

Alphabet: endliche, nicht-leere Menge

- die Elemente werden **Zeichen (Symbole)** genannt
- typische Bezeichner für Alphabete:
 Σ "Sigma", Γ "Gamma", ...

endliches Wort über dem Alphabet Σ : endliche Folge

$$w = a_1 a_2 \dots a_i a_{i+1} \dots a_k \quad \text{von Zeichen } a_i \in \Sigma$$

$|w| = k$ Länge von w

falls $k = 0$: $w = \varepsilon$ "epsilon", leeres Wort

Präfix von w : jedes Wort $a_1 a_2 \dots a_i$ oder ε

Suffix von w : jedes Wort $a_i a_{i+1} \dots a_k$ oder ε

Alphabet: endliche, nicht-leere Menge

Wort über dem Alphabet Σ : endliche Folge

$$w = a_1 a_2 \dots a_k \text{ von Zeichen } a_i \in \Sigma$$

Σ^* = Menge aller Wörter über Σ

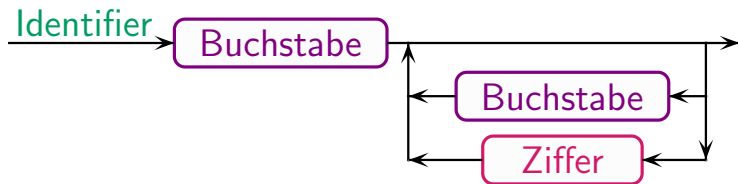
Σ^+ = $\Sigma^* \setminus \{\epsilon\}$ Menge aller nichtleeren Wörter

formale Sprache (kurz Sprache) über Σ :

$$L \subseteq \Sigma^*, \quad \text{d.h., Menge von Wörtern über } \Sigma$$

Jeder **Identifier** besteht aus einem **Buchstaben**, gefolgt von beliebig vielen **Buchstaben** und **Ziffern**.

Syntaxdiagramm:



formale Sprache über dem Alphabet $\Sigma = \mathcal{B} \cup \mathcal{Z}$,

wobei $\mathcal{B} = \{a, b, c, \dots\}$, $\mathcal{Z} = \{0, 1, 2, \dots, 9\}$

$L_{\text{Idf}} = \{w \in \Sigma^+ : \text{erstes Zeichen von } w \text{ ist in } \mathcal{B}\}$

Alphabet: endliche, nicht-leere Menge

Wort über dem Alphabet Σ :

endliche Folge $w = a_1 a_2 \dots a_k$ von Zeichen $a_i \in \Sigma$

Sprache über Σ : $L \subseteq \Sigma^*$, d.h., Menge von Wörtern

Σ^* und alle Sprachen L über Σ sind **abzählbar**

$$\Sigma^* = \text{Menge aller Wörter über } \Sigma = \{\varepsilon\} \cup \underbrace{\bigcup_{k \geq 1} \Sigma^k}_{\text{abzählbar}}$$

Also ist jede Teilmenge L von Σ^* abzählbar.

Seien $L, L_1, L_2 \subseteq \Sigma^*$ Sprachen.

$\varepsilon \in L^0 \subseteq L^*$
für alle Sprachen L

- Vereinigung: $L_1 \cup L_2$
- Durchschnitt: $L_1 \cap L_2$
- Komplement: $\bar{L} = \Sigma^* \setminus L$
- Konkatenation:

$$L_1 \circ L_2 = L_1 L_2 = \{ w_1 w_2 : w_1 \in L_1, w_2 \in L_2 \}$$

- Kleeneabschluss, Stern-Operator: $L^* = \bigcup_{n \geq 0} L^n$
- Plus-Operator:

$$L^+ = \bigcup_{n \geq 1} L^n = \begin{cases} L^* & : \text{ falls } \varepsilon \in L \\ L^* \setminus \{\varepsilon\} & : \text{ falls } \varepsilon \notin L \end{cases}$$

$L = \emptyset$ leere Sprache

$L^* = \{\varepsilon\}$ = Sprache bestehend aus dem leeren Wort

$L = \{aa\}$ = Sprache bestehend aus dem Wort aa

$L^* = \{a^{2n} : n \in \mathbb{N}\} = \{\varepsilon, aa, aaaa, aaaaaa, \dots\}$

$L =$ Menge aller Wörter über $\Sigma = \{a, b\}$, für die gilt:
"Anzahl an a 's \neq Anzahl an b 's"

$L^* = \Sigma^* =$ Menge aller Wörter über $\{a, b\}$

Seien $L_1 \subseteq \Sigma_1^*$, $L_2 \subseteq \Sigma_2^*$, $K \subseteq \Gamma^*$ Sprachen.

Gesetze für Konkatenation, z.B.:

$$(L_1 \cup L_2)K = L_1K \cup L_2K$$

$$K(L_1 \cup L_2) = KL_1 \cup KL_2$$

$$\{\varepsilon\}K = K\{\varepsilon\} = K$$

$$\emptyset K = K\emptyset = \emptyset$$

$w\varepsilon = \varepsilon w = w$
für alle Wörter w

Gesetze für Kleeneabschluss und Plusoperator, z.B.:

$$K^* = (K \setminus \{\varepsilon\})^* = K^+ \cup \{\varepsilon\}$$

$$K^+ = K^*K = KK^*$$

Beispiel: Grammatik für "Identifizier"

116

Jeder **Identifizier** besteht aus einem **Buchstaben**, gefolgt von beliebig vielen **Buchstaben** und **Ziffern**.

$$(1) \text{ Idf} \rightarrow \text{BA}$$

$$(2) \text{ A} \rightarrow \varepsilon \mid \text{BA} \mid \text{ZA}$$

$$(3) \text{ B} \rightarrow \text{a} \mid \text{b} \mid \text{c} \mid \text{d} \mid \dots \mid \text{z}$$

$$(4) \text{ Z} \rightarrow \text{0} \mid \text{1} \mid \dots \mid \text{9}$$

Beispiel für den Wortgenerierungsvorgang:

$$\begin{array}{ccccccc} \text{Idf} & \xrightarrow{(1)} & \text{BA} & \xrightarrow{\text{A} \rightarrow \text{ZA} \quad (2)} & \text{BZA} & \xrightarrow{\text{A} \rightarrow \text{BA} \quad (2)} & \text{BZBA} & \xrightarrow{\text{A} \rightarrow \varepsilon \quad (2)} & \text{BZB} \\ & & & & \xrightarrow{\text{B} \rightarrow \text{a} \quad (3)} & & \xrightarrow{\text{Z} \rightarrow \text{8} \quad (4)} & & \xrightarrow{\text{B} \rightarrow \text{d} \quad (3)} \\ & & & & \text{aZB} & & \text{a8B} & & \text{a8d} \end{array}$$

Eine Grammatik ist ein 4-Tupel $G = (V, \Sigma, \mathcal{P}, S)$

- V endliche Menge von Variablen/Nichtterminalen
- Σ Alphabet, dessen Elemente auch Terminale oder Terminalzeichen genannt werden
- $S \in V$ Startsymbol
- \mathcal{P} Produktionssystem, d.h., endliche Menge von Paaren, auch Regeln genannt

Forderung:
 $V \cap \Sigma = \emptyset$

(x, y) mit $x, y \in (V \cup \Sigma)^*$, wobei $x \notin \Sigma^*$

Schreibweise: $x \rightarrow y$ statt $(x, y) \in \mathcal{P}$

$x \rightarrow y_1 \mid \dots \mid y_n$ statt $\{(x, y_1), \dots, (x, y_n)\} \subseteq \mathcal{P}$

$$G = (V, \Sigma, \mathcal{P}, S)$$

- V Variablenmenge
- Σ Terminalalphabet
- $S \in V$ Startsymbol
- \mathcal{P} Produktionssystem, d.h., endliche Menge von Regeln $x \rightarrow y$ mit $x, y \in (V \cup \Sigma)^*$, $x \notin \Sigma^*$

erzeugte Sprache: Menge aller Wörter w über Σ , die in G aus S hergeleitet werden können:

$$\mathcal{L}(G) \stackrel{\text{def}}{=} \{ w \in \Sigma^* : S \Rightarrow^* w \}$$

↑
Herleitungsrelation in G

Sei $G = (V, \Sigma, \mathcal{P}, S)$ eine Grammatik.

1-Schritt-Herleitungsrelation \Rightarrow

für Wörter $u, v \in (V \cup \Sigma)^*$:

$u \Rightarrow v$ gdw v entsteht aus u durch Anwenden
genau einer Regel

gdw es gibt Wörter $w, z \in (V \cup \Sigma)^*$
und eine Regel $x \rightarrow y$ in G mit

$$u = wxz \quad \text{und} \quad v = wyz$$

Herleitungsrelation \Rightarrow^* : transitive, reflexive Hülle von \Rightarrow

$$\begin{array}{ll} S \rightarrow BA & B \rightarrow b \\ A \rightarrow BA \mid CA \mid \varepsilon & C \rightarrow c \end{array}$$

wobei S, A, B, C Variablen, b, c Terminale

Behauptung: $\mathcal{L}(G) = \{ bx : x \in \{b, c\}^* \}$

Beweis:

“ \supseteq ”: durch Angabe einer Herleitung $S \Rightarrow^* bx$

“ \subseteq ”: zeige, dass jedes aus G herleitbare Wort über dem Terminalalphabet $\Sigma = \{b, c\}$ die Gestalt bx mit $x \in \{b, c\}^*$

hat.