5.3 Alphabet, Wort und Sprache

Ihre Handy-PIN besteht aus vier Ziffern, und 0000 zählt genauso wie 4711. Erlaubt sind die zehn Zeichen 0 bis 9, jede PIN ist eine Folge aus genau vier davon. In der Sprache dieses Abschnitts: Das Ziffernfeld ist das Alphabet, jede PIN ein Wort der Länge 4, und die PINs, die Ihr Handy entsperren, bilden eine sehr kleine Sprache. Wie viele PINs es gibt, rechnen Sie mit dem Satz weiter unten in einer Zeile aus.

Definition (Alphabet). Ein **Alphabet** A ist eine endliche, geordnete Menge von Zeichen.

Beispiele: $A = \{0, 1\}$ für Dualzahlen, $A = \{a, \ldots, z, 0, \ldots, 9\}$ für Bezeichner, für Pascal die Menge aller erlaubten Zeichen samt Schlüsselwörtern.

Definition (Wort). Ein **Wort** über A ist eine endliche Folge von Zeichen aus A. Das **leere Wort** wird mit ε bezeichnet und hat die Länge 0.

Für $A = \{a, b, c\}$ sind a, ab, cba, $aaaa$ und ε Wörter.

Bezeichnung. Die Menge aller endlich langen Wörter über A heißt A^*, die Menge aller Wörter der Länge n heißt A^n.

$$A^0 = \{\varepsilon\} \qquad A^1 = \{a, b, c\} \qquad A^2 = \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$$

Satz. Ist A ein Alphabet mit k Zeichen, so hat A^n genau k^n Elemente. A^* hat unendlich viele Elemente — aber abzählbar unendlich viele, denn man kann die Wörter der Länge nach und innerhalb gleicher Länge alphabetisch aufzählen.

Definition (formale Sprache). Jede Teilmenge $L \subseteq A^*$ heißt **formale Sprache** über A.

Diese Definition ist großzügig: Jede beliebige Auswahl von Wörtern ist eine Sprache. Interessant werden nur die Sprachen, die sich *endlich beschreiben* lassen — durch eine Regel, eine Grammatik oder einen Automaten.

Arbeitsauftrag 3

7. $A = \{0, 1\}$. Geben Sie A^2 und A^3 vollständig an. Wie viele Elemente hat A^8?
8. $A = \{x, y, z\}$. Notieren Sie A^1 und die ersten zehn Wörter von A^3.
9. $A = \{a, b, c\}$. Welche der folgenden Mengen sind formale Sprachen über A? $L_4 = \{\}$, $L_5 = \{\varepsilon\}$, $L_6 = \{x\}$, $L_7 = \{a\}$. Begründen Sie.
10. Beschreiben Sie in Worten: $L_{10} = \{b, ab, aab, aaab, \ldots\}$ und $L_{11} = \{b, ba, baa, baaa, \ldots\}$.
11. $A = \{a, b, c\}$. Geben Sie die Sprache aller Wörter der Länge 3 an, deren mittleres Zeichen ein c ist. Wie viele Wörter sind das?