5.8 Grammatik und Automat sind zwei Seiten derselben Sache

Eine Bauanleitung für ein Klemmbausteinmodell sagt, wie man es Schritt für Schritt *aufbaut*. Wer ein fertiges Modell vor sich hat, kann umgekehrt Stein für Stein *prüfen*, ob es nach dieser Anleitung gebaut ist. Beides beschreibt dasselbe Modell, einmal erzeugend, einmal prüfend. Für reguläre Sprachen sind Grammatik und Automat genau dieses Paar.

Zwischen regulären Grammatiken und endlichen Automaten lässt sich mechanisch hin- und herübersetzen.

Von der Grammatik zum Automaten. Jedes Nichtterminal wird ein Zustand, das Startsymbol der Startzustand. Aus $A \to xB$ wird der Übergang $f(A, x) = B$. Aus $A \to x$ wird ein Übergang in einen zusätzlichen Endzustand E.

Beispiel mit $P = \{S \to 1A,\ A \to 0A \mid 1A \mid 0\}$ (gerade Dualzahlen). Die Regeln $A \to 0A$ und $A \to 0$ beginnen beide mit einer 0 — der Zustand A hat bei 0 also **zwei** Nachfolger:

Zustand	bei 0	bei 1	Endzustand
S	–	$\{A\}$	nein
A	$\{A, E\}$	$\{A\}$	nein
E	–	–	**ja**

Stolperstein

Die Konstruktion liefert im Allgemeinen einen **NEA**, keinen DEA — immer dann, wenn zu einem Nichtterminal zwei Regeln mit demselben Terminalzeichen gehören. Das ist kein Mangel: Mit der Teilmengenkonstruktion aus 5.6.2 wird daraus ein DEA.

Hier von Hand nachgebaut: Aus $\{A, E\}$ wird ein DEA-Zustand, den wir E' nennen. Er ist Endzustand, weil er E enthält, und verhält sich sonst wie A:

Zustand	bei 0	bei 1	Endzustand
S	–	A	nein
A	E'	A	nein
E'	E'	A	**ja**

Probe: 1010 läuft $S \to A \to E' \to A \to E'$ — akzeptiert. 1011 endet in A und wird abgelehnt, richtig, denn 1011 ist ungerade.

Stolperstein

In der Tabelle steht bei S und der 0 ein Strich: Dieser Übergang kommt in der Grammatik nicht vor, weil keine Dualzahl mit 0 beginnt. Streng genommen verlangt die Definition aus 5.6 aber eine **auf ganz** $Z \times A$ **definierte** Übergangsfunktion. Man ergänzt deshalb einen zusätzlichen **Fehlerzustand** z_F, der nie Endzustand ist und alle Zeichen auf sich selbst abbildet; jeder Strich führt dorthin. Gezeichnet wird er meist weggelassen, weil er das Bild unübersichtlich macht. Im Programm später entspricht ihm der Zweig „Übergang nicht in der Tabelle → abweisen".

Vom Automaten zur Grammatik. Umgekehrt wird jeder Zustand ein Nichtterminal, der Startzustand das Startsymbol. Aus $f(Z_1, x) = Z_2$ wird die Produktion $Z_1 \to xZ_2$ — und ist Z_2 ein **Endzustand**, so kommt zusätzlich die abschließende Regel $Z_1 \to x$ hinzu.

Die naheliegende Abkürzung $Z \to \varepsilon$ für jeden Endzustand wäre bequem, ist nach unserer Definition aus 5.5 aber nicht erlaubt: Dort darf rechts nur xB oder x stehen, niemals das leere Wort. Deshalb der Umweg über die zusätzliche Regel.

Merke

Reguläre Grammatik, endlicher Automat und regulärer Ausdruck (Abschnitt 5.11) beschreiben genau dieselbe Klasse von Sprachen — die regulären Sprachen. Drei Schreibweisen, ein Begriff.

Ein Sonderfall bleibt: Nach unserer Definition erzeugt jede Produktion mindestens ein Zeichen, das leere Wort ε ist also in keiner so erzeugten Sprache enthalten. Ein Automat kann es dagegen akzeptieren — nämlich dann, wenn sein Startzustand ein Endzustand ist. Für diesen einen Fall erlaubt man zusätzlich die Regel $S \to \varepsilon$ für das Startsymbol. Mit ihr stimmen die drei Beschreibungen dann ohne Ausnahme überein.

Gymnasium Ottweiler