5.5 Reguläre Grammatiken

Wer ein Kennzeichen wie WND-AB 123 vorliest, geht Zeichen für Zeichen von links nach rechts. Merken muss er sich dabei nur eines: ob er gerade im Kreiskürzel, bei den Buchstaben oder bei den Ziffern ist. Er blättert nie zurück und muss nichts Späteres vorwegnehmen. Grammatiken, die so gebaut sind — ein Zeichen anhängen, dann weiter —, heißen regulär.

Definition. Eine Grammatik heißt **regulär**, wenn jede Produktion die Form

$$A \to xB \qquad \text{oder} \qquad A \to x$$

hat, mit $A, B \in V_N$ und $x \in V_T$. Rechts steht also **genau ein Terminalzeichen** und dahinter höchstens *ein* Nichtterminal — und wenn, dann *ganz rechts*. Die erzeugte Sprache heißt **reguläre Sprache**.

Stolperstein

Die Definition ist enger, als sie beim Querlesen wirkt. Nicht erlaubt sind $A \to xyB$ (zwei Terminalzeichen), $A \to Bx$ (Nichtterminal links) und $A \to \varepsilon$ (leeres Wort). Umgekehrt gilt: Eine Grammatik, die diese Form *nicht* hat, kann durchaus eine reguläre Sprache erzeugen — man kann sie dann in die reguläre Form umschreiben. „Regulär" ist also eine Eigenschaft der *Grammatik*; „reguläre Sprache" eine Eigenschaft der *Sprache*. Beides auseinanderzuhalten lohnt sich.

Die Klammergrammatik aus 5.4 ist nach dieser Definition nicht regulär — bei $S \to (S)$ steht das Nichtterminal in der Mitte. Das ist kein Zufall, wie sich zeigen wird.

Beispiel: gerade Dualzahlen. Gesucht ist eine Grammatik für alle Dualzahlen ohne führende Null, die gerade sind — die also mit 1 beginnen und auf 0 enden.

$$V_T = \{0, 1\}, \quad V_N = \{S, A\}, \quad P = \{\, S \to 1A, \quad A \to 0A \mid 1A \mid 0 \,\}$$

Ableitung von 1010: $S \Rightarrow 1A \Rightarrow 10A \Rightarrow 101A \Rightarrow 1010$

> In der Webfassung steht hier ein Bedienfeld zum Ausprobieren: **Ableitungswerkzeug**.

Geben Sie im Bedienfeld auch 101 ein — eine ungerade Zahl. Es gibt keine Ableitung, und das Werkzeug sagt es.

Beispiel: Bezeichner. In vielen Sprachen gilt: Ein Bezeichner beginnt mit einem Buchstaben, danach folgen beliebig viele Buchstaben oder Ziffern. Mit $V_T = \{bu, zi\}$ (Buchstabe, Ziffer):

$$V_N = \{S, R\}, \qquad P = \{\, S \to bu \mid bu\,R, \quad R \to bu\,R \mid zi\,R \mid bu \mid zi \,\}$$

Arbeitsauftrag 5

17. Prüfen Sie, ob die Grammatik $P = \{S \to aS \mid bS \mid aaA,\ A \to aA \mid bA \mid \varepsilon\}$ die obige Definition erfüllt — prüfen Sie jede Produktion einzeln. Beschreiben Sie anschließend $L(G)$ in Worten und geben Sie eine Grammatik an, die dieselbe Sprache erzeugt und die Definition erfüllt.
18. Geben Sie eine reguläre Grammatik für alle ungeraden Dualzahlen an und leiten Sie 1011 und 1001 ab.
19. Gibt es eine reguläre Grammatik für $L = \{10, 100, 110, 1000, 1010, 1100, 1110, \ldots\}$? Beschreiben Sie die Sprache zuerst in Worten.
20. Begründen Sie, warum $S \to aSb$ nicht regulär ist.