# 7 FORMALE SPRACHEN

Den Begriff der formalen Sprache haben wir schon in Kapitel 4 eingeführt. Eine formale Sprache über einem Alphabet $A$ ist eine Teilmenge $L \subseteq A^*$. In diesem kurzen Kapitel definieren wir noch einige nützliche Operationen auf formalen Sprachen.

## 7.1 OPERATIONEN AUF FORMALEN SPRACHEN

### 7.1.1 Produkt oder Konkatenation formaler Sprachen

Wir haben schon definiert, was die Konkatenation zweier Wörter ist. Das erweitern wir nun auf eine übliche Art und Weise auf Mengen von Wörtern: Für zwei formale Sprachen $L_1$ und $L_2$ heißt

$$L_1 \cdot L_2 = \{w_1 w_2 \mid w_1 \in L_1 \text{ und } w_2 \in L_2\}$$

das *Produkt der Sprachen $L_1$ und $L_2$.* *Produkt von Sprachen*

**7.1 Lemma.** Für jede formale Sprache $L$ ist

$$L \cdot \{\varepsilon\} = L = \{\varepsilon\} \cdot L\,.$$

**7.2 Beweis.** Einfaches Nachrechnen:

$$\begin{aligned} L \cdot \{\varepsilon\} &= \{w_1 w_2 \mid w_1 \in L \text{ und } w_2 \in \{\varepsilon\}\} \\ &= \{w_1 w_2 \mid w_1 \in L \text{ und } w_2 = \varepsilon\} \\ &= \{w_1 \varepsilon \mid w_1 \in L\} \\ &= \{w_1 \mid w_1 \in L\} \\ &= L \end{aligned}$$

Analog zeigt man $L = \{\varepsilon\} \cdot L$. ■

In Abschnitt 4.4.3 haben wir ein erstes Mal über den Aufbau von E-Mails gesprochen. Produkte formaler Sprachen könnte man nun benutzen, um folgende Festlegung zu treffen:

- Die formale Sprache $L_{email}$ der syntaktisch korrekten E-Mails ist

$$L_{email} = L_{header} \cdot L_{leer} \cdot L_{body}$$

- Dabei sei

- $L_{header}$ die formale Sprache der syntaktisch korrekten E-Mail-Köpfe,
- $L_{leer}$ die formale Sprache, die nur die Leerzeile enthält, also $L_{leer} = \{ \boxed{\texttt{CR}}\boxed{\texttt{LF}} \}$ und
- $L_{body}$ die formale Sprache der syntaktisch korrekten E-Mail-Rümpfe.

$L_{header}$ und $L_{body}$ muss man dann natürlich auch noch definieren. Eine Möglichkeit, das bequem zu machen, werden wir in einem späteren Kapitel kennenlernen.

*Potenzen $L^k$*

Wie bei Wörtern will man *Potenzen* $L^k$ definieren. Der „Trick" besteht darin, für den Fall $k = 0$ etwas Sinnvolles zu finden — Lemma 7.1 gibt einen Hinweis. Die Definition geht wieder induktiv:

$$\begin{aligned} L^0 &= \{\varepsilon\} \\ \text{für jedes } k \in \mathbb{N}_0 : \; L^{k+1} &= L \cdot L^k \end{aligned}$$

Wie auch schon bei der Konkatenation einzelner Wörter kann man auch hier wieder nachrechnen, dass z. B. gilt:

$$\begin{aligned} L^1 &= L \\ L^2 &= L \cdot L \\ L^3 &= L \cdot L \cdot L \end{aligned}$$

Genau genommen hätten wir in der dritten Zeile $L \cdot (L \cdot L)$ schreiben müssen. Aber Sie dürfen glauben (oder nachrechnen), dass sich die Assoziativität vom Produkt von Wörtern auf das Produkt von Sprachen überträgt.

Als einfaches Beispiel betrachte man $L = \{\texttt{aa}, \texttt{b}\}$. Dann ist

$$\begin{aligned} L^0 &= \{\varepsilon\} \\ L^1 &= \{\texttt{aa}, \texttt{b}\} \\ L^2 &= \{\texttt{aa}, \texttt{b}\} \cdot \{\texttt{aa}, \texttt{b}\} = \{\texttt{aa} \cdot \texttt{aa}, \texttt{aa} \cdot \texttt{b}, \texttt{b} \cdot \texttt{aa}, \texttt{b} \cdot \texttt{b}\} \\ &= \{\texttt{aaaa}, \texttt{aab}, \texttt{baa}, \texttt{bb}\} \\ L^3 &= \{\texttt{aa} \cdot \texttt{aa} \cdot \texttt{aa},\ \texttt{aa} \cdot \texttt{aa} \cdot \texttt{b},\ \texttt{aa} \cdot \texttt{b} \cdot \texttt{aa},\ \texttt{aa} \cdot \texttt{b} \cdot \texttt{b}, \\ &\qquad \texttt{b} \cdot \texttt{aa} \cdot \texttt{aa},\ \texttt{b} \cdot \texttt{aa} \cdot \texttt{b},\ \texttt{b} \cdot \texttt{b} \cdot \texttt{aa},\ \texttt{b} \cdot \texttt{b} \cdot \texttt{b}\} \\ &= \{\texttt{aaaaaa}, \texttt{aaaab}, \texttt{aabaa}, \texttt{aabb}, \texttt{baaaa}, \texttt{baab}, \texttt{bbaa}, \texttt{bbb}\} \end{aligned}$$

In diesem Beispiel ist $L$ endlich. Man beachte aber, dass die Potenzen auch definiert sind, wenn $L$ unendlich ist. Betrachten wir etwa den Fall

$$L = \{\texttt{a}^n\texttt{b}^n \mid n \in \mathbb{N}_+\} ,$$

es ist also (angedeutet)

$$L = \{\mathtt{ab}, \mathtt{aabb}, \mathtt{aaabbb}, \mathtt{aaaabbbb}, \dots\} .$$

Welche Wörter sind in $L^2$? Die Definion besagt, dass man alle Produkte $w_1 w_2$ von Wörtern $w_1 \in L$ und $w_2 \in L$ bilden muss. Man erhält also (erst mal ungenau hingeschrieben)

$$\begin{aligned} L^2 = \quad & \{\mathtt{ab} \cdot \mathtt{ab}, \mathtt{ab} \cdot \mathtt{aabb}, \mathtt{ab} \cdot \mathtt{aaabbb}, \dots\} \\ \cup & \{\mathtt{aabb} \cdot \mathtt{ab}, \mathtt{aabb} \cdot \mathtt{aabb}, \mathtt{aabb} \cdot \mathtt{aaabbb}, \dots\} \\ \cup & \{\mathtt{aaabbb} \cdot \mathtt{ab}, \mathtt{aaabbb} \cdot \mathtt{aabb}, \mathtt{aaabbb} \cdot \mathtt{aaabbb}, \dots\} \\ & \vdots \end{aligned}$$

Mit anderen Worten ist

$$L^2 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2} \mid n_1 \in \mathbb{N}_+ \text{ und } n_2 \in \mathbb{N}_+\} .$$

Man beachte, dass bei die Exponenten $n_1$ „vorne“ und $n_2$ „hinten“ verschieden sein dürfen.

Für ein Alphabet $A$ und für $i \in \mathbb{N}_0$ hatten wir auch die Potenzen $A^i$ definiert. Und man kann jedes Alphabet ja auch als eine formale Sprache auffassen, die genau alle Wörter der Länge 1 über $A$ enthält. Machen Sie sich klar, dass die beiden Definitionen für Potenzen konsistent sind, d. h. $A^i$ ergibt immer die gleiche formale Sprache, egal, welche Definition man zu Grunde legt.

### 7.1.2 Konkatenationsabschluss einer formalen Sprache

Bei Alphabeten hatten wir neben den $A^i$ auch noch $A^*$ definiert und darauf hingewiesen, dass für ein Alphabet $A$ gilt:

$$A^* = \bigcup_{i \in \mathbb{N}_0} A^i .$$

Das nehmen wir zum Anlass nun den *Konkatenationsabschluss $L^*$ von $L$* und den *ε-freien Konkatenationsabschluss $L^+$ von $L$* definieren:

*Konkatenationsabschluss $L^*$ von $L$*
*ε-freier Konkatenationsabschluss $L^+$ von $L$*

$$L^+ = \bigcup_{i \in \mathbb{N}_+} L^i \qquad \text{und} \qquad L^* = \bigcup_{i \in \mathbb{N}_0} L^i$$

Wie man sieht, ist $L^* = L^0 \cup L^+$. In $L^*$ sind also alle Wörter, die sich als Produkt einer beliebigen Zahl (einschließlich 0) von Wörtern schreiben lassen, die alle Element von $L$ sind.

Als Beispiel betrachten wieder $L = \{\mathtt{a}^n\mathtt{b}^n \mid n \in \mathbb{N}_+\}$. Weiter vorne hatten wir schon gesehen:

$$L^2 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2} \mid n_1 \in \mathbb{N}_+ \text{ und } n_2 \in \mathbb{N}_+\} .$$

Analog ist

$$L^3 = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\mathtt{a}^{n_2}\mathtt{b}^{n_2}\mathtt{a}^{n_3}\mathtt{b}^{n_3} \mid n_1 \in \mathbb{N}_+ \text{ und } n_2 \in \mathbb{N}_+ \text{ und } n_3 \in \mathbb{N}_+\} .$$

Wenn wir uns erlauben, Pünktchen zu schreiben, dann ist allgemein

$$L^i = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\cdots\mathtt{a}^{n_i}\mathtt{b}^{n_i} \mid n_1 \ldots, n_i \in \mathbb{N}_+\} .$$

Und für $L^+$ könnte man vielleicht notieren:

$$L^+ = \{\mathtt{a}^{n_1}\mathtt{b}^{n_1}\cdots\mathtt{a}^{n_i}\mathtt{b}^{n_i} \mid i \in \mathbb{N}_+ \text{ und } n_1 \ldots, n_i \in \mathbb{N}_+\} .$$

Aber man merkt (hoffentlich!) an dieser Stelle doch, dass uns $^+$ und $^*$ die Möglichkeit geben, etwas erstens präzise und zweitens auch noch kürzer zu notieren, als wir es sonst könnten.

Zum Abschluss wollen wir noch darauf hinweisen, dass die Bezeichnung $\varepsilon$-freier Konkatenationsabschluss für $L^+$ leider (etwas?) irreführend ist. Wie steht es um das leere Wort bei $L^+$ und $L^*$? Klar sollte inzwischen sein, dass für *jede* formale Sprache $L$ gilt: $\varepsilon \in L^*$. Das ist so, weil ja $\varepsilon \in L^0 \subseteq L^*$ ist. Nun läuft zwar in der Definition von $L^+$ die Vereinigung der $L^i$ nur ab $i = 1$. Es kann aber natürlich sein, dass $\varepsilon \in L$ ist. In diesem Fall ist dann aber offensichtlich $\varepsilon \in L = L^1 \subseteq L^+$. Also kann $L^+$ sehr wohl das leere Wort enthalten.

Außerdem sei erwähnt, dass die Definition von $L^*$ zur Folge hat, dass gilt:

$$\{\}^* = \{\varepsilon\}$$

## 7.2 ZUSAMMENFASSUNG

In diesem Kapitel wurden *Produkt* und der *Konkatenationsabschluss formaler Sprachen* eingeführt.

Wir haben gesehen, dass man damit jedenfalls manche formalen Sprachen kurz und verständlich beschreiben kann. Dass diese Notationsmöglichkeiten auch in der Praxis Verwendung finden, wird später noch deutlich werden. Manchmal reicht das, was wir bisher an Notationsmöglichkeiten haben, aber noch nicht. Deshalb werden wir in späteren Kapiteln auch mächtigere Hilfsmittel kennenlernen.