

## Grammatiken zur Beschreibung formaler Sprachen

Die Wörter, die zu einer formalen Sprache gehören, können wir angeben, indem wir alle Wörter aufzählen. Es gibt jedoch auch formale Sprachen, die sehr viele oder sogar unendlich viele Wörter umfassen. In diesen Fällen können wir versuchen Regeln zu finden, nach denen die Wörter einer Sprache aufgebaut sind, um sie zu beschreiben. Je komplexer die Regeln sind, desto schwieriger wird es jedoch, den Aufbau der Sprache in unserer Alltagssprache zu formulieren. Eine umgangssprachliche Formulierung der Regeln ist zudem für einen Computer schwer zu verstehen und daher als Grundlage für einen Algorithmus, der überprüft, ob ein Wort die Regeln einer formalen Sprache erfüllt, ungeeignet. Im Folgenden lernen wir mit den *Grammatiken* daher ein Instrument der theoretischen Informatik kennen, mit denen eine formale Sprache eindeutig beschrieben werden kann.

**Aufgabe 1:** Nennen Sie jeweils Beispiele:

- für formale Sprachen, die sich durch Aufzählen aller Wörter angeben lassen
- für formale Sprachen, die aus unendlich vielen Wörtern bestehen
- für formale Sprachen, bei denen sich der Aufbau der Wörter leicht umgangssprachlich beschreiben lässt
- für formale Sprachen, bei denen die Regeln, nach denen die Wörter aufgebaut sind, komplex sind

## Grammatiken zum Erzeugen formaler Sprachen

Papier, Hefte, Briefumschläge usw. haben einheitliche Größen, die in verschiedenen DIN-Formaten festgelegt sind. Die Menge aller DIN-Formate können wir als formale Sprache auffassen:

$L_{\text{DIN}} = \{A0, A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, B0, B1, B2, B3, B4, B5, B6, B7, B8, B9, B10, C0, C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, D0, D1, D2, D3, D4, D5, D6, D7, D8, D9, D10\}$

Die Sprache der DIN-Formate wurde hier beschrieben, indem alle Formate aufgezählt wurden.

**Aufgabe 2:**

- Formulieren Sie die Regeln, nach denen die Wörter der Sprache der DIN-Formate aufgebaut sind, umgangssprachlich.
- Geben Sie das Alphabet  $\Sigma_{\text{DIN}}$  für die formale Sprache der DIN-Formate an.

Schauen wir uns nun für das Beispiel der formalen Sprache der DIN-Formate an, wie eine Grammatik zur Beschreibung einer formalen Sprache aussieht. Eine *Grammatik* besteht aus vier Komponenten: einer endlichen Menge  $V$  von *Nichtterminalsymbolen*, einer endlichen Menge  $\Sigma$  von *Terminalsymbolen*, den *Produktionen*  $P$  und einem *Startsymbol*. Die *Terminalsymbole* sind die Zeichen, aus denen die Wörter der Sprache aufgebaut sind. Die Menge der Terminalsymbole entspricht also dem Alphabet  $\Sigma$  der Sprache. Die *Nichtterminalsymbole* sind Zeichen, die nicht zum Alphabet  $\Sigma$  gehören. Eines dieser Nichtterminalsymbole ist das *Startsymbol*. Die *Produktionen* sind Ersetzungsregeln, die angeben, durch welche Zeichen ein Nichtterminalsymbol ersetzt werden darf. Um ein Wort der formalen Sprache zu erzeugen, beginnen wir mit dem Startsymbol und wenden so lange Ersetzungsregeln an, bis das Wort nur noch aus Terminalsymbolen besteht.

Eine Grammatik  $G_{\text{Din}}$ , welche die formale Sprache der DIN-Formate beschreibt, könnte also so aussehen:  $G_{\text{Din}} = (V, \Sigma, P, S)$  mit

$$V = \{S, T, Z\}$$

$$\Sigma = \{A, B, C, D, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Produktionen P:

$$S \rightarrow TZ \mid T10$$

$$T \rightarrow A \mid B \mid C \mid D$$

$$Z \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$

Der senkrechte Strich  $|$  ist dabei als *oder* zu lesen. Die Produktion  $S \rightarrow TZ \mid T10$  ist also eine Kurzschreibweise für die Produktionen

$$S \rightarrow TZ$$

$$S \rightarrow T10$$

Wenn wir jetzt beginnend mit dem Startsymbol  $S$  alle Kombinationen von Ersetzungen, die gemäß der Produktionsregeln erlaubt sind, nacheinander durchführen, erhalten wir alle Wörter, die zur formalen Sprache der DIN-Formate gehören:

$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A0$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B0$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C0$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D0$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A1$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B1$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C1$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D1$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A2$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B2$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C2$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D2$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A3$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B3$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C3$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D3$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A4$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B4$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C4$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D4$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A5$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B5$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C5$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D5$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A6$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B6$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C6$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D6$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A7$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B7$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C7$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D7$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A8$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B8$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C8$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D8$
$S \Rightarrow TZ \Rightarrow AZ \Rightarrow A9$	$S \Rightarrow TZ \Rightarrow BZ \Rightarrow B9$	$S \Rightarrow TZ \Rightarrow CZ \Rightarrow C9$	$S \Rightarrow TZ \Rightarrow DZ \Rightarrow D9$
$S \Rightarrow T10 \Rightarrow A10$	$S \Rightarrow T10 \Rightarrow B10$	$S \Rightarrow T10 \Rightarrow C10$	$S \Rightarrow T10 \Rightarrow D10$

Wichtig ist, dass eine Grammatik eine formale Sprache nur dann erzeugt, wenn sich mit den Produktionen der Grammatik genau die Wörter vom Startsymbol ableiten lassen, die zu der formalen Sprache gehören. Das heißt einerseits, dass man jedes Wort, das zur Sprache gehört, aus dem Startsymbol ableiten kann. Das heißt aber auch, dass es nicht möglich sein darf, zusätzliche Wörter aus dem Startsymbol abzuleiten, die nicht zur Sprache gehören.

Dieser Zusammenhang wird so ausgedrückt: Die formale Sprache  $\mathcal{L}(G)$ , die von der Grammatik erzeugt wird, ist die Menge aller Wörter über dem Alphabet  $\Sigma$ , die sich in endlich vielen Schritten durch das Anwenden von Ersetzungsregeln (Produktionen) aus dem Startsymbol  $S$  ableiten lassen.

Formal schreibt man:  $\mathcal{L}(G) = \{\omega \in \Sigma^* \mid S \Rightarrow^* \omega\}$

Dabei kann es für eine formale Sprache unterschiedliche Grammatiken geben, ähnlich wie es zum Lösen eines Problems unterschiedliche Algorithmen gibt.

**Aufgabe 3:** Entscheiden Sie für die folgenden Grammatiken jeweils, ob es sich um eine Grammatik handelt, welche die formale Sprache der DIN-Formate erzeugt.

- a)  $G_a = (V, \Sigma, P, S)$  mit  
 $V = \{S, T, Z\}$   
 $\Sigma = \{A, B, C, D, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$   
Produktionen P:  
 $S \rightarrow TZ \mid TZZ$   
 $T \rightarrow A \mid B \mid C \mid D$   
 $Z \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$
- b)  $G_b = (V, \Sigma, P, S)$  mit  
 $V = \{S, Z\}$   
 $\Sigma = \{A, B, C, D, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$   
Produktionen P:  
 $S \rightarrow AZ \mid BZ \mid CZ \mid DZ$   
 $Z \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$
- c)  $G_c = (V, \Sigma, P, S)$  mit  
 $V = \{E, S, Z\}$   
 $\Sigma = \{A, B, C, D, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$   
Produktionen P:  
 $S \rightarrow AZ \mid BZ \mid CZ \mid DZ$   
 $Z \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \mid 1E$   
 $E \rightarrow 0$

Schauen wir uns als zweites Beispiel die formale Sprache der positiven Dezimalzahlen an. Als Trennzeichen zwischen dem ganzzahligen Teil und dem gebrochenen Teil verwenden wir ein Komma. Außerdem sollen führende Nullen vermieden werden.

- $G_{\text{Dezi}} = (V, \Sigma, P, A)$  mit  
 $V = \{A, B, N, Z\}$   
 $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ', '\}$   
Produktionen P:  
 $A \rightarrow Z \mid Z,N \mid 0,N \mid 0$   
 $Z \rightarrow 1B \mid 2B \mid 3B \mid 4B \mid 5B \mid 6B \mid 7B \mid 8B \mid 9B$   
 $N \rightarrow 0B \mid 1B \mid 2B \mid 3B \mid 4B \mid 5B \mid 6B \mid 7B \mid 8B \mid 9B$   
 $B \rightarrow BB \mid 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \mid \varepsilon$

Neu ist hier das Symbol  $\varepsilon$  in der letzten Ersetzungsregel  $B \rightarrow \varepsilon$ . Das Symbol  $\varepsilon$  steht für das leere Wort. Die Ersetzungsregel besagt also, dass B ersatzlos gestrichen werden kann.

Die Zahl 1,782 können wir daher wie folgt aus dem Startsymbol ableiten:

$A \Rightarrow Z,N \Rightarrow 1B,N \Rightarrow 1,N \Rightarrow 1,7B \Rightarrow 1,7BB \Rightarrow 1,78B \Rightarrow 1,782$

**Aufgabe 4:**

- a) Geben Sie für die Ableitung des Wortes 1,782 an, welche Ersetzungsregel der Grammatik  $G_{\text{Dezi}}$  jeweils angewendet wurde.
- b) Leiten Sie die folgenden Wörter aus dem Startsymbol A ab:  
1) 123                      2) 0,08                      3) 40,3
- c) Begründen Sie, dass sich das Wort 03,6 nicht ableiten lässt.
- d) Bei einer Dezimalzahl können vor und hinter dem Komma beliebig viele, theoretisch sogar unendlich viele Ziffern stehen. Erläutern Sie, wie diese Eigenschaft in den Produktionen abgebildet wird.
- e) Ändern Sie die Produktionen so, dass das Symbol  $\varepsilon$  nicht benötigt wird.

**Aufgabe 5:** Deutsche Postleitzahlen bestehen aus genau fünf Ziffern. Geben Sie eine Grammatik an, welche die formale Sprache der deutschen Postleitzahlen erzeugt. Die Sprache soll dabei alle syntaktisch korrekten Postleitzahlen umfassen, unabhängig davon, ob diese Postleitzahl vergeben ist.

**Aufgabe 6:** Die Raumnummern einer Firma, die auf drei Gebäudekomplexe verteilt ist, werden von der folgenden Grammatik  $G_{NR}$  erzeugt.

$G_{NR} = (V, \Sigma, P, S)$  mit

$V = \{D, S, Z\}$

$\Sigma = \{A, B, C, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Produktionen P:

$S \rightarrow DZZZ$

$D \rightarrow A \mid B \mid C$

$Z \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

- Geben Sie drei Beispiele für gültige Raumnummern inklusive ihrer Ableitung vom Startsymbol an.
- Entscheiden Sie für die folgenden Wörter, ob es sich um gültige Raumnummern handelt:
  - B012
  - A13
  - D263
- Beschreiben Sie die Regeln, nach denen die Raumnummern aufgebaut sind, in Worten.

**Aufgabe 7:** Abbildung 1 zeigt den Aufbau der Kennzeichnungen von Hühnereiern. Die siebenstellige Betriebsnummer beinhaltet eine Kennziffer des Bundeslandes und eine Stallnummer.

Entwerfen Sie eine Grammatik, welche die formale Sprache der Kennzeichnungen von Hühnereiern aus Deutschland erzeugt.

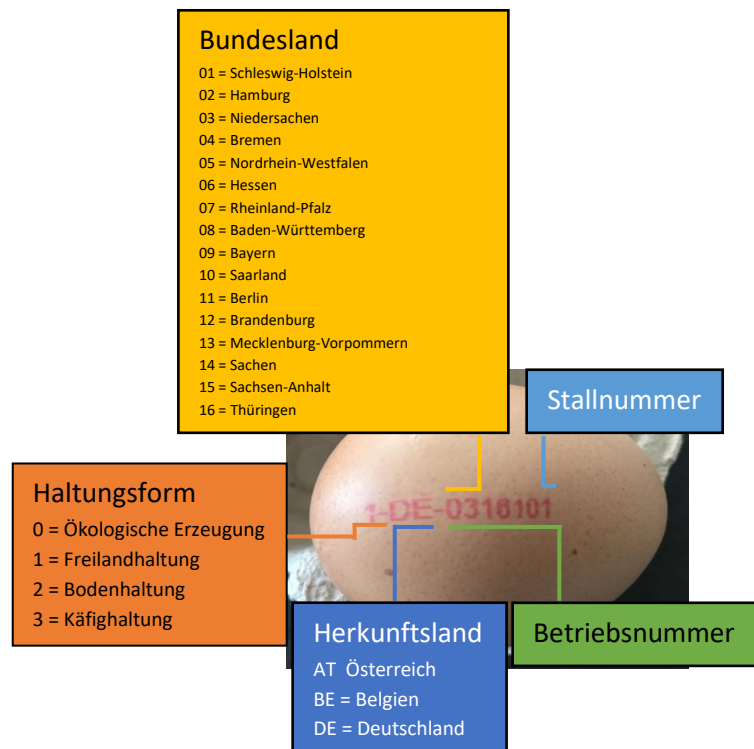


Abbildung 1: Kennzeichnung von Hühnereiern

**Aufgabe 8:** Tabelle 1 zeigt den Anfang der homologen Reihe der Alkene.<sup>1</sup>

- Beschreiben Sie den Aufbau der Halbstrukturformel der Alkene in Worten.
- Entwerfen Sie eine Grammatik, welche die formale Sprache der Halbstrukturformeln der Alkene erzeugt.

Name	Halbstrukturformel	Summenformel
Ethen	$\text{CH}_2=\text{CH}_2$	$\text{C}_2\text{H}_4$
Propen	$\text{CH}_2=\text{CH}-\text{CH}_3$	$\text{C}_3\text{H}_6$
Buten	$\text{CH}_2=\text{CH}-\text{CH}_2-\text{CH}_3$	$\text{C}_4\text{H}_8$
Penten	$\text{CH}_2=\text{CH}-\text{CH}_2-\text{CH}_2-\text{CH}_3$	$\text{C}_5\text{H}_{10}$
Hexen	$\text{CH}_2=\text{CH}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3$	$\text{C}_6\text{H}_{12}$
Hepten	$\text{CH}_2=\text{CH}-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_3$	$\text{C}_7\text{H}_{14}$

Tabelle 1: Chemische Formeln der Alkene

- Ändern sie Ihre Grammatik so, dass nur die Halbstrukturformeln der Alkene *Ethen* bis *Decen* ( $\text{C}_{10}\text{H}_{20}$ ) in der formalen Sprache enthalten sind.
- Geben Sie eine allgemeine mathematische Formel an, die den Aufbau der Summenformel der Alkene beschreibt. Begründen Sie, dass es deutlich schwieriger ist, eine entsprechende Grammatik anzugeben.
- Als Vorstufe soll die Summenformel der Ethene als Reihe aus Cs und Hs dargestellt werden. Beispiel:  $\text{C}_2\text{H}_4$  wird zu CCHHHH. Entwerfen sie eine entsprechende Grammatik.

**Aufgabe 9:** Bei dem Spiel „Onkel Otto sitzt in der Badewanne und trinkt Tee“ wird dieser Satz auf ein Blatt geschrieben und so gefaltet, dass man immer nur ein Wort sehen kann. Der Zettel wird nun im Kreis herumgereicht und jeder Mitspieler bzw. jede Mitspielerin schreibt ein beliebiges Wort der gleichen Wortart auf. Spielerperson 1 schreibt also unter *Onkel* einen beliebigen Titel, der vor einem Namen stehen kann, z. B. Tante, Professor, Frau und klappt die Spalte anschließend um. Spielerperson 2 schreibt unter *Otto* einen beliebigen Namen und klappt auch diese Spalte um. Spielerperson 3 schreibt unter *sitzt* ein beliebiges Verb in der dritten Person Singular usw. Dadurch entstehen dann mehr oder weniger sinnvolle, aber dafür vielleicht umso lustigere Fantasiessätze.

- Entwerfen Sie eine Grammatik, mit der sich Sätze nach dem Muster „Onkel Otto sitzt in der Badewanne und trinkt Tee“ erzeugen lassen. Da es sich bei dem deutschen Satz um ein Wort der formalen Sprache handelt, die von der Grammatik erzeugt wird, dürfen Sie in diesem Fall ein deutsches Wort als ein Terminalsymbol und damit als ein Zeichen des Alphabets betrachten. Beschränken Sie sich für jede Wortart auf drei Möglichkeiten.
- Zeigen sie an zwei Beispielen, dass ihre Grammatik Sätze nach dem Muster „Onkel Otto sitzt in der Badewanne und trinkt Tee“ erzeugt. Welche Probleme treten hier ggf. auf?
- Erläutern Sie anhand Ihrer Grammatik den Unterschied zwischen Syntax und Semantik.
- Implementieren Sie Ihre Grammatik als Programm, das Zufallssätze nach dem Muster „Onkel Otto sitzt in der Badewanne und trinkt Tee“ erzeugt.
- Diskutieren Sie, welche Schwierigkeiten bei der Beschreibung einer natürlichen Sprache durch eine Grammatik der Form  $G = (V, \Sigma, P, S)$  auftreten.

<sup>1</sup> Die Tabelle und die Fortsetzung der Reihe können unter <https://www.msa-berlin.de/chemie/kohlenwasserstoffe-alkane-alkene-alkine-1/> eingesehen werden. [Datum des Zugriffs: 17.08.2020]

## Kontextfreie vs. Kontextsensitive Grammatiken

Die Produktionen der Grammatiken, die wir uns bislang angeschaut haben, waren immer so aufgebaut, dass auf der linken Seite ein einzelnes Nichtterminalsymbol steht. Das heißt, die Ersetzungsregeln gelten unabhängig davon, welche Zeichen gerade links und rechts im Wort neben dem Nichtterminalsymbol stehen. Grammatiken, deren Ersetzungsregeln auf der linken Seite immer nur ein einzelnes Nichtterminalsymbol enthalten, heißen daher *kontextfrei*. Die formalen Sprachen, die von *kontextfreien Grammatiken* erzeugt werden können, heißen entsprechend *kontextfreie Sprachen*.

In Aufgabe 9 lassen sich die Zufallssätze mit einer kontextfreien Grammatik beschreiben. Um nur grammatikalisch korrekte Sätze zu erhalten, benötigt man deutlich mehr Regeln, welche nur Kombinationen von Wörtern erzeugen, die zulässig sind: beispielweise den grammatikalisch richtigen Artikel zu einem Substantiv. Betrachtet man beliebige Sätze einer natürlichen Sprache, hängen Teile eines Satzes manchmal auch von anderen Satzteilen ab und es ergeben sich komplexe ineinander verschränkte Strukturen, die sich nur mit Regeln beschreiben lassen, die auch berücksichtigen, was links und rechts von einem Nichtterminal steht, so dass eine kontextfreie Grammatik zur Beschreibung nicht mehr ausreicht.

Um den Unterschied zwischen kontextfreien und kontextsensitiven Grammatiken zu verdeutlichen, betrachten wir hier eine überschaubare Sprache, für die es nicht möglich ist, eine kontextfreie Grammatik zu finden: Einfachzucker, die Monosaccharide, haben die Summenformel  $C_nH_{2n}O_n$  mit  $n \in \mathbb{N}$ . Schreibt man die Summenformeln wie in Aufgabe 8e aus, z. B.  $C_2H_4O_2$  als CCHHHOO, so wird es nicht gelingen für die formale Sprache aller ausgeschriebenen Summenformeln der Einfachzucker eine kontextfreie Grammatik zu finden. Probieren Sie es aus!

Die Sprache lässt sich nur mit einer *kontextsensitiven Grammatik* beschreiben, die z. B. so aussehen könnte:  $G_{\text{Zucker}} = (V, \Sigma, P, S)$  mit

$$V = \{S, 1, 2, 3\}^2$$

$$\Sigma = \{C, H, O\}$$

Produktionen P:

$$S \rightarrow S1223 \mid CHHO$$

$$31 \rightarrow 13$$

$$O1 \rightarrow 1O$$

$$C1 \rightarrow CC$$

$$21 \rightarrow 12$$

$$O2 \rightarrow 2O$$

$$H2 \rightarrow HH$$

$$32 \rightarrow 23$$

$$H1 \rightarrow 1H$$

$$O3 \rightarrow OO$$

### Aufgabe 10:

- Leiten Sie das Wort CCCHHHHHOOO aus dem Startsymbol *S* ab.
- Erläutern Sie, worin sich die Produktionsregeln einer kontextsensitiven Grammatik von den Produktionsregeln einer kontextfreien Grammatik unterscheiden.

Die meisten formalen Sprachen, die uns im Alltag begegnen und sogar Programmiersprachen, wie Java, Processing, Python oder C++, sind kontextfrei. Wie Aufgabe 9 gezeigt hat, lassen sich sogar viele grammatikalische Konstrukte natürlicher Sprachen mit kontextfreien Produktionen beschreiben. Daher werden wir die kontextsensitiven Sprachen hier nicht weiter thematisieren.

<sup>2</sup> Bitte beachten Sie, dass hier als Nichtterminalsymbole Zahlen verwendet werden, um sie leichter von den Terminalsymbolen unterscheiden zu können.

## Reguläre Grammatiken

Innerhalb der kontextfreien Grammatiken gibt es eine Teilmenge, die *regulären Grammatiken*, für deren Produktionen man noch genauere Vorgaben machen kann. Auf der rechten Seite der Produktionen steht entweder nur ein Terminal oder ein Terminal gefolgt von einem Nichtterminal. Dadurch erfolgt eine Veränderung des Wortes beim Ableiten immer nur ganz rechts. Man nennt diese Grammatiken daher auch rechtslinear. Die formalen Sprachen, die von einer regulären bzw. rechtslinearen Grammatik erzeugt werden, nennt man *reguläre Sprachen*. Bei der Grammatik aus Aufgabe 3c) zur Beschreibung der formalen Sprache der DIN-Formate handelt es sich z. B. um eine reguläre (rechtslineare) Grammatik:  $G_c = (V, \Sigma, P, S)$  mit

$$V = \{S, Z\}$$

$$\Sigma = \{A, B, C, D, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Produktionen P:

$$S \rightarrow AZ \mid BZ \mid CZ \mid DZ$$

$$Z \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \mid 1E$$

$$E \rightarrow 0$$

Die formale Sprache der DIN-Formate ist somit eine reguläre Sprache. Die erste Grammatik  $G_{\text{DIN}}$ , die wir für diese Sprache betrachtet haben, war nicht regulär. Daran sehen wir, dass es ausreicht, wenn wir eine reguläre Grammatik finden können, welche die Sprache erzeugt, damit eine Sprache zu den regulären Sprachen zählt.

Manchmal werden die Grammatiken regulärer Sprachen auch als linkslineare Grammatiken angegeben. Dann besteht die rechte Seite aus einem Nichtterminal gefolgt von einem Terminalsymbol oder nur aus einem Terminalsymbol. Entscheidend ist, dass eine reguläre Grammatik nicht gleichzeitig rechts- und linkslineare Produktionen enthalten darf.

**Aufgabe 11:** Wandeln Sie die rechtslineare Grammatik für die reguläre Sprache der DIN-Formate in eine linkslineare Grammatik um.

**Aufgabe 12:** Zeigen Sie, dass es sich bei der Sprache der Raumnummern aus Aufgabe 6, um eine reguläre Sprache handelt, indem Sie die Produktionen so verändern, dass es sich um eine reguläre (rechtslineare) Grammatik handelt.

**Aufgabe 13:** In der Mathematik werden Klammern verwendet, um die Reihenfolge, in der die Rechenoperationen angewendet werden sollen, festzulegen.

**Beispiel:**  $3 \cdot (4 + 8 - (6 + 2) + 5) - (8 + 9)$

- Entwerfen Sie eine kontextfreie Grammatik, welche die formale Sprache der geklammerten Rechenausdrücke erzeugt. Gehen Sie vereinfachend davon aus, dass alle Zahlen einstellig sind und nur die Operatoren  $+$ ,  $-$  und  $\cdot$  verwendet werden.
- Nicht nur bei Rechenausdrücken, auch bei vielen Programmiersprachen ist es wichtig, dass Klammern korrekt geschachtelt werden und zu jeder öffnenden eine schließende Klammer existiert. Im Folgenden konzentrieren wir uns daher auf die Klammern. Gesucht ist eine kontextfreie Grammatik, welche die formale Sprache der korrekten Klammerungen erzeugt. Beispiele für korrekte Klammerungen:  $()()()$ ,  $((()))$ ,  $((()))$   
Beispiele für nicht korrekte Klammerungen:  $)$ ,  $((()$ ,  $()()$   
Entwerfen Sie eine entsprechende Grammatik.
- Können Sie für die formale Sprache der korrekten Klammerungen eine reguläre Grammatik entwerfen? Erläutern Sie, welche Probleme hier auftreten.

**Aufgabe 14:** Gegeben ist die folgende Grammatik  $G_K = (V, \Sigma, P, A)$  mit

$$V = \{A, B\}$$

Produktionen P:

$$A \rightarrow [B$$

$$T = \{[, ]\}$$

$$B \rightarrow A] \mid ]$$

- Leiten Sie aus dem Startsymbol A drei verschiedene Wörter ab.
- Beschreiben Sie die Sprache, welche die Grammatik  $G_K$  erzeugt, in Worten.
- Handelt es sich bei der formalen Sprache, welche die Grammatik  $G_K$  erzeugt, um eine reguläre Sprache? Begründen Sie.

**Aufgabe 15:** Entscheiden Sie für die folgenden formalen Sprachen jeweils, ob es sich um eine reguläre Sprache handelt.

- Deutsche Postleitzahlen
- Kennzeichnung von Hühnereiern
- Halbstrukturformel der Alkene
- Summenformeln der Ethene als Reihe aus Cs und Hs

### Exkurs: Ableitungsbäume

Die Ableitung eines Wortes aus dem Startsymbol lässt sich auch als Baumstruktur darstellen. Dabei steht das Startsymbol in der Wurzel. Die inneren Knoten enthalten jeweils ein Nichtterminalsymbol. Die Terminalsymbole stehen in den Blättern und ergeben von links nach rechts aneinandergereiht das abgeleitete Wort. Abbildung 2 zeigt die Ableitung des Wortes A103 für die Grammatik  $G_{NR}$  aus Aufgabe 6 als Ableitungsbaum.

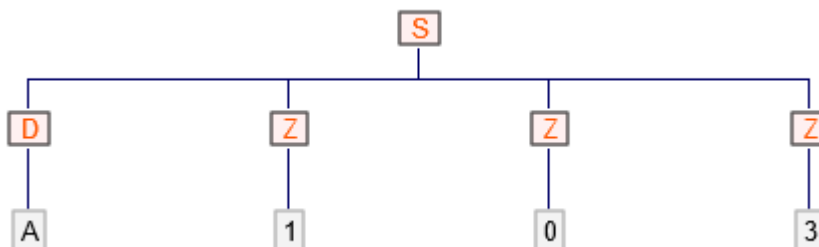


Abbildung 2: Ableitungsbaum des Wortes A103 entsprechend der Produktionen der Grammatik  $G_{NR}$

**Aufgabe 16:** Stellen Sie die Ableitung der Dezimalzahl 34,234 für die Grammatik  $G_{Dezi}$  (s. S. 3) als Ableitungsbaum dar.

**Aufgabe 17:** Abbildung 3 zeigt den Ableitungsbaum für das Wort B10 gemäß den Regeln der Grammatik  $G_C$  für die formale Sprache der DIN-Formate (s. S. 3, Aufg. 3).

- Zeichnen Sie den Ableitungsbaum, der sich ergibt, wenn das Wort B10 mithilfe der Grammatik  $G_{DIN}$  erzeugt wird.
- Vergleichen Sie die Ableitungsbäume für das Wort B10 und erläutern Sie mithilfe des Ableitungsbaums aus Abbildung 3 den Begriff *rechtslineare Grammatik*.

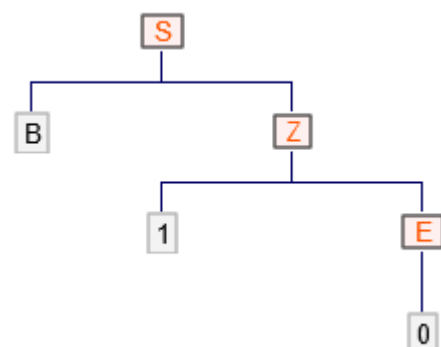


Abbildung 3: Ableitungsbaum für das Wort B10 und die Grammatik  $G_C$

## Hinweis

Die Materialien erheben keinen Anspruch auf Vollständigkeit hinsichtlich der für die Abiturprüfung erwarteten Kompetenzen. Verbindlich für das Abitur in Niedersachsen sind allein das niedersächsische Kerncurriculum für die gymnasiale Oberstufe sowie die ergänzenden Hinweise in der jeweils aktuellen Fassung.

## Lizenz

Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen 4.0 International Lizenz](#). Von der Lizenz ausgenommen ist das InfSII-Logo.

**Abbildungsnachweise:** Abbildung 1 wurde von der Autorin selbst erstellt. Abbildungen 2 und 3 wurden mithilfe der Software FLACI.com erzeugt. FLACI.com wird von der Pädagogischen Hochschule Schwyz kostenlos für Bildungszwecke zur Verfügung gestellt: <https://flaci.com>

## Quellen

D. W. Hoffmann (2011). *Theoretische Informatik*. (2. Aufl.) Carl Hanser Verlag München.

Die Liste der DIN-Formate basiert auf den Angaben auf: R. Sachs. *Übersicht über die Maße der Papierformate DIN A, DIN B, DIN C und DIN D, Größen in Millimeter mm und Zentimeter cm* <https://www.din-formate.de/uebersicht-tabelle-masse-din-a-b-c-d-din-lang-groessen-zb-a4-b4-c4-d4-zusammenfassung-papierformate-in-mm-cm.html> (Abgerufen: 3. Juli 2020)

Die Informationen in Abbildung 1 basieren auf: A. Waschneck (2020). Ostern aus dem Supermarkt: Bunte Eier lieber nicht kaufen. Kennzeichnung von Hühnereiern. <https://web.de/magazine/ratgeber/essen-trinken/ostern-supermarkt-bunte-eier-lieber-nicht-kaufen-34591082> (Abgerufen: 3. Juli 2020)

Die Informationen in Aufgabe 8 und Tabelle 1 basieren auf: A. Angelov (Hrsg.) Mittleren Schulabschluss (MSA) nachholen: *Kohlenwasserstoffe: Alkane, Alkene und Alkine*. <https://www.msa-berlin.de/chemie/kohlenwasserstoffe-alkane-alkene-alkine-1/> (Abgerufen: 3. Juli 2020)

Summenformel der Monosaccharide: Seite „Monosaccharide“. In: Wikipedia, Die freie Enzyklopädie. Bearbeitungsstand: 22. Juni 2020. URL: <https://de.wikipedia.org/w/index.php?title=Monosaccharide&oldid=201212018> (Abgerufen: 3. Juli 2020)