dass ϵ nicht als rechte Seite vorkommen kann.

- 5 Eine Chomsky-Grammatik ist vom **Typ 3** (auch: **regulär**, **rechtslinear**), falls sie monoton ist und für alle Regeln $\alpha \rightarrow \beta$ in P mit $\beta \neq \epsilon$ gilt:

$$\alpha \in V \text{ und } \beta \in \Sigma^+ \cup \Sigma^*V .$$

Auch hier gilt die entsprechende Bemerkung zur Monotonie-Bedingung.