Regel- und Symbolkomplexität kontextfreier Sprachen unter ausgewählten Operationen

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von Dipl.-Inform. Ronny Harbich

geb. am 06.03.1985 in Stendal

Gutachterinnen/Gutachter

Prof. Dr. Jürgen Dassow
Prof. Dr. Martin Kutrib
Prof. Dr. Friedrich Otto

Magdeburg, den 08.11.2019

Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik

Dissertation

Regel- und Symbolkomplexität kontextfreier Sprachen unter ausgewählten Operationen

Dipl.-Inform. Ronny Harbich

8. November 2019

Betreut von Prof. Dr. rer. nat. habil. Jürgen Dassow

Prosodia – Verlag für Musik und Literatur

Autor:	Dipl.-Inform. Ronny Harbich `https://ronny-harbich.de/`, `info@ronny-harbich.de`
Betreuer:	Prof. Dr. rer. nat. habil. Jürgen Dassow `http://theo.cs.ovgu.de/dassow_eng.html`
Literaturverlag:	Prosodia – Verlag für Musik und Literatur `https://prosodia.de/`, `info@prosodia.de`
Typographie:	Digitale Dienstleistungen Ronny Harbich `https://falconiform.de/`, `info@falconiform.de`
Satzsysteme:	LaTeX, KOMA-Script `https://www.latex-project.org/`, `https://www.komascript.de/`
Druck, Bindung:	Druckmanufaktur Hartmut Holz e. K. `https://druckmanufaktur.biz/`, `info@druckmanufaktur.biz`

1. Auflage (2020)

Printed in Germany
ISBN: 978-3-945469-08-8

Vorwort

Neben der gedruckten Fassung dieser Arbeit liegen drei Versionen in digitaler Form unter den folgenden Adressen und QR-Codes vor:

Tablet-Version:	`https://ronny-harbich.de/dissertation-tablet.pdf`
Desktop-Version:	`https://ronny-harbich.de/dissertation-desktop.pdf`
A4-Druck-Version:	`https://ronny-harbich.de/dissertation-a4-druck.pdf`

Tablet-Version   Desktop-Version   A4-Druck-Version

Der Arbeit liegt ein loses tabellarisches Satz-/Lemma-/Definitions-Verzeichnis bei, welches der Leserin / dem Leser zum schnellen Nachschlagen parallel zum eigentlichen Text dient. Es ist zudem im Anhang aufgeführt.

Zusammenfassung

Wir untersuchen *kontextfreie Sprachen* in Bezug auf die Beschreibungskomplexitäten *Regel-* und *Symbolkomplexität* nach Anwendung ausgewählter Sprachoperationen (z. B. *Vereinigung*).

Kontextfreie Sprachen werden durch *kontextfreie Grammatiken* allein durch *Platzhalter/Variablen*, *Buchstaben*, *Ersetzungsregeln* und eine für kontextfreie Grammatiken festgelegte Vorgabe der Regelanwendung erzeugt.

Eine *regel-minimale (kontextfreie) Grammatik* ist eine kontextfreie Grammatik, die die geringste Regelanzahl für die kontextfreie Sprache L, die sie erzeugt, hat, bezogen auf alle kontextfreien Grammatiken, die L erzeugen.

Bei Konstruktion einer regel-minimalen Grammatik G, die eine Sprache L erzeugt, gibt es die Möglichkeit, L in andere Sprachen zu zerlegen, die unter Anwendung gewisser Operationen wieder L ergeben. So könnte L derart in Sprachen L_1 und L_2 zerlegt werden, dass mithilfe der Operation Vereinigung gerade $L = L_1 \cup L_2$ gelte. Anschließend werden für L_1 und L_2 regel-minimale Grammatiken G_1 bzw. G_2 konstruiert. Die Regelkomplexitäten von G, G_1 und G_2 können nun verglichen werden: So könnte die addierte Anzahl der Regeln von G_1 und G_2 größer sein als die von G – oder umgekehrt oder gleich.

Wir zeigen in dieser Arbeit, welche Komplexitätsänderungen für regel-minimale Grammatiken bei der Sprachzerlegung prinzipiell möglich sind: Es sei $\mathrm{Prod}(L)$ die Anzahl an Regeln, die eine regel-minimale Grammatik benötigt, um eine kontextfreie Sprache L zu erzeugen. Für zwei natürliche Zahlen n und m ist dann

$$\begin{aligned}\mathrm{C}_{\cup}^{\mathrm{Prod}}(n,m) = \{&\mathrm{Prod}(L_1 \cup L_2) \mid \mathrm{Prod}(L_1) = n, \mathrm{Prod}(L_2) = m,\\ &\qquad L_1 \text{ und } L_2 \text{ sind kontextfreie Sprachen}\}\end{aligned}$$

die Menge möglicher Regelkomplexitäten, die zwei beliebige kontextfreie Sprachen mit der minimalen Regelanzahl n bzw. m unter Vereinigung $\cup$ haben können. Dies lässt sich analog für weitere Operationen definieren. Für $m \leq n$ sind die folgenden Aussagen ein Teil der wesentlichen Ergebnisse dieser Arbeit:

- Unter der Operation *Spiegelbild* gilt

$$\mathrm{C}_{()^{\mathrm{R}}}^{\mathrm{Prod}}(n) = \{n\}.$$

- Unter der Operation *Vereinigung* gilt

$$\begin{aligned}6, \ldots, n+m+2 \in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,m) &= \mathrm{C}_{\cup}^{\mathrm{Prod}}(m,n)\\ &\not\ni n+m+3, \ldots \text{ für } n \geq 2, m \geq 2.\end{aligned}$$

- Unter der Operation *Konkatenation* gilt

$$
\begin{aligned}
n+2,\dots,n+m+1 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m) &= \mathrm{C}^{\mathrm{Prod}}_{\cdot}(m,n) \\
&\not\ni n+m+2,\dots \text{ für } n \geq 5, m \geq 5.
\end{aligned}
$$

- Unter der Operation *Kleene-Abschluss* gilt

$$
\mathrm{C}^{\mathrm{Prod}}_{()^*}(n) = \begin{cases} \{1\} & \text{für } n = 0 \\ \{1,2\} & \text{für } n = 1 \\ \{2,\dots,n+2\} & \text{sonst.} \end{cases}
$$

- Unter der Operation *Homomorphismus* (Menge aller Homomorphismen Hom) gilt

$$
\mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}}(n) = \begin{cases} \{0\} & \text{für } n = 0 \\ \{1,\dots,n\} & \text{sonst.} \end{cases}
$$

- Unter der Operation *Inverser Homomorphismus* (Menge aller inversen Homomorphismen Hom^{-1}) gelten

$$
\begin{aligned}
\{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}^{-1}}(0), \\
\mathbb{N} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}^{-1}}(n) \text{ für } n \geq 1.
\end{aligned}
$$

- Unter der Operation *Schnitt mit regulärer Sprache* (Menge aller Schnitte mit regulärer Sprache Reg) gelten

$$
\begin{aligned}
\{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(0), \\
\{0,1\} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(1), \\
\mathbb{N}_0 &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(n) \text{ für } n \geq 2.
\end{aligned}
$$

Zuvor Genanntes untersuchen wir zudem analog für *symbol-minimale Grammatiken* – hier werden die Anzahl der Platzhalter/Variablen und Buchstaben in den Regeln als Maß verwendet. Wir erhalten für die Mengen möglicher Symbolkomplexitäten unter Sprachoperationen im Wesentlichen – bezogen auf die Mengen möglicher Regelkomplexitäten unter Sprachoperationen – ähnliche Ergebnisse.

Inhaltsverzeichnis

1. Einführung

In diesem Kapitel erläutern wir zunächst in Abschnitt 1.1 die Beweggründe dieser Arbeit. Hierbei führen wir die Relevanz dieser Arbeit auf praktische Gegebenheiten zurück. Im Abschnitt 1.2 zeigen wir, wie diese Arbeit in die aktuelle Forschung einzuordnen ist. Außerdem wird dargelegt, welche Fragen sie offen lassen wird.

1.1. Motivation

Programme für Computer werden in mathematisch exakt definierten Programmiersprachen geschrieben – denn im Gegensatz zu natürlichen Sprachen sind hier nicht-eindeutige Ausdrucksweisen unerwünscht. Eine Möglichkeit der Beschreibung dieser formalen Sprachen bietet das mathematische Modell der formalen Grammatiken. Diese legen eine formale Sprache allein durch Platzhalter/Variablen, Buchstaben, Ersetzungsregeln und eine für kontextfreie Grammatiken fixierte Vorgabe der Regelanwendung fest. Formale Grammatiken, die zur Beschreibung von Programmiersprachen oder Teilen von ihnen dienen, können üblicherweise zu sogenannten kontextfreien Grammatiken eingeschränkt werden, das heißt, die Ersetzungsregeln haben dann eine gewisse, einfachere Form als im allgemeinen Fall. Lässt sich eine formale Sprache durch eine kontextfreie Grammatik erzeugen, so wird sie als kontextfrei bezeichnet. Kontextfreie Grammatiken sind derart einfach, dass es effiziente Algorithmen gibt ([Sch01], [Ear83]), die bei Eingabe einer kontextfreien Grammatik G und eines geschriebenen Programms P entscheiden, ob P syntaktisch korrekt ist, es also tatsächlich nur solche sprachlichen Konstruktionen und Ausdrücke enthält, die in der Programmiersprache (die durch G beschriebene kontextfreie Sprache) erlaubt sind. Diese Algorithmen lösen damit das sogenannte Wort- oder Mitgliedsproblem für kontextfreie Sprachen.

Neben der zuvor betrachteten Anwendung kontextfreier Grammatiken finden wir sie in der Computerlinguistik als Hilfsmittel zur Beschreibung natürlicher Sprachen. Wo auch immer Grammatiken verwendet werden, müssen diese gespeichert werden, genauer: die Platzhalter/Variablen, Buchstaben und Ersetzungsregeln. Da praktische Speicher – wie zum Beispiel der Arbeitsspeicher eines Computers – begrenzt sind, gelangt man schnell zur Frage, ob es vielleicht eine andere Grammatik geben könnte, welche dieselbe Sprache beschriebe, aber deren Ersetzungsregeln insgesamt aus weniger Zeichen/Symbolen bestünden, sodass weniger Speicher benötigt würde. Eventuell hängt auch die Laufzeit eines Algorithmus, der eine kontextfreie Grammatik als Eingabe hat, von der Anzahl der Ersetzungsregeln ab – dann wäre es sicher gut zu wissen, ob nicht eine äquivalente Grammatik existiert, die mit weniger Ersetzungsregeln auskäme, und sich damit die Laufzeit des Algorithmus verringerte. Um das Optimum zu erhalten, wird man daher versuchen, eine Grammatik zu finden, die die geringste Symbolanzahl in den Regeln oder die geringste Regelanzahl hat – eine sogenannte symbol-minimale bzw.

regel-minimale Grammatik. Wie in [Gru72] gezeigt wurde, kann unerfreulicherweise kein Algorithmus existieren, der zu einer gegebenen kontextfreien Grammatik eine weitere kontextfreie Grammatik findet, die dieselbe Sprache erzeugt und symbol-minimal oder regel-minimal ist – das Problem ist unentscheidbar.

Bei der Konstruktion einer Grammatik G, die eine Sprache L erzeugt, ist es möglicherweise hilfreich, L in andere Sprachen zu zerlegen, die unter Anwendung gewisser Operationen wieder L ergeben. So könnte etwa L derart in Sprachen L_1 und L_2 zerlegt werden, dass mithilfe der Operation Vereinigung gerade $L = L_1 \cup L_2$ gelte. Anschließend werden für L_1 und L_2 Grammatiken G_1 bzw. G_2 konstruiert. Im Fall der kontextfreien Grammatiken kann die oben erwähnte Überprüfung, ob ein Programm syntaktisch korrekt ist, dann so gelöst werden, dass erst die Grammatik G_1 in den Algorithmus eingegeben wird und anschließend G_2. Das Programm ist nur dann korrekt, wenn der Algorithmus für eine der beiden Grammatiken oder für beide Grammatiken angibt, dass das Programm syntaktisch korrekt ist. Trotz der Zerlegung von L können wir also den Algorithmus – leicht verändert – weiterhin verwenden.

Das Festlegen einer Sprache durch mehrere Sprachen und Operationen wird praktisch besonders im Rahmen der regulären Ausdrücke angewendet. Diese erzeugen Sprachen, welche durch reguläre Grammatiken (weitere Form einfacherer formaler Grammatiken) definiert sind ([Sch01]). Reguläre Ausdrücke werden in vielen Programmiersprachen und Texteditoren verwendet, um Zeichenketten innerhalb eines Texts zu finden (pattern matching).

Betrachten wir nun die symbol- oder regel-minimalen Grammatiken von zerlegten Sprachen, so stellt sich die Frage, ob durch die Zerlegung nicht deutlich mehr Speicherplatz verbraucht würde oder Algorithmen nicht deutlich länger liefen als ohne. So könnte etwa eine Sprache L durch regel-minimale kontextfreie Grammatiken G_1, G_2 und Vereinigung leichter beschrieben werden, aber G_1 und G_2 könnten zusammen deutlich mehr Regeln als eine regel-minimale kontextfreie Grammatik G haben, die L direkt erzeugt. Damit wir entscheiden können, ob wir entweder G oder G_1 und G_2 verwenden sollen, müsste die Minimalität händisch bewiesen werden – es ist nämlich offen, ob ein Algorithmus existiert, der feststellen kann, ob eine kontextfreie Grammatik symbol- oder regel-minimal ist ([Gru72]). Um diese nicht zufriedenstellende Situation etwas zu verbessern, zeigen wir in dieser Arbeit anhand ausgewählter Operationen und konkreter Sprachen, welche Einsparungen oder Verschlechterungen für kontextfreie Grammatiken bei Zerlegung kontextfreier Sprachen prinzipiell möglich sind.

Wir stellen die Kernaufgabe dieser Arbeit im Folgenden etwas formaler dar: Es seien L eine kontextfreie Sprache und G eine regel-minimale kontextfreie Grammatik, die L erzeugt. Dann ist $\mathrm{Prod}(L)$ – genannt Regelkomplexität von L – als die Anzahl der Regeln in G definiert. Der Ausdruck $\mathrm{Prod}(L)$ ist also die minimale Anzahl an Regeln, die eine kontextfreie Grammatik benötigt, um L zu erzeugen. Weiterhin sei für zwei natürliche Zahlen n und m

$$\begin{aligned}\mathrm{C}_{\cup}^{\mathrm{Prod}}(n,m) = \{\mathrm{Prod}(L_1 \cup L_2) \mid {} & \mathrm{Prod}(L_1) = n, \mathrm{Prod}(L_2) = m, \\ & L_1 \text{ und } L_2 \text{ sind kontextfreie Sprachen}\}\end{aligned}$$

die Menge möglicher Regelkomplexitäten, die zwei beliebige kontextfreie Sprachen mit der minimalen Regelanzahl n bzw. m unter Vereinigung $\cup$ haben können. Kennten wir die Menge $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ für alle natürlichen Zahlen n und m, so wüssten wir für zwei kontextfreie Sprachen L_1 und L_2 immerhin, welche Regelkomplexitäten für $L_1 \cup L_2$ infrage kämen. Im weiteren Verlauf werden wir unter anderem

$$6, \ldots, n+m+2 \in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m) \not\ni 0, 1, n+m+3, \ldots$$

für $n \geq 2$ und $m \geq 2$ erhalten – es ist noch nicht bekannt, ob $2, \ldots, 5 \in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ gilt.

In dieser Arbeit beschränken wir uns nicht auf die Bestimmung der Menge möglicher Regelkomplexitäten $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$, ferner suchen wir analog diese Menge mit den Operationen: Spiegelbild, Konkatenation, Kleene-Abschluss, Homomorphismus, inverser Homomorphismus und Schnitt mit regulärer Sprache, statt, wie zuvor definiert, mit der Operation Vereinigung. Weiterhin betrachten wir diese Mengen nicht nur in Bezug auf die Regelkomplexität, sondern entsprechend auch in Bezug auf die Symbolkomplexität $\mathrm{Symb}(L)$ – die Symbolanzahl der Regeln, die eine symbol-minimale kontextfreie Grammatik hat, um L zu erzeugen. Im Abschnitt 5 sind hierzu die Ergebnisse zusammengefasst.

Wählen wir aus $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ die größte Zahl $\max(\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m))$ – sofern der Raum der Regelkomplexität $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ endlich und nicht leer ist –, so kennen wir die maximale Regelkomplexität (obere Schranke), die nach Vereinigung zweier kontextfreien Sprachen mit den Regelkomplexitäten n bzw. m auftreten kann. Für die Vereinigung werden wir später konkret

$$\max\left(\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)\right) = n+m+2$$

für $n \geq 2$ und $m \geq 2$ herausfinden. Für die zuvor erwähnten Operationen und die Symbolkomplexität werden wir ebenso Untersuchungen bezüglich der oberen Schranken anstellen.

Wie gesehen, betrachten wir hier jeweils nur eine Operation. Relevanter wäre jedoch das Beachten beliebig – aber endlich – vieler kombinierter Operationen. So wird das Beschreiben einer Programmiersprache allein durch zwei kontextfreie Sprachen und Vereinigung kaum sinnvoll sein. Selbiges gilt für die oben genannten regulären Ausdrücke – sie bestehen in praktischen Systemen aus vielen Operationen. Diese Arbeit liefert also nur Ergebnisse für die einfache Anwendung. Wir werden sehen, dass bereits bei der Verwendung genau einer Operation recht komplexe Beweise verwendet werden, um die Regel- und Symbolkomplexitäten zu bestimmen. Vermutlich lässt sich diese Arbeit nicht einfach für beliebig viele Operationen erweitern, obschon sie einige allgemeine Resultate liefert.

1.2. Einordnung

Wir haben bereits erwähnt, dass reguläre Sprachen, welche eine echte Teilmenge der kontextfreien Sprachen sind ([Sch01]), in Form regulärer Ausdrücke in der Software-Entwicklung eingesetzt werden. Damit können zum Beispiel Texteingaben von Benutzern auf syntaktische Gültigkeit überprüft werden. Insbesondere kann mithilfe des regulären

Ausdrucks (nach POSIX-Spezifikation, [IEE04])

```
^[a-zA-Z0-9+-._]+@[a-zA-Z0-9-.]+$
```

durch Lösung des Wortproblems überprüft werden, ob eine eingegebene E-Mail-Adresse das korrekte Format hat (nur approximativ). Hierzu wird der reguläre Ausdruck in einen (deterministischen) endlichen Automaten ([Sch01]) überführt und anschließend das Wortproblem für endliche Automaten gelöst, indem die Buchstaben des Eingabewortes Stück für Stück in den Automaten eingeben werden und sich damit der Zustand des Automaten solange ändert (oder nicht), bis alle Buchstaben eingeben wurden – befindet sich der Automat dann in einem sogenannten Endzustand, ist das Eingabewort, die E-Mail-Adresse, korrekt, ansonsten nicht.

Wie [Yu00] begründet, ist es im Falle endlicher Automaten sinnvoll, die Anzahl der Zustände als Komplexitätsmaß zu verwenden. Genauer: Die Zustandskomplexität $\mathrm{Zust}(R)$ einer regulären Sprache R ist die Anzahl der Zustände, die ein endlicher Automat mit möglichst wenigen Zuständen hat, der R akzeptiert. Die Gründe, dieses Komplexitätsmaß zu verwenden, sind im Wesentlichen die, die wir bereits für die Symbol- und Regelkomplexität kontextfreier Sprachen angeben haben.

Im Abschnitt 1.1 haben wir schon erläutert, weswegen die Komplexitäten in Bezug auf Operationen untersucht werden sollten. Und in der Tat finden wir im regulären Ausdruck zur E-Mail-Adressprüfung von oben indirekt die Operationen positiver Kleene-Abschluss – durch `+`-Zeichen ausgedrückt – und Konkatenation – durch kein Zeichen dargestellt (die Symbole werden einfach hintereinander weggeschrieben und das Zeichen für die Konkatenation wird weggelassen). Für die Zustandskomplexität lässt sich also analog die Menge

$$\begin{aligned}\mathrm{C}_{\cup}^{\mathrm{Zust}}(n,m) = \{\mathrm{Zust}(R_1 \cup R_2) \mid \mathrm{Zust}(R_1) = n, \mathrm{Zust}(R_2) = m,\\ R_1 \text{ und } R_2 \text{ sind reguläre Sprachen}\}\end{aligned}$$

für zwei natürliche Zahlen n und m definieren. Statt der Vereinigung $\cup$ können entsprechend Mengen möglicher Zustandskomplexität unter weiteren Operationen festgelegt werden.

Wir geben im Folgenden die maximalen Zustandskomplexitäten unter einigen Operationen an:

- Spiegelbild in [SWY04]: $\max(\mathrm{C}_{()^{\mathrm{R}}}^{\mathrm{Zust}}(n)) = 2^n$,
- Vereinigung in [YZS94]: $\max(\mathrm{C}_{\cup}^{\mathrm{Zust}}(n,m)) = nm$,
- Konkatenation in [YZS94]: $\max(\mathrm{C}_{\cdot}^{\mathrm{Zust}}(n,m)) = n2^m - 2^{m-1}$ für $m \geq 2$,
- Kleene-Abschluss in [YZS94]: $\max(\mathrm{C}_{()^*}^{\mathrm{Zust}}(n)) = 2^{n-1} + 2^{n-2}$ für $n \geq 2$.

Darüber hinaus werden maximale Zustandskomplexitäten unter weiteren Operationen in den Übersichtsartikeln [Yu00] und [GMRY16] angegeben. Die vollständige Ermittlung der Zustandskomplexitäten unter einigen Operationen ist in der Vergangenheit gelungen. Wir erwähnen hier:

- Spiegelbild in [Jir14]: $\log_2(n), \ldots, 2^n \in \mathrm{C}^{\mathrm{Zust}}_{()^{\mathrm{R}}}(n) \not\ni 2^n+1, \ldots$ für $n \geq 3$,
- Vereinigung in [HJS05]: $\mathrm{C}^{\mathrm{Zust}}_{\cup}(n,m) = \{1, \ldots, nm\}$,
- Konkatenation in [Jir09]: $\mathrm{C}^{\mathrm{Zust}}_{\cdot}(n,m) = \{1, \ldots, n2^m - 2^{m-1}\}$ für $m \geq 2$,
- Kleene-Abschluss in [Jir14]: $\mathrm{C}^{\mathrm{Zust}}_{()^*}(n) = \{1, \ldots, 2^{n-1} + 2^{n-2}\}$ für $n \geq 2$.

Am Ende des Abschnitts 1.1 haben wir angedeutet, dass die Anwendung mehrerer Operationen relevanter ist als die einer einzelnen. Im obigen regulären Ausdruck zur E-Mail-Adressvalidierung wird dies deutlich, da Konkatenation und positiver Kleene-Abschluss mehrfach verwendet werden. Die maximalen Zustandskomplexitäten im Falle kombinierter Operationen sind zusammenfassend in [Gao10] und [CGKY12] dargelegt.

Bezüglich der kontextfreien Sprachen wurden bisher im Grunde nur in [DS08] umfangreiche Untersuchungen angestellt. Dort wird als Komplexitätsmaß $\mathrm{Var}(L)$ – die geringste Anzahl an Variablen, die eine kontextfreie Grammatik benötigt, um die kontextfreie Sprache L zu erzeugen – verwendet. Bis auf ein Ergebnis sind die Ergebnisse der Variablenkomplexität vollständig und nicht auf die maximalen Komplexitäten beschränkt:

- Vereinigung:

$$\begin{aligned} \mathrm{C}^{\mathrm{Var}}_{\cup}(n,m) &= \{1, \ldots, n+m+1\}, \\ \max\left(\mathrm{C}^{\mathrm{Var}}_{\cup}(n,m)\right) &= n+m+1, \end{aligned}$$

- Konkatenation:

$$\begin{aligned} \mathrm{C}^{\mathrm{Var}}_{\cdot}(n,m) &\ni \max(\{n,m\}), \ldots, n+m+1, \\ \max\left(\mathrm{C}^{\mathrm{Var}}_{\cdot}(n,m)\right) &= n+m+1, \end{aligned}$$

- Kleene-Abschluss:

$$\begin{aligned} \mathrm{C}^{\mathrm{Var}}_{()^*}(n) &= \{1, \ldots, n+1\}, \\ \max\left(\mathrm{C}^{\mathrm{Var}}_{()^*}(n)\right) &= n+1, \end{aligned}$$

- Homomorphismus:

$$\begin{aligned} \mathrm{C}^{\mathrm{Var}}_{\mathrm{h}}(n) &= \{1, \ldots, n\}, \\ \max\left(\mathrm{C}^{\mathrm{Var}}_{\mathrm{h}}(n)\right) &= n, \end{aligned}$$

  für einen konkreten Homomorphismus h,

- inverser Homomorphismus:

$$\begin{aligned} \mathrm{C}^{\mathrm{Var}}_{\mathrm{h}^{-1}}(n) &= \{1, \ldots\} = \mathbb{N}, \\ \max\left(\mathrm{C}^{\mathrm{Var}}_{\mathrm{h}^{-1}}(n)\right) &\text{ nicht existent} \end{aligned}$$

für einen konkreten inversen Homomorphismus h^{-1},

- Schnitt mit regulären Sprachen:

$$\begin{gathered}\mathrm{C}^{\mathrm{Var}}_{\mathrm{Reg}}(n) = \{1, \dots\} = \mathbb{N}, \\ \max\left(\mathrm{C}^{\mathrm{Var}}_{\mathrm{Reg}}(n)\right) \text{ nicht existent}\end{gathered}$$

  für unendlich viele von n abhängigen regulären Sprachen

für natürliche Zahlen $n \geq 1$ und $m \geq 1$.

Diese Arbeit setzt die Arbeit von [DS08] fort, indem statt $\mathrm{Var}(L)$ die Maße $\mathrm{Prod}(L)$ und $\mathrm{Symb}(L)$ verwendet werden. Außerdem untersuchen wir zusätzlich die Operation Spiegelbild.

2. Grundlagen

Wir klären anfangs im Abschnitt 2.1 die verwendete allgemeine mathematische Notation. Danach geben wir im Abschnitt 2.2 eine kurze Einführung in die Theorie formaler Sprachen. Wir schließen dieses Kapitel mit Abschnitt 2.3 ab, welcher die Grundlagen der Regel- und Symbolkomplexität behandelt.

2.1. Allgemeines

In dieser Arbeit bezeichnen wir die Menge der positiven natürlichen Zahlen mit $\mathbb{N}$ und setzen $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathbb{N}^{\not\ni 1} = \mathbb{N} \setminus \{1\}$ sowie $\mathbb{N}_0^{\not\ni 1} = \mathbb{N}_0 \setminus \{1\}$. Außerdem legen wir fest, dass wir für die Variablen i, j, k stets $i, j, k \in \mathbb{N}_0$ annehmen, sofern diese lokal in einem Absatz oder mehreren nicht anders definiert werden. Für $x, y \in \mathbb{N}_0$, $x \leq y$, bedeutet $x, \ldots, y$ das Auflisten der natürlichen Zahlen von einschließlich x bis einschließlich y ausschließlich in Einerschritten.

Wir nennen zwei Mengen X und Y *disjunkt*, falls $X \cap Y = \emptyset$ zutrifft. Die Mächtigkeit einer Menge X drücken wir mit $|X|$ aus. Die Potenzmenge (Menge aller Teilmengen) einer Menge X bezeichnen wir mit $\mathcal{P}(X)$. Wenn wir nicht wissen, ob ein Objekt x Element einer Menge X ist, schreiben wir $x \in^? X$.

2.2. Formale Sprachen

Im Folgenden geben wir einige Definitionen und Lemmata aus der Theorie formaler Sprachen an, die wir später benötigen werden.

Definition 2.1 – Buchstabe, Alphabet, Wort, Sprache

Ein *Alphabet* V ist eine endliche, nicht leere Menge ($|V| \in \mathbb{N}$), die das sogenannte leere Wort λ (Definition nachstehend) nicht enthält. Die Elemente aus V bezeichnen wir als *Buchstaben*.

Ein *Wort* w über einem Alphabet V ist ein n-Tupel $(v_1, \ldots, v_n)$ für $n \in \mathbb{N}_0$ und $v_i \in V$, $1 \leq i \leq n$, wobei $(v_1, \ldots, v_0) = () = \lambda$ das *leere Wort* ist. Anstelle des Tupels $(v_1, \ldots, v_n)$ schreiben wir $v_1 \cdots v_n$, da im Nachstehenden Alphabete stets so gewählt werden, dass diese Darstellung eindeutig ist (z. B. wäre im Fall von $(a, aa) \neq (aa, a)$ für $a, aa \in V$ die Darstellung aaa nicht eindeutig). Die Menge $\mathrm{Alph}(w)$ ist die Menge aller Buchstaben, die in einem Wort w vorkommen.

Mit V^* bezeichnen wir die Menge aller Wörter über V – inklusive λ. Wir setzen $V^+ = V^* \setminus \{\lambda\}$. Eine *Sprache* L ist eine Teilmenge von V^*, und $\mathscr{L}$ ist die Menge aller Sprachen. Die Menge $\mathrm{Alph}(L)$ enthält alle Buchstaben, die in Wörtern der Sprache L vorkommen und nur diese.

Die *Länge* $|w|$ eines Wortes $w \in V^*$ ist n für $w = v_1 \cdots v_n$, $n \in \mathbb{N}_0, v_i \in V, 1 \leq i \leq n$, und damit 0 für $w = \lambda$. Die Schreibweise $|w|_v$ drückt die Anzahl der Vorkommen von $v \in V$ im Wort $w \in V^*$ aus. Weiterhin ersetzt $\operatorname{subst}(w, x, y)$ für $x \in V$, $y \in V^*$ jedes Vorkommen des Buchstabens x in w durch das Wort y. ◊

Definition 2.2 – Konkatenation zweier Wörter, Potenz eines Wortes
Die *Konkatenation* zweier Wörter $x = x_1 \cdots x_n$, $n \in \mathbb{N}_0$, $x_i \in V$, $1 \leq i \leq n$, und $y = y_1 \cdots y_m$, $m \in \mathbb{N}_0$, $y_j \in V$, $1 \leq j \leq m$, über einem Alphabet V ist die binäre Operation $\cdot \colon V^* \times V^* \to V^*$ mit

$$x \cdot y = x_1 \cdots x_n \cdot y_1 \cdots y_m = x_1 \cdots x_n y_1 \cdots y_m.$$

Meist werden wir das Operationssymbol $\cdot$ einfach fortlassen.

Für $i \in \mathbb{N}_0$ ist die i-te *Potenz* eines Wortes w über einem Alphabet V durch

$$\begin{aligned} w^0 &= \lambda, \\ w^i &= w \cdot w^{i-1}, i \geq 1, \end{aligned}$$

definiert. ◊

Definition 2.3 – Grammatik, kontextfreie und reguläre
Eine *Grammatik* G ist ein 4-Tupel $G = (N, T, P, S)$, wobei

- N und T zwei Alphabete mit $N \cap T = \emptyset$ oder leere Mengen sind,
- $P \subseteq ((N \cup T)^* \setminus T^*) \times (N \cup T)^*$, P endlich ist und
- es für alle $A \in N$ ein $(w, w') \in P$, $w, w' \in (N \cup T)^*$, mit $|ww'|_A \geq 1$ gibt sowie kein $(w'', w'') \in P$ für alle $w'' \in (N \cup T)^*$ existiert und
- $S \in N$ gilt, wenn $P \neq \emptyset$ ist, und $S = \lambda$ gilt, wenn $P = \emptyset$ ist.

Die Elemente aus N heißen *Nichtterminale*, die aus T *Terminale* und die aus P *Regeln*. Wir nennen S *Startsymbol*. Regeln $(w, w') \in P$, $w, w' \in (N \cup T)^*$, schreiben wir meist einfach als $w \to w'$.

Eine *kontextfreie Grammatik* $G = (N,\ T,\ P,\ S)$ ist eine Grammatik, nur dass einschränkend $P \subseteq N \times (N \cup T)^*$ gilt. Die *Menge aller kontextfreien Grammatiken* nennen wir CF.

Eine *reguläre Grammatik* $G = (N, T, P, S)$ ist eine Grammatik mit der Einschränkung $P \subseteq N \times (T^* \cdot (N \cup \{\lambda\}))$. Die *Menge aller regulären Grammatiken* bezeichnen wir mit REG. ◊

Definition 2.4 – Ableitungsrelation, Ableitung, Schleifenabl., Initialschleifenabl.
Es sei $G = (N,\ T,\ P,\ S)$ eine Grammatik. Wir legen die von G abhängige, binäre *Ableitungsrelation* $\Longrightarrow_G \subseteq (N \cup T)^* \times (N \cup T)^*$ so fest, dass die *Ableitung* $w \Longrightarrow_G w'$

genau dann gilt, wenn für $x_1, x_2, a, b \in (N \cup T)^*$ und $w = x_1 a x_2$ sowie $w' = x_1 b x_2$ die Regel $a \to b \in P$ existiert oder $w = w' = \lambda$ gilt.

Es sei weiterhin $\Longrightarrow_G^* \subseteq (N \cup T)^* \times (N \cup T)^*$ der reflexive und transitive Abschluss von $\Longrightarrow_G$ – es gilt also $w \Longrightarrow_G^* w'$ für $w, w' \in (N \cup T)^*$, genau dann, wenn es ein $n \in \mathbb{N}_0$ und Wörter $w_0, w_1, \ldots, w_n \in (N \cup T)^*$ so gibt, dass die Ableitung

$$w = w_0 \Longrightarrow_G w_1 \Longrightarrow_G \cdots \Longrightarrow_G w_n = w'$$

existiert.

Ist aus dem Zusammenhang zu erkennen, um welche Grammatik G es sich handelt, schreiben wir schlicht $\Longrightarrow$ statt $\Longrightarrow_G$ und $\Longrightarrow^*$ statt $\Longrightarrow_G^*$.

Wir nennen eine Ableitung der Form

$$A \Longrightarrow^* xAy$$

für $A \in N$ und $x, y \in (N \cup T)^*$, $xy \neq \lambda$, *Schleifenableitung* oder genauer *A-Schleifenableitung*. Im Falle einer S-Schleifenableitung verwenden wir auch den Begriff *Initialschleifenableitung*. $\diamondsuit$

Definition 2.5 – Sprache einer Grammatik, Sprachmengen

Es sei $G = (N, T, P, S)$ eine Grammatik. Die von einem *Nichtterminal $A \in N$ oder $A = \lambda$ erzeugte Sprache* ist

$$\mathrm{L}(G, A) = \begin{cases} \emptyset & A = \lambda \\ \{w \mid A \Longrightarrow_G^* w, w \in T^*\} & \text{sonst.} \end{cases}$$

Die *Sprache von G* ist

$$\mathrm{L}(G) = \mathrm{L}(G, S).$$

Die *Menge aller kontextfreien Sprachen* ist

$$\mathcal{L}(\mathrm{CF}) = \{\mathrm{L}(G) \mid G \in \mathrm{CF}\}$$

und die *Menge aller regulären Sprachen* analog

$$\mathcal{L}(\mathrm{REG}) = \{\mathrm{L}(G) \mid G \in \mathrm{REG}\}.$$

$\diamondsuit$

Definition 2.6 – Operationen auf Sprachen

Es seien X und X' zwei Alphabete sowie $L \subseteq X^*$, $L \in \mathscr{L}$, und $L' \subseteq X'^*$, $L' \in \mathscr{L}$, zwei Sprachen. Für L und L' sind

- das *Spiegelbild* $()^{\mathrm{R}} \colon \mathscr{L} \to \mathscr{L}$ mit

$$L^{\mathrm{R}} = \{w^{\mathrm{R}} \mid w \in L\},$$

wobei $\lambda^{\mathrm{R}} = \lambda$, $a^{\mathrm{R}} = a$ für $a \in X$ und $(uv)^{\mathrm{R}} = v^{\mathrm{R}}u^{\mathrm{R}}$ für $u, v \in X^*$ sind (die Bezeichnung $()^{\mathrm{R}}$ wird mehrfach verwendet),

- die *Vereinigung* $\cup\colon \mathscr{L} \times \mathscr{L} \to \mathscr{L}$ mit
$$L \cup L' = \{w \mid w \in L \text{ oder } w \in L'\},$$
- die *Konkatenation* $\cdot\colon \mathscr{L} \times \mathscr{L} \to \mathscr{L}$ mit
$$L \cdot L' = \{u \cdot v \mid u \in L \text{ und } v \in L'\},$$
(die Bezeichnung $\cdot$ wird mehrfach verwendet)
- der *Kleene-Abschluss* $()^*\colon \mathscr{L} \to \mathscr{L}$ mit
$$L^* = \bigcup_{i \in \mathbb{N}_0} L^i,$$
wobei $L^0 = \{\lambda\}$, $L^i = L^{i-1} \cdot L$ gilt,
- ein *Homomorphismus* $\mathrm{h}\colon \mathscr{L} \to \mathscr{L}$ mit
$$\mathrm{h}(L) = \{\mathrm{h}(w) \mid w \in L\},$$
wobei $\mathrm{h}\colon X^* \to X'^*$ – ebenfalls *Homomorphismus* genannt – mit
$$\mathrm{h}(u \cdot v) = \mathrm{h}(u) \cdot \mathrm{h}(v) \text{ für alle } u, v \in X^*$$
ist (die Bezeichnung h wird mehrfach verwendet),
- ein *inverser Homomorphismus* $\mathrm{h}^{-1}\colon \mathscr{L} \to \mathscr{L}$ mit
$$\mathrm{h}^{-1}(L') = \{w \mid w \in X^* \text{ und } \mathrm{h}(w) \in L'\}$$
für einen Homomorphismus h (Definition zuvor),
- ein *Schnitt mit regulärer Sprache* $R \in \mathcal{L}(\mathrm{REG})$ $\cap_{\mathrm{R}}\colon \mathscr{L} \to \mathscr{L}$ mit
$$\cap_{\mathrm{R}}(L) = L \cap R$$

einige Operationen auf Sprachen.

Wir bezeichnen mit Hom, Hom^{-1} und Reg die *Menge aller Homomorphismen*, die *Menge aller inversen Homomorphismen* bzw. die *Menge aller Schnitte mit regulärer Sprache.* $\Diamond$

Für einen Homomorphismus h und $w \in X^*$ gilt $\mathrm{h}(w) = \mathrm{h}(w \cdot \lambda) = \mathrm{h}(w) \cdot \mathrm{h}(\lambda)$. Daraus folgt unmittelbar $\mathrm{h}(\lambda) = \lambda$. Es sei $w = a_1 \cdots a_n$ für $a_i \in X$, $1 \leq i \leq n$, und $n \in \mathbb{N}$. Offenbar gilt $\mathrm{h}(w) = \mathrm{h}(a_1) \cdots \mathrm{h}(a_n)$. Damit können wir für jedes Wort $w' \in X^*$ das

Bild $\mathrm{h}(w')$ bestimmen, wenn wir das Bild $\mathrm{h}(x)$ für alle $x \in X$ angeben – h ist dann vollständig definiert.

Lemma 2.7 – Fakta über den Kleene-Abschluss
Für L^* einer beliebigen kontextfreien Sprache L sind die nachstehenden Eigenschaften stets wahr:

i. Es gelten $\lambda \in L^*$ und $L^* \neq \emptyset$.

ii. Es gilt $|L^*| = 1$ (und damit $L^* = \{\lambda\}$) genau dann, wenn $L = \emptyset$ oder $L = \{\lambda\}$ ist. Aus $|L^*| \neq 1$ folgt stets $|L^*| = |\mathbb{N}|$.

Beweis: Zu i. Die Eigenschaften folgen sofort aus der Definition 2.6 des Kleene-Abschlusses.

Zu ii. Wenn $L = \{\lambda\}$ oder $L = \emptyset$ ist, gilt nach Definition 2.6 des Kleene-Abschlusses $L^* = \{\lambda\}$, und damit $|L^*| = 1$. Andererseits folgen aus $|L^*| = 1$, also $|L^*| = |\bigcup_{i\in\mathbb{N}_0} L^i| = 1$, wegen $L^0 = \{\lambda\}$ die einzigen Möglichkeiten $L = \emptyset$ oder $L = \{\lambda\}$ – ansonsten gäbe es aufgrund der Potenzbildung mehr als nur ein Wort in L^*.

Falls $L \neq \{\lambda\}$ und $L \neq \emptyset$ ist (also $|L^*| \neq 1$), existiert ein $w \in L$ mit $w \neq \lambda$. Für $L^* = \bigcup_{i\in\mathbb{N}_0} L^i$ und $w^i \in L^i$ erhalten wir dann $w^i \in L^*$ für alle $i \in \mathbb{N}_0$. Mithin gilt $|L^*| \geq |\mathbb{N}|$. Da entweder $|L| \in \mathbb{N}_0$ gilt oder $|L| = |\mathbb{N}|$, folgt auch entweder $|L^i| \in \mathbb{N}_0$ oder $|L^i| = |\mathbb{N}|$, und die Vereinigung abzählbar vieler abzählbarer Mengen/Sprachen $\bigcup_{i\in\mathbb{N}_0} L^i = L^*$ liefert weiterhin $|L^*| \leq |\mathbb{N}|$. Schließlich erhalten wir $|\mathbb{N}| \leq |L^*| \leq |\mathbb{N}|$, mithin $|L^*| = |\mathbb{N}|$. □

Wir werden im Beweis zum folgenden Lemma über die Abgeschlossenheit kontextfreier Sprachen unter ausgewählten Operationen kontextfreie Grammatiken, welche die Ergebnissprachen der Operationen erzeugen, angeben, da wir diese in kommenden Beweisen verwenden werden. Für die Operationen inversen Homomorphismus und Schnitt mit regulärer Sprache konnten wir keine Konstruktionen für die entsprechenden kontextfreien Grammatiken finden.

Lemma 2.8 – Abgeschlossenheit kontextfreier Sprachen unter Operationen
Die kontextfreien Sprachen sind unter den Operationen aus Definition 2.6 abgeschlossen, das heißt, für $\circ \in \{()^{\mathrm{R}}, \cup, \cdot, ()^*\} \cup \mathrm{Hom} \cup \mathrm{Hom}^{-1} \cup \mathrm{Reg}$ und alle $L_1, L_2 \in \mathcal{L}(\mathrm{CF})$ gilt $\circ(L_1) \in \mathcal{L}(\mathrm{CF})$ für die 1-stelligen Operationen bzw. $L_1 \circ L_2 \in \mathcal{L}(\mathrm{CF})$ für die 2-stelligen.

Beweis: Es seien R eine reguläre Sprache, X und Y zwei Alphabete, $L \subseteq X^*$ und $L' \subseteq Y^*$ zwei kontextfreie Sprachen sowie $G = (N, T, P, S) \in \mathrm{CF}$ mit $\mathrm{L}(G) = L$ und $G' = (N', T', P', S') \in \mathrm{CF}$ mit $\mathrm{L}(G') = L'$. Außerdem gelten $N \cap N' = \emptyset$ und $S'' \notin N \cup N'$ ohne Beschränkung der Allgemeinheit. Dann lässt sich für

- L^{R} die kontextfreie Grammatik

$$G_{()^{\mathrm{R}}} = (N, T, \{A \to w^{\mathrm{R}} \mid A \to w \in P, A \in N, w \in (N \cup T)^*\}, S)$$

angeben, für die sicher $\mathrm{L}(G_{()^{\mathrm{R}}}) = L^{\mathrm{R}}$ gilt.

- $L \cup L'$ die kontextfreie Grammatik

$$G_{\cup} = \begin{cases} G & \text{für } P' = \emptyset \\ G' & \text{für } P = \emptyset \\ (N \cup N' \cup \{S''\}, T \cup T', P \cup P' \cup P'', S'') & \text{sonst} \end{cases}$$

  mit

$$P'' = \{S'' \to S, S'' \to S'\}$$

  angeben, für die sicher $\mathrm{L}(G_{\cup}) = L \cup L'$ gilt.

- $L \cdot L'$ die kontextfreie Grammatik

$$G_{\cdot} = \begin{cases} G & \text{für } P' = \emptyset \\ G' & \text{für } P = \emptyset \\ (N \cup N' \cup \{S''\}, T \cup T', P \cup P' \cup P'', S'') & \text{sonst} \end{cases}$$

  mit

$$P'' = \{S'' \to SS'\}$$

  angeben, für die sicher $\mathrm{L}(G_{\cdot}) = L \cdot L'$ gilt.

- L^* die kontextfreie Grammatik

$$G_{()^*} = \begin{cases} (\{S''\}, \emptyset, \{S'' \to \lambda\}, S'') & \text{für } P = \emptyset \\ (N \cup \{S''\}, T, P \cup P', S'') & \text{sonst} \end{cases}$$

  mit

$$P' = \{S'' \to SS'', S'' \to \lambda\}$$

  angeben, für die sicher $\mathrm{L}(G_{()^*}) = L^*$ gilt.

- $\mathrm{h}(L)$ und einen Homomorphismus $\mathrm{h}\colon X^* \to Y^*$ die kontextfreie Grammatik

$$G_{\mathrm{h}} = (N, Y, \{A \to \mathrm{h}'(w) \mid A \in N, w \in (N \cup X)^*, A \to w \in P\}, S)$$

  mithilfe eines Homomorphismus $\mathrm{h}'\colon (X \cup N)^* \to (Y \cup N)^*$ mit

$$\mathrm{h}'(A) = A \text{ für } A \in N,$$
$$\mathrm{h}'(a) = \mathrm{h}(a) \text{ für } a \in X$$

  angeben, für die sicher $\mathrm{L}(G_{\mathrm{h}}) = \mathrm{h}(L)$ gilt.

- $\mathrm{h}^{-1}(L)$ und einen Homomorphismus $\mathrm{h}\colon X^* \to Y^*$ die Kellerautomaten-Konstruktion aus Theorem 7.30 in [HMU00] verwenden, um $\mathrm{h}^{-1}(L) \in \mathcal{L}(\mathrm{CF})$ zu konstruieren.

- $\cap_{\mathrm{R}}(L)$ die Kellerautomaten-Konstruktion aus Theorem 7.27 in [HMU00] verwenden, um $\cap_{\mathrm{R}}(L) \in \mathcal{L}(\mathrm{CF})$ zu konstruieren.

Somit entsteht nach Anwendung der genannten Operationen auf kontextfreie Sprachen stets eine kontextfreie Sprache, womit das Lemma bewiesen ist (siehe auch [HMU00]). □

Lemma 2.9 – Separate Regeln für nicht gemeinsam vorkommende Buchstaben
Es seien L eine kontextfreie Sprache,

$$\text{Wört}_L\colon \text{Alph}(L) \to \mathcal{P}(L)$$

mit

$$\text{Wört}_L(v) = \{w \mid w \in L, |w|_v \geq 1\}, v \in \text{Alph}(L)$$

und $X \subseteq \text{Alph}(L)$ eine Menge von Buchstaben, für die gilt: Aus $u, v \in X$, $u \neq v$, folgt $\text{Wört}_L(u) \cap \text{Wört}_L(v) = \emptyset$. Unter diesen Bedingungen enthält jede kontextfreie Grammatik, die L erzeugt, für jedes $v \in X$ eine separate Regel mit v auf der rechten Seite.

Beweis: Es sei $G = (N, T, P, S)$ eine kontextfreie Grammatik mit $\text{L}(G) = L$. Für eine Ableitung $S \Longrightarrow^* w_v$ mit $w_v \in \text{Wört}_L(v)$ und $v \in X$ muss es mindestens eine Regel p_v mit einem Vorkommen von v auf der rechten Seite geben. Die Regel p_v findet keine Anwendung in den Ableitungen $S \Longrightarrow^* w$ für $w \in L \setminus \text{Wört}_L(v)$, da w kein v enthält. Mithin gibt es für jeden Buchstaben aus $v \in X$ mindestens eine separate Regel mit v auf der rechten Seite in P. □

Lemma 2.10 – Separate Regeln für allein vorkommende Buchstaben
Es seien L eine kontextfreie Sprache und

$$X = \{v \mid v \in \text{Alph}(L), \{v\}^+ \cap L \neq \emptyset\}.$$

Dann enthält jede kontextfreie Grammatik (N, T, P, S), die L erzeugt, für jedes $v \in X$ eine Regel $A_v \to u_v$ für gewisse $A_v \in N$ und $u_v \in (N \cup \{v\})^+$ sowie $|u_v|_v \geq 1$.

Beweis: Es sei $G = (N, T, P, S)$ eine kontextfreie Grammatik mit $\text{L}(G) = L$. Offenbar existiert eine Ableitung $S \Longrightarrow^* w_v \in L$ für $v \in X$ und ein gewisses $w_v \in \{v\}^+$. Es muss eine Regel $A_v \to u_v$ für $A_v \in N$ und $u_v \in (N \cup \{v\})^+$, $|u_v|_v \geq 1$, geben – anderenfalls könnte w_v nicht erzeugt werden, da es stets von v verschiedene Terminale enthielte oder keine. Diese Aussage stimmt für jedes $v \in X$. Demnach existiert für jeden Buchstaben aus X mindestens eine separate Regel in der Menge P. □

2.3. Regelkomplexität und Symbolkomplexität

In diesem Abschnitt geben wir Definitionen und Lemmata rund um die Regel- und Symbolkomplexität an. Zudem legen wir einen Satz dar, der sich als ein wichtiges Hilfsmittel für die späteren Untersuchungen herausstellen wird.

Definition 2.11 – Regel-/Symbolanzahl, Regel-/Symbolkomplexität
Die *Anzahl der Regeln* einer kontextfreien Grammatik $G = (N, T, P, S)$ ist

$$\mathrm{Prod}(G) = |P|,$$

und die *Anzahl der Symbole* von G ist

$$\mathrm{Symb}(G) = \sum_{\substack{A \to w \in P,\\ A \in N,\\ w \in (N \cup T)^*}} (|w| + 2)$$

(wir addieren 2, da A und $\to$ mitgezählt werden sollen).

Die *Regelkomplexität* einer kontextfreien Sprache L ist

$$\mathrm{Prod}(L) = \min(\{\mathrm{Prod}(G) \mid G \in \mathrm{CF}, \mathrm{L}(G) = L\}).$$

Wir nennen G *regel-minimal*, wenn $\mathrm{Prod}(G) = \mathrm{Prod}(\mathrm{L}(G))$ gilt.

Die *Symbolkomplexität* von L ist analog

$$\mathrm{Symb}(L) = \min(\{\mathrm{Symb}(G) \mid G \in \mathrm{CF}, \mathrm{L}(G) = L\}).$$

Wir nennen G *symbol-minimal*, wenn $\mathrm{Symb}(G) = \mathrm{Symb}(\mathrm{L}(G))$ zutrifft. $\Diamond$

Häufig werden wir G nur als *minimal* bezeichnen, wenn aus dem Kontext hervorgeht, ob in Bezug auf die Regel- oder Symbolanzahl.

Definition 2.12 – Räume der Regel-/Symbolkomplexität

Es seien $n \in \mathbb{N}$, $\mathrm{K} \in \{\mathrm{Prod}, \mathrm{Symb}\}$ und $x_i \in \mathbb{N}_0$, $1 \leq i \leq n$. Für eine n-stellige Operation $\circ_n$, unter der die kontextfreien Sprachen abgeschlossen sind, legen wir

$$\mathrm{C}^{\mathrm{K}}_{\circ_n} \colon \underbrace{\mathbb{N}_0 \times \cdots \times \mathbb{N}_0}_{n\text{-mal}} \to \mathcal{P}(\mathbb{N}_0)$$

mit

$$\mathrm{C}^{\mathrm{K}}_{\circ_n}(x_1, \ldots, x_n) = \{\mathrm{K}(\circ_n(L_1, \ldots, L_n)) \mid \mathrm{K}(L_i) = x_i, L_i \in \mathcal{L}(\mathrm{CF}), 1 \leq i \leq n\}$$

als den *Raum der Regelkomplexität unter* $\circ_n$ im Fall von $\mathrm{K} = \mathrm{Prod}$ bzw. als den *Raum der Symbolkomplexität unter* $\circ_n$ im Fall von $\mathrm{K} = \mathrm{Symb}$ fest. Häufig werden wir nur von *Raum der Regelkomplexität* oder *Raum der Symbolkomplexität* sprechen, wenn die Operation aus dem Kontext klar wird.

Für eine Menge Ω_n von n-stelligen Operationen, unter denen die kontextfreien Sprachen abgeschlossen sind, bezeichnen wir

$$\mathrm{C}^{\mathrm{K}}_{\Omega_n} \colon \underbrace{\mathbb{N}_0 \times \cdots \times \mathbb{N}_0}_{n\text{-mal}} \to \mathcal{P}(\mathbb{N}_0)$$

mit

$$\mathrm{C}^{\mathrm{K}}_{\Omega_n}(x_1, \ldots, x_n) = \bigcup_{\circ_n \in \Omega_n} \mathrm{C}^{\mathrm{K}}_{\circ_n}(x_1, \ldots, x_n)$$

als den *Raum der Regelkomplexität unter* Ω_n im Fall von $\mathrm{K} = \mathrm{Prod}$ bzw. als den *Raum der Symbolkomplexität unter* Ω_n im Fall von $\mathrm{K} = \mathrm{Symb}$. $\lozenge$

Definition 2.13 – Reduzierte kontextfreie Grammatik
Eine kontextfreie Grammatik $G = (N, T, P, S)$ heißt *reduziert*, wenn für alle $A \in N$ sowohl eine Ableitung $S \Longrightarrow^* w_1 A w_2$ für gewisse $w_1, w_2 \in (N \cup T)^*$ existiert als auch eine Ableitung $A \Longrightarrow^* w$ für ein gewisses $w \in T^*$. $\lozenge$

Bemerkung: Für alle Nichtterminale $A \in N$ einer reduzierten kontextfreien Grammatik $G = (N, T, P, S)$ gelten offenbar $S \Longrightarrow^* xAy \Longrightarrow^* w \in \mathrm{L}(G)$ für gewisse $x, y, w \in T^*$. Weiterhin gibt es zu jedem $A \in N$ eine Regel $A \to w \in P$ für ein gewisses $w \in (N \cup T)^*$. Darüber hinaus ist G auch reduziert, wenn $P = \emptyset$ gilt (nach Definition 2.3 folgt aus $P = \emptyset$ auch $N = \emptyset$). $\triangle$

Lemma 2.14 – Minimale Grammatiken sind reduziert
Jede regel- oder symbol-minimale kontextfreie Grammatik ist reduziert.

Beweis: Es sei $G = (N, T, P, S)$ eine kontextfreie Grammatik mit $\mathrm{K}(G) = \mathrm{K}(\mathrm{L}(G))$, $\mathrm{K} \in \{\mathrm{Prod}, \mathrm{Symb}\}$. Gäbe es für ein Nichtterminal $A \in N$ keine Ableitungen $S \Longrightarrow^* w_1 A w_2$ für gewisse $w_1, w_2 \in (N \cup T)^*$ oder $A \Longrightarrow^* w$ für ein gewisses $w \in T^*$, so stünde A in keiner Ableitung der Wörter aus $\mathrm{L}(G)$. Daher käme jede Regel mit A auf der linken Seite und jede Regel mit einem Vorkommen von A auf der rechten Seite in keiner Ableitung der Wörter aus $\mathrm{L}(G)$ vor und könnte somit gestrichen werden, ohne dass sich die erzeugte Sprache änderte. Mithin wäre G weder regel- noch symbol-minimal, was G aber nach Annahme sein soll. Folglich muss G reduziert sein. $\square$

Lemma 2.15 – Anzahl der von Nichtterminalen erzeugten Wörter in regel-minimalen Grammatiken
Es sei $G = (N, T, P, S)$ eine regel-minimale kontextfreie Grammatik mit $|\mathrm{L}(G)| \geq 2$. Dann gilt $|\mathrm{L}(G, A)| \geq 2$ für jedes $A \in N$.

Beweis: Aufgrund der Regelminimalität von G, ist G reduziert (Lemma 2.14). Somit gilt $|\mathrm{L}(G, A)| \geq 1$ für jedes $A \in N$, da stets eine Ableitung $S \Longrightarrow^* xAy \Longrightarrow^* w \in \mathrm{L}(G)$ für gewisse $x, y, w \in T^*$ existiert. Im Fall von $A = S$ erhalten wir wegen $\mathrm{L}(G) = \mathrm{L}(G, S)$ und $|\mathrm{L}(G)| \geq 2$ direkt $\mathrm{L}(G, S) \geq 2$.

Wir behandeln nun den Fall $A \neq S$. Angenommen, es sei $|\mathrm{L}(G, A)| = 1$, und damit $\mathrm{L}(G, A) = \{w\}$, $w \in T^*$. Dann könnten wir A auf der rechten Seite jeder Regel aus P durch w ersetzen und jede Regel mit A auf der linken Seite streichen. Die derart konstruierte kontextfreie Grammatik erzeugte $\mathrm{L}(G)$ und enthielte weniger Regeln als G. Allerdings soll G minimal sein – ein Widerspruch. Daher muss $|\mathrm{L}(G, A)| \geq 2$ gelten. $\square$

Lemma 2.16 – Sprachmächtigkeit und Schleifenableitungen in reduzierten Grammatiken
Es sei $G = (N, T, P, S)$ eine reduzierte kontextfreie Grammatik.

i. Ist $\mathrm{L}(G)$ endlich, erzeugt G *keine* Schleifenableitung.

ii. Ist $\mathrm{L}(G)$ unendlich, erzeugt G *mindestens eine* Schleifenableitung.

Beweis: Zu i. Angenommen, es gibt eine Schleifenableitung der Form $A \Longrightarrow^* xAy$ für $A \in N$ und $x, y \in T^*$, $xy \neq \lambda$. Dann existieren aufgrund der Reduziertheit von G auch die Ableitung $S \Longrightarrow^* x'Ay'$ für $x', y' \in T^*$ und die Ableitung $A \Longrightarrow^* w$ für ein $w \in T^*$. Offenbar können wir damit für $p \in \mathbb{N}_0$ mittels

$$S \Longrightarrow^* x'Ay' \Longrightarrow^* x'xAyy' \Longrightarrow^* \cdots \Longrightarrow^* x'x^pAy^py' \Longrightarrow^* x'x^pwy^py' \in T^*$$

unendlich viele Terminalwörter erzeugen. Dies widerspricht der Endlichkeit von $\mathrm{L}(G)$. Folglich gibt es keine Schleifenableitungen in G.

Zu ii. Angenommen, es gibt keine Schleifenableitung der Form $A \Longrightarrow^* xAy$ für $A \in N$ und $x, y \in T^*$, $xy \neq \lambda$. Dann existieren nur endlich viele Ableitungen, die zu einem Terminalwort führen. Mithin ist $\mathrm{L}(G)$ endlich – ein Widerspruch zur Voraussetzung, dass $\mathrm{L}(G)$ unendlich ist. $\square$

Lemma 2.17 – Ableitungseigenschaften in minimalen Grammatiken
Es sei $G = (N, T, P, S)$ eine regel- oder symbol-minimale kontextfreie Grammatik.

i. Existiert unter der Bedingung $\mathrm{L}(G) \neq \{\lambda\}$ die Ableitung $A \Longrightarrow^* \lambda$ für ein $A \in N$, so muss es eine (weitere) Ableitung $A \Longrightarrow^* w \in T^+$ geben.

ii. Existiert unter der Prämisse $\mathrm{L}(G) \neq \{x\}$ für alle $x \in T$ die Ableitung $A \Longrightarrow^* a$ für $A \in N$ und $a \in T$, so muss es eine (weitere) Ableitung $A \Longrightarrow^* w \in T^*$, $w \neq a$, geben.

Beweis: Die Grammatik G ist nach Lemma 2.14 reduziert. Für $\mathrm{L}(G) = \emptyset$ folgen die beiden Aussagen sofort. Wir nehmen daher im Folgenden $\mathrm{L}(G) \neq \emptyset$ an.

Zu i. Es sei zunächst $A = S$. Da nach Voraussetzung $\mathrm{L}(G) \neq \{\lambda\}$ und $\mathrm{L}(G) \neq \emptyset$ gelten, muss es eine Ableitung $S \Longrightarrow^* w \in T^+$ geben.

Es sei nun $A \neq S$. Angenommen, es gibt nur die Ableitung $A \Longrightarrow^* \lambda$ für A. Dann können wir eine Grammatik $G' = (N \setminus \{A\}, T, P', S)$ so angeben, dass P' das gleiche wie P mit folgender Modifizierung ist: Jedes Vorkommen von A auf der rechten Seite aller Regeln aus P wird durch λ ersetzt und sämtliche Regeln mit A auf der linken Seite werden gestrichen. G' erzeugt dann $\mathrm{L}(G)$ mit weniger Regeln und weniger Symbolen als G – ein Widerspruch zur Minimalität von G. Daher kann $A \Longrightarrow^* \lambda$ nicht die einzige Ableitung für A sein.

Zu ii. Der Beweis kann analog zum Anstrich i geführt werden. Hierbei wird statt λ der Buchstabe a ersetzt. Die Bedingung $\mathrm{L}(G) \neq \{x\}$ ist notwendig, da sonst $S \to x$ die einzige und kürzeste Regel in G wäre und somit $S \Longrightarrow^* x$ die einzige Ableitung für S wäre. $\square$

Definition 2.18 – Kontextfreies Nichtterminal
Ein Nichtterminal $A \in N$, $A \neq S$, einer kontextfreien Grammatik $G = (N, T, P, S)$ heißt *kontextfrei*, falls in G aus $S \Longrightarrow^* xAy$ für $x, y \in (N \cup T)^*$ stets $xy = \lambda$ folgt. $\Diamond$

Bemerkung: Für ein kontextfreies Nichtterminal A gilt nur $S \Longrightarrow^* A$. △

Lemma 2.19 – Keine kontextfreien Nichtterminale in minimalen Grammatiken
In einer regel- oder symbol-minimalen kontextfreien Grammatik existieren keine kontextfreien Nichtterminale.

Beweis: Es sei $G = (N,\ T,\ P,\ S)$ eine reduzierte kontextfreie Grammatik mit der Eigenschaft $\mathrm{Prod}(G) = \mathrm{Prod}(\mathrm{L}(G))$ oder $\mathrm{Symb}(G) = \mathrm{Symb}(\mathrm{L}(G))$ (Lemma 2.14). Wir nehmen an, dass für ein $A \in N$, $A \neq S$, aus $S \Longrightarrow^* uAv$ für $u, v \in (N \cup T)^*$ stets $uv = \lambda$ folgt, A also ein kontextfreies Nichtterminal ist. Da nach Annahme $S \Longrightarrow^* A$ gilt, muss es eine gewisse Regel mit A auf der rechten Seite geben: $X \to xAy \in P$ mit $X \in N$ und $x, y \in (N \cup T)^*$. Wegen der Reduziertheit von G gibt es ferner eine Ableitung $S \Longrightarrow^* x'Xy'$ für gewisse $x', y' \in T^*$. Da $S \Longrightarrow^* x'Xy' \Longrightarrow^* x'xAyy'$ gilt und links und rechts von A kein Terminal vorkommen darf, muss $x'xyy' = \lambda$ sein, und folglich $X \to A \in P$. Da $S \Longrightarrow^* A$ bereits mit $X \to A$ erzeugt wird, wäre jede weitere Regel mit A auf der rechten Seite nutzlos und minimalitätsverletzend.

Wir erstellen nun eine Regelmenge R, indem wir zunächst $R = P$ festlegen. Anschließend streichen wir die Regel $X \to A$ in R und alle Regeln in R, die sich in

$$Q = \{A \to u \mid A \to u \in R, u \in (N \cup T)^*\}$$

befinden. Nun fügen wir alle Regeln aus

$$O = \{X \to u \mid A \to u \in Q, u \in (N \cup T)^*\}$$

in R ein (es gilt $|O| = |Q|$). Die Regelmenge R besitzt nun genau eine Regel weniger als P und besteht aus genau drei Symbolen weniger.

Die kontextfreie Grammatik $G' = (N \setminus \{A\}, T, R, S)$ generiert $\mathrm{L}(G)$. Grund dafür ist die Tatsache, dass in G' im Prinzip alle Regeln aus G enthalten sind, allerdings mit der Einschränkung, dass die Regel $X \to A$ getilgt wurde und jeweils eine Regel mit A auf der linken Seite durch genau eine neue Regel mit X auf der linken Seite ersetzt wurde. Zu jeder Ableitung

$$S \Longrightarrow_G^* X \Longrightarrow_G^* A \Longrightarrow_G^* w$$

für gewisse $w \in \mathrm{L}(G)$ existiert folglich eine Ableitung

$$S \Longrightarrow_{G'}^* X \Longrightarrow_{G'}^* w,$$

in der kein A vorkommt. Alle anderen Ableitungen bleiben unverändert.

Die Grammatik G' besteht aus $\mathrm{Prod}(G) - 1$ Regeln und $\mathrm{Symb}(G) - 3$ Symbolen. Dies widerspricht der Voraussetzung $\mathrm{Prod}(G) = \mathrm{Prod}(\mathrm{L}(G))$ oder $\mathrm{Symb}(G) = \mathrm{Symb}(\mathrm{L}(G))$. Mithin existieren in G doch keine kontextfreien Nichtterminale. □

Definition 2.20 – Isoliertes Nichtterminal
Ein Nichtterminal $A \in N$, $A \neq S$, einer kontextfreien Grammatik $G = (N, T, P, S)$ heißt *isoliert*, falls in G aus $S \Longrightarrow^* xAy$ und $S \Longrightarrow^* x'Ay'$ für $x, y, x', y' \in (N \cup T)^*$ stets $xAy = x'Ay'$ folgt. ♢

Bemerkung: Falls A nicht isoliert ist, gibt es entweder mindestens zwei Ableitungen $S \Longrightarrow^* xAy$ und $S \Longrightarrow^* x'Ay'$ mit $x, y, x', y' \in (N \cup T)^*$, wobei $xAy \neq x'Ay'$ gilt, oder es gilt $S \not\Longrightarrow^* xAy$ für alle $x, y \in (N \cup T)^*$. △

Lemma 2.21 – Regel-minimale Grammatiken enthalten keine isolierten Nichtterminale

Jede regel-minimale kontextfreie Grammatik G enthält keine isolierten Nichtterminale.

Beweis: Es sei $G = (N,\ T,\ P,\ S)$ eine reduzierte kontextfreie Grammatik mit der Eigenschaft $\mathrm{Prod}\,(G) = \mathrm{Prod}\,(\mathrm{L}\,(G))$ (Lemma 2.14). Angenommen, es gibt ein isoliertes Nichtterminal $A \in N$, $A \neq S$. Dann folgt aus $S \Longrightarrow^* xAy$ und $S \Longrightarrow^* x'Ay'$ für $x, y, x', y' \in (N \cup T)^*$ stets $xAy = x'Ay'$. Wir dürfen davon ausgehen, dass es nur eine Regel $X \to w \in P$ mit $X \in N$ und $w \in (N \cup T)^*$ mit mindestens einem Vorkommen von A auf der rechten Seite ($|w|_A \geq 1$) in P gibt – ansonsten könnten wir andere Regeln mit A auf der rechten Seite streichen, da $S \Longrightarrow^* xAy$ als einzige Ableitung mit A dennoch gewährleistet wäre. Das Nichtterminal X kann im Übrigen nicht A sein ($X \in N \setminus \{A\}$), weil andererseits die Regel $X \to w = A \to w$ mit A auf der rechten Seite ($|w|_A \geq 1$) nie Anwendung fände (da sie die einzige Regel mit A auf der rechten Seite ist) und es folglich keine Ableitung der Form $S \Longrightarrow^* xAy$ gäbe.

Wir konstruieren nun eine Regelmenge R und setzen zu Beginn $R = P$. Als Nächstes streichen wir die Regel $X \to w$ in R und alle Regeln in R, die sich in

$$Q = \{A \to u \mid A \to u \in R, u \in (N \cup T)^*\}$$

befinden. Anschließend fügen wir alle Regeln aus

$$O = \{X \to \mathrm{subst}\,(w, A, u) \mid A \to u \in Q, u \in (N \cup T)^*\}$$

in R ein (es gilt $|O| = |Q|$). Die Regelmenge R besitzt nun mindestens eine Regel weniger als P.

Die kontextfreie Grammatik $(N \setminus \{A\},\ T,\ R,\ S)$ erzeugt schließlich $\mathrm{L}\,(G)$, benötigt allerdings höchstens $\mathrm{Prod}\,(G) - 1$ Regeln. Dies widerspricht der Voraussetzung $\mathrm{Prod}\,(G) = \mathrm{Prod}\,(\mathrm{L}\,(G))$. Mithin existieren in G doch keine isolierten Nichtterminale. □

Lemma 2.21 lässt sich derart nicht auf die Symbolkomplexität übertragen. Grund dafür ist die Substitution $\mathrm{subst}\,(w,\ A,\ u)$ beim Festlegen der Menge O, die allgemein auch die Anzahl der Symbole der neukonstruierten Grammatik erhöhen kann (wenn mehrere A in w ersetzt werden).

Der nachstehende Satz ist ein wichtiges Werkzeug für die künftigen Untersuchungen.

Satz 2.22 – Partition einer kontextfreien Sprache bezüglich der Regel-/Symbolkomplexität

Es seien L_1 und L_2 zwei kontextfreie Sprachen, für die

$$\mathrm{Alph}\,(L_1) \cap \mathrm{Alph}\,(L_2) = \emptyset \text{ und } \lambda \notin L_1, \lambda \notin L_2$$

gelten. Außerdem seien $i \in \{1, 2\}$, $\mathrm{K} \in \{\mathrm{Prod}, \mathrm{Symb}\}$ *sowie* $G_i^{\mathrm{K}} = (N_i^{\mathrm{K}}, T_i, P_i^{\mathrm{K}}, S_i)$ *eine kontextfreie Grammatik mit* $\mathrm{L}(G_i^{\mathrm{K}}) = L_i$, $\mathrm{K}(G_i^{\mathrm{K}}) = \mathrm{K}(L_i)$, *für die ohne Beschränkung der Allgemeinheit* $N_1^{\mathrm{K}} \cap N_2^{\mathrm{K}} = \emptyset$ *und* $S \notin N_1^{\mathrm{K}} \cup N_2^{\mathrm{K}}$ *gelten. Des Weiteren soll* G_i^{K} *im Fall von* $|L_i| = |\mathbb{N}|$ *nur dann eine Initialschleifenableitung haben, wenn jede minimale kontextfreie Grammatik* $G_\#^{\mathrm{K}}$ *mit* $\mathrm{K}(G_\#^{\mathrm{K}}) = \mathrm{K}(G_i^{\mathrm{K}})$, *die* L_i *erzeugt, eine Initialschleifenableitung hat (ansonsten wird* G_i^{K} *durch eine ohne Initialschleifenableitung ersetzt).*

Unter diesen Bedingungen gibt es eine kontextfreie Grammatik $G_{\mathrm{K}} = (N_{\mathrm{K}}, T_1 \cup T_2, P_{\mathrm{K}}, S)$ *mit* $\mathrm{L}(G_{\mathrm{K}}) = L_1 \cup L_2$ *und* $\mathrm{K}(G_{\mathrm{K}}) = \mathrm{K}(L_1 \cup L_2)$ *mit nachstehenden Eigenschaften: Für* $Q_i^{\mathrm{K}} = \{S_i \to w \mid S_i \to w \in P_i^{\mathrm{K}},\ w \in (N_i^{\mathrm{K}} \cup T_i^{\mathrm{K}})^*\}$ *und* $Q^{\mathrm{K}} = Q_1^{\mathrm{K}} \cup Q_2^{\mathrm{K}}$ *sowie* $x_i, y_i \in {T_i}^*$, $x_i y_i \neq \lambda$, *erhalten wir:*

i.

$$\begin{aligned}
N_{\mathrm{K}} &= \left(\left(N_1^{\mathrm{K}} \cup N_2^{\mathrm{K}}\right) \setminus \{S_1, S_2\}\right) \cup \begin{cases} \{S\} & \text{für } P_1^{\mathrm{K}} \neq \emptyset, P_2^{\mathrm{K}} \neq \emptyset \\ \emptyset & \text{sonst,} \end{cases} \\
P_{\mathrm{K}} &= \left(\left(P_1^{\mathrm{K}} \cup P_2^{\mathrm{K}}\right) \setminus Q^{\mathrm{K}}\right) \cup \\
&\quad \Big\{S \to w \,\Big|\, X \to w \in Q^{\mathrm{K}}, X \in \{S_1, S_2\}, \\
&\qquad\qquad w \in \left(N_1^{\mathrm{K}} \cup N_2^{\mathrm{K}} \cup T_1^{\mathrm{K}} \cup T_2^{\mathrm{K}}\right)^*\Big\}, \\
\mathrm{K}(G_{\mathrm{K}}) &= \mathrm{K}\left(G_1^{\mathrm{K}}\right) + \mathrm{K}\left(G_2^{\mathrm{K}}\right),
\end{aligned}$$

falls keine Initialschleifenableitungen $S_1 \Longrightarrow^*_{G_1^{\mathrm{K}}} x_1 S_1 y_1$ *und* $S_2 \Longrightarrow^*_{G_2^{\mathrm{K}}} x_2 S_2 y_2$ *existieren, oder falls* $\mathrm{L}(G_1^{\mathrm{K}}) = \emptyset$ *oder* $\mathrm{L}(G_2^{\mathrm{K}}) = \emptyset$ *gelten,*

ii.

$$\begin{aligned}
N_{\mathrm{Prod}} &= N_1^{\mathrm{Prod}} \cup N_2^{\mathrm{Prod}} \cup \{S\}, \\
P_{\mathrm{Prod}} &= P_1^{\mathrm{Prod}} \cup P_2^{\mathrm{Prod}} \cup \\
&\quad \{S \to S_1, S \to S_2\}, \\
\mathrm{Prod}(G_{\mathrm{Prod}}) &= \mathrm{Prod}\left(G_1^{\mathrm{Prod}}\right) + \mathrm{Prod}\left(G_2^{\mathrm{Prod}}\right) + 2, \\
\mathrm{Symb}(G_{\mathrm{Symb}}) &\geq \mathrm{Symb}\left(G_1^{\mathrm{Symb}}\right) + \mathrm{Symb}\left(G_2^{\mathrm{Symb}}\right) + 2,
\end{aligned}$$

falls Initialschleifenableitungen $S_1 \Longrightarrow^*_{G_1^{\mathrm{K}}} x_1 S_1 y_1$ *und* $S_2 \Longrightarrow^*_{G_2^{\mathrm{K}}} x_2 S_2 y_2$ *existieren,*

iii.

$$\begin{aligned}
N_{\mathrm{Prod}} &= \left(\left(N_1^{\mathrm{Prod}} \cup N_2^{\mathrm{Prod}}\right) \setminus \{S_j\}\right) \cup \{S\}, \\
P_{\mathrm{Prod}} &= \left(\left(P_1^{\mathrm{Prod}} \cup P_2^{\mathrm{Prod}}\right) \setminus Q_j^{\mathrm{Prod}}\right) \cup \\
&\quad \Big\{S \to w \,\Big|\, S_j \to w \in Q_j^{\mathrm{Prod}}, w \in \left(N_j^{\mathrm{Prod}} \cup T_j^{\mathrm{Prod}}\right)^*\Big\} \cup \\
&\quad \{S \to S_k\}, \\
\mathrm{Prod}(G_{\mathrm{Prod}}) &= \mathrm{Prod}\left(G_1^{\mathrm{Prod}}\right) + \mathrm{Prod}\left(G_2^{\mathrm{Prod}}\right) + 1, \\
\mathrm{Symb}(G_{\mathrm{Symb}}) &\geq \mathrm{Symb}\left(G_1^{\mathrm{Symb}}\right) + \mathrm{Symb}\left(G_2^{\mathrm{Symb}}\right) + 1
\end{aligned}$$

für $j \in \{1,2\}$ und $k \in \{1,2\} \setminus \{j\}$, falls keine Initialschleifenableitung $S_j \Longrightarrow^*_{G_j^{\mathrm{K}}} x_j S_j y_j$ existiert, es aber $S_k \Longrightarrow^*_{G_k^{\mathrm{K}}} x_k S_k y_k$ gibt.

Beweis: Wir bemerken zunächst, dass es im Fall von $|L_i| \in \mathbb{N}_0$ nie eine Initialableitung $S_i \Longrightarrow^*_{G_i^{\mathrm{K}}} x_i S_i y_i$ für $x_i, y_i \in T_i{}^*$, $x_i y_i \neq \lambda$, in G_i^{K} geben kann (Lemma 2.16, Anstrich i).

Wir gehen in diesem Beweis nicht weiter auf die Korrektheit von $\mathrm{L}(G_{\mathrm{K}}) = L_1 \cup L_2$ ein. Die Konstruktion von G_{K} aus Satz 2.22 liefert entsprechend der Fallunterscheidung $\mathrm{Prod}(L_1 \cup L_2) \leq \mathrm{Prod}(L_1) + \mathrm{Prod}(L_2) + p$ für $p \in \{0,\, 1,\, 2\}$ und $\mathrm{Symb}(L_1 \cup L_2) \leq \mathrm{Symb}(L_1) + \mathrm{Symb}(L_2)$ (für den ersten Fall). Im Folgenden werden wir zeigen, dass $L_1 \cup L_2$ mit nicht weniger Regeln bzw. Symbolen erzeugt werden kann.

Es sei $G'_{\mathrm{K}} = (N'_{\mathrm{K}}, T', P'_{\mathrm{K}}, S')$ eine minimale kontextfreie Grammatik mit $\mathrm{L}(G'_{\mathrm{K}}) = L_1 \cup L_2$ und $\mathrm{K}(G'_{\mathrm{K}}) = \mathrm{K}(L_1 \cup L_2)$. Wir dürfen G'_{K} wegen Lemma 2.14 und Lemma 2.19 als reduziert bzw. frei von kontextfreien Nichtterminalen betrachten. Es gibt für ein beliebiges Nichtterminal $A \in N'_{\mathrm{K}}$ keine zwei Ableitungen

$$\begin{aligned} S' &\Longrightarrow^*_{G'_{\mathrm{K}}} x_1 A y_1 \Longrightarrow^*_{G'_{\mathrm{K}}} w_1 \in L_1, \\ S' &\Longrightarrow^*_{G'_{\mathrm{K}}} x_2 A y_2 \Longrightarrow^*_{G'_{\mathrm{K}}} w_2 \in L_2 \end{aligned} \tag{3.1}$$

für $x_i, y_i \in \mathrm{Alph}(L_i)^*$ und $x_1 y_1 x_2 y_2 \neq \lambda$ (Lemma 2.19) in G'_{K} – denn wenn es sie gäbe, existierten auch die Ableitungen

$$\begin{aligned} A &\Longrightarrow^*_{G'_{\mathrm{K}}} u \text{ wegen } S \Longrightarrow^*_{G'_{\mathrm{K}}} x_1 A y_1 \Longrightarrow^*_{G'_{\mathrm{K}}} x_1 u y_1 = w_1 \in L_1, \\ A &\Longrightarrow^*_{G'_{\mathrm{K}}} v \text{ wegen } S \Longrightarrow^*_{G'_{\mathrm{K}}} x_2 A y_2 \Longrightarrow^*_{G'_{\mathrm{K}}} x_2 v y_2 = w_2 \in L_2 \end{aligned}$$

für $u \in \mathrm{Alph}(L_1)^*$ und $v \in \mathrm{Alph}(L_2)^*$. Infolgedessen wären Ableitungen

$$\begin{aligned} S &\Longrightarrow^*_{G'_{\mathrm{K}}} x_1 A y_1 \Longrightarrow^*_{G'_{\mathrm{K}}} x_1 v y_1, \\ S &\Longrightarrow^*_{G'_{\mathrm{K}}} x_2 A y_2 \Longrightarrow^*_{G'_{\mathrm{K}}} x_2 u y_2 \end{aligned}$$

möglich (u und v wurden „gewechselt“). Wir beachten, dass wir aufgrund von Lemma 2.17 $uv \neq \lambda$ annehmen dürfen und dass wegen $\lambda \notin L_i$ die Aussagen $x_i u y_i \neq \lambda$ und $x_i v y_i \neq \lambda$ stimmen. Als Schlussfolgerung erhalten wir $x_1 v y_1 \notin L_1 \cup L_2$ oder $x_2 u y_2 \notin L_1 \cup L_2$ – es gibt kein Wort in $L_1 \cup L_2$, in dem sowohl Buchstaben aus $\mathrm{Alph}(L_1)$ als auch Buchstaben aus $\mathrm{Alph}(L_2)$ vorkommen. Mithin gibt es die Ableitungen (3.1) in G'_{K} tatsächlich nicht.

Angesichts der vorherigen Betrachtung dürfen wir davon ausgehen, dass keine Regel in G'_{K}, die für die Erzeugung eines Wortes aus L_1 verantwortlich ist, für die Erzeugung der Wörter aus L_2 verwendet wird – und umgekehrt. Folglich gibt es insbesondere keine Initialschleifenableitung in G'_{K}. Damit ist klar, dass wir eine kontextfreie Grammatik $G_i'^{\mathrm{K}} = (N_i'^{\mathrm{K}},\, T'_i,\, P_i'^{\mathrm{K}},\, S')$, $i \in \{1,\, 2\}$, mit $\mathrm{L}(G_i'^{\mathrm{K}}) = L_i$ basierend auf G'_{K} entwerfen können, indem wir alle Regeln, die für die Erzeugung der Wörter aus L_i in G'_{K} benötigt werden, in $G_i'^{\mathrm{K}}$ aufnehmen – $\mathrm{K}(G'_{\mathrm{K}}) = \mathrm{K}(G_1'^{\mathrm{K}}) + \mathrm{K}(G_2'^{\mathrm{K}})$. Für $G_i'^{\mathrm{K}}$ ergibt sich die Aussage $\mathrm{K}(G_i'^{\mathrm{K}}) \geq \mathrm{K}(G_i^{\mathrm{K}})$, da anderenfalls $\mathrm{K}(L_i) \leq \mathrm{K}(G_i'^{\mathrm{K}}) < \mathrm{K}(G_i^{\mathrm{K}}) = \mathrm{K}(L_i)$ widersprüchlich wäre.

- Zu i. Falls $S_1 \Longrightarrow^*_{G_1^{\mathrm{K}}} x_1 S_1 y_1$ und $S_2 \Longrightarrow^*_{G_2^{\mathrm{K}}} x_2 S_2 y_2$ nicht gelten, oder falls $\mathrm{L}(G_1^{\mathrm{K}}) = \emptyset$ oder $\mathrm{L}(G_2^{\mathrm{K}}) = \emptyset$ sind, gilt $\mathrm{K}(G_i'^{\mathrm{K}}) \leq \mathrm{K}(G_i^{\mathrm{K}})$ – wäre dies nicht der Fall, könnten wir die Regeln von G_i^{K} statt derer von $G_i'^{\mathrm{K}}$ in G'^{K} verwenden und erhielten schließlich eine Grammatik mit weniger Regeln oder Symbolen. Mit Obigem ergibt sich $\mathrm{K}(G_i'^{\mathrm{K}}) = \mathrm{K}(G_i^{\mathrm{K}})$, und folglich $\mathrm{K}(G_{\mathrm{K}}) = \mathrm{K}(G'_{\mathrm{K}}) = \mathrm{K}(G_1^{\mathrm{K}}) + \mathrm{K}(G_2^{\mathrm{K}})$. Mithin ist die Konstruktion von G_{K} aus Satz 2.22 korrekt.

- Zu ii. Falls $S_1 \Longrightarrow^*_{G_1^{\mathrm{K}}} x_1 S_1 y_1$ und $S_2 \Longrightarrow^*_{G_2^{\mathrm{K}}} x_2 S_2 y_2$ gelten, enthält nach Voraussetzung jede kontextfreie Grammatik $G_\#^{\mathrm{K}}$ mit $\mathrm{L}(G_\#^{\mathrm{K}}) = L_i$ und $\mathrm{K}(G_\#^{\mathrm{K}}) = \mathrm{K}(L_i)$ eine Initialschleifenableitung. Wäre $\mathrm{K}(G_i'^{\mathrm{K}}) = \mathrm{K}(L_i) = \mathrm{K}(G_i^{\mathrm{K}})$, so gäbe es eine Initialschleifenableitung $S' \Longrightarrow^*_{G_i'^{\mathrm{K}}} x_i' S' y_i'$ für $x_i', y_i' \in T_i'^*$ und $x_i' y_i' \neq \lambda$. Solch eine Initialschleifenableitung gäbe es folglich auch in G'^{K}, was allerdings der Tatsache widerspricht, dass G'^{K} keine Initialschleifenableitung erzeugt. Mithin ist $\mathrm{K}(G_i'^{\mathrm{K}}) \geq \mathrm{K}(G_i^{\mathrm{K}}) + 1$. Mit Obigem folgt

  $$\mathrm{K}(G_{\mathrm{K}}) = \mathrm{K}(G'_{\mathrm{K}}) = \mathrm{K}\left(G_1'^{\mathrm{K}}\right) + \mathrm{K}\left(G_2'^{\mathrm{K}}\right) \geq \mathrm{K}\left(G_1^{\mathrm{K}}\right) + 1 + \mathrm{K}\left(G_2^{\mathrm{K}}\right) + 1.$$

  Die Konstruktion von G_{K} aus Satz 2.22 liefert

  $$\mathrm{Prod}(G_{\mathrm{Prod}}) = \mathrm{Prod}(G'_{\mathrm{Prod}}) \leq \mathrm{Prod}\left(G_1^{\mathrm{Prod}}\right) + \mathrm{Prod}\left(G_2^{\mathrm{Prod}}\right) + 2.$$

  Insgesamt erhalten wir

  $$\mathrm{Prod}(G_{\mathrm{Prod}}) = \mathrm{Prod}(G'_{\mathrm{Prod}}) = \mathrm{Prod}\left(G_1^{\mathrm{Prod}}\right) + \mathrm{Prod}\left(G_2^{\mathrm{Prod}}\right) + 2$$

  sowie

  $$\mathrm{Symb}(G_{\mathrm{Symb}}) = \mathrm{Symb}\left(G'_{\mathrm{Symb}}\right) \geq \mathrm{Symb}\left(G_1^{\mathrm{Symb}}\right) + \mathrm{Symb}\left(G_2^{\mathrm{Symb}}\right) + 2.$$

- Zu iii. Falls $S_j \Longrightarrow^*_{G_j^{\mathrm{K}}} x_j S_j y_j$ für $j \in \{1, 2\}$ nicht gilt und $S_k \Longrightarrow^*_{G_k^{\mathrm{K}}} x_k S_k y_k$ für $k \in \{1, 2\} \setminus \{j\}$ gilt sowie $\mathrm{L}(G_j^{\mathrm{K}}) \neq \emptyset$ ist, erhalten wir durch analoge Beweisführung zum vorherigen Fall die Aussagen

  $$\mathrm{Prod}(G_{\mathrm{Prod}}) = \mathrm{Prod}(G'_{\mathrm{Prod}}) = \mathrm{Prod}\left(G_1^{\mathrm{Prod}}\right) + \mathrm{Prod}\left(G_2^{\mathrm{Prod}}\right) + 1$$

  sowie

  $$\mathrm{Symb}(G_{\mathrm{Symb}}) = \mathrm{Symb}\left(G'_{\mathrm{Symb}}\right) \geq \mathrm{Symb}\left(G_1^{\mathrm{Symb}}\right) + \mathrm{Symb}\left(G_2^{\mathrm{Symb}}\right) + 1.$$

Hier hat nur G_k^{K} eine Initialschleifenableitung, weswegen genau eine zusätzliche Regel im Fall der Regelminimalität genügt. Für die Symbolminimalität wissen wir lediglich, dass mindestens ein weiteres Symbol notwendig ist.

$\square$

Lemma 2.23 – Mindestanzahl von Regeln und Symbolen unendlicher kontextfreier Sprachen
Es sei L eine unendliche kontextfreie Sprache und $G = (N, T, P, S)$ eine reduzierte kontextfreie Grammatik, die L erzeugt. Dann existieren mindestens 2 Regeln in G, die eine Schleifenableitung der Form $A \Longrightarrow^* xAy$ für ein $A \in N$ und $x, y \in T^*$, $xy \neq \lambda$, und eine Ableitung der Form $A \Longrightarrow^* w$, $w \in T^*$, ermöglichen. Zusammen bestehen diese beiden Regeln aus mindestens 5 Symbolen der Art Pfeil und Nichtterminal.

Beweis: Dass Ableitungen der Form $A \Longrightarrow^* xAy$ und der Form $A \Longrightarrow^* w$ durch G erzeugt werden, ist wegen der Reduziertheit und des Lemmas 2.16 klar.

Die Schleifenableitung der Form $A \Longrightarrow^* xAy$ und die Ableitung der Form $A \Longrightarrow^* w$ können nicht von genau einer Regel $A \to z$ für $z \in (N \cup T)^*$ erzeugt werden. Die Regel $A \to z$ allein generierte im Fall $z \in T^*$ nur das Terminalwort z, und im anderen Fall ($|z|_B \geq 1$ für ein $B \in N$) könnte kein Terminalwort durch $A \to z$ allein generiert werden.

Die beiden Regeln $A \to zA$, $z \in T$, und $A \to \lambda$ erzeugen offenbar eine Schleifenableitung der Form $A \Longrightarrow^* xAy$ und eine Ableitung der Form $A \Longrightarrow^* w$. Das Entfernen nur eines Symbols aus diesen Regeln – es kommen nur z oder A auf der rechten Seite von $A \to zA$ infrage – führte dazu, dass die besagte Schleifenableitung nicht erzeugt werden könnte. Auch generierte die Verwendung eines von A verschiedenen Nichtterminals die Ableitungen nicht.

Mithin müssen insgesamt mindestens 2 Regeln und mindestens 5 Symbole der Art Pfeil und Nichtterminal für die Generierung der Schleifenableitung der Form $A \Longrightarrow^* xAy$ und der Ableitung der Form $A \Longrightarrow^* w$ in G vorhanden sein. □

3. Regelkomplexität

Zu Beginn bestimmen wir im Abschnitt 3.1 die Regelkomplexität einiger Sprachen, die wir in den darauffolgenden Abschnitten verwenden werden. In diesen bestimmen wir (in Teilen) die Räume der Regelkomplexität unter den Operationen Spiegelbild (in 3.2), Vereinigung (in 3.3), Konkatenation (in 3.4), Kleene-Abschluss (in 3.5), Homomorphismus (in 3.6), inverser Homomorphismus (in 3.7) und Schnitt mit regulärer Sprache (in 3.8).

3.1. Allgemeines

Wir geben die Regelkomplexitäten einiger Sprachen an, die wir später benötigen werden.

Lemma 3.1 – Kontextfreie Sprachen mit der Regelkomplexität höchstens 2
Die folgenden Aussagen sind für kontextfreie Sprachen L wahr:

i. Es gilt $\mathrm{Prod}(L) = 0$ genau dann, wenn $L = \emptyset$ ist.

ii. Es gilt $\mathrm{Prod}(L) = 1$ genau dann, wenn $|L| = 1$ ist.

iii. Es gilt entweder $|L| = 2$ oder $|L| = |\mathbb{N}|$, wenn $\mathrm{Prod}(L) = 2$ ist.

iv. Es gilt $\mathrm{Prod}(L) = 2$, wenn $|L| = 2$ ist.

Beweis: Es sei $G = (N, T, P, S)$ mit $T = \mathrm{Alph}(L)$ für $L = \mathrm{L}(G)$ eine regel-minimale kontextfreie Grammatik. Die Grammatik G ist nach Lemma 2.14 reduziert.

Zu i. Es gilt die Voraussetzung

$$\mathrm{Prod}(\mathrm{L}(G)) = \mathrm{Prod}(G) = 0.$$

Es existiert also keine Regel in G. Damit gibt es kein $v \in T^*$, für das $S \Longrightarrow^* v$ gilt, woraus unmittelbar $L = \mathrm{L}(G) = \emptyset$ folgt.

Wir geben direkt eine regel-minimale kontextfreie Grammatik an, welche die Sprache $L = \emptyset$ erzeugt:

$$(\emptyset, \emptyset, \emptyset, \lambda).$$

(Eine Grammatik mit weniger als 0 Regeln existiert nicht.)

Zu ii. Es sei

$$\mathrm{Prod}(\mathrm{L}(G)) = \mathrm{Prod}(G) = 1.$$

Nach Voraussetzung gibt es nur eine Regel in G. Dies muss eine Regel für S sein, da sonst $\mathrm{L}(G) = \emptyset$ wäre, wobei $\mathrm{Prod}(\emptyset)$ allerdings 0 ist (Anstrich i). Es sei nun $S \to w$, $w \in (N \cup T)^*$, diese Regel. Das Wort w darf kein Nichtterminal $X \in N$ enthalten, ansonsten enthielte das Wort $v \in (N \cup T)^*$ nach einer beliebigen Ableitung $S \Longrightarrow^* v$ immer das Nichtterminal X, und $\mathrm{L}(G)$ wäre damit die leere Menge. Folglich besteht w nur aus Terminalen aus T. Im Ganzen existiert damit nur die Ableitung $S \Longrightarrow w$, woraus wir sofort $L = \mathrm{L}(G) = \{w\}$ erhalten.

Für $L = \{w\}$, $w \in T^*$, konstruieren wir die kontextfreie Grammatik

$$(\{S\}, T, \{S \to w\}, S),$$

die sicher $\{w\} = L$ erzeugt. Außerdem ist sie regel-minimal, da für jede kontextfreie Grammatik G' mit $\mathrm{Prod}(G') = 0$ stets $\mathrm{L}(G') = \emptyset$ gilt (Anstrich i).

Zu iii. Es sei

$$\mathrm{Prod}(\mathrm{L}(G)) = \mathrm{Prod}(G) = 2.$$

Aufgrund der Voraussetzung $\mathrm{Prod}(G) = 2$ existieren in G die Regeln $X_1 \to w_1$ und $X_2 \to w_2$ für $X_1, X_2 \in N, w_1, w_2 \in (N \cup T)^*$ und $X_1 \to w_1 \neq X_2 \to w_2$. Das Nichtterminal S muss das einzige in N sein: Da es zu jedem Nichtterminal mindestens eine Regel geben muss (sonst wäre es überflüssig) und G genau 2 Regeln haben soll, kann es nicht mehr als 2 Nichtterminale geben. Es seien nun $N = \{S, A\}$ und ohne Beschränkung der Allgemeinheit $S \to w_1, A \to w_2$. Dann darf in w_1 kein S vorkommen – sonst enthielte jedes ableitbare Wort ein S, und es gäbe kein einziges Terminalwort ($\mathrm{L}(G) = \emptyset$) –, dann wäre aber $\mathrm{Prod}(\mathrm{L}(G)) = 0$. Analog enthält w_2 kein A. Allerdings muss in w_1 mindestens ein A vorkommen, anderenfalls könnte die Regel $A \to w_2$ nie angewendet werden. Wir könnten dann aber jedes Vorkommen von A in w_1 durch w_2 ersetzen, wodurch $A \to w_2$ überflüssig wäre. Mithin kann es das Nichtterminal A nicht geben, und wir erhalten $N = \{S\}$ sowie $S \to w_1, S \to w_2$.

Im Fall von $w_1, w_2 \in T^*$ folgt sofort $\mathrm{L}(G) = \{w_1, w_2\}$, womit wir $|L| = |\mathrm{L}(G)| = 2$ gezeigt haben. Die Wörter w_1 und w_2 können nicht beide ein S enthalten, da sonst keine Ableitung zu einem Terminalwort führte, und daher $\mathrm{L}(G)$ die leere Menge wäre sowie $\mathrm{Prod}(\mathrm{L}(G)) = 0$. Auch ist eine Regel $S \to S$ nicht möglich (bereits nach Definition 2.3), da $\mathrm{L}(G)$ sonst nur ein Wort enthielte und $\mathrm{Prod}(\mathrm{L}(G)) = 1$ resultierte. Infolgedessen nehmen wir ohne Beschränkung der Allgemeinheit an, dass in $w_1 = xSy$ mit $x, y \in (N \cup T)^*$, $xy \neq \lambda$, mindestens ein S vorkommt und dass $w_2 \in T^*$ gilt. Wir erhalten dann

$$S \Longrightarrow xSy \Longrightarrow xxSyy \Longrightarrow \cdots \Longrightarrow x^n S y^n, n \in \mathbb{N}.$$

Wenden wir nun (in eventuell mehreren Schritten) die Regel $S \to w_2$ auf alle S in $x^n S y^n$ an, so entsteht aufgrund der Reduziertheit von G ein Terminalwort $x'^n w_2 y'^n \in \mathrm{L}(G)$ mit $x', y' \in T^*$. Aufgrund von $n \in \mathbb{N}$ und der Tatsache, dass immer nur abzählbar viele Wörter erzeugt werden können ($|\mathrm{L}(G)| \leq |\mathbb{N}|$), folgt $|L| = |\mathrm{L}(G)| = |\mathbb{N}|$.

Zu iv. Es sei $L = \mathrm{L}(G) = \{w_1, w_2\}$ mit $w_1, w_2 \in T^*$ und $w_1 \neq w_2$ gegeben ($|L| = 2$). Dann können wir wegen Ausschlusses der Fälle in Anstrich i und ii direkt eine regel-minimale kontextfreie Grammatik angeben, die $L = \{w_1, w_2\}$ mittels zweier Regeln erzeugt:

$$(\{S\}, T, \{S \to w_1, S \to w_2\}, S).$$

□

Lemma 3.2 – Regelkomplexität von $\{a^{2^i} \mid n \leq i \leq m\}$
Für ein Alphabet T, $a \in T$ und $L_{n,m} = \{a^{2^i} \mid n \leq i \leq m\}$ mit $n, m \in \mathbb{N}_0$, $n \leq m$ gilt $\mathrm{Prod}(L_{n,m}) = m - n + 1$.

Beweis: Die obere Schranke $\mathrm{Prod}(L_{n,m}) \leq m - n + 1$ erhalten wir mittels

$$G_{n,m} = \left(\{S\}, \{a\}, \left\{S \to a^{2^i} \;\middle|\; n \leq i \leq m\right\}, S\right), \mathrm{L}(G_{n,m}) = L_{n,m}.$$

Die Existenz der unteren Schranke $\mathrm{Prod}(L_{n,m}) \geq m - n + 1$ zeigen wir im Folgenden, indem wir im Wesentlichen G r u s k a s Beweis des Satzes 6.3 aus [Gru69] wiederholen (der dortige Beweis gilt exakt für $n = 0$, $m \geq 1$). Wir ergründen, dass für die Erzeugung der Wörter $a^{2^i} \in L_{n,m}$, $n \leq i \leq m$, mindestens $m-n+1$ Regeln in einer regel-minimalen kontextfreien Grammatik

$$G_{n,m} = (N_{n,m}, \{a\}, P_{n,m}, S)$$

mit $\mathrm{L}(G_{n,m}) = L_{n,m}$ benötigt werden. Nach Lemma 2.14 ist $G_{n,m}$ reduziert.

Es seien im Folgenden $i, j \in \mathbb{N}$. Zuerst betrachten wir den Fall, dass es nur ein Wort a^{2^i} gibt, also $i = n = m$. Da $S \Longrightarrow^* a^{2^i}$ gilt, muss es mindestens eine Regel in $G_{n,m}$ geben (Lemma 3.1, Anstrich ii). Damit ist $\mathrm{Prod}(G_{n,m}) \geq m - n + 1 = n - n + 1 = 1$ erfüllt.

Es sei nun $n < m$ ($|L_{n,m}| \geq 2$). An dieser Stelle werden wir durch einen Widerspruch feststellen, dass in einem Wort in einer Ableitung $S \Longrightarrow \cdots \Longrightarrow a^{2^i}$ keine zwei Nichtterminale vorkommen können. Wir nehmen dazu an, es gibt eine Ableitung der Form

$$S \Longrightarrow^* xAyBz \Longrightarrow^* a^{2^i}$$

für $x, y, z \in \{a\}^*$, $A, B \in N_{n,m}$ und $n \leq i \leq m$, in der zwei Nichtterminale vorkommen. Nach Lemma 2.15 müssen die folgenden Ableitungen für $p_1 < p_2$ und $q_1 < q_2$ und gewisse $p_1, p_2, q_1, q_2 \in \mathbb{N}_0$ existieren:

$$\begin{aligned} A &\Longrightarrow^* a^{p_1} & \qquad B &\Longrightarrow^* a^{q_1} \\ A &\Longrightarrow^* a^{p_2} & \qquad B &\Longrightarrow^* a^{q_2}. \end{aligned}$$

Anschließend substituieren wir A und B in $xAyBz$ durch a^{p_1} oder a^{p_2} bzw. a^{q_1} oder a^{q_2}, und erhalten damit die vier Wörter

$$\begin{aligned} w_1 &= xa^{p_1}ya^{q_1}z \in L_{n,m}, & \qquad w_3 &= xa^{p_1}ya^{q_2}z \in L_{n,m}, \\ w_2 &= xa^{p_2}ya^{q_1}z \in L_{n,m}, & \qquad w_4 &= xa^{p_2}ya^{q_2}z \in L_{n,m}. \end{aligned}$$

Wir betrachten nun die Anzahl s_j der Buchstaben a in den Wörtern w_j, $1 \leq j \leq 4$, und setzen dazu $s = |xyz|$:

$$\begin{aligned} s_1 &= s + p_1 + q_1, & \qquad s_3 &= s + p_1 + q_2, \\ s_2 &= s + p_2 + q_1, & \qquad s_4 &= s + p_2 + q_2. \end{aligned}$$

Offenbar gelten $s_1 < s_2$, $s_1 < s_3$ sowie

$$s_4 - s_3 = s_2 - s_1.$$

Jedes s_j, $1 \leq j \leq 4$, muss nach Voraussetzung eine Zweierpotenz sein. Daher setzen wir $s_j = 2^{t_j}$ für $t_j \in \mathbb{N}_0$. Ersetzen wir die s_j in $s_4 - s_3 = s_2 - s_1$ durch die 2^{t_j}, erhalten wir

$$\begin{aligned} 2^{t_4} - 2^{t_3} &= 2^{t_2} - 2^{t_1} \\ 2^{t_4} &= 2^{t_1} \cdot \left(2^{t_2 - t_1} + 2^{t_3 - t_1} - 1\right), \end{aligned}$$

was zu einem Widerspruch führt, da die Zweierpotenz 2^{t_4} keinen ungeraden Teiler $2^{t_2-t_1} + 2^{t_3-t_1} - 1$ haben kann – die Faktorisierung ist wegen $s_1 < s_2$ und $s_1 < s_3$ möglich. Somit entstünde doch ein Wort der Form a^j für ein $j \in \mathbb{N}_0$, $n \leq j \leq m$, und $j \neq 2^i$, welches es nach Voraussetzung allerdings nicht geben darf. Letztendlich existiert also keine Ableitung der Form

$$S \Longrightarrow^* xAyBz \Longrightarrow^* a^{2^i},$$

die zwei Nichtterminale enthält. Folglich können auf der rechten Seite von Regeln aus $P_{n,m}$ nicht zwei oder mehr Nichtterminale vorkommen – sämtliche Regeln haben also die Form $A \to w$ oder $A \to uBv$ für $A, B \in N_{n,m}$ und $w, u, v \in \{a\}^*$.

Wir nehmen nun an, dass es mindestens zwei verschiedene Regeln $B_1 \to u_1Av_1 \in P_{n,m}$ und $B_2 \to u_2Av_2 \in P_{n,m}$ für $u_1, v_1, u_2, v_2 \in (N_{n,m} \cup \{a\})^*$ und $B_1, B_2, A \in N_{n,m}$ mit A auf der rechten Seite gibt ($B_1 \to u_1Av_1 \neq B_2 \to u_2Av_2$). Aufgrund der Reduziertheit von $G_{n,m}$ existieren die Ableitungen

$$\begin{aligned} S &\Longrightarrow^* x_1B_1y_1 \Longrightarrow^* x_1u_1Av_1y_1, & A &\Longrightarrow^* a^{p_1}, \\ S &\Longrightarrow^* x_2B_2y_2 \Longrightarrow^* x_2u_2Av_2y_2, & A &\Longrightarrow^* a^{p_2} \end{aligned}$$

für $x_1, y_1, x_2, y_2 \in \{a\}^*$ und $p_1, p_2 \in \mathbb{N}_0$. Die Aussage $x_1u_1v_1y_1 = x_2u_2v_2y_2$ kann nicht stimmen, da wir ansonsten eine der beiden Regeln $B_1 \to u_1Av_1$ und $B_2 \to u_2Av_2$ aus $P_{n,m}$ streichen könnten, ohne dass sich die von $G_{n,m}$ erzeugte Sprache änderte, wir aber – die angenommene Minimalität verletzend – weniger Regeln hätten. Es gilt daher $x_1u_1v_1y_1 \neq x_2u_2v_2y_2$. Ferner gilt aufgrund von Lemma 2.15 die Aussage $a^{p_1} \neq a^{p_2}$, also auch $p_1 \neq p_2$. Als Nächstes substituieren wir A und erhalten die Wörter

$$\begin{aligned} w_1 &= x_1u_1a^{p_1}v_1y_1 \in L_{n,m}, & w_3 &= x_1u_1a^{p_2}v_1y_1 \in L_{n,m}, \\ w_2 &= x_2u_2a^{p_1}v_2y_2 \in L_{n,m}, & w_4 &= x_2u_2a^{p_2}v_2y_2 \in L_{n,m}. \end{aligned}$$

Wir setzen $q_j = |x_ju_jv_jy_j|$, $j \in \{1, 2\}$, und ohne Beschränkung der Allgemeinheit nehmen wir $p_1 < p_2$ und $q_1 < q_2$ an. Es sei weiterhin 2^{t_j} die Anzahl der Buchstaben im Wort w_j, $1 \leq j \leq 4$, welches nur aus Vorkommen von a besteht und dessen Länge nach Voraussetzung eine Zweierpotenz ist. Aus den vorherigen Wörtern ergeben sich daher die Längen

$$2^{t_1} = q_1 + p_1, \qquad\qquad 2^{t_3} = q_1 + p_2,$$

$$2^{t_2} = q_2 + p_1, \qquad 2^{t_4} = q_2 + p_2.$$

Wir können hier wie oben einen Widerspruch herleiten, denn es gelten $2^{t_1} < 2^{t_2}$ und $2^{t_1} < 2^{t_3}$ sowie $2^{t_4} - 2^{t_3} = 2^{t_2} - 2^{t_1}$. Somit kommt jedes Nichtterminal $A \in N$ auf der rechten Seite höchstens einer Regel vor, und auf dieser rechten Seite höchstens einmal (obiger Fakt). Da dies für alle $A \in N$ gilt, folgt, dass $A \neq S$ ein isoliertes Nichtterminal ist. Nach Lemma 2.21 existiert ein solches allerdings in der regel-minimalen kontextfreien Grammatiken $G_{n,m}$ nicht. Folglich ist $S \in N$ das einzige Nichtterminal in $G_{n,m}$. Daher und infolge der Nichtexistenz von Schleifenableitungen in $G_{n,m}$ darf S auf keiner rechten Seite einer Regel vorkommen (Lemma 2.16). Folglich gibt es keine Regeln in $P_{n,m}$, in denen ein Nichtterminal auf den rechten Seiten vorkommt.

Es kommen letztlich nur Terminale a auf der rechten Seite einer jeden Regel aus $P_{n,m}$ vor. Daher werden für die $m-n+1$ Wörter in $L_{n,m}$ auch mindestens $m-n+1$ Regeln mit S auf der linken Seite in $P_{n,m}$ benötigt. Es folgt $\operatorname{Prod}(L_{n,m}) = \operatorname{Prod}(G_{n,m}) \geq m-n+1$. Zusammen mit der obigen Aussage $\operatorname{Prod}(L_{n,m}) \leq m-n+1$ erhalten wir schlussendlich $\operatorname{Prod}(L_{n,m}) = m-n+1$. $\square$

Lemma 3.3 – Regelkomplexität von $\{a^{iq} \mid i \geq p\}$
Für ein Alphabet T, $a \in T$ und $L_{p,q} = \{a^{iq} \mid i \geq p\}$ mit $p \in \mathbb{N}_0$ sowie $q \in \mathbb{N}$ gilt $\operatorname{Prod}(L_{p,q}) = 2$. Außerdem erzeugt jede regel-minimale kontextfreie Grammatik für $L_{p,q}$ eine Initialschleifenableitung und enthält nur ein Nichtterminal.

Beweis: Für die kontextfreie Grammatik

$$G_{p,q} = (\{S\}, \{a\}, P_{p,q}, S)$$

mit

$$P_{p,q} = \{S \to a^q S, S \to a^{pq}\}$$

gilt $\operatorname{L}(G_{p,q}) = L_{p,q}$, und folglich erhalten wir $\operatorname{Prod}(L_{p,q}) \leq \operatorname{Prod}(G_{p,q}) = 2$. Lemma 3.1, Anstrich iii und $|L_{p,q}| = |\mathbb{N}|$ führen zu $\operatorname{Prod}(L_{p,q}) \geq 2$. Die beiden Tatsachen liefern unmittelbar $\operatorname{Prod}(L_{p,q}) = 2$.

Dem Beweis zum Lemma 3.1, Anstrich iii entnehmen wir, dass es nur ein Nichtterminal S – nämlich das Startsymbol – in einer regel-minimalen kontextfreien Grammatik $G_{p,q}$ mit $\operatorname{L}(G_{p,q}) = L_{p,q}$ und $\operatorname{Prod}(G_{p,q}) = \operatorname{Prod}(L_{p,q}) = 2$ geben kann. Da $|L_{p,q}| = |\mathbb{N}|$ gilt, folgt dann $S \Longrightarrow^* xSy$ für gewisse $x, y \in \{a\}^*$, $xy \neq \lambda$, für $G_{p,q}$ (Lemma 2.16). $\square$

Lemma 3.4 – Regelkomplexität von $\{a^1, a^3, a^4, \dots\}$
Für ein Alphabet T, $a \in T$ und $L_3 = \{a\} \cup \{a^i \mid i \geq 3\}$ gilt $\operatorname{Prod}(L_3) = 3$. Jede regel-minimale kontextfreie Grammatik für L_3 erzeugt eine Initialschleifenableitung.

Beweis: Für die kontextfreie Grammatik

$$G_3 = (\{S\}, \{a\}, P_3, S)$$

mit

$$P_3 = \{S \to aaS, S \to aaaS, S \to a\}$$

gilt $\mathrm{L}(G_3) = L_3$ und folglich $\mathrm{Prod}(L_3) \leq \mathrm{Prod}(G_3) = 3$.

Aus Lemma 3.1, Anstrich iii und $|L_3| = |\mathbb{N}|$ folgt $\mathrm{Prod}(L_3) \geq 2$. Wir zeigen nun, dass $\mathrm{Prod}(L_3) = 2$ nicht gilt. Angenommen, es gibt eine regel-minimale kontextfreie Grammatik $G = (N, T, P, S)$ mit $T = \{a\}$, $\mathrm{L}(G) = L_3$ und $\mathrm{Prod}(G) = \mathrm{Prod}(L_3) = 2$. Die Grammatik G ist nach Lemma 2.14 reduziert. Es kann nur ein Nichtterminal, nämlich S, in N geben – anderenfalls wäre die Bedingung $|\mathrm{L}(G, A)| \geq 2$ für ein $S \neq A \in N$ (Lemma 2.15) nicht gegeben, da genau eine Regel mit S und genau eine mit A auf der linken Seite existierte.

Aufgrund von $a \in L_3$ muss es die Regel $S \to a \in P$ geben. Weiterhin existiert nach Lemma 2.16, Anstrich ii eine Schleifenableitung der Form $A \Longrightarrow^* xAy$ für ein $A \in N$ und $x, y \in T^*$, $xy \neq \lambda$, genauer: $S \Longrightarrow^* xSy$ für $x, y \in \{a\}^*$. Die Regel $S \to a$ generiert diese Schleifenableitung nicht. Daher muss es eine Regel der Form $S \to uSv$ für $u, v \in \{S, a\}^*$ und $uv \neq \lambda$ geben (eine Regel $S \to S$ ist nach Definition 2.3 sinnvollerweise nicht erlaubt). Die Fälle $uv = a$ oder $uv = S$ kommen nicht infrage, da sonst $a^2 \notin L_3$ ableitbar wäre. Des Weiteren kann $|uv| \geq 4$ nicht gelten – anderenfalls könnte $a^3 \in L_3$ nicht erzeugt werden. Es gibt nun keine andere Möglichkeit, als $a^3 \in L_3$ über $S \to uSv$ mit $|uv| = 2$ zu erzeugen. Wir erhalten bis auf Veränderung der Reihenfolge von a und S die Fälle $S \to aaS$, $S \to aSS$ und $S \to SSS$ – die Reihenfolge von a und S schränkt die Allgemeinheit nicht ein. Wie wir sehen, lässt sich $a^4 \in L$ mit den beiden Regeln $S \to a$ und $S \to uSv$, $|uv| = 2$, nicht erzeugen, da $S \Longrightarrow uSv \Longrightarrow u^2Sv^2 \Longrightarrow^* a^5$ die nächstkürzere Ableitung ist. Unsere Annahme muss somit fallen gelassen werden, und wir erhalten insgesamt $\mathrm{Prod}(L_3) = 3$.

Wir nehmen an, es gibt eine regel-minimale kontextfreie Grammatik $G = (N, T, P, S)$ mit $T = \{a\}$, $\mathrm{L}(G) = L_3$ und $\mathrm{Prod}(G) = \mathrm{Prod}(L_3) = 3$, für die keine Ableitung der Form $S \Longrightarrow^* x'Sy'$ mit $x', y' \in T^*$ und $x'y' \neq \lambda$ existiert. Die Grammatik G ist nach Lemma 2.14 reduziert. Da L_3 eine unendliche Sprache ist, existiert nach Lemma 2.16, Anstrich ii eine Schleifenableitung der Form $A \Longrightarrow^* xAy$ für ein $A \in N$ und $x, y \in T^*$, $xy \neq \lambda$. Es gilt $A \neq S$. Es kann neben S und A kein weiteres Nichtterminal $B \in N$ geben, da letztendlich die Bedingung $|\mathrm{L}(G, A)| \geq 2$ oder $|\mathrm{L}(G, B)| \geq 2$ aus Lemma 2.15 nicht erfüllt wäre.

Es darf wegen der Reduziertheit von G keine Regel mit S auf der rechten Seite geben, da sonst $S \Longrightarrow^* x'Sy'$ möglich wäre. Aufgrund der Reduziertheit muss A außerdem auf der rechten Seite einer Regel von S vorkommen. Diese Regel sei $S \to x_1Ay_1$ mit $x_1, y_1 \in (\{A\} \cup T)^*$. Hieraus folgt, dass Regeln $A \to x_2Ay_2$ mit $x_2, y_2 \in (\{A\} \cup T)^*$ und $A \to w$ mit $w \in T^*$ existieren müssen ($S \to w'$, $w' \in T^*$, ist nicht möglich). Wäre $|x_1Ay_1| \geq 2$, könnte $a \in L_3$ nicht erzeugt werden. Es folgt $x_1y_1 = \lambda$, also $S \to A \in P$. Der Aussage $\mathrm{Prod}(L) = 3$ widersprechend, könnte nun die kontextfreie Grammatik

$$(\{A\}, \{a\}, \{A \to x_2Ay_2, A \to w\}, A)$$

konstruiert werden, die L_3 mit 2 Regeln erzeugt. □

Lemma 3.5 – Regelkomplexität von $\{a^{iq} \mid n \leq i \leq m\}$

Für ein Alphabet T, $a \in T$ und $L_{n,m,q} = \{a^{iq} \mid n \leq i \leq m\}$ mit $q \in \mathbb{N}$ und $n, m \in \mathbb{N}_0$, $n \leq m$ sowie $m - n + 1 \geq 3$ gilt $\mathrm{Prod}(L_{n,m,q}) = 3$.

Beweis: Für die Grammatik

$$G_{n,m,q} = (\{S, A\}, \{a\}, P_{n,m,q}, S)$$

mit

$$P_{n,m,q} = \{S \to a^{nq}A^{m-n}, A \to a^q, A \to \lambda\}$$

gilt $\mathrm{L}(G_{n,m,q}) = L_{n,m,q}$. Damit haben wir $\mathrm{Prod}(L_{n,m,q}) \leq \mathrm{Prod}(G_{n,m,q}) = 3$ gezeigt.
Aufgrund von Lemma 3.1 und $|L_{n,m,q}| \geq 3$ (wegen $m - n + 1 \geq 3$) sowie $|L_{n,m,q}| \in \mathbb{N}$ erhalten wir $\mathrm{Prod}(L_{n,m,q}) \geq 3$. Mit der obigen Aussage folgt sofort $\mathrm{Prod}(L_{n,m,q}) = 3$. □

Lemma 3.6 – Regelkomplexität von $\{a^{2^i} \mid n \leq i \leq m\}$ mit Erweiterungen
Für ein Alphabet T, $a \in T$ und

i. $L_{n,m} = \{a^{2^i} \mid n \leq i \leq m\} \cup \{a^{i2^{m+1}} \mid i \geq 1\}$,

ii. $L_{n,m,p,q} = \{a^{2^i} \mid n \leq i \leq m\} \cup \{a^{i2^{m+1}} \mid p \leq i \leq q\}$

mit $n, m, p, q \in \mathbb{N}_0$ und $n \leq m$, $p \leq q$ sowie $q - p + 1 \geq 3$ gilt $\mathrm{Prod}(L_{n,m}) = \mathrm{Prod}(L_{n,m,p,q}) = m - n + 4$. Des Weiteren erzeugt keine regel-minimale kontextfreie Grammatik für $L_{n,m}$ eine Initialschleifenableitung.

Beweis: Zu i. Es sei $G_{n,m} = (N_{n,m}, \{a\}, P_{n,m}, S)$ eine regel-minimale kontextfreie Grammatik mit der Eigenschaft $\mathrm{L}(G_{n,m}) = L_{n,m}$. Die Grammatik $G_{n,m}$ ist nach Lemma 2.14 reduziert. Reichel hat in Lemma 4.3 in [Rei90] gezeigt, dass in $G_{0,k-1}$, $k \in \mathbb{N}$, die Regeln $S \to a^{2^i}$ für $0 \leq i \leq k-1$, $i \in \mathbb{N}$, existieren müssen. Der zugehörige Beweis lässt sich für $L_{n,m}$ leicht erweitern und begründet dann die Existenz von $m - n + 1$ Regeln $S \to a^{2^i}$, $n \leq i \leq m$, $i \in \mathbb{N}$ in $P_{n,m}$.
Aus Lemma 2.16, Anstrich ii lesen wir das Vorkommen einer Schleifenableitung $A \Longrightarrow^* a^iAa^j$ mit $A \in N_{n,m}$ und $i, j \in \mathbb{N}_0$, $i+j \geq 1$, ab. Angenommen, das Nichtterminal A ist S. Dann existieren die Ableitungen

$$S \Longrightarrow^* a^iSa^j \Longrightarrow a^ia^{2^m}a^j = a^{k2^{m+1}}, k \in \mathbb{N}_0, \tag{1.1}$$

$$S \Longrightarrow^* a^iSa^j \Longrightarrow^* a^{2i}Sa^{2j} \Longrightarrow a^{2i}a^{2^m}a^{2j} = a^{k'2^{m+1}}, k' \in \mathbb{N}_0. \tag{1.2}$$

Wir betrachten die Länge des Terminalwortes in Ableitung (1.1):

$$\begin{aligned} i + 2^m + j &= k2^{m+1} = 2k2^m \\ i + j &= 2k2^m - 2^m = 2^m(2k-1). \end{aligned}$$

Als Nächstes schauen wir uns die Länge des Terminalwortes in Ableitung (1.2) an:

$$\begin{aligned} 2i + 2^m + 2j &= k'2^{m+1} \\ i + 2^{m-1} + j &= k'2^m \\ i + j &= k'2^m - 2^{m-1} = 2^m\left(k' - \frac{1}{2}\right). \end{aligned}$$

Schließlich setzen wir die beiden vorher erwähnten Gleichungen gleich:

$$\begin{aligned} 2^m (2k-1) &= 2^m \left(k' - \frac{1}{2}\right) \\ 2k-1 &= k' - \frac{1}{2} \\ 2k - \frac{1}{2} &= k' \notin \mathbb{N}_0. \end{aligned}$$

Folglich müssen wir die Annahme $A = S$ fallen lassen – es gibt demnach keine Initialschleifenableitung in $G_{n,m}$.

Da $A \neq S$ gilt, muss es neben den $m-n+1$ Regeln $S \to a^{2^i}$, $n \leq i \leq m$, $i \in \mathbb{N}$, eine weitere Regel geben, die $S \Longrightarrow^* xAy$ für $x, y \in \{a\}^*$ ermöglicht. Darüber hinaus müssen eine zusätzliche Regel für $A \Longrightarrow^* a^i A a^j$ für $i, j \in \mathbb{N}_0$ und $i + j \geq 1$ existieren sowie eine Regel für $A \Longrightarrow^* w$ für $w \in \{a\}^*$ (Lemma 2.23). Wir zählen insgesamt drei weitere Regeln, wodurch wir zu $\mathrm{Prod}(G_{n,m}) \geq m-n+4$ und $\mathrm{Prod}(L_{n,m}) \geq m-n+4$ gelangen.

Wir setzen nun (siehe auch Lemma 3.3)

$$\begin{aligned} G_{n,m} &= (\{S, A\}, \{a\}, P_{n,m}, S), \\ P_{n,m} &= \left\{S \to a^{2^i} \,\middle|\, n \leq i \leq m\right\} \cup \left\{S \to A, A \to a^{2^{m+1}} A, A \to a^{2^{m+1}}\right\} \end{aligned}$$

und erhalten $\mathrm{L}(G_{n,m}) = L_{n,m}$ sowie $\mathrm{Prod}(G_{n,m}) = m-n+4$. Es folgt $\mathrm{Prod}(L_{n,m}) \leq m-n+4$. Mit Obigem ergibt sich daraus $\mathrm{Prod}(L_{n,m}) = m-n+4$.

Zu ii. Analog zu Anstrich i sind in einer regel-minimalen kontextfreien Grammatik $G_{n,m,p,q} = (N_{n,m,p,q}, \{a\}, P_{n,m,p,q}, S)$ mit $\mathrm{L}(G_{n,m,p,q}) = L_{n,m,p,q}$ die Regeln $S \to a^{2^i}$ für $n \leq i \leq m$ enthalten. Da $L_{n,m,p,q}$ endlich ist, existiert nach Lemma 2.16, Anstrich i keine Ableitung der Form $S \Longrightarrow^* xSy$ für $x, y \in \{a\}^*$, $xy \neq \lambda$. Folglich gibt es keine Regel mit S auf der rechten Seite. Ergo finden die Regeln $S \to a^{2^i}$ keine Anwendung bei der Erzeugung der Wörter $a^{i2^{m+1}}$ für $p \leq i \leq q$, $i \in \mathbb{N}$. Wir argumentieren nun wie im Beweis zu Lemma 3.5, um zu zeigen, dass wir drei Regeln für die Generierung dieser Wörter benötigen. Wir erhalten damit zusammen, dass die Aussage $\mathrm{Prod}(G_{n,m,p,q}) \geq m-n+4$ gilt.

Schlussendlich sehen wir mit Obigem $\mathrm{Prod}(L_{n,m,p,q}) = m-n+4$ wegen der $L_{n,m,p,q}$ erzeugenden kontextfreien Grammatik

$$\begin{aligned} G_{n,m,p,q} &= (\{S, A\}, \{a\}, P_{n,m,p,q}, S), \\ P_{n,m,p,q} &= \left\{S \to a^{2^i} \,\middle|\, n \leq i \leq m\right\} \cup \left\{S \to a^{p2^{m+1}} A^{q-p}, A \to a^{2^{m+1}}, A \to \lambda\right\} \end{aligned}$$

ein. □

Lemma 3.7 – Regelkomplexität von $\{a_i \,|\, 1 \leq i \leq n\}^+$ und $\{a_i \,|\, 1 \leq i \leq n\}^*$
Für $n \in \mathbb{N}$, ein Alphabet T_n mit mindestens n Buchstaben und paarweise verschiedene $a_i \in T_n$, $1 \leq i \leq n$, sowie

i. die kontextfreie Sprache $L_n = \{a_i \,|\, 1 \leq i \leq n\}^+$,

ii. die kontextfreie Sprache $J_n = \{a_i \,|\, 1 \leq i \leq n\}^*$

gilt $\mathrm{Prod}\,(L_n) = \mathrm{Prod}\,(J_n) = n+1$. Außerdem erzeugt jede regel-minimale kontextfreie Grammatik für L_n oder J_m eine Initialschleifenableitung.

Beweis: Zu i. Die kontextfreie Grammatik

$$G_n = (\{S\}, \{a_i \mid 1 \leq i \leq n\}, P_n, S)$$

mit

$$P_n = \{S \to SS\} \cup \{S \to a_i \mid 1 \leq i \leq n\}$$

erzeugt die Sprache L_n mit $\mathrm{Prod}\,(G_n) = n+1$ Regeln. Es folgt $\mathrm{Prod}\,(L_n) \leq n+1$.

Es sei nun $G_n = (N_n,\ T_n,\ P_n,\ S)$ mit $T_n = \{a_i \,|\, 1 \leq i \leq n\}$ eine regel-minimale kontextfreie Grammatik, für die $\mathrm{L}\,(G_n) = L_n$ gilt. Diese Grammatik ist nach Lemma 2.14 reduziert. Für das in Lemma 2.10 angegebene X gilt $X = T_n$. Daher existiert eine Regel $A_i \to w_i \in P_n$ mit $A_i \in N_n$, $w_i \in (\{a_i\} \cup N_n)^*$ und $|w_i|_{a_i} = 1$ für jedes a_i, $1 \leq i \leq n$, in P_n. Es gibt also mindestens n Regeln.

Für alle $1 \leq i \leq n$ muss es eine Ableitung $B_i \Longrightarrow^* x_i B_i y_i \Longrightarrow^* a_i^j$ mit $B_i \in N_n$ und $x_i, y_i \in \{a_i\}^*$, $x_i y_i \neq \lambda$, geben, die das Erzeugen unendlich vieler Wörter a_i^j für alle $j \in \mathbb{N}$ ermöglicht (Lemma 2.16, Anstrich ii). Angenommen, das w_i aus $A_i \to w_i$ enthält neben a_i nur von A_i verschiedene Nichtterminale oder keines. Dann benötigen wir mindestens eine weitere Regel in P_n, damit $B_i \Longrightarrow^* x_i B_i y_i$ existiert – die Regeln $A_i \to w_i$ ($|w_i|_{A_i} = 0$) und $A_j \to w_j$ für $1 \leq j \leq n$ mit $j \neq i$ erzeugen eine solche Schleifenableitung nicht. Enthält w_i andererseits ein Nichtterminal A_i, kann damit die Ableitung $B_i \Longrightarrow^* x_i B_i y_i$ erzeugt werden, allerdings muss es dann mindestens eine weitere Regel $A_i \to w_i'$, $w_i' \in ((N_n \setminus \{A_i\}) \cup T_n)^*$, in P_n geben, um eine Terminierung dieser Schleifenableitung zu ermöglichen. Mithin gelten $\mathrm{Prod}\,(G_n) = \mathrm{Prod}\,(L_n) \geq n+1$. Insgesamt bekommen wir $\mathrm{Prod}\,(L_n) = n+1$ heraus.

Angenommen, es gibt eine kontextfreie Grammatik $G_n = (N_n,\ T_n,\ P_n,\ S)$, $T_n = \{a_i \,|\, 1 \leq i \leq n\}$, mit $\mathrm{Prod}\,(G_n) = \mathrm{Prod}\,(L_n) = n+1$, für die es keine Ableitung $S \Longrightarrow^* xSy$ für gewisse $x, y \in {T_n}^*$, $xy \neq \lambda$, gibt. Dann existiert aufgrund von Lemma 2.16, Anstrich ii ein weiteres – von S verschiedenes – Nichtterminal $A \in N_n \setminus \{S\}$, für das es dann eine Schleifenableitung $A \Longrightarrow^* x'Ay'$, $x', y' \in {T_n}^*$, $x'y' \neq \lambda$, gibt. Aufgrund der Reduziertheit von G_n (Lemma 2.14) existieren weiterhin Ableitungen $S \Longrightarrow^* x''Ay''$, $x'', y'' \in {T_n}^*$, und $A \Longrightarrow^* w$, $w \in {T_n}^*$. In P_n müssen $n+1$ Regeln enthalten sein. Wir benötigen wie oben n Regeln der Form $A_i \to w_i \in P_n$ mit $A_i \in N_n$, $w_i \in (\{a_i\} \cup N_n)^*$ und $|w_i|_{a_i} = 1$ für $1 \leq i \leq n$. Des Weiteren muss es wie zuvor eine Ableitung

$$S \Longrightarrow^* x_i' B_i y_i' \Longrightarrow^* x_i' x_i B_i y_i y_i' \Longrightarrow^* a_i^j \text{ für } x_i', y_i' \in {T_n}^* \tag{1.3}$$

geben, allerdings ist B_i nicht S. Da ausschließlich das Terminal a_i erzeugt wird, wird keine Regel $A_k \to w_k$ für $1 \leq k \leq n$ und $k \neq i$ angewendet. Für die Realisierung der Ableitung (1.3) benötigen wir mindestens drei Regeln: eine für $S \Longrightarrow^* x_i' B_i y_i'$, eine für $B_i \Longrightarrow^* x_i B_i y_i$ und eine für $B_i \Longrightarrow^* a_i^l, l \in \mathbb{N}_0$. Selbst wenn wir $A_i \to w_i$ verwenden, muss es mindestens zwei weitere Regeln geben. Somit gilt $\mathrm{Prod}\,(G_n) \geq n+2$ – ein Widerspruch zur Annahme $\mathrm{Prod}(G_n) = n + 1$. Daher muss es doch eine Initialschleifenableitung $S \Longrightarrow^* xSy$ geben.

Zu ii. Die kontextfreie Grammatik

$$G_n = (\{S\}, \{a_i \mid 1 \leq i \leq n\}, P_n, S)$$

mit

$$P_n = \{S \to \lambda\} \cup \{S \to a_i S \mid 1 \leq i \leq n\}$$

generiert J_n mit $\mathrm{Prod}(G_n) = n + 1$ Regeln. Es folgt $\mathrm{Prod}(J_n) \leq n + 1$.

Dass wir n Regeln für die Erzeugung der $a_i \in J_n$, $1 \leq i \leq n$, und eine Regel für die Erzeugung oder das Terminieren einer Schleifenableitung nicht entbehren können, können wir analog zum Beweis von Anstrich i zeigen. Damit gilt $\mathrm{Prod}(J_n) \geq n + 1$. Schlussendlich erhalten wir $\mathrm{Prod}(J_n) = n + 1$.

Analog zum Beweis von Anstrich i können wir außerdem zeigen, dass es für $G_n = (N_n, T_n, P_n, S)$, $T_n = \{a_i \mid 1 \leq i \leq n\}$, mit $\mathrm{Prod}(G_n) = \mathrm{Prod}(J_n) = n + 1$ immer eine Initialschleifenableitung $S \Longrightarrow^* xSy$ für gewisse $x, y \in T_n{}^*$, $xy \neq \lambda$ gibt. □

Lemma 3.8 – Regelkomplexität von $\{a_i \mid 1 \leq i \leq n\}$
Für $n \in \mathbb{N}$, ein Alphabet T_n mit mindestens n Buchstaben und paarweise verschiedene $a_i \in T_n$ sowie $L_n = \{a_i \mid 1 \leq i \leq n\}$ gilt $\mathrm{Prod}(L_n) = n$.

Beweis: Die Sprache L_n lässt sich als

$$L_n = \bigcup_{i=1}^{n} \{a_i\}$$

darstellen. Nach Lemma 3.1 gilt $\mathrm{Prod}(\{a_i\}) = 1, 1 \leq i \leq n$. Außerdem sind die Alphabete der Sprachen $\{a_i\}$ und $\{a_j\}$ für $1 \leq i \leq n$, $1 \leq j \leq n$ und $i \neq j$ disjunkt. Durch die mehrmalige Anwendung des Satzes 2.22 resultiert dann die nachstehende Aussage:

$$\mathrm{Prod}(L_n) = \mathrm{Prod}\left(\bigcup_{i=1}^{n} \{a_i\}\right) = \sum_{i=1}^{n} \mathrm{Prod}(\{a_i\}) = \sum_{i=1}^{n} 1 = n.$$

Damit ist das Lemma bewiesen. □

Lemma 3.9 – Regelkomplexität von $\{a^i x b^i \mid i \geq 0, x \in \{c_1, \ldots, c_{n-1}\}\}$
Für $n \in \mathbb{N}$, $n \geq 2$, ein Alphabet T_n mit mindestens $n + 1$ Buchstaben, paarweise verschiedene $a, b, c_i \in T_n$, $1 \leq i \leq n - 1$, und

$$L_n = \left\{a^i x b^i \,\middle|\, i \geq 0, x \in \{c_1, \ldots, c_{n-1}\}\right\}$$

gilt $\mathrm{Prod}(L_n) = n$.

Beweis: Es sei $G_n = (N_n, T_n, P_n, S)$ mit $\mathrm{L}(G_n) = L_n$ eine regel-minimale kontextfreie Grammatik. Nach Lemma 2.14 ist G_n reduziert. Da L_n eine unendliche Sprache ist, existiert nach Lemma 2.16 eine Schleifenableitung $A \Longrightarrow^* uAv$ für $A \in N_n$ und $u, v \in {T_n}^*$, $uv \neq \lambda$. Die Wörter u und v dürfen kein c_i, $1 \leq i \leq n-1$, enthalten – sonst gäbe es eine Ableitung $A \Longrightarrow^* uAv \Longrightarrow^* u^2Av^2$, in der das Wort u^2Av^2 mehr als nur ein c_i enthielte, und somit ein Wort in L_n existierte, in dem mehr als nur ein c_i vorkäme. Infolgedessen muss es für die Erzeugung der Ableitung $A \Longrightarrow^* uAv$ mindestens eine Regel in P_n geben, in der kein c_i vorkommt.

Es sei $X = \{c_1, \dots, c_{n-1}\}$. Offenbar gelten für X die Voraussetzungen in Lemma 2.9. Folglich gibt es für jedes $c \in X$ eine separate Regel mit c auf der rechten Seite. Mit Obigem befinden sich in P_n folglich mindestens $1 + (n-1) = n$ Regeln. Daher ist $\mathrm{Prod}(L_n) \geq n$.

Die kontextfreie Grammatik

$$(\{S\}, T_n, \{S \to aSb\} \cup \{S \to c_i \mid 1 \leq i \leq n-1\}, S)$$

mit

$$T_n = \{a, b\} \cup \{c_i \mid 1 \leq i \leq n-1\}$$

generiert L_n mit genau n Regeln. Mithin ergeben sich $\mathrm{Prod}(L_n) \leq n$ und insgesamt $\mathrm{Prod}(L_n) = n$. □

Lemma 3.10 – Regelkomplexität von $\{a^ixb^i \mid i \geq 0, x \in \{c_1, \dots, c_{n-2}\}\} \cdot \{\alpha\}$
Für $n \in \mathbb{N}$, $n \geq 3$, ein Alphabet T_n mit mindestens $n+1$ Buchstaben, paarweise verschiedene $a, b, c_i, \alpha \in T_n$, $1 \leq i \leq n-2$, und

$$L_n = \left\{ a^ixb^i \;\middle|\; i \geq 0, x \in \{c_1, \dots, c_{n-2}\} \right\} \cdot \{\alpha\}$$

gilt $\mathrm{Prod}(L_n) = n$.

Beweis: Es sei $G_n = (N_n, T_n, P_n, S)$ mit $\mathrm{L}(G_n) = L_n$ eine regel-minimale kontextfreie Grammatik und $C_n = \{c_1, \dots, c_{n-2}\}$. Nach Lemma 2.14 ist G_n reduziert. Da L_n eine unendliche Sprache ist, existiert nach Lemma 2.16 eine Schleifenableitung $A \Longrightarrow^* uAv$ für $A \in N_n$ und $u, v \in {T_n}^*$, $uv \neq \lambda$. Außerdem gibt es wegen der Reduziertheit von G_n die Ableitungen $S \Longrightarrow^* u'Av'$ mit $u', v' \in {T_n}^*$ und $A \Longrightarrow^* w$ für $w \in {T_n}^*$. Die Wörter u und v dürfen keinen Buchstaben $c \in C_n$ und nicht α enthalten – sonst gäbe es eine Ableitung $A \Longrightarrow^* uAv \Longrightarrow^* u^2Av^2$, in der das Wort u^2Av^2 mehr als ein c bzw. α enthielte, und somit ein Wort in L_n existierte, in dem mehr als ein c bzw. α vorkäme. Es gilt also $u, v \in \{a, b\}^*$, $uv \neq \lambda$. Infolgedessen muss es für die Erzeugung der Ableitung $A \Longrightarrow^* uAv$ mindestens eine Regel p_A in P_n geben, in der kein $c \in C_n$ und kein α vorkommt.

Es existiert eine Ableitung

$$S \Longrightarrow^* u'Av' \Longrightarrow^* u'u^pAv^pv' \Longrightarrow^* u'u^pwv^pv' = a^mxb^m\alpha \in L_n$$

für $p, m \in \mathbb{N}_0$ sowie $x \in C_n$. Wir zeigen nun alle möglichen Fälle für u und v ($uv \neq \lambda$):

- $u, v \in \{a, b\}^* \setminus (\{a\}^+ \cup \{b\}^+)$:

  - Aus $u \neq \lambda$ folgen $|u|_a \geq 1$ und $|u|_b \geq 1$. Somit enthält $u'u^p wv^p v'$ ein unmögliches Teilwort ab oder ba, was $u'u^p wv^p v' \notin L_n$ impliziert. Aus $u = \lambda$ folgt wegen $uv \neq \lambda$ die Tatsache $v \neq \lambda$. Ergo gelten $|v|_a \geq 1$ und $|v|_b \geq 1$, was wiederum zu einem unmöglichen Teilwort führt und $u'u^p wv^p v' \notin L_n$ zur Folge hat.
  - Analoges ergibt sich für v.

- $u \in \{a\}^+ \cup \{b\}^+$, $v = \lambda$:
  - Aus $u \in \{a\}^+$ folgt

$$\begin{aligned} S &\Longrightarrow^* u'u^{p+1}Av^{p+1}v' = u'u^{p+1}A\lambda^{p+1}v' \\ &\Longrightarrow^* a^{m+|u|}xb^m\alpha \notin L_n. \end{aligned}$$

  - Der Fall $u \in \{b\}^+$ führt analog zu einem Wort, das nicht in L_n enthalten ist.

- Für $u = \lambda$, $v \in \{a\}^+ \cup \{b\}^+$ zeigen wir analog zum vorherigen Fall, dass ungültige Wörter entstünden.

- Aus $u \in \{b\}^+$, $v \in \{a\}^+$ folgt

$$\begin{aligned} S &\Longrightarrow^* u'Av' \Longrightarrow u'uAvv' = u'b^{|u|}Aa^{|v|}v' \\ &\Longrightarrow^* u'b^{|u|}wa^{|v|}v' \notin L_n. \end{aligned}$$

- Aus $u \in \{a\}^+$, $v \in \{a\}^+$ folgt

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^{m+|u|+|v|}xb^m\alpha \notin L_n.$$

- Der Fall $u \in \{b\}^+$, $v \in \{b\}^+$ impliziert analog die Entstehung ungültiger Wörter.

- $u \in \{a\}^+$, $v \in \{b\}^+$:

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^{m+|u|}xb^{m+|v|}\alpha.$$

  Offenbar gilt $a^{m+|u|}xb^{m+|v|}\alpha \in L_n$ nur dann, wenn $|u| = |v|$ ist.

Nach Obigem gibt es eine Ableitung

$$S \Longrightarrow^* u'Av' \Longrightarrow^* u'uAvv' \Longrightarrow^* a^m xb^m\alpha \in L_n$$

für ein $m \in \mathbb{N}$ und ein $x \in C_n$. Da $u \in \{a\}^+$ und $v \in \{b\}^+$ gelten, muss $|v'|_\alpha = 1$ sein. Wir folgern $A \neq S$, sonst wäre $S \Longrightarrow^* u'Av'$ das gleiche wie $S \Longrightarrow^* u'Sv'$, und es gäbe eine weitere Ableitung $S \Longrightarrow^* u'^2Sv'^2$, in der bereits 2 Buchstaben α vorkämen. In v' und u' darf kein Buchstabe aus C_n vorkommen, da $S \Longrightarrow^* u'uAvv'$ möglich ist. Neben der Regel p_A muss es folglich mindestens eine Regel p_S ohne Buchstaben aus C_n geben, die die Ableitung $S \Longrightarrow^* u'Av'$ ermöglicht.

Offenbar gilt für C_n die Voraussetzung in Lemma 2.9 ($X = C_n$). Folglich gibt es für jedes $c \in C_n$ eine separate Regel mit c auf der rechten Seite. Da in den Regeln p_S und p_A kein c vorkommt, befinden sich in P_n folglich mindestens $2 + (n - 2) = n$ Regeln. Mithin gilt $\mathrm{Prod}(L_n) \geq n$.

Die kontextfreie Grammatik

$$(\{S, S'\}, T_n, \{S \to S'\alpha, S' \to aS'b\} \cup \{S' \to c_i \mid 1 \leq i \leq n-2\}, S)$$

mit

$$T_n = \{a, b, \alpha\} \cup \{c_i \mid 1 \leq i \leq n-2\}$$

erzeugt L_n mit genau n Regeln. Mithin ergibt sich insgesamt $\mathrm{Prod}(L_n) = n$. □

Lemma 3.11 – Regelkomplexität von $\{a^ixb^i \mid i \geq 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-4}\}\}$
Für $n \in \mathbb{N}$, $n \geq 4$, ein Alphabet T_n mit mindestens $n - 1$ Buchstaben, paarweise verschiedene $a, b, c, d_i \in T_n$, $1 \leq i \leq n - 4$, und

$$L_n = \left\{a^ixb^i \,\middle|\, i \geq 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-4}\}\right\}$$

gilt $\mathrm{Prod}(L_n) = n$.

Beweis: Es seien $G_n = (N_n, T_n, P_n, S)$ eine regel-minimale kontextfreie Grammatik, die L_n erzeugt, und $C = \{c\}^+$ sowie $D_n = \{d_1, \ldots, d_{n-4}\}$. Nach Lemma 2.14 ist G_n reduziert. Da L_n eine unendliche Sprache ist, muss es nach Lemma 2.16 eine Schleifenableitung $A \Longrightarrow^* uAv$ für $A \in N_n$ und $u, v \in {T_n}^*$, $uv \neq \lambda$, geben. Die Wörter u und v dürfen kein $d \in D_n$ enthalten – anderenfalls existierte eine Ableitung $A \Longrightarrow^* uAv \Longrightarrow^* u^2Av^2$, in der das Wort u^2Av^2 mehr als ein d enthielte und folglich nicht Teil eines Wortes aus L_n sein kann. Weiterhin können in u und v jeweils keine zwei verschiedenen Buchstaben vorkommen, da u^2Av^2 dann ein ungültiges Teilwort generierte. Für u und v gilt also $u, v \in B$ für $B \in \{\{a\}^*, \{b\}^*, \{c\}^*\}$. Aufgrund der Reduziertheit von G_n gibt es die Ableitung

$$S \Longrightarrow^* u'Av' \Longrightarrow^* u'u^pAv^pv' \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^ixb^i \in L_n$$

für $u', v' \in {T_n}^*$ und $p, i \in \mathbb{N}_0$ sowie $x \in C \cup D_n$. Wir untersuchen nun alle möglichen Fälle für u und v ($uv \neq \lambda$):

- $u \in B$, $v = \lambda$:
  - Aus $u \in \{a\}^*$ folgt

    $$\begin{aligned} S &\Longrightarrow^* u'u^pAv^pv' = u'u^pA\lambda^pv' \\ &\Longrightarrow^* a^{i+|u|}xb^i \notin L_n. \end{aligned}$$
  - Der Fall $u \in \{b\}^*$ führt analog zu einem nicht in L_n vorhandenen Wort.
  - Der Fall $u \in \{c\}^*$ kann zu einem Wort in L_n führen.

- Für $u = \lambda$, $v \in B$ zeigen wir analog zum vorherigen Fall, dass ungültige Wörter erzeugt würden und $v \in \{c\}^*$ gültige hervorbringen kann.
- $(u, v) \in B'$ für

$$B' \in \ \{(\{b\}^+, \{a\}^+)\} \cup \\ \{(\{c\}^+, \{a\}^+), (\{b\}^+, \{c\}^+)\}$$

  (Fälle mit „falscher" Buchstabenreihenfolge):
  - Aus $u \in \{b\}^+$, $v \in \{a\}^+$ folgt

$$S \Longrightarrow^* u'Av' \Longrightarrow u'uAvv' = u'b^{|u|}Aa^{|v|}v' \\ \Longrightarrow^* u'b^{|u|}wa^{|v|}v' \notin L_n$$

    für $A \Longrightarrow^* w$, $w \in {T_n}^*$ (Reduziertheit).
  - Die restlichen Fälle erzeugen analog ungültige Wörter.
- $u, v \in B'$ für $B' \in B$ (Fälle mit gleichen Buchstaben):
  - Aus $u \in \{a\}^+$, $v \in \{a\}^+$ folgt

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^{i+|u|+|v|}xb^i \notin L_n.$$

  - Für $u \in \{b\}^+$, $v \in \{b\}^+$ entstehen analog ungültige Wörter.
  - Im Fall von $u \in \{c\}^+$, $v \in \{c\}^+$ können Wörter aus L_n generiert werden.
- $u \in \{a\}^+$, $v \in \{b\}^+$:

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^{i+|u|}xb^{i+|v|}.$$

  Offenbar gilt $a^{i+|u|}xb^{i+|v|} \in L_n$ genau dann, wenn $|u| = |v|$ ist.
- $u \in \{c\}^+$, $v = \lambda$:

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^i c^{(p+1)\cdot|u|}xb^i.$$

  Offenbar gilt $a^i c^{(p+1)\cdot|u|}xb^i \in L_n$ genau dann, wenn $x \notin D_n$ ist. Analoges gilt für $u \in \lambda$, $v \in \{c\}^+$.
- $u, v \in \{c\}^+$:

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^i c^{(p+1)\cdot|u|}xc^{(p+1)\cdot|v|}b^i.$$

  Offenbar gilt $a^i c^{(p+1)\cdot|u|}xc^{(p+1)\cdot|v|}b^i \in L_n$ genau dann, wenn $x \notin D_n$ ist.

In G_n müssen beide zuvor ermittelten Schleifenableitungen

$$A_{ab} \Longrightarrow^* u_a A_{ab} v_b \text{ und } A_c \Longrightarrow^* u_c A_c v_c$$

mit $A_{ab}, A_c \in N_n$ und $u_a \in \{a\}^+$, $v_b \in \{b\}^+$, wobei $|u_a| = |v_a|$ gilt, sowie $u_c, v_c \in \{c\}^*$, wobei $u_c v_c \neq \lambda$ gilt, vorkommen, da sonst nicht alle unendlich vielen Wörter mit beliebig vielen Vorkommen von a, b und c generiert werden könnten.

Offenbar ist $A_{ab} = A_c$ nicht möglich, da sonst zum Beispiel $A_{ab} \Longrightarrow^* u_c A_{ab} v_c \Longrightarrow^* u_c u_a A_{ab} v_b v_c$ ableitbar wäre, was zu einem ungültigem Teilwort führte. Des Weiteren gilt $S \neq A_c$: Wäre $A_c = S$, gäbe es die Ableitung $S \Longrightarrow^* u_c S v_c$. Außerdem gibt es wegen der Reduziertheit eine Ableitung $S \Longrightarrow^* u' A_{ab} v'$ für gewisse $u', v' \in {T_n}^*$. Mithin existierte eine Ableitung $S \Longrightarrow^* u_c S v_c \Longrightarrow^* u_c u' u_a A_{ab} v_b v' v_c$, die zu einem nicht validen Wort führen würde (kein Wort aus L_n endet oder beginnt mit c). Analoges gilt *nicht* für den Fall $A_{ab} = S$.

Um $A_{ab} \Longrightarrow^* u_a A_{ab} v_b$ zu realisieren, benötigen wir mindestens eine Regel $p_{ab} \in P_n$ mit A_{ab} auf der linken Seite und einem Nichtterminal auf der rechten. Für $A_c \Longrightarrow^* u_c A_c v_c$ ist mindestens eine Regel $p_c \in P_n$ mit A_c auf der linken Seite und einem Nichtterminal auf der rechten notwendig. Da $A_{ab} \neq A_c$ gilt, gilt auch $p_{ab} \neq p_c$. Weder p_{ab} noch p_c dürfen auf der rechten Seite ein $d \in D_n$ enthalten, da sonst Teilwörter in den Schleifenableitungen entstünden, in denen mehr als nur ein d vorkäme.

Verwenden wir in Lemma 2.9 für das dortige X die Menge D_n, folgt, dass es für jeden Buchstaben $d_i \in D_n$, $1 \leq i \leq n-4$ eine separate Regel $p_{d_i} \in P_n$ mit mindestens einem Vorkommen von d_i auf der rechten Seite geben muss. Mit dem zuvor Betrachteten gilt daher $p_{ab} \neq p_{d_i} \neq p_c$. Ferner darf auf der rechten Seite von p_{d_i} kein c vorkommen – es gibt kein Wort in L_n, das sowohl ein d_i als auch ein c enthält, welches dann allerdings ableitbar wäre.

Aufgrund der Reduziertheit von G_n muss es eine Ableitung $A_c \Longrightarrow^* w_c$ mit $w_c \in {T_n}^*$ geben. Hierbei findet eine weitere Regel p'_c Anwendung, die auf der rechten Seite kein Nichtterminal enthält (sonst könnte kein Terminalwort erzeugt werden) und die keine der Regeln p_{ab} (erzeugt immer ein Nichtterminal), p_c (erzeugt immer ein Nichtterminal) oder p_{d_i} (in w_c darf kein d_i vorkommen) ist.

Offenbar muss es aufgrund der Reduziertheit von G_n eine Ableitung $S \Longrightarrow^* u A_c v$ mit $u, v \in {T_n}^*$ geben. Es sei das erste Mal, dass das Nichtterminal A_c in dieser Ableitung auftaucht (dies können wir ohne Beschränkung der Allgemeinheit annehmen). Folglich muss es eine Regel p mit A_c auf der rechten Seite geben. Die Regel p_{ab} enthält rechts kein A_c – sonst könnte wegen $A_c \Longrightarrow^* u_c A_c v_c$ der Buchstabe c in der Ableitung $A_{ab} \Longrightarrow^* u_a A_{ab} v_b$ vorkommen (p_{ab} wird bei dieser Ableitung verwendet). Daher gilt $p \neq p_{ab}$. Dass $p \neq p_{d_i}$ für $1 \leq i \leq n-4$ korrekt ist, folgern wir daraus, dass das in p_{d_i} auf der rechten Seite vorkommende d_i nicht in $S \Longrightarrow^* u A_c v$ vorkommen kann, da sonst in einem resultierenden Wort zugleich c und d_i verkämen. Zudem gilt $p \neq p_c$, da A_c auf der linken Seite von p_c vorhanden ist und derart nicht zur Anwendung kommt – A_c kommt zum ersten Mal in $S \Longrightarrow^* u A_c v$ vor. Zuletzt gilt $p \neq p'_c$, da in rechts in p'_c kein Nichtterminal vorkommt. Mithin müssen in G_n mindestens $4 + (n-4)$ Regeln vorhanden sein, also $\mathrm{Prod}(L_n) = \mathrm{Prod}(G_n) \geq n$.

Die kontextfreie Grammatik

$$(\{S, S'\}, T_n, P_n, S)$$

mit

$$\begin{aligned} P_n = &\{S \to aSb\} \cup \{S \to ad_ib \mid 1 \le i \le n-4\} \cup \\ &\{S \to aS'b\} \cup \{S' \to cS', S' \to c\} \end{aligned}$$

erzeugt die Sprache L_n. Wir zählen $4 + (n-4)$ Regeln, also $\mathrm{Prod}(L_n) \le n$. Mit Obigem folgt $\mathrm{Prod}(L_n) = n$. □

Lemma 3.12 – Regelkomplexität von $\{a^ixb^i \mid i \ge 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-5}\}\} \cdot \{\alpha\}$
Für $n \in \mathbb{N}$, $n \ge 5$, ein Alphabet T_n mit mindestens $n-1$ Buchstaben, paarweise verschiedene $a, b, c, \alpha, d_i \in T_n$, $1 \le i \le n-5$, und

$$L_n = \left\{a^ixb^i \,\middle|\, i \ge 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-5}\}\right\} \cdot \{\alpha\}$$

gilt $\mathrm{Prod}(L_n) = n$.

Beweis: Der Beweis von Lemma 3.11 lässt sich derart abwandeln, dass wir den zusätzlichen Buchstaben α einbeziehen und analog zum Beweis von Lemma 3.10 sehen, dass es zu den $4 + (n-5)$ Regeln aufgrund von α in einer regel-minimalen Grammatik, die L_n erzeugt, eine zusätzliche geben muss. Hieraus folgt $\mathrm{Prod}(L_n) \ge 5 + (n-5) = n$.

Die kontextfreie Grammatik

$$(\{S, S', S''\}, T_n, P_n, S)$$

mit

$$\begin{aligned} P_n = &\{S \to S'\alpha, S' \to aS'b\} \cup \{S' \to ad_ib \mid 1 \le i \le n-5\} \cup \\ &\{S' \to aS''b\} \cup \{S'' \to cS'', S'' \to c\} \end{aligned}$$

erzeugt die Sprache L_n. Wir zählen $5 + (n-5)$ Regeln, also $\mathrm{Prod}(L_n) \le n$. Mit Obigem folgt $\mathrm{Prod}(L_n) = n$. □

3.2. Spiegelbild

Wir untersuchen den Raum der Regelkomplexität unter Spiegelbild.

Satz 3.13 – Raum der Regelkomplexität unter Spiegelbild
Für $n \in \mathbb{N}_0$ gilt

$$\{n\} = \mathrm{C}^{\mathrm{Prod}}_{()^{\mathrm{R}}}(n).$$

Beweis: Es seien L eine kontextfreie Sprache und $G = (N, T, P, S)$ eine regel-minimale kontextfreie Grammatik mit $\mathrm{L}(G) = L$. Die kontextfreie Grammatik

$$(N, T, \{A \to w^{\mathrm{R}} \mid A \to w \in P, A \in N, w \in (N \cup T)^*\}, S)$$

generiert offenbar L^{R} mittels $\mathrm{Prod}(L)$ Regeln (Lemma 2.8). Daher gilt $\mathrm{Prod}(L^{\mathrm{R}}) \leq \mathrm{Prod}(L)$. Analog erhalten wir $\mathrm{Prod}(L) = \mathrm{Prod}((L^{\mathrm{R}})^{\mathrm{R}}) \leq \mathrm{Prod}(L^{\mathrm{R}})$. Zusammen folgt $\mathrm{Prod}(L^{\mathrm{R}}) = \mathrm{Prod}(L)$. □

3.3. Vereinigung

Wir untersuchen den Raum der Regelkomplexität unter Vereinigung.

Lemma 3.14 – Kommutativität von $\mathrm{C}_{\cup}^{\mathbf{Prod}}$
Für $n, m \in \mathbb{N}_0$ gilt
$$\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m) = \mathrm{C}_{\cup}^{\mathrm{Prod}}(m, n).$$

Beweis: Es sei $k \in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$. Dann existieren zwei kontextfreie Sprachen L und J mit $\mathrm{Prod}(L) = n$, $\mathrm{Prod}(J) = m$ und $\mathrm{Prod}(L \cup J) = k$. Da $\cup$ eine kommutative Operation ist, also stets $L \cup J = J \cup L$ stimmt, gilt auch $\mathrm{Prod}(J \cup L) = k$ und mithin $k \in \mathrm{C}_{\cup}^{\mathrm{Prod}}(m, n)$. Somit ist $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m) \subseteq \mathrm{C}_{\cup}^{\mathrm{Prod}}(m, n)$. Dass umgekehrt $\mathrm{C}_{\cup}^{\mathrm{Prod}}(m, n) \subseteq \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ korrekt ist, können wir analog zeigen. Folglich ergibt sich $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m) = \mathrm{C}_{\cup}^{\mathrm{Prod}}(m, n)$. □

Wir haben $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m) = \mathrm{C}_{\cup}^{\mathrm{Prod}}(m, n)$ für $n, m \in \mathbb{N}_0$ insbesondere dafür gezeigt, dass wir im Folgenden nur $\mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ für $m \leq n$ bestimmen müssen, um ohne Zutun auch $\mathrm{C}_{\cup}^{\mathrm{Prod}}(m, n)$ zu erhalten (und damit den vollständigen Raum).

Lemma 3.15 – Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Vereinigung
Unter den Voraussetzungen $n, m \in \mathbb{N}_0$ gelten

i. $0 \notin \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ für $n \geq 1$,

ii. $1 \notin \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$ für $n \geq 2$,

iii. $\{1, 2\} = \mathrm{C}_{\cup}^{\mathrm{Prod}}(1, 1)$,

iv. $\{n\} = \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, 0)$,

v. $n + 3, \ldots \notin \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, 1)$ für $n \geq 1$,

vi. $n + m + 3, \ldots \notin \mathrm{C}_{\cup}^{\mathrm{Prod}}(n, m)$.

Beweis: Es seien L und J zwei kontextfreie Sprachen, für die $\mathrm{Prod}(L) = n$, $n \in \mathbb{N}_0$, und $\mathrm{Prod}(J) = m$, $m \in \mathbb{N}_0$, gelten.

Zu i. Es sei $\mathrm{Prod}(L \cup J) = 0$. Dann muss nach Lemma 3.1 $L \cup J = \emptyset$ gelten. Dies impliziert $L = \emptyset$ und $J = \emptyset$. Ebenfalls wegen Lemma 3.1 folgen sofort $\mathrm{Prod}(L) = 0$ und $\mathrm{Prod}(J) = 0$, und damit $n = m = 0$.

Zu ii. Es sei $\mathrm{Prod}(L \cup J) = 1$. Dann gilt wegen Lemma 3.1 $L \cup J = \{w\}$, $w \in T^*$, für ein beliebiges Alphabet T. Ohne Beschränkung der Allgemeinheit dürfen wir $L = \{w\}$ und

$J = \emptyset$ bzw. $L = \{w\}$ und $J = \{w\}$ annehmen. Lemma 3.1 liefert direkt $\mathrm{Prod}(L) = n = 1$ und $\mathrm{Prod}(J) = m = 0$ bzw. $\mathrm{Prod}(L) = n = 1$ und $\mathrm{Prod}(J) = m = 1$.

Zu iii. Lemma 3.1 entsprechend müssen $L = \{w\}$ und $J = \{w'\}$ für ein Alphabet T und $w, w' \in T^*$ sein, wenn $\mathrm{Prod}(L) = 1$ und $\mathrm{Prod}(J) = 1$ gelten sollen. Der Fall $w = w'$ impliziert $L \cup J = \{w\}$, was nach Lemma 3.1 die Tatsache $\mathrm{Prod}(L \cup J) = 1$ zur Konsequenz hat. Dasselbe Lemma begründet $\mathrm{Prod}(L \cup J) = 2$ für den anderen Fall $w \neq w'$, $L \cup J = \{w, w'\}$.

Zu iv. Es sei $\mathrm{Prod}(J) = m = 0$. Aufgrund von Lemma 3.1 gilt $J = \emptyset$. Infolgedessen erhalten wir $L \cup J = L \cup \emptyset = L$, und damit $\mathrm{Prod}(L \cup J) = \mathrm{Prod}(L) = n$.

Zu v. Es seien $G = (N,\ T,\ P,\ S)$ mit $\mathrm{L}(G) = L$, $\mathrm{Prod}(G) = \mathrm{Prod}(L) = n \geq 1$ und $G' = (N',\ T',\ P',\ S')$ mit $\mathrm{L}(G') = J$, $\mathrm{Prod}(G') = \mathrm{Prod}(J) = 1$ regel-minimale kontextfreie Grammatiken, für die wir ohne Beschränkung der Allgemeinheit $S'' \notin N$ und $S'' \notin N'$ voraussetzen. Nach Lemma 3.1 gilt $J = \{w'\}$ mit $w' \in T'^*$. Wir können nun stets die kontextfreie Grammatik

$$(N \cup \{S''\}, T \cup T', P \cup \{S'' \to S, S'' \to w'\}, S'')$$

angeben, die $L \cup J$ mit genau $\mathrm{Prod}(G) + 2$ Regeln erzeugt. Mithin erhalten wir $\mathrm{Prod}(L \cup J) \leq \mathrm{Prod}(G) + 2 = n + 2$.

Zu vi. Wählen wir die kontextfreien Grammatiken G mit $\mathrm{L}(G) = L$ und G' mit $\mathrm{L}(G') = J$ aus dem Beweis des Lemmas 2.8 so, dass $\mathrm{Prod}(G) = \mathrm{Prod}(L) = n$ und $\mathrm{Prod}(G') = \mathrm{Prod}(J) = m$ gelten, dann zählen wir in $G_\cup$ mit $\mathrm{L}(G_\cup) = L \cup J$ höchstens $n + m + 2$ Regeln (Beweis zu Lemma 2.8). Mithin gilt $\mathrm{Prod}(L \cup J) \leq \mathrm{Prod}(G_\cup) = n + m + 2$. □

Lemma 3.16 – Ergebnisse zum Raum der Regelkomplexität unter Vereinigung
Unter den Voraussetzungen $n, m \in \mathbb{N}_0$ gelten

i. $2 \in \mathrm{C}_\cup^{\mathrm{Prod}}(3, m)$ für $1 \leq m \leq 3$,

ii. $2 \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, 2)$ für $n \geq 2$,

iii. $2 \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, 3)$ für $n \geq 3$,

iv. $3 \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, 2)$ für $n \geq 2$,

v. $m - 1 \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, m)$ für $n \geq 5$ und $5 \leq m \leq n$,

vi. $n - 1 \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, m)$ für $n \geq 5$ und $1 \leq m \leq n$,

vii. $4, \ldots, n \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, 2)$ für $n \geq 4$,

viii. $6, \ldots, n \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, m)$ für $n \geq 6$ und $4 \leq m \leq n$.

ix. $m, \ldots, n \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, m)$ für $n \geq 3$ und $3 \leq m \leq n$,

x. $n, \ldots, n + m \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, m)$ für $n \geq 0$ und $0 \leq m \leq n$,

xi. $n + m + 1 \in \mathrm{C}_\cup^{\mathrm{Prod}}(n, m)$ für $n \geq 2$ und $1 \leq m \leq n$,

xii. $n+m+2 \in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,m)$ für $n \geq 2$ und $2 \leq m \leq n$,

Beweis: Es seien T ein Alphabet und $a, b \in T$.

Zu i. Für $1 \leq m \leq 3$ setzen wir

$$L_3 = \{a\} \cup \left\{ a^i \;\middle|\; i \geq 3 \right\},$$
$$J_m = \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq m \right\}.$$

Es gelten $\mathrm{Prod}(L_3) = 3$ nach Lemma 3.4 und $\mathrm{Prod}(J_m) = m$ nach Lemma 3.2, zudem

$$L_3 \cup J_m = \{a\}^+.$$

Lemma 3.3 liefert schließlich $\mathrm{Prod}(L_3 \cup J_m) = 2$.

Zu ii. Für $n \geq 2$ seien

$$L_n = \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq n \right\},$$
$$J_2 = \{a\}^+.$$

Nach Lemma 3.2 gilt $\mathrm{Prod}(L_n) = n$ und nach Lemma 3.3 $\mathrm{Prod}(J_2) = 2$. Des Weiteren ergibt sich aus

$$L_n \cup J_2 = J_2 = \{a\}^+$$

und Lemma 3.3 $\mathrm{Prod}(L_n \cup J_2) = 2$.

Zu iii. Für $n \geq 3$ setzen wir

$$L_n = \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq n \right\},$$
$$J_3 = \{a\} \cup \left\{ a^i \;\middle|\; i \geq 3 \right\}.$$

Wegen von Lemma 3.2 stimmt $\mathrm{Prod}(L_n) = n$ und wegen Lemma 3.4 $\mathrm{Prod}(J_3) = 3$. Es gilt

$$L_n \cup J_3 = \{a\}^+,$$

wobei sich $\mathrm{Prod}(L_n \cup J_3) = 2$ nach Lemma 3.3 ergibt.

Zu iv. Für $n \geq 2$ legen wir

$$L_n = \{a\} \cup \left\{ a^{2^i} \;\middle|\; 2 \leq i \leq n \right\},$$
$$J_2 = \left\{ a^i \;\middle|\; i \geq 3 \right\}$$

fest. Die Aussage $\mathrm{Prod}(L_n) = n$ kann analog zum Beweis von Lemma 3.2 geführt werden – es wird lediglich a^2 aus der Sprache entfernt. Lemma 3.3 entnehmen wir $\mathrm{Prod}(J_2) = 2$. Für die Vereinigung der beiden Sprachen erhalten wir

$$L_n \cup J_2 = \{a\} \cup \left\{ a^i \;\middle|\; i \geq 3 \right\}.$$

Lemma 3.4 liefert $\mathrm{Prod}(L_n \cup J_2) = 3$.

Zu v. Für $n \geq 5$ und $5 \leq m \leq n$ seien

$$\begin{aligned} L_n &= \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq n \right\}, \\ J_m &= \{a\} \cup \left\{ a^i \;\middle|\; i \geq 3 \right\} \cup \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq m-4 \right\}. \end{aligned}$$

Nach Lemma 3.2 gilt $\text{Prod}(L_n) = n$. Wir können die Sprache J_m als $A_3 \cup B_m$ für

$$\begin{aligned} A_3 &= \{a\} \cup \left\{ a^i \;\middle|\; i \geq 3 \right\}, \\ B_m &= \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq m-4 \right\} \end{aligned}$$

ausdrücken. Aufgrund von Lemma 3.4 gilt $\text{Prod}(A_3) = 3$, und nach Lemma 3.2 gilt $\text{Prod}(B_m) = m - 4$. Wir dürfen Satz 2.22 auf A_3 und B_m anwenden und erhalten

$$\begin{aligned} \text{Prod}(J_m) &= \text{Prod}(A_3 \cup B_m) = \text{Prod}(A_3) + 1 + \text{Prod}(B_m) \\ &= 3 + 1 + (m-4) = m \end{aligned}$$

(1 muss addiert werden, da jede regel-minimale kontextfreie Grammatik, die A_3 erzeugt, eine Initialschleifenableitung hat). Weiterhin ergibt sich

$$L_n \cup J_m = \{a\}^+ \cup \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq m-4 \right\}.$$

Wie zuvor zerlegen wir die Sprache $L_n \cup J_m$ in $\{a\}^+$ und $\{b^{2^i} \mid 1 \leq i \leq m-4\}$, und verifizieren mit Lemma 3.3, Lemma 3.2 und Satz 2.22

$$\begin{aligned} \text{Prod}(L_n \cup J_m) &= \text{Prod}(\{a\}^+) + 1 + \text{Prod}\left(\left\{ b^{2^i} \;\middle|\; 1 \leq i \leq m-4 \right\}\right) \\ &= 2 + 1 + (m-4) = m - 1 \end{aligned}$$

(da $\{a\}^+$ nur mit Initialschleifenableitung generiert werden kann, muss 1 addiert werden).

Zu vi. Für $n \geq 5$ und $1 \leq m \leq n$ legen wir

$$\begin{aligned} L_n &= \{a\} \cup \left\{ a^i \;\middle|\; i \geq 3 \right\} \cup \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq n-4 \right\}, \\ J_m &= \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq m \right\} \end{aligned}$$

fest. Wir zerlegen die Sprache L_n, indem wir $L_n = A_3 \cup B_n$ für

$$\begin{aligned} A_3 &= \{a\} \cup \left\{ a^i \;\middle|\; i \geq 3 \right\}, \\ B_n &= \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq n-4 \right\} \end{aligned}$$

wählen. Wegen Lemma 3.4 gilt $\text{Prod}(A_3) = 3$ und wegen Lemma 3.2 $\text{Prod}(B_n) = n - 4$. Die Anwendung des Satzes 2.22 auf A_3 und B_n ergibt

$$\begin{aligned} \text{Prod}(L_n) &= \text{Prod}(A_3 \cup B_n) = \text{Prod}(A_3) + 1 + \text{Prod}(B_n) \\ &= 3 + 1 + (n-4) = n \end{aligned}$$

(1 muss addiert werden, da jede regel-minimale kontextfreie Grammatik, die A_3 erzeugt, eine Initialschleifenableitung hat). Nach Lemma 3.2 gilt $\mathrm{Prod}(J_m) = m$. Es gilt die Aussage

$$L_n \cup J_m = \{a\}^+ \cup \left\{ b^{2^i} \,\middle|\, 1 \le i \le n-4 \right\}.$$

Eine Zerlegung von $L_n \cup J_m$ in $\{a\}^+$ und $\{b^{2^i} \mid 1 \le i \le n-4\}$ ergibt nach Lemma 3.3, Lemma 3.2 und Satz 2.22

$$\begin{aligned} \mathrm{Prod}(L_n \cup J_m) &= \mathrm{Prod}(\{a\}^+) + 1 + \mathrm{Prod}\left(\left\{ b^{2^i} \,\middle|\, 1 \le i \le n-4 \right\}\right) \\ &= 2 + 1 + (n-4) = n-1 \end{aligned}$$

(1 muss wegen der notwendigen Initialschleifenableitung zur Generierung von $\{a\}^+$ addiert werden).

Zu vii. Für $n \ge 4$ und $4 \le k \le n$ wählen wir

$$\begin{aligned} k' &= k-3, \\ L_{n,k} &= \left\{ a^{2^i} \,\middle|\, 1 \le i \le k' \right\} \cup \left\{ a^{2^i} \,\middle|\, k'+1 \le i \le n \right\}, \\ J_2 &= \left\{ a^{i2^{k'+1}} \,\middle|\, i \ge 1 \right\}. \end{aligned}$$

Lemma 3.2 liefert $\mathrm{Prod}(L_{n,k}) = n$ und Lemma 3.3 $\mathrm{Prod}(J_2) = 2$. Für die nachstehende Vereinigung

$$L_{n,k} \cup J_2 = \left\{ a^{2^i} \,\middle|\, 1 \le i \le k' \right\} \cup \left\{ a^{i2^{k'+1}} \,\middle|\, i \ge 1 \right\}.$$

ergibt sich nach Anwendung von Lemma 3.6 die Beschreibungskomplexität $\mathrm{Prod}(L_{n,k} \cup J_2) = k' + 3 = k$.

Zu viii. Für die Variablen $n \ge 6$, $4 \le m \le n$ und $6 \le k \le n$ verwenden wir die Variablen und Sprachen

$$\begin{aligned} n' &= n-3, m' = m-3, k' = k-6, \\ L_{n,k} &= \left\{ a^{2^i} \,\middle|\, 1 \le i \le k' \right\} \cup \left\{ b^{2^i} \,\middle|\, k'+1 \le i \le n' \right\} \cup \{c\}^+, \\ J_m &= \{b\}^+ \cup \left\{ c^{2^i} \,\middle|\, 1 \le i \le m' \right\}. \end{aligned}$$

Wir zerlegen die Sprache $L_{n,k} = A_k \cup B_{n,k} \cup C$ in 3 Sprachen, indem wir die folgenden Festlegungen vornehmen:

$$\begin{aligned} A_k &= \left\{ a^{2^i} \,\middle|\, 1 \le i \le k' \right\}, \\ B_{n,k} &= \left\{ b^{2^i} \,\middle|\, k'+1 \le i \le n' \right\}, \\ C &= \{c\}^+. \end{aligned}$$

Nach Lemma 3.2 erhalten wir $\mathrm{Prod}(A_k) = k'$ und $\mathrm{Prod}(B_{n,k}) = n' - k'$ sowie $\mathrm{Prod}(C) = 2$ nach Lemma 3.3. Durch zweimalige Anwendung des Satzes 2.22 ergibt sich dann

$$\begin{aligned}\mathrm{Prod}(L_{n,k}) &= \mathrm{Prod}(A_k) + \mathrm{Prod}(B_{n,k}) + \mathrm{Prod}(C) + 1\\ &= k' + n' - k' + 3\\ &= (k-6) + (n-3) - (k-6) + 3\\ &= n.\end{aligned}$$

Analog errechnen wir für J_m die Regelanzahl $\mathrm{Prod}(J_m) = m$. Die Vereinigung der Sprachen $L_{n,k}$ und J_m liefert

$$L_{n,k} \cup J_m = \left\{a^{2^i} \;\middle|\; 1 \leq i \leq k'\right\} \cup \{b\}^+ \cup \{c\}^+.$$

Durch Verwendung derselben Lemmata und Sätze wie im vorherigen Absatz ergibt sich

$$\begin{aligned}\mathrm{Prod}(L_{n,k} \cup J_m) &= \mathrm{Prod}\left(\left\{a^{2^i} \;\middle|\; 1 \leq i \leq k'\right\}\right) +\\ &\quad \mathrm{Prod}(\{b\}^+) + 1 + \mathrm{Prod}(\{c\}^+) + 1\\ &= k' + 3 + 3\\ &= (k-6) + 6\\ &= k.\end{aligned}$$

Zu ix. Für $n \geq 3$, $3 \leq m \leq n$ und $m \leq k \leq n$ seien

$$\begin{aligned}&m' = m - 3, k' = k - 3,\\ &L_n = \left\{a^{2^i} \;\middle|\; 1 \leq i \leq k'\right\} \cup \left\{b^{2^i} \;\middle|\; 1 \leq i \leq n - k'\right\},\\ &J_m = \left\{a^{2^i} \;\middle|\; 1 \leq i \leq m'\right\} \cup \left\{b^{i} \;\middle|\; 1 \leq i \leq 2^{n-k'}\right\}.\end{aligned}$$

Wir setzen $A_k = \{a^{2^i} \mid 1 \leq i \leq k'\}$ und $B_{n,k} = \{b^{2^i} \mid 1 \leq i \leq n - k'\}$ für $L_n = A_k \cup B_{n,k}$. Nach Satz 2.22 und Lemma 3.2 erhalten wir

$$\begin{aligned}\mathrm{Prod}(L_n) &= \mathrm{Prod}(A_k \cup B_{n,k})\\ &= \mathrm{Prod}(A_k) + \mathrm{Prod}(B_{n,k})\\ &= k' + (n - k') = n\end{aligned}$$

(regel-minimale kontextfreie Grammatiken für A_k oder $B_{n,k}$ haben keine Initialschleifen-ableitungen). Unter Zuhilfenahme von Lemma 3.5 ergibt sich analog

$$\begin{aligned}\mathrm{Prod}(J_m) &= \mathrm{Prod}\left(\left\{a^{2^i} \;\middle|\; 1 \leq i \leq m'\right\}\right) + \mathrm{Prod}\left(\left\{b^{i} \;\middle|\; 1 \leq i \leq 2^{n-k'}\right\}\right)\\ &= m' + 3\\ &= (m-3) + 3 = m.\end{aligned}$$

Die Vereinigung der Sprachen L_n und J_m ist

$$L_n \cup K_m = \left\{a^{2^i} \;\middle|\; 1 \leq i \leq k'\right\} \cup \left\{b^{i} \;\middle|\; 1 \leq i \leq 2^{n-k'}\right\}.$$

Durch abermaliges Anwenden von Satz 2.22 und Lemma 3.2 erhalten wir

$$\begin{aligned}\operatorname{Prod}(L_n \cup J_m) &= \operatorname{Prod}\left(\left\{a^{2^i} \,\middle|\, 1 \le i \le k'\right\}\right) + \operatorname{Prod}\left(\left\{b^i \,\middle|\, 1 \le i \le 2^{n-k'}\right\}\right) \\ &= k' + 3 = (k-3) + 3 = k.\end{aligned}$$

Zu x. Für $n \ge 0$, $0 \le m \le n$ und $n \le k \le n+m$ legen wir

$$\begin{aligned}L_n &= \left\{a^{2^i} \,\middle|\, 1 \le i \le n\right\}, \\ J_{n,m,k} &= \left\{a^{2^{(n-p)+i}} \,\middle|\, p = n+m-k, 1 \le i \le m\right\}\end{aligned}$$

fest. Aufgrund von Lemma 3.2 folgern wir sofort $\operatorname{Prod}(L_n) = n$ sowie $\operatorname{Prod}(J_{n,m,k}) = m$. Es sind immer genau $p = n+m-k$ Elemente von $J_{n,m,k}$ in L_n enthalten. Daher gehen bei der Vereinigung von L_n und $J_{n,m,k}$ stets p Elemente „verloren“. Für $L_n \cup J_{n,m,k}$ benötigen wir – abermals wegen Lemma 3.2 – n Regeln für die Wörter a^{2^i}, $1 \le i \le n$, und $m-p$ Regeln für die Wörter $a^{2^{n+i}}$, $1 \le i \le m-p$, die nicht in L_n aber in $J_{n,m,k}$ enthalten sind. Damit erhalten wir

$$\operatorname{Prod}(L_n \cup J_{n,m,k}) = n + (m-p) = n + (m - (n+m-k)) = k.$$

Zu xi. Für $n \ge 2$ und $1 \le m \le n$ seien

$$\begin{aligned}L_n &= \{a_i \mid 1 \le i \le n-1\}^+, \\ J_m &= \left\{b^{2^i} \,\middle|\, 1 \le i \le m\right\}.\end{aligned}$$

Die Aussagen $\operatorname{Prod}(L_n) = n$ bzw. $\operatorname{Prod}(J_m) = m$ entnehmen wir dem Lemma 3.7 bzw. dem Lemma 3.2. Die Alphabete der Sprachen L_n und J_m sind disjunkt. Folglich können wir Satz 2.22 anwenden und erhalten $\operatorname{Prod}(L_n \cup J_m) = n+m+1$ (1 muss addiert werden, da jede regel-minimale kontextfreie Grammatik, die L_n generiert, eine Initialschleifenableitung hat).

Zu xii. Für $n \ge 2$ und $2 \le m \le n$ legen wir

$$\begin{aligned}L_n &= \{a_i \mid 1 \le i \le n-1\}^+, \\ J_m &= \{b_i \mid 1 \le i \le m-1\}^+\end{aligned}$$

fest. Aufgrund von Lemma 3.7 sind $\operatorname{Prod}(L_n) = n$ und $\operatorname{Prod}(J_m) = m$ korrekt. Die Alphabete von L_n und J_m sind disjunkt. Somit gilt $\operatorname{Prod}(L_n \cup J_m) = n+m+2$ nach Anwendung des Satzes 2.22 (2 muss addiert werden, da jede regel-minimale kontextfreie Grammatik, die L_n oder J_m generiert, eine Initialschleifenableitung hat). □

Satz 3.17 – Teilbestimmung des Raums der Regelkomplexität unter Vereinigung
Für $n, m \in \mathbb{N}_0$ und $m \le n$ gelten

- $$\{n\} = C_{\cup}^{\operatorname{Prod}}(n, 0),$$
- $$\{1, 2\} = C_{\cup}^{\operatorname{Prod}}(1, 1),$$

$$\begin{aligned}
\{2,3,4\} &= \mathrm{C}_{\cup}^{\mathrm{Prod}}(2,1), \\
\{2,3,4,5\} &= \mathrm{C}_{\cup}^{\mathrm{Prod}}(3,1), \\
4,5,6 &\in \mathrm{C}_{\cup}^{\mathrm{Prod}}(4,1) \not\ni 0,1,7,\dots, \\
4,5,6,7 &\in \mathrm{C}_{\cup}^{\mathrm{Prod}}(5,1) \not\ni 0,1,8,\dots, \\
n-1,\dots,n+2 &\in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,1) \not\ni 0,1,n+3,\dots \textit{ für } n \geq 6,
\end{aligned}$$

- $$\begin{aligned}
\{2,\dots,n+4\} &= \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,2) \textit{ für } n \geq 2, \\
\{2,\dots,n+5\} &= \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,3) \textit{ für } n \geq 3, \\
2,4,\dots,n+6 &\in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,4) \not\ni 0,1,n+7,\dots \textit{ für } n \geq 4, \\
4,\dots,n+7 &\in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,5) \not\ni 0,1,n+8,\dots \textit{ für } n \geq 5, \\
5,\dots,n+8 &\in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,6) \not\ni 0,1,n+9,\dots \textit{ für } n \geq 6, \\
6,\dots,n+m+2 &\in \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,m) \not\ni 0,1,n+m+3,\dots \textit{ für } n \geq 7, m \geq 7,
\end{aligned}$$

- $$\mathrm{C}_{\cup}^{\mathrm{Prod}}(n,m) = \mathrm{C}_{\cup}^{\mathrm{Prod}}(m,n).$$

Beweis: Der Satz ist eine Zusammenfassung der Lemmata 3.15 und 3.16 unter Zuhilfenahme des Lemmas 3.14. □

Bemerkung: Es seien $n, m \in \mathbb{N}_0$ und $m \leq n$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 3.17 vollständig dargestellt:

- $$\begin{aligned}
2,3 &\in^{?} \mathrm{C}_{\cup}^{\mathrm{Prod}}(4,1), \\
2,3 &\in^{?} \mathrm{C}_{\cup}^{\mathrm{Prod}}(5,1), \\
2,\dots,n-2 &\in^{?} \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,1) \text{ für } n \geq 6,
\end{aligned}$$

- $$\begin{aligned}
3 &\in^{?} \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,4) \text{ für } n \geq 4, \\
2,3 &\in^{?} \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,5) \text{ für } n \geq 5, \\
2,3,4 &\in^{?} \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,6) \text{ für } n \geq 6, \\
2,3,4,5 &\in^{?} \mathrm{C}_{\cup}^{\mathrm{Prod}}(n,m) \text{ für } n \geq 7, m \geq 7.
\end{aligned}$$

△

3.4. Konkatenation

Wir untersuchen den Raum der Regelkomplexität unter Konkatenation.

Lemma 3.18 – Kommutativität von $\mathrm{C}_{\cdot}^{\mathrm{Prod}}$
Für $n, m \in \mathbb{N}_0$ gilt

$$\mathrm{C}_{\cdot}^{\mathrm{Prod}}(n,m) = \mathrm{C}_{\cdot}^{\mathrm{Prod}}(m,n).$$

Beweis: Es sei $k \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m)$. Dann existieren zwei kontextfreie Sprachen L und J mit $\mathrm{Prod}(L) = n$, $\mathrm{Prod}(J) = m$ und $\mathrm{Prod}(L \cdot J) = k$. Nach Satz 3.13 gilt $\mathrm{Prod}((L \cdot J)^{\mathrm{R}}) = \mathrm{Prod}(L \cdot J)$. Weiterhin kann das Faktum $(L \cdot J)^{\mathrm{R}} = J^{\mathrm{R}} \cdot L^{\mathrm{R}}$ aus der Definition 2.6 abgeleitet werden. Die Kombination der beiden Tatsachen ergibt

$$k = \mathrm{Prod}(L \cdot J) = \mathrm{Prod}\big((L \cdot J)^{\mathrm{R}}\big) = \mathrm{Prod}(J^{\mathrm{R}} \cdot L^{\mathrm{R}}).$$

Da $\mathrm{Prod}(J^{\mathrm{R}}) = \mathrm{Prod}(J) = m$ und $\mathrm{Prod}(L^{\mathrm{R}}) = \mathrm{Prod}(L) = n$ gelten, erhalten wir $k \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(m, n)$. Somit ist $\mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, m) \subseteq \mathrm{C}^{\mathrm{Prod}}_{\cdot}(m, n)$. Dass andersherum $\mathrm{C}^{\mathrm{Prod}}_{\cdot}(m, n) \subseteq \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, m)$ korrekt ist, können wir analog zeigen. Folglich ergibt sich $\mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, m) = \mathrm{C}^{\mathrm{Prod}}_{\cdot}(m, n)$. □

Lemma 3.19 – Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Konkatenation

Unter den Voraussetzungen $n, m \in \mathbb{N}_0$ gelten

i. $0 \notin \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m)$ für $n \geq 1$, $m \geq 1$,

ii. $1 \notin \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m)$ für $n \geq 2$,

iii. $\{1\} = \mathrm{C}^{\mathrm{Prod}}_{\cdot}(1,1)$,

iv. $\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,0)$,

v. $n + 2, \ldots \notin \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,1)$ für $n \geq 1$,

vi. $n + m + 2, \ldots \notin \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m)$.

Beweis: Es seien L und J zwei kontextfreie Sprachen, für die $\mathrm{Prod}(L) = n$, $n \in \mathbb{N}_0$, und $\mathrm{Prod}(J) = m$, $m \in \mathbb{N}_0$, gelten.

Zu i. Es seien $n \geq 1$, $m \geq 1$ und $\mathrm{Prod}(L \cdot J) = 0$. Dann muss nach Lemma 3.1 $L \cdot J = \emptyset$ gelten. Dies impliziert $L = \emptyset$ oder $J = \emptyset$. Ebenfalls wegen Lemma 3.1 folgen sofort $\mathrm{Prod}(L) = 0$ oder $\mathrm{Prod}(J) = 0$, und damit $n = 0$ oder $m = 0$, was den Voraussetzungen widerspricht.

Zu ii. Es sei $\mathrm{Prod}(L \cdot J) = 1$. Dann gilt wegen Lemma 3.1 $L \cdot J = \{w\}$, $w \in T^*$, für ein beliebiges Alphabet T. Die einzigen Möglichkeit ist $L = \{w_L\}$, $w_L \in T^*$, und $J = \{w_J\}$, $w_J \in T^*$, für $w_L w_J = w$. Lemma 3.1 liefert direkt $\mathrm{Prod}(L) = n = 1$ und $\mathrm{Prod}(J) = m = 1$.

Zu iii. Es seien $\mathrm{Prod}(L) = 1$ und $\mathrm{Prod}(J) = 1$. Dann gelten wegen Lemma 3.1 $L = \{w_L\}$, $w_L \in T^*$, und $J = \{w_J\}$, $w_J \in T^*$, für ein beliebiges Alphabet T. Wir erhalten $L \cdot J = \{w_L w_J\}$, was nach demselben Lemma $\mathrm{Prod}(L \cdot J) = \mathrm{Prod}(\{w_L w_J\}) = 1$ zur Folge hat.

Zu iv. Es sei $\mathrm{Prod}(J) = m = 0$. Aufgrund von Lemma 3.1 gilt $J = \emptyset$. Infolgedessen erhalten wir $L \cdot J = L \cdot \emptyset = \emptyset$, und damit $\mathrm{Prod}(L \cdot J) = \mathrm{Prod}(\emptyset) = 0$ (Lemma 3.1).

Zu v. Es seien $G = (N, T, P, S)$ mit $\mathrm{L}(G) = L$, $\mathrm{Prod}(G) = \mathrm{Prod}(L) = n \geq 1$ und $G' = (N', T', P', S')$ mit $\mathrm{L}(G') = J$, $\mathrm{Prod}(G') = \mathrm{Prod}(J) = 1$ regel-minimale kontextfreie Grammatiken, für die wir ohne Beschränkung der Allgemeinheit $S'' \notin N$

und $S'' \notin N'$ voraussetzen. Nach Lemma 3.1 gilt $J = \{w'\}$ mit $w' \in T'^*$. Wir können nun stets die kontextfreie Grammatik

$$(N \cup \{S''\}, T \cup T', P \cup \{S'' \to Sw'\}, S'')$$

angeben, die $L{\cdot}J$ mit genau $\operatorname{Prod}(G){+}1$ Regeln erzeugt. Mithin erhalten wir $\operatorname{Prod}(L{\cdot}J) \leq \operatorname{Prod}(G) + 1 = n + 1$.

Zu vi. Wählen wir die Grammatiken G mit $\mathrm{L}(G) = L$ und G' mit $\mathrm{L}(G') = J$ aus dem Beweis des Lemmas 2.8 so, dass $\operatorname{Prod}(G) = \operatorname{Prod}(L) = n$ und $\operatorname{Prod}(G') = \operatorname{Prod}(J) = m$ gelten, sind in $G_.$ mit $\mathrm{L}(G_.) = L \cdot J$ höchstens $n + m + 1$ Regeln enthalten (Beweis zu Lemma 2.8). Folglich gilt $\operatorname{Prod}(L \cdot J) \leq \operatorname{Prod}(G_.) = n + m + 1$. □

Lemma 3.20 – Es gilt $\mathbf{2, 3, 4 \in C_.^{Prod}(2, 2)}$

Es gilt $2, 3, 4 \in \mathrm{C}_.^{\mathrm{Prod}}(2, 2)$.

Beweis: Es seien T ein Alphabet mit mindestens 4 Buchstaben und paarweise verschiedene $a, b, c, d \in T$.

Es sei $L = \{a\}^*$. Nach Lemma 3.3 gilt $\operatorname{Prod}(L) = 2$. Weiterhin ist die Aussage $L \cdot L = \{a\}^* \cdot \{a\}^* = \{a\}^* = L$ korrekt, also auch $\operatorname{Prod}(L \cdot L) = \operatorname{Prod}(L) = 2$.

Es seien $L = \{a\}^+$ und $J = \{b\}^+$. Nach Lemma 3.3 gelten $\operatorname{Prod}(L) = 2$ und $\operatorname{Prod}(J) = 2$. Die Konkatenation der beiden Sprachen ist $L \cdot J = \{a\}^+ \cdot \{b\}^+$. Es sei $G = (N, T, P, S)$ eine regel-minimale kontextfreie Grammatik, die $L \cdot J$ erzeugt. Da $|L{\cdot}J| = |\mathbb{N}|$ gilt, folgt nach Lemma 2.23 die Regelkomplexität $\operatorname{Prod}(L{\cdot}J) \geq 2$ und zudem die Existenz einer Schleifenableitung $A \Longrightarrow^* xAy$ für $A \in N$ und $x, y \in T^*$, $xy \neq \lambda$ sowie die Ableitung $A \Longrightarrow^* w$, $w \in T^*$, die mithilfe von 2 Regeln realisiert werden. Existierte nur eine Schleifenableitung, so müssten $x \in \{a\}^+$ und $y \in \{b\}^+$ gelten. Dann aber könnten die Buchstaben a und b nicht unabhängig in Bezug auf die Anzahl ihrer Vorkommen in den generierten Wörtern erzeugt werden. Folglich „fehlten“ Wörter, die in $L \cdot J$ vorkommen. Infolgedessen gibt es eine weitere Schleifenableitung. Diese bedingt mindestens eine weitere Regel. Mithin gilt $\operatorname{Prod}(L) = \operatorname{Prod}(G) \geq 3$. Andererseits erzeugt die kontextfreie Grammatik

$$(\{S\}, \{a, b\}, \{S \to aS, S \to Sb, S \to ab\}, S)$$

die Sprache $L{\cdot}J$ mit genau 3 Regeln, also $\operatorname{Prod}(L{\cdot}J) \leq 3$. Zusammen gilt $\operatorname{Prod}(L{\cdot}J) = 3$.

Es seien $L = \{a,\ b\}$ und $J = \{c,\ d\}$. Nach Lemma 3.8 gelten $\operatorname{Prod}(L) = 2$ und $\operatorname{Prod}(J) = 2$. Als Konkatenation der beiden Sprachen erhalten wir $L \cdot J = \{a, b\} \cdot \{c, d\} = \{ac,\ ad,\ bc,\ bd\}$. Aus Lemma 3.1 folgt $\operatorname{Prod}(L \cdot J) \geq 3$. Durch Ausprobieren aller kontextfreien Grammatiken, die genau 3 Regeln haben, die als Terminale nur a, b, c und d enthalten (dürfen nur begrenzt häufig vorkommen), und damit höchstens 3 Nichtterminale (jedes Nichtterminal bedingt mindestens eine Regel, dürfen nur begrenzt häufig in Regeln vorkommen), erkennen wir, dass es mit genau 3 Regeln nicht möglich ist, die Sprache $L \cdot J$ zu erzeugen. Es folgt $\operatorname{Prod}(L \cdot J) \geq 4$. Die kontextfreie Grammatik

$$(\{S\}, \{a, b, c, d\}, \{S \to ac, S \to ad, S \to bc, S \to bd\}, S)$$

erzeugt die Sprache $L \cdot J$ mit genau 4 Regeln. Mit Obigem folgt $\mathrm{Prod}(L \cdot J) = 4$. $\square$

Lemma 3.21 – Es gilt $n \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, 2)$
Für $n \in \mathbb{N}_0$, $n \geq 2$, gilt $n \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, 2)$.

Beweis: Es seien T_n ein Alphabet mit mindestens $n-1$ Buchstaben, paarweise verschiedene $b, a_i \in T_n$, $1 \leq i \leq n-2$, sowie

$$\begin{aligned} L_n &= \{a_1, \ldots, a_{n-2}\}^* \cdot \{b\}^*, \\ J &= \{b\}^*. \end{aligned}$$

Für das in Lemma 2.10 definierte X erhalten wir im Fall von L_n die Menge $X = \{a_1, \ldots, a_{n-2}, b\}$. Folglich muss es in einer regel-minimalen kontextfreien Grammatik, die L_n generiert, mindestens $n-1$ Regeln geben. Jede dieser Regeln enthält mindestens ein Terminal auf der rechten Seite. Daher benötigen wir zudem eine Regel mit nur λ auf der rechten Seite, um auch das Wort $\lambda \in L_n$ erzeugen zu können. Mithin gilt $\mathrm{Prod}(L_n) \geq n$. Die kontextfreie Grammatik

$$(\{S\}, T_n, P_n, S)$$

mit

$$P_n = \{S \to \lambda, S \to Sb\} \cup \{S \to a_i S \mid 1 \leq i \leq n-2\},$$

die L_n mittels n Regeln generiert. Mit Obigem folgt $\mathrm{Prod}(L_n) = n$.

Aufgrund von Lemma 3.3 gilt $\mathrm{Prod}(J) = 2$.

Die Konkatenation von L_n und J ergibt

$$L_n \cdot J = \{a_1, \ldots, a_{n-2}\}^* \cdot \{b\}^* \cdot \{b\}^* = \{a_1, \ldots, a_{n-2}\}^* \cdot \{b\}^* = L_n.$$

Von oben entnehmen wir daher das Ergebnis $\mathrm{Prod}(L_n \cdot J) = \mathrm{Prod}(L_n) = n$. $\square$

Lemma 3.22 – Es gilt $n + m + 1 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, m)$
Für $n \in \mathbb{N}_0$, $n \geq 2$, und $m \in \mathbb{N}_0$, $m \geq 2$, gilt $n + m + 1 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, m)$.

Beweis: Es seien $T_{n,m}$ ein Alphabet mit mindestens $n + m + 2$ Buchstaben, paarweise verschiedene $a, b, d, e, c_i, f_j \in T_{n,m}$, $1 \leq i \leq n-1$, $1 \leq j \leq m-1$, und $C_n = \{c_1, \ldots, c_{n-1}\}$, $F_m = \{f_1, \ldots, f_{m-1}\}$ sowie

$$\begin{aligned} L_n &= \left\{a^i x b^i \;\middle|\; i \geq 0, x \in C_n\right\}, \\ J_m &= \left\{d^i x e^i \;\middle|\; i \geq 0, x \in F_m\right\}. \end{aligned}$$

Aufgrund von Lemma 3.9 ergeben sich $\mathrm{Prod}(L_n) = n$ und $\mathrm{Prod}(J_m) = m$.

Wir zeigen im Folgenden, dass die Regelkomplexität für

$$L_n \cdot J_m = \left\{a^i x b^i d^j y e^j \;\middle|\; i \geq 0, j \geq 0, x \in C_n, y \in F_m\right\}$$

$n + m + 1$ beträgt.

Es sei $G_{n,m} = (N_{n,m}, T_{n,m}, P_{n,m}, S)$ eine regel-minimale kontextfreie Grammatik, die $L_n \cdot J_m$ erzeugt. Nach Lemma 2.14 ist $G_{n,m}$ reduziert. Da $L_n \cdot J_m$ eine unendliche Sprache ist, muss es nach Lemma 2.16 eine Schleifenableitung $A \Longrightarrow^* uAv$ für $A \in N_{n,m}$ und $u, v \in T_{n,m}{}^*$, $uv \neq \lambda$, geben. Die Wörter u und v dürfen kein c_i, $1 \leq i \leq n-1$, oder f_j, $1 \leq j \leq m-1$, enthalten – anderenfalls existierte eine Ableitung $A \Longrightarrow^* uAv \Longrightarrow^* u^2Av^2$, in der das Wort u^2Av^2 mehr als ein c_i bzw. f_j enthielte, und folglich nicht Teil eines Wortes aus $L_n \cdot J_m$ sein kann. Weiterhin können in u und v keine zwei verschiedenen Buchstaben vorkommen, da u^2Av^2 dann ein ungültiges Teilwort generierte. Für u und v gilt also $u, v \in B$ für $B \in \{\{a\}^*, \{b\}^*, \{d\}^*, \{e\}^*\}$. Aufgrund der Reduziertheit von $G_{n,m}$ gibt es die Ableitung

$$S \Longrightarrow^* u'Av' \Longrightarrow^* u'u^pAv^pv' \Longrightarrow^* a^ixb^id^jye^j \in L_n \cdot J_m$$

für $u', v' \in T_{n,m}{}^*$ und $p, i, j \in \mathbb{N}_0$ sowie $x \in C_n$ und $y \in F_m$. Wir untersuchen nun alle möglichen Fälle für u und v ($uv \neq \lambda$):

- $u \in B$, $v = \lambda$:
  - Aus $u \in \{a\}^*$ folgt

    $$\begin{aligned} S &\Longrightarrow^* u'u^{p+1}Av^{p+1}v' = u'u^{p+1}A\lambda^{p+1}v' \\ &\Longrightarrow^* a^{i+|u|}xb^id^jye^j \notin L_n \cdot J_m. \end{aligned}$$

  - Die restlichen Fälle $u \in B \setminus \{a\}^*$ führen analog zu einem nicht in $L_n \cdot J_m$ vorhandenen Wort.

- Für $u = \lambda$, $v \in B$ zeigen wir analog zum vorherigen Fall, dass ungültige Wörter erzeugt würden.

- Für

  $$\begin{aligned} (u, v) \in\ & \{b\}^+ \times \{a\}^+ \cup \\ & \{d\}^+ \times \{a\}^+ \cup \{d\}^+ \times \{b\}^+ \cup \\ & \{e\}^+ \times \{a\}^+ \cup \{e\}^+ \times \{b\}^+ \cup \{e\}^+ \times \{d\}^+ \end{aligned}$$

  (Fälle mit „falscher“ Buchstabenreihenfolge):
  - Aus $u \in \{b\}^+$, $v \in \{a\}^+$ folgt

    $$\begin{aligned} S &\Longrightarrow^* u'Av' \Longrightarrow u'uAvv' = u'b^{|u|}Aa^{|v|}v' \\ &\Longrightarrow^* u'b^{|u|}wa^{|v|}v' \notin L_n \cdot J_m \end{aligned}$$

    für $A \Longrightarrow^* w$, $w \in T_{n,m}{}^*$ (Reduziertheit).
  - Die restlichen Fälle erzeugen analog ungültige Wörter.

- $u, v \in C$ für $C \in B$ (Fälle mit gleichen Buchstaben):

- Aus $u \in \{a\}^+$, $v \in \{a\}^+$ folgt

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^{i+|u|+|v|}xb^id^jye^j \notin L_n \cdot J_m.$$

- In den übrigen Fällen entstehen analog ungültige Wörter.

• Für

$$(u, v) \in \{a\}^+ \times \{d\}^+ \cup \{a\}^+ \times \{e\}^+ \cup \{b\}^+ \times \{d\}^+ \cup \{b\}^+ \times \{e\}^+$$

(Fälle mit „falschem“ Buchstabenpaar):

- Aus $u \in \{a\}^+$, $v \in \{d\}^+$ folgt

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^{i+|u|}xb^id^{j+|v|}ye^j \notin L_n \cdot J_m.$$

- In den anderen Fällen entstehen analog Wörter, die nicht in $L_n \cdot J_m$ enthalten sind.

• $u \in \{a\}^+$, $v \in \{b\}^+$:

$$S \Longrightarrow^* u'u^{p+1}Av^{p+1}v' \Longrightarrow^* a^{i+|u|}xb^{i+|v|}d^jye^j.$$

Offenbar gilt $a^{i+|u|}xb^{i+|v|}d^jye^j \in L_n \cdot J_m$ genau dann, wenn $|u| = |v|$ ist.

• $u \in \{d\}^+$, $v \in \{e\}^+$: analog zum vorherigen Fall.

In $G_{n,m}$ müssen beide zuvor ermittelten Schleifenableitungen $A_k \Longrightarrow^* u_kA_kv_k$, $k \in \{1, 2\}$, mit $A_k \in N_{n,m}$ und $u_1 \in \{a\}^+$, $v_1 \in \{b\}^+$ sowie $u_2 \in \{d\}^+$, $v_2 \in \{e\}^+$, wobei $|u_k| = |v_k|$ gilt, vorkommen, da sonst nicht alle unendlich vielen Wörter mit beliebig vielen Vorkommen von a, b und d, e generiert werden könnten. Offenbar ist $A_1 = A_2$ nicht möglich, da sonst beispielsweise $A_1 \Longrightarrow^* u_1A_1v_1 \Longrightarrow^* u_1u_2A_1v_2v_1$ ableitbar wäre, was zu einem ungültigem Teilwort führte. Des Weiteren gilt $S \neq A_k$: Wäre $A_1 = S$, gäbe es die Ableitung $S \Longrightarrow^* u_1Sv_1$. Außerdem gibt es wegen der Reduziertheit eine Ableitung $S \Longrightarrow^* u'A_2v'$ für gewisse $u', v' \in T_{n,m}{}^*$. Mithin existierte eine Ableitung $S \Longrightarrow^* u_1Sv_1 \Longrightarrow^* u_1u'u_2A_2v_2v'v_1$, die zu einem nicht validen Wort führte (kein Wort aus $L_n \cdot J_m$ endet auf $v_1 \in \{b\}^+$). Analoges gilt für den Fall $A_2 = S$. Insgesamt sind in $P_{n,m}$ mindestens 3 Regeln für die Ableitungen $S \Longrightarrow^* u'_kA_kv'_k$, $u'_k, v'_k \in T_{n,m}{}^*$, und $A_k \Longrightarrow^* u_kA_kv_k$ für $k \in \{1, 2\}$ vorhanden – nämlich Regeln p_S, p_{A_1} und p_{A_2} mit S, A_1 bzw. A_2 auf der linken Seite –, da $S \neq A_1 \neq A_2 \neq S$ gilt.

Die Existenz der Ableitung $A_1 \Longrightarrow^* u_1A_1v_1$ ($u_1 \neq \lambda$, $v_1 \neq \lambda$) bedingt, dass es eine Ableitung der Form $A_1 \Longrightarrow^* w_1$ mit $w_1 \in \{a\}^* \cdot C_n \cdot \{b\}^*$ für jeden Buchstaben aus C_n geben muss, da in den Wörtern aus $L_n \cdot J_m$ nur zwischen a und b genau ein Buchstabe aus C_n vorkommen darf. Analog gibt es eine Ableitung der Form $A_2 \Longrightarrow^* w_2$ mit $w_2 \in \{d\}^* \cdot F_m \cdot \{e\}^*$ für jeden Buchstaben aus F.

In der Regel p_S darf kein Buchstabe aus $C_n \cup F_m$ vorkommen, da diese Regel bei der Ableitung $S \Longrightarrow^* u'_kA_kv'_k$ angewendet wird und dort $A_k \Longrightarrow^* w_k$ abgeleitet werden darf – w_k enthält bereits einen Buchstaben aus $C_n \cup F_m$. Analoges gilt auch für die Regeln p_{A_k}. Keine der 3 Regeln p_S, p_{A_1} und p_{A_2} enthält einen Buchstaben aus $C_n \cup F_m$. Darüber hinaus existiert keine Regel, auf deren rechter Seite mehr als ein Buchstabe aus $C_n \cup F_m$ vorkommt. Somit muss es für jeden der $(n-1)+(m-1)$ Buchstaben aus $C_n \cup F_m$ mindestens eine separate Regel geben. Insgesamt folgt $\mathrm{Prod}(L_n \cdot J_m) \geq 3+(n-1)+(m-1) = n+m+1$.

Die kontextfreie Grammatik

$$(\{S, S_1, S_2\}, T_{n,m}, P_{n,m}, S)$$

mit

$$\begin{aligned} P_{n,m} = {} & \{S \to S_1S_2, S_1 \to aS_1b, S_2 \to dS_2e\} \cup \\ & \{S_1 \to c_i \mid 1 \leq i \leq n-1\} \cup \{S_2 \to f_i \mid 1 \leq i \leq m-1\} \end{aligned}$$

erzeugt die Sprache $L_n \cdot J_m$. Wir zählen $n+m+1$ Regeln. Mit Obigem folgt $\mathrm{Prod}(L_n \cdot J_m) = n+m+1$. □

Lemma 3.23 – Es gilt $n+1 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,2)$

Für $n \in \mathbb{N}_0$, $n \geq 4$, gilt $n+1 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,2)$.

Beweis: Es seien T_n ein Alphabet mit mindestens $n-1$ Buchstaben, paarweise verschiedene $a, b, c, d_i \in T_n$, $1 \leq i \leq n-4$, sowie

$$\begin{aligned} L_n &= \left\{a^i x b^i \,\middle|\, i \geq 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-4}\}\right\}, \\ J &= \{c\}^+. \end{aligned}$$

Aufgrund von Lemma 3.11 erhalten wir $\mathrm{Prod}(L_n) = n$ und wegen Lemma 3.3 die Aussage $\mathrm{Prod}(J) = 2$. Die Konkatenation der beiden Sprachen ergibt $L_n \cdot J = \{a^i x b^i \mid i \geq 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-4}\}\} \cdot \{c\}^+$. Analog zum Beweis von Lemma 3.11 muss eine regel-minimale kontextfreie Grammatik, die $L_n \cdot J$ generiert, auch die dort ermittelten Regeln enthalten. Da die Erzeugung das zusätzlichen $\{c\}^+$ eine „unabhängige" Schleife bedingt, kommt mindestens eine weitere Regel hinzu. In der Tat genügt die kontextfreie Grammatik

$$(\{S, S_1, S_2\}, T_n, P_n, S)$$

mit

$$\begin{aligned} P_n = {} & \{S \to S_1S_2\} \cup \\ & \{S_1 \to aS_1b\} \cup \{S_1 \to ad_ib \mid 1 \leq i \leq n-4\} \cup \{S_1 \to aS_2b\} \cup \\ & \{S_2 \to cS_2, S_2 \to c\}, \end{aligned}$$

die $L_n \cdot J$ mittels $n+1$ Regeln erzeugt. Mithin gilt $\mathrm{Prod}(L_n \cdot J) = n+1$. □

Lemma 3.24 – Es gilt $n+m \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m)$

Für $n \in \mathbb{N}_0$, $n \geq 3$, und $m \in \mathbb{N}_0$, $m \geq 2$ gilt $n+m \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m)$.

Beweis: Es seien $T_{n,m}$ ein Alphabet mit mindestens $n+m+2$ Buchstaben, paarweise verschiedene $a, b, d, e, \alpha, c_i, f_j \in T_{n,m}$, $1 \le i \le n-2$, $1 \le j \le m-1$, und

$$\begin{aligned} L_n &= \left\{ a^i x b^i \,\middle|\, i \ge 0, x \in \{c_1, \ldots, c_{n-2}\} \right\} \cdot \{\alpha\}\,, \\ J_m &= \left\{ d^i x e^i \,\middle|\, i \ge 0, x \in \{f_1, \ldots, f_{m-1}\} \right\}. \end{aligned}$$

Die Regelkomplexität von L_n lesen wir aus Lemma 3.10 ab und die von J_m aus Lemma 3.9 und erhalten $\mathrm{Prod}(L_n) = n$ bzw. $\mathrm{Prod}(J_m) = m$.

Für die Konkatenation von L_n und J_m erhalten wir

$$L_n \cdot J_m = \left\{ a^i x b^i \alpha d^j y e^j \,\middle|\, i \ge 0, j \ge 0, x \in \{c_1, \ldots, c_{n-2}\}, y \in \{f_1, \ldots, f_{m-1}\} \right\}.$$

Wir können den Beweis des Lemmas 3.22 so abwandeln, dass er $\mathrm{Prod}(L_n \cdot J_m) \ge n+m$ liefert – es muss lediglich α eingefügt werden. Eine kontextfreie Grammatik, die $L_n \cdot J_m$ generiert, ist

$$(\{S, S_1, S_2\}, T_{n,m}, P_{n,m}, S)$$

mit

$$\begin{aligned} P_{n,m} = {} & \{S \to S_1 \alpha S_2, S_1 \to a S_1 b, S_2 \to d S_2 e\} \cup \\ & \{S_1 \to c_i \mid 1 \le i \le n-2\} \cup \{S_2 \to f_i \mid 1 \le i \le m-1\}\,, \\ T_{n,m} = {} & \{a, b, d, e, \alpha\} \cup \{c_1, \ldots, c_{n-2}\} \cup \{f_1, \ldots, f_{m-1}\}\,. \end{aligned}$$

Wir zählen $n+m$ Regeln. Insgesamt resultiert $\mathrm{Prod}(L_n \cdot J_m) = n+m$. □

Lemma 3.25 – Es gilt $n+m-1 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n,m)$

Für $n \in \mathbb{N}_0$, $n \ge 3$, und $m \in \mathbb{N}_0$, $m \ge 3$, gilt $n+m-1 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n,m)$.

Beweis: Es seien $T_{n,m}$ ein Alphabet mit mindestens $n+m+1$ Buchstaben, paarweise verschiedene $a, b, d, e, \alpha, c_i, f_j \in T_{n,m}$, $1 \le i \le n-2$, $1 \le j \le m-2$, und

$$\begin{aligned} L_n &= \left\{ a^i x b^i \,\middle|\, i \ge 0, x \in \{c_1, \ldots, c_{n-2}\} \right\} \cdot \{\alpha\}\,, \\ K_m &= \left\{ d^i x e^i \,\middle|\, i \ge 0, x \in \{f_1, \ldots, f_{m-2}\} \right\} \cdot \{\alpha\}\,. \end{aligned}$$

Wir zeigen $\mathrm{Prod}(L_n) = n$, $\mathrm{Prod}(K_m) = m$ und $\mathrm{Prod}(L_n \cdot K_m) = n+m-1$: Lemma 3.10 liefert die Aussagen $\mathrm{Prod}(L_n) = n$ und $\mathrm{Prod}(K_m) = m$. Durch eine leichte Variation des Beweises zum Lemma 3.24 lässt sich $\mathrm{Prod}(L_n \cdot K_m) \ge n+m-1$ ermitteln – der Buchstabe f_{m-1} muss dort getilgt und ein weiteres α hinzugefügt werden. Die kontextfreie Grammatik

$$(\{S, S_1, S_2\}, T_{n,m}, P_{n,m}, S)$$

mit

$$\begin{aligned} P_{n,m} = {} & \{S \to S_1 \alpha S_2 \alpha, S_1 \to a S_1 b, S_2 \to d S_2 e\} \cup \\ & \{S_1 \to c_i \mid 1 \le i \le n-2\} \cup \{S_2 \to f_i \mid 1 \le i \le m-2\}\,, \end{aligned}$$

$$T_{n,m} = \{a, b, d, e, \alpha\} \cup \{c_1, \ldots, c_{n-2}\} \cup \{f_1, \ldots, f_{m-2}\}$$

erzeugt $L_n \cdot K_m$ mittels $n + m - 1$ Regeln. Daher gilt $\mathrm{Prod}(L_n \cdot K_m) = n + m - 1$. $\square$

Lemma 3.26 – Es gilt $n + m - 2 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, m)$
Für $n \in \mathbb{N}_0$, $n \geq 5$, und $m \in \mathbb{N}_0$, $m \geq 4$, gilt $n + m - 2 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, m)$.

Beweis: Es seien $T_{n,m}$ ein Alphabet mit mindestens $n + m - 3$ Buchstaben, paarweise verschiedene $a, b, c, e, f, \alpha, d_i, g_j \in T_{n,m}$, $1 \leq i \leq n - 5$, $1 \leq j \leq m - 4$, und

$$L_n = \left\{ a^i x b^i \,\middle|\, i \geq 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-5}\} \right\} \cdot \{\alpha\},$$
$$K_m = \left\{ e^i x f^i \,\middle|\, i \geq 1, x \in \{c\}^+ \cup \{g_1, \ldots, g_{m-4}\} \right\}.$$

Die Komplexitäten $\mathrm{Prod}(L_n) = n$ und $\mathrm{Prod}(K_m) = m$ entnehmen wir Lemma 3.12 bzw. Lemma 3.11. Um $\mathrm{Prod}(L_n \cdot K_m) = n + m - 2$ zu zeigen, können wir strukturell so wie im Beweis zum Lemma 3.24 verfahren. Dabei muss beachtet werden, dass die Regeln zur Erzeugung der Wörter aus $\{c\}^+$, die zweimal als Teilwörter in einem Wort aus $L_n \cdot K_m$ vorkommen können, nicht „doppelt“ vorhanden sein müssen. Dadurch „sparen“ wir 2 Regeln, bezogen auf die Summe der Anzahl der Regeln von regel-minimalen Grammatiken für L_n und K_m. Die kontextfreie Grammatik

$$(\{S, S_1, S_2, S_c\}, T_{n,m}, P_{n,m}, S)$$

mit

$$\begin{aligned}
P_{n,m} = {} & \{S \to S_1 \alpha S_2, S_1 \to a S_1 b, S_2 \to e S_2 f\} \cup \\
& \{S_1 \to d_i \mid 1 \leq i \leq n - 5\} \cup \{S_2 \to g_i \mid 1 \leq i \leq m - 4\} \cup \\
& \{S_1 \to a S_c b, S_2 \to e S_c f, S_c \to c S_c, S_c \to c\}, \\
T_{n,m} = {} & \{a, b, c, e, f, \alpha\} \cup \{d_1, \ldots, d_{n-5}\} \cup \{g_1, \ldots, g_{m-4}\}
\end{aligned}$$

erzeugt $L_n \cdot K_m$ mittels $n + m - 2$ Regeln. Es folgt $\mathrm{Prod}(L_n \cdot K_m) = n + m - 2$. $\square$

Lemma 3.27 – Es gilt $n + 2, \ldots, n + m - 3 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, m)$
Für $n \in \mathbb{N}_0$, $n \geq 5$, und $m \in \mathbb{N}_0$, $m \geq 5$, gilt $n + 2, \ldots, n + m - 3 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, m)$.

Beweis: Es seien $k \in \mathbb{N}_0$, $n + 2 \leq k \leq n + m - 3$, und $p = (n + m) - k$. Ferner seien $T_{n,m,k}$ ein Alphabet mit mindestens $k - 4$ Buchstaben, paarweise verschiedene $a, a', b, b', d, e, c_i, c'_{i'}, f_j \in T_{n,m,k}$, $1 \leq i \leq n - p - 1$, $1 \leq i' \leq m - p - 1$, $1 \leq j \leq p - 2$, und $C_{n,k} = \{c_1, \ldots, c_{n-p-1}\}$, $C'_{m,k} = \{c'_1, \ldots, c'_{m-p-1}\}$, $F_k = \{f_1, \ldots, f_{p-2}\}$ sowie

$$R_{n,m,k} = \left\{ d^i x e^i \,\middle|\, i \geq 0, x \in F_k \right\},$$
$$L_{n,m,k} = \left\{ a^i x b^i \,\middle|\, i \geq 0, x \in C_{n,k} \right\} \cdot R_{n,m,k},$$
$$K_{n,m,k} = \left\{ a'^i x b'^i \,\middle|\, i \geq 0, x \in C'_{m,k} \right\} \cdot R_{n,m,k}.$$

Dann gelten nach dem Beweis von Lemma 3.22 die Aussagen $\mathrm{Prod}(L_{n,m,k}) = n$ und $\mathrm{Prod}(K_{n,m,k}) = m$.

Es sei $G_{n,m,k} = (N_{n,m,k}, T_{n,m,k}, P_{n,m,k}, S)$ eine regel-minimale kontextfreie Grammatik, die $L_{n,m,k} \cdot K_{n,m,k}$ erzeugt. Dass $\mathrm{Prod}(G_{n,m,k}) = \mathrm{Prod}(L_{n,m,k} \cdot K_{n,m,k}) \geq k$ korrekt ist, lässt sich zeigen, indem analog zum Beweis von Lemma 3.22 argumentiert wird. Derart existieren in $P_{n,m,k}$ für jeden der $k-4$ Buchstaben aus $C'_{m,k}$, $C_{n,k}$ und F_k mindestens eine separate Regel und darüber hinaus mindestens 3 weitere Regeln, die Schleifenableitungen der Form $A_i \Longrightarrow^* u_i A_i v_i$ mit $A_i \in N_{n,m,k}$ und $u_1 \in \{a\}^+$, $v_1 \in \{b\}^+$, $u_2 \in \{a'\}^+$, $v_2 \in \{b'\}^+$, $v_3 \in \{d\}^+$ sowie $v_3 \in \{e\}^+$ für $i \in \{1,2,3\}$ erzeugen. Schließlich wird analog zum Beweis von Lemma 3.22 noch mindestens 1 zusätzliche Regel mit S auf der linken Seite benötigt. Insgesamt sind es folglich mindestens $(k-4)+3+1 = k$ Regeln. In der Tat genügen genau k Regeln, um $L_{n,m,k} \cdot K_{n,m,k}$ zu generieren:

$$(\{S, S_L, S_K, S_K\}, T_{n,m,k}, P_{n,m,k}, S)$$

mit

$$\begin{aligned}
P_{n,m,k} = {} & \{S \to S_L S_R S_K S_R, S_L \to a S_L b, S_K \to a' S_K b', S_R \to d S_R e\} \cup \\
& \{S_L \to c_i \mid 1 \leq i \leq n-p-1\} \cup \\
& \{S_K \to c'_i \mid 1 \leq i \leq m-p-1\} \cup \\
& \{S_R \to f_i \mid 1 \leq i \leq p-2\}, \\
T_{n,m,k} = {} & \{a, b, a', b', d, e\} \cup C_{n,k} \cup C'_{m,k} \cup F_k
\end{aligned}$$

Somit gilt letztendlich $\mathrm{Prod}(L_{n,m,k} \cdot K_{n,m,k}) = k$. □

Satz 3.28 – Teilbestimmung des Raums der Regelkomplexität unter Konkatenation
Für $n, m \in \mathbb{N}_0$ und $m \leq n$ gelten

- $$\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n, 0),$$

- $$\begin{aligned} \{1\} &= \mathrm{C}^{\mathrm{Prod}}_{\cdot}(1,1), \\ & \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,1) \not\ni 0, 1, n+2, \ldots \text{ für } n \geq 2, \end{aligned}$$

- $$\begin{aligned} \{2, \ldots, 5\} &= \mathrm{C}^{\mathrm{Prod}}_{\cdot}(2,2), \\ n, n+1, n+2, n+3 &\in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,2) \not\ni 0, 1, n+4, \ldots \text{ für } n \geq 3, \end{aligned}$$

- $$\begin{aligned} & n+m-1, n+m, n+m+1 \\ & \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,3) \not\ni 0, 1, n+m+2, \ldots \text{ für } n \geq 3, \end{aligned}$$

-

$$\begin{aligned}7, \ldots, 9 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(4,4) \not\ni 0, 1, 10, \ldots, \\ n+m-2, n+m-1, n+m, n+m+1 \\ \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,4) \not\ni 0, 1, n+m+2, \ldots \textit{ für } n \geq 5,\end{aligned}$$

- $$\begin{aligned}&n+2, \ldots, n+m+1 \\ &\in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m) \not\ni 0, 1, n+m+2, \ldots \textit{ für } n \geq 5, m \geq 5,\end{aligned}$$

- $$\mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m) = \mathrm{C}^{\mathrm{Prod}}_{\cdot}(m,n).$$

Beweis: Der Satz ist eine Zusammenfassung der Lemmata 3.19, 3.20, 3.21, 3.22, 3.23, 3.24, 3.25, 3.26 und 3.27 unter Zuhilfenahme des Lemmas 3.18. □

Bemerkung: Es seien $n, m \in \mathbb{N}_0$ und $m \leq n$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 3.28 vollständig dargestellt:

- $$2, \ldots, n+1 \in^? \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,1) \text{ für } n \geq 2,$$

- $$2, \ldots, n-1 \in^? \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,2) \text{ für } n \geq 3,$$

- $$2, \ldots, n+m-2 \in^? \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,3) \text{ für } n \geq 3,$$

- $$\begin{aligned}2, \ldots, 6 \in^? \mathrm{C}^{\mathrm{Prod}}_{\cdot}(4,4), \\ 2, \ldots, n+m-3 \in^? \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,4) \text{ für } n \geq 5,\end{aligned}$$

- $$2, \ldots, n+1 \in^? \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m) \text{ für } n \geq 5, m \geq 5.$$

Wie in [DS08] bezüglich der Variablenkomplexität bei der Konkatenation ist also auch hier im Wesentlichen unklar, ob die Zahlen 2, …, n im Raum der Regelkomplexität enthalten sind oder nicht. △

3.5. Kleene-Abschluss

Wir untersuchen den Raum der Regelkomplexität unter Kleene-Abschluss.

Lemma 3.29 – Ergebnisse zum Raum der Regelkomplexität unter Kleene-Abschluss
Unter den Voraussetzungen $n \in \mathbb{N}_0$ gelten

i. $\{1\} = \mathrm{C}^{\mathrm{Prod}}_{()^*}(0)$,

ii. $\{1, 2\} = \mathrm{C}^{\mathrm{Prod}}_{()^*}(1)$,

iii. $0 \notin \mathrm{C}_{()^*}^{\mathrm{Prod}}(n)$,

iv. $1 \notin \mathrm{C}_{()^*}^{\mathrm{Prod}}(n)$ für $n \geq 2$,

v. $2, \ldots, n+1 \in \mathrm{C}_{()^*}^{\mathrm{Prod}}(n)$ für $n \geq 2$,

vi. $n+2 \in \mathrm{C}_{()^*}^{\mathrm{Prod}}(n)$ für $n \geq 2$,

vii. $n+3, \ldots \notin \mathrm{C}_{()^*}^{\mathrm{Prod}}(n)$.

Beweis: Zu i. Aufgrund des Lemmas 3.1 folgt aus $\mathrm{Prod}(L) = 0$ die Tatsache $L = \emptyset$, und damit $L^* = \emptyset^* = \{\lambda\}$ (Lemma 2.7). Dasselbe Lemma liefert dann $\mathrm{Prod}(L^*) = \mathrm{Prod}(\{\lambda\}) = 1$.

Zu ii. Wegen Lemma 3.1 erhalten wir aus $\mathrm{Prod}(L) = 1$, dass $L = \{w\}$ mit $w \in T^*$ für ein Alphabet T ist. Falls $w = \lambda$ gilt, folgt $L^* = \{\lambda\}^* = \{\lambda\}$. Lemma 3.1 entnehmen wir dann sofort $\mathrm{Prod}(L^*) = \mathrm{Prod}(\{\lambda\}) = 1$. Für den Fall $w \neq \lambda$ geben wir die kontextfreie Grammatik

$$G = (N, T, \{S \to wS, S \to \lambda\}, S)$$

an, die sicher $L^* = \{w\}^*$ erzeugt. Dies impliziert $\mathrm{Prod}(L^*) \leq \mathrm{Prod}(G) = 2$. Bezug nehmend auf Lemma 2.23, benötigen wir für eine regel-minimale kontextfreie Grammatik, die $L^* = \{w\}^*$ ($|L^*| = |\mathbb{N}|$) erzeugt, mindestens zwei Regeln. Damit gilt $\mathrm{Prod}(L^*) \geq 2$ und insgesamt $\mathrm{Prod}(L^*) = 2$.

Zu iii. Aufgrund des Lemmas 2.7 Anstrich i gilt $L^* \neq \emptyset$ für alle kontextfreien Sprachen L. Wegen Lemma 3.1 folgt dann immer $\mathrm{Prod}(L^*) \geq 1$.

Zu iv. Aus $\mathrm{Prod}(L^*) = 1$ folgt aufgrund des Lemmas 3.1 $L^* = \{w\}$ mit $w \in T^*$ für ein Alphabet T. Für w kommt wegen Lemma 2.7, Anstrich ii nur λ infrage. Damit erhalten wir als einzige Möglichkeiten entweder $L = \emptyset$ oder $L = \{\lambda\}$ und schließlich wegen Lemma 3.1 $\mathrm{Prod}(L) = 0$ bzw. $\mathrm{Prod}(L) = 1$.

Zu v. Es seien für $n \geq 2$ und $2 \leq k \leq n+1$ die Variable T_k ein Alphabet mit mindestens k' Buchstaben, paarweise verschiedene $a_i, b \in T_k$, $1 \leq i \leq k'-1$, sowie

$$\begin{aligned} k' &= k-1, \\ L_{n,k} &= \{a_i \mid 1 \leq i \leq k'-1\} \cup \left\{ b^{2^i} \,\middle|\, 0 \leq i \leq n-k' \right\}. \end{aligned}$$

Die Sprache $L_{n,k}$ lässt sich als $L_{n,k} = A_k \cup B_{n,k}$ mit

$$A_k = \{a_i \mid 1 \leq i \leq k'-1\}$$

und

$$B_{n,k} = \left\{ b^{2^i} \,\middle|\, 0 \leq i \leq n-k' \right\}$$

darstellen. Die Alphabete der Sprachen A_k und $B_{n,k}$ sind disjunkt und nach den Lemmata 3.8 und 3.2 gelten $\mathrm{Prod}(A_k) = k'-1 = k-2$ bzw. $\mathrm{Prod}(B_{n,k}) = n-k'+1 = n-k+2$. Mithin ergibt sich unter Berücksichtigung von Satz 2.22 $\mathrm{Prod}(L_{n,k}) = \mathrm{Prod}(A_k \cup B_{n,k}) = (k-2) + (n-k+2) = n$.

Die Anwendung des Kleene-Abschlusses auf $L_{n,k}$ bringt

$$L_{n,k}{}^* = \left(\{a_i \mid 1 \leq i \leq k' - 1\} \cup \{b\}\right)^*,$$

wodurch wir Lemma 3.7 anwenden können und $\mathrm{Prod}(L_{n,k}{}^*) = k' + 1 = k$ erhalten.

Zu vi. Es seien für $n \geq 2$ die Variable T_n ein Alphabet mit mindestens $n+1$ Buchstaben, paarweise verschiedene $a, b, c_i \in T_n$, $1 \leq i \leq n-1$, und

$$L_n = \left\{ a^i x b^i \,\middle|\, i \geq 0, x \in \{c_1, \ldots, c_{n-1}\} \right\}.$$

Nach Lemma 3.9 erhalten wir $\mathrm{Prod}(L_n) = n$.

Es sei $G_n = (N_n, T_n, P_n, S)$ mit $T_n = \{a, b\} \cup \{c_1, \ldots, c_{n-1}\}$ und $\mathrm{L}(G_n) = L_n{}^*$ eine regel-minimale kontextfreie Grammatik. Nach Lemma 2.14 ist G_n reduziert. Die Existenz der Ableitung $S \Longrightarrow^* \lambda \in L_n{}^*$ bedingt eine Regel $B \to \lambda \in P_n$ für $B \in N_n$. Wenden wir Lemma 2.10 auf $L_n{}^*$ an, erhalten wir $\{c_i \mid 1 \leq i \leq n-1\}$ für das dort definierte X. Es folgt $C_i \to w_{c_i} \in P_n$ für gewisse $C_i \in N_n$ und $w_{c_i} \in (N_n \cup \{c_i\})^+$, $|w_{c_i}|_{c_i} \geq 1$, und alle i, $1 \leq i \leq n-1$. Aufgrund von $ac_1b \in L_n{}^*$ muss es eine Regel geben, in der ein Buchstabe a vorkommt. In den Regeln $B \to \lambda$ und $C_i \to w_{c_i}$ kommt kein a vor. Folglich existiert eine weitere Regel $A \to w_A \in P_n$ für $A \in N_n$ und $w_A \in (N_n \cup T_n)^+$ mit der Bedingung $|w_A|_a \geq 1$. Außerdem folgt aus $S \Longrightarrow^* ac_1b$ die Tatsache $|w_A|_a \leq 1$, also $|w_A|_a = 1$. Insgesamt enthält P_n somit mindestens $2 + (n-1)$ Regeln.

Wir werden nun zeigen, dass G_n, enthielte sie nur $2 + (n-1)$ Regeln, $L_n{}^*$ nicht erzeugen könnte. Die Ableitung $S \Longrightarrow^* c_i \in L_n{}^*$ impliziert $|w_{c_i}|_{c_i} = 1$. Da w_A ein a enthält, kann die Regel $A \to w_A$ nicht bei $S \Longrightarrow^* c_i$ angewendet worden sein. Deshalb bleibt als einzige Möglichkeit für C_i nur S übrig. Analog leiten wir $B = S$, also $S \to \lambda$, her.

Wenn $A \neq S$ und $|w_{c_i}|_A \geq 1$ gälten, wäre $S \Longrightarrow^* c_i$ nicht realisierbar – es würde immer auch ein a erzeugt werden. Gälten $A \neq S$ und $|w_{c_i}|_A = 0$, fände die Regel $A \to w_A$ nie Anwendung. Wir schlussfolgern hieraus $A = S$, also $S \to w_A$.

Für alle $m \in \mathbb{N}_0$ existiert die Ableitung $S \Longrightarrow^* a^m c_1 b^m$. Die Buchstaben a^m können nicht mithilfe von $S \to w_{c_1}$ erzeugt werden, da sonst $|w_{c_1}|_S \geq 1$ gelten müsste und mit jeder neuen Anwendung von $S \to w_{c_1}$ ein weiteres c_1 zur Ableitung hinzukäme. Daher bleibt nur $|w_A|_S \geq 1$ übrig.

Das Terminalwort w'_A in der Ableitung $S \Longrightarrow w_A \Longrightarrow^* w'_A \in T_n{}^*$ enthält, wenn $C_i \to w_{c_i} = S \to w_{c_i}$ nicht angewendet wird, sondern nur $S \to \lambda$, kein c_i, und ist somit kein Wort von $L_n{}^*$. Folglich gilt $|w_A|_S = 0$. Dieser Widerspruch zur obigen Aussage $|w_A|_S \geq 1$ lässt sich nur auflösen, indem wir die Annahme, dass G_n nur $2 + (n-1)$ Regeln enthält, fallen lassen. Es gilt $\mathrm{Prod}(G_n) \geq n + 2$. Insgesamt erhalten wir unter Beachtung der kontextfreien Grammatik

$$(\{S, S'\}, T_n, P'_n, S)$$

mit

$$P'_n = \{S \to SS', S \to \lambda, S' \to aS'b\} \cup \{S' \to c_i \mid 1 \leq i \leq n-1\},$$

die $L_n{}^*$ mittels $n+2$ Regeln erzeugt, die Aussage $\mathrm{Prod}(L_n{}^*) = n + 2$.

Zu vii. Wählen wir die Grammatik G mit $\mathrm{L}(G) = L$ aus dem Beweis des Lemmas 2.8 so, dass $\mathrm{Prod}(G) = \mathrm{Prod}(L) = n$ gilt, sind in $G_{()^*}$ mit $\mathrm{L}(G_{()^*}) = L^*$ höchstens $n+2$ Regeln enthalten (Beweis zu Lemma 2.8). Folglich gilt $\mathrm{Prod}(L^*) \leq \mathrm{Prod}(G_{()^*}) = n+2$. $\square$

Satz 3.30 – Raum der Regelkomplexität unter Kleene-Abschluss
Für $n \in \mathbb{N}_0$ gelten

$$\{1\} = \mathrm{C}^{\mathrm{Prod}}_{()^*}(0),$$
$$\{1, 2\} = \mathrm{C}^{\mathrm{Prod}}_{()^*}(1),$$
$$\{2, \ldots, n+2\} = \mathrm{C}^{\mathrm{Prod}}_{()^*}(n) \text{ für } n \geq 2.$$

Beweis: Der Satz ist eine Zusammenfassung des Lemmas 3.29. $\square$

Bemerkung: Satz 3.30 lässt keine Fälle offen. $\triangle$

3.6. Homomorphismus

Wir untersuchen den Raum der Regelkomplexität unter Homomorphismus.

Lemma 3.31 – Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Homomorphismus
Unter den Voraussetzungen $n \in \mathbb{N}_0$ und einem beliebigen Homomorphismus h gelten die Aussagen

i. $\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}}(0)$,

ii. $0 \notin \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}}(n)$ für $n \geq 1$,

iii. $n+1, \ldots \notin \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}}(n)$.

Beweis: Es sei L eine kontextfreie Sprache, auf die der Homomorphismus h angewendet werden kann.

Zu i. Das Lemma 3.1 liefert für $\mathrm{Prod}(L) = 0$ als einzige Möglichkeit $L = \emptyset$, womit sofort $\mathrm{h}(L) = \mathrm{h}(\emptyset) = \emptyset$ und $\mathrm{Prod}(\mathrm{h}(L)) = \mathrm{Prod}(\emptyset) = 0$ folgen.

Zu ii. Soll $\mathrm{Prod}(\mathrm{h}(L)) = 0$ sein, so kommt wegen Lemma 3.1 nur $\mathrm{h}(L) = \emptyset$ infrage, was unmittelbar zu $L = \emptyset$ und $\mathrm{Prod}(L) = \mathrm{Prod}(\emptyset) = 0$ führt. Vorausgesetzt wird aber $\mathrm{Prod}(L) = n \geq 1$.

Zu iii. Wählen wir die kontextfreie Grammatik G aus dem Beweis des Lemmas 2.8 so, dass $\mathrm{L}(G) = L$ und $\mathrm{Prod}(G) = \mathrm{Prod}(L) = n$ gelten, sind in G_{h} mit $\mathrm{L}(G_{\mathrm{h}}) = \mathrm{h}(L)$ genau $\mathrm{Prod}(L)$ Regeln enthalten (Beweis zu Lemma 2.8). Folglich gilt $\mathrm{Prod}(\mathrm{h}(L)) \leq \mathrm{Prod}(G_{\mathrm{h}}) = \mathrm{Prod}(L) = n$. $\square$

Lemma 3.32 – Es gibt ein h, sodass $1, \ldots, n \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n)$ gilt
Für ein Alphabet T mit mindestens 3 Buchstaben, paarweise verschiedene $a, b, c \in T$ und $n \in \mathbb{N}_0$, $n \geq 1$, sowie den (nicht-löschenden) Homomorphismus

$$\mathrm{h}\colon \{a, b, c\}^* \to \{a, b\}^* \text{ mit } \mathrm{h}(a) = a, \mathrm{h}(b) = b, \mathrm{h}(c) = b$$

gilt $1, \ldots, n \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n)$.

Beweis: Es seien $n \geq 1$ und $1 \leq k \leq n$ sowie

$$L_{n,k} = \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq k-1 \right\} \cup \left\{ b^{2^n - 2^i} c^{2^i} \;\middle|\; k \leq i \leq n \right\}$$

eine kontextfreie Sprache. Wir können die Sprache $L_{n,k}$ in

$$K_k = \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq k-1 \right\}$$

und

$$Q_{n,k} = \left\{ b^{2^n - 2^i} c^{2^i} \;\middle|\; k \leq i \leq n \right\}$$

zerlegen ($L_{n,k} = K_k \cup Q_{n,k}$). Die Alphabete der Teilsprachen K_k und $Q_{n,k}$ sind disjunkt. Lemma 3.2 liefert $\mathrm{Prod}(K_k) = k - 1$. Die Regelanzahl $\mathrm{Prod}(Q_{n,k})$ bestimmen wir im Folgenden.

Es sei $G_{n,k} = (N,\ T,\ P_{n,k},\ S)$ eine regel-minimale kontextfreie Grammatik mit $\mathrm{L}(G_{n,k}) = Q_{n,k}$. Wir dürfen $T = \{b, c\}$ annehmen. Im Falle von $k = n$, also $|Q_{n,k}| = 1$, lesen wir aus Lemma 3.1 $\mathrm{Prod}(Q_{n,k}) = 1$ ab. Für die übrigen Fälle gilt: Nach Lemma 2.14 ist $G_{n,k}$ reduziert, und nach Lemma 2.21 existieren in $G_{n,k}$ keine isolierten Nichtterminale. Es muss daher wenigstens zwei verschiedene Ableitungen

$$\begin{aligned} S &\Longrightarrow^* xAy \text{ mit } x, y \in T^*, \\ S &\Longrightarrow^* pAq \text{ mit } p, q \in T^* \end{aligned}$$

für jedes $A \in N$, $A \neq S$, geben ($xAy \neq pAq$). Nach Lemma 2.15 gelten ferner

$$\begin{aligned} A &\Longrightarrow^* u \in T^*, \\ A &\Longrightarrow^* v \in T^*, \end{aligned}$$

wobei $u \neq v$ ist. Substituieren wir nun A in den obigen Ableitungen, entstehen

$$\begin{aligned} S &\Longrightarrow^* xuy \in Q_{n,k}, \\ S &\Longrightarrow^* xvy \in Q_{n,k}, \\ S &\Longrightarrow^* puq \in Q_{n,k}, \\ S &\Longrightarrow^* pvq \in Q_{n,k}. \end{aligned}$$

Aufgrund der Tatsache, dass alle Wörter aus $Q_{n,k}$ die Länge 2^n haben, ergeben sich hieraus die verschärfenden Eigenschaften $|u| = |v|$ und $|xy| = |pq|$.

Wir betrachten nachstehend konkret die drei vollständigen Fälle für das Ableiten von

Wörtern aus $Q_{n,k}$. Aus

$$S \Longrightarrow^* xAy = b^j A b^{j'} c^{2^i} \Longrightarrow^* b^j u b^{j'} c^{2^i} = b^{2^n - 2^i} c^{2^i} \in Q_{n,k}$$

für $j \in \mathbb{N}_0$ und $j', i \in \mathbb{N}$ folgt $u = b^{2^n - 2^i - j - j'}$. Für v muss dann $v \in \{b\}^*$ wegen

$$S \Longrightarrow^* xAy \Longrightarrow^* b^j v b^{j'} c^{2^i} \in Q_{n,k}$$

gelten. Es existiert allerdings kein $v \in \{b\}^*$, für das gleichzeitig $|v| = |u|$ und $v \neq u = b^{2^n - 2^i - j - j'}$ wahr ist. Somit kann dieser Fall nicht eintreten (Analoges gilt für $S \Longrightarrow^* pAq$).

Der Fall

$$S \Longrightarrow^* xAy = b^{2^n - 2^i} c^j A c^{j'} \Longrightarrow^* b^{2^n - 2^i} c^j u c^{j'} = b^{2^n - 2^i} c^{2^i} \in Q_{n,k}$$

für $j' \in \mathbb{N}_0$ und $j, i \in \mathbb{N}$ sowie $u \in \{c\}^*$ kommt in analoger Weise nicht infrage (Analoges gilt für $S \Longrightarrow^* pAq$).

Es verbleibt der Fall

$$\begin{aligned}
S &\Longrightarrow^* xAy = b^{j_1'} A c^{j_1} \Longrightarrow^* b^{j_1'} u c^{j_1} = b^{2^n - 2^{i_1}} c^{2^{i_1}} \in Q_{n,k}, \\
S &\Longrightarrow^* xAy = b^{j_1'} A c^{j_1} \Longrightarrow^* b^{j_1'} v c^{j_1} = b^{2^n - 2^{i_2}} c^{2^{i_2}} \in Q_{n,k}, \\
S &\Longrightarrow^* pAq = b^{j_2'} A c^{j_2} \Longrightarrow^* b^{j_2'} u c^{j_2} = b^{2^n - 2^{i_3}} c^{2^{i_3}} \in Q_{n,k}, \\
S &\Longrightarrow^* pAq = b^{j_2'} A c^{j_2} \Longrightarrow^* b^{j_2'} v c^{j_2} = b^{2^n - 2^{i_4}} c^{2^{i_4}} \in Q_{n,k}
\end{aligned} \tag{6.1}$$

für $j_1, j_1', j_2, j_2' \in \mathbb{N}_0$, $i_t \in \mathbb{N}$ und $t \in \{1,\ 2,\ 3,\ 4\}$. Er impliziert $u = b^{r_1} c^{s_1}$ und $v = b^{r_2} c^{s_2}$ für $r_t, s_t \in \mathbb{N}_0$, $r_t + s_t \neq 0$ und $t \in \{1,\ 2\}$. Darüber hinaus muss $s_1 < s_2$ ohne Beschränkung der Allgemeinheit wegen $|b^{r_1} c^{s_1}| = |b^{r_2} c^{s_2}|$ und $b^{r_1} c^{s_1} \neq b^{r_2} c^{s_2}$ sein. Analog muss ohne Beschränkung der Allgemeinheit $j_1 < j_2$ gelten, da $|b^{j_1'} A c^{j_1}| = |b^{j_2'} A c^{j_2}|$ und $b^{j_1'} A c^{j_1} \neq b^{j_2'} A c^{j_2}$ eingehalten werden müssen. Wir ersetzen nun u und v in den Terminalwörtern der Ableitungen in (6.1) und betrachten die Anzahl der Buchstaben c:

$$\begin{aligned}
|c^{s_1} c^{j_1}| &= s_1 + j_1 = 2^{i_1}, \\
|c^{s_2} c^{j_1}| &= s_2 + j_1 = 2^{i_2}, \\
|c^{s_1} c^{j_2}| &= s_1 + j_2 = 2^{i_3}, \\
|c^{s_2} c^{j_2}| &= s_2 + j_2 = 2^{i_4}.
\end{aligned} \tag{6.2}$$

Offenbar resultieren $2^{i_4} > 2^{i_3}$, $2^{i_2} > 2^{i_1}$ und $2^{i_3} > 2^{i_1}$ ohne Beschränkung der Allgemeinheit. Durch Gleichsetzen und Faktorisierung der Gleichungen in (6.2) kommen wir auf

$$\begin{aligned}
2^{i_4} + 2^{i_1} &= 2^{i_3} + 2^{i_2} \\
2^{i_4} &= 2^{i_1} \left(-1 + 2^{i_3 - i_1} + 2^{i_2 - i_1}\right),
\end{aligned}$$

was zu einem Widerspruch führt, da $-1 + 2^{i_3 - i_1} + 2^{i_2 - i_1}$ nicht durch 2 teilbar ist ($i_3 = i_1$

und $i_2 = i_1$ sind oben ausgeschlossen worden). Damit ist auch der letzte Fall unmöglich.

Aufgrund der Reduziertheit von $G_{n,k}$ – Lemma 2.14 – können wir jeden Ableitungsschritt in Ableitungen der Wörter aus $Q_{n,k}$, der mehr als ein Nichtterminal enthält, auf einen solchen mit einem Nichtterminal vereinfachen. Da die oben betrachteten drei Fälle nicht möglich sind, kann es keine zwei verschiedenen Ableitungen $S \Longrightarrow^* xAy$ und $S \Longrightarrow^* pAq$ mit einem Nichtterminal $A \in N \setminus \{S\}$ geben. Damit wäre A ein isoliertes Nichtterminal, das allerdings nach Lemma 2.21 in $G_{n,k}$ nicht existiert. Mithin befindet sich ausschließlich S in N. Für jedes Wort in $Q_{n,k}$ muss es genau eine Regel in $P_{n,k}$ gegeben, da $Q_{n,k}$ endlich ist und es somit keine Regel der Form $S \to xSy$ mit $x, y \in (T \cup \{S\})^*$ geben darf. Wir konstruieren nun eine regel-minimale kontextfreie Grammatik $G_{n,k}$, die $Q_{n,k}$ erzeugt:

$$\begin{aligned} G_{n,k} &= (\{S\}, \{b, c\}, P_{n,k}, S), \\ P_{n,k} &= \left\{ S \to b^{2^n - 2^i} c^{2^i} \,\middle|\, k \le i \le n \right\}. \end{aligned}$$

Damit ergibt sich $\mathrm{Prod}(Q_{n,k}) = n - k + 1$.

Wie oben erwähnt, sind die Alphabete von K_k und $Q_{n,k}$ disjunkt. Zusammen mit Satz 2.22 erhalten wir

$$\mathrm{Prod}(L_{n,k}) = \mathrm{Prod}(K_k \cup Q_{n,k}) = (k-1) + (n-k+1) = n.$$

Wenden wir den Homomorphismus h auf $L_{n,k}$ an, erhalten wir die Sprache

$$\mathrm{h}(L_{n,k}) = \left\{ a^{2^i} \,\middle|\, 1 \le i \le k-1 \right\} \cup \left\{ b^{2^n} \right\}.$$

Die Alphabete der Teilsprachen $\{a^{2^i} \mid 1 \le i \le k-1\}$ und $\{b^{2^n}\}$ sind disjunkt, woraus sich letztendlich

$$\mathrm{Prod}(\mathrm{h}(L_{n,k})) = (k-1) + 1 = k$$

ergibt. □

Satz 3.33 – Teilbestimmung des Raums der Regelkomplexität unter Homomorphismus

Es sei $n \in \mathbb{N}_0$.

i. Für einen beliebigen Homomorphismus h *gelten*

$$\begin{aligned} \{0\} &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(0), \\ & \mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n) \not\ni 0, n+1, \ldots \text{ für } n \ge 1. \end{aligned}$$

ii. Es existiert ein Homomorphismus h*, für den die folgenden Aussagen korrekt sind:*

$$\begin{aligned} \{0\} &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(0), \\ \{1, \ldots, n\} &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n) \text{ für } n \ge 1. \end{aligned}$$

iii. Für die Menge aller Homomorphismen Hom *gelten*

$$\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}}(0),$$
$$\{1, \ldots, n\} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}}(n) \text{ für } n \geq 1.$$

Beweis: Anstrich i und ii sind eine Zusammenfassung der Lemmata 3.31 und 3.32. Der Anstrich iii ergibt sich aus den Anstrichen i und ii. □

Bemerkung: Es sei $n \in \mathbb{N}_0$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 3.33 vollständig dargelegt:

- Es sei ein beliebiger Homomorphismus h gegeben. Gilt $1, \ldots, n \in \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}}(n)$ für $n \geq 1$?

△

3.7. Inverser Homomorphismus

Wir untersuchen den Raum der Regelkomplexität unter inversem Homomorphismus.

Lemma 3.34 – Allgemeine Ergebnisse zum Raum der Regelkomplexität unter inversem Homomorphismus
Für einen beliebigen Homomorphismus h gilt $\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}^{-1}}(0)$.

Beweis: Es sei L eine kontextfreie Sprache, auf die der inverse Homomorphismus h^{-1} angewendet werden kann. Lemma 3.1 entnehmen wir, dass für $\mathrm{Prod}(L) = 0$ nur $L = \emptyset$ infrage kommt. Es folgen $\mathrm{h}^{-1}(L) = \mathrm{h}^{-1}(\emptyset) = \emptyset$ und somit $\mathrm{Prod}(\mathrm{h}^{-1}(L)) = \mathrm{Prod}(\emptyset) = 0$ aufgrund desselben Lemmas. □

Lemma 3.35 – Es gibt ein $\mathbf{h}$, sodass $\mathbf{0, \ldots, n \in C^{Prod}_{h^{-1}}(n)}$ gilt
Für ein Alphabet T mit mindestens 3 Buchstaben, paarweise verschiedene $a, b, c \in T$ und $n \in \mathbb{N}_0$ sowie den Homomorphismus

$$\mathrm{h}\colon \{a, b\}^* \to \{a, b, c\}^* \text{ mit } \mathrm{h}(a) = a, \mathrm{h}(b) = c$$

gilt $0, \ldots, n \in \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}^{-1}}(n)$.

Beweis: Es sei

$$L_{n,k} = \left\{a^{2^i} \;\middle|\; 1 \leq i \leq k\right\} \cup \left\{b^{2^i} \;\middle|\; k+1 \leq i \leq n\right\}$$

eine kontextfreie Sprache für $n \in \mathbb{N}_0$ und $k \in \mathbb{N}_0$, $0 \leq k \leq n$. Wir setzen $K_k = \{a^{2^i} \mid 1 \leq i \leq k\}$ und $Q_{n,k} = \{b^{2^i} \mid k+1 \leq i \leq n\}$. Es gilt $L_{n,k} = K_k \cup Q_{n,k}$. Nach den Lemmata 3.2 und 3.1 (im Fall $P_k = \emptyset$ oder $Q_{n,k} = \emptyset$) sind $\mathrm{Prod}(K_k) = k$ sowie $\mathrm{Prod}(Q_{n,k}) = n - k$ gültig. Die Alphabete der Sprachen K_k und $Q_{n,k}$ sind disjunkt. Mithin gilt

$$\mathrm{Prod}(K_k \cup Q_{n,k}) = k + (n - k) = n$$

unter Zuhilfenahme des Satzes 2.22.

Für den inversen Homomorphismus auf $L_{n,k}$ erhalten wir

$$\mathrm{h}^{\text{-}1}(L_{n,k}) = \left\{ a^{2^i} \,\middle|\, 1 \leq i \leq k \right\} = K_k,$$

und damit $\mathrm{Prod}(\mathrm{h}^{\text{-}1}(L_{n,k})) = \mathrm{Prod}(K_k) = k$. □

Lemma 3.36 – Es gibt ein $\mathbf{h}_{n,k}$, sodass $k \in \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}_{n,k}{}^{\text{-}1}}(n)$ gilt

Es sei $T_{x,y} = \{a_i \,|\, x \leq i \leq y\}$ für $x, y \in \mathbb{N}_0$, $x \leq y$, ein Alphabet. Unter den Voraussetzungen $n \in \mathbb{N}$ und $k \in \mathbb{N}$ sowie Homomorphismen

$$\mathrm{h}_{n,k} \colon {T_{1,k}}^* \to {T_{1,n}}^* \text{ mit } \mathrm{h}_{n,k}(a_i) = a_1, 1 \leq i \leq k,$$

gilt $k \in \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h_{n,k}}{}^{\text{-}1}}(n)$.

Beweis: Es sei

$$L_n = T_{1,n}$$

eine kontextfreie Sprache. Mittels Lemma 3.8 erhalten wir für die Menge $T_{x,y}$ für $x, y \in \mathbb{N}_0$, $x \leq y$, als Anzahl der Regeln $\mathrm{Prod}(T_{x,y}) = y - x + 1$. Daher gilt $\mathrm{Prod}(L_n) = \mathrm{Prod}(T_{1,n}) = n - 1 + 1 = n$.

Wenden wir den inversen Homomorphismus auf L_n an, so erhalten wir

$$\mathrm{h}_{n,k}{}^{\text{-}1}(L_n) = \{a_i \,|\, 1 \leq i \leq k\} = T_{1,k}.$$

Mit Obigem schlussfolgern wir

$$\mathrm{Prod}(\mathrm{h}_{n,k}{}^{\text{-}1}(L_n)) = \mathrm{Prod}(T_{1,k}) = k - 1 + 1 = k.$$

□

Satz 3.37 – Teilbestimmung des Raums der Regelkomplexität unter inversem Homomorphismus

Es sei $n \in \mathbb{N}_0$.

i. Für einen beliebigen inversen Homomorphismus $\mathrm{h}^{\text{-}1}$ gilt

$$\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}^{\text{-}1}}(0).$$

ii. Es existiert ein inverser Homomorphismus $\mathrm{h}^{\text{-}1}$, für den die folgenden Aussagen korrekt sind:

$$\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}^{\text{-}1}}(0),$$
$$0, \ldots, n \in \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}^{\text{-}1}}(n) \textit{ für } n \geq 1.$$

iii. *Für die Menge aller inversen Homomorphismen* Hom^{-1} *gelten*

$$\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}^{-1}}(0),$$
$$\mathbb{N} = \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}^{-1}}(n) \textit{ für } n \geq 1.$$

Beweis: Anstrich i und ii sind eine Zusammenfassung der Lemmata 3.34 und 3.35. Anstrich iii ergibt sich aus Anstrich ii und Lemma 3.36. □

Bemerkung: Es seien $n, k \in \mathbb{N}_0$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 3.37 vollständig dargelegt:

- Es sei ein beliebiger inverser Homomorphismus h^{-1} gegeben. Gilt $k \in \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}^{-1}}(n)$ für $n \geq 1$?

△

3.8. Schnitt mit regulärer Sprache

Wir untersuchen den Raum der Regelkomplexität unter Schnitt mit regulärer Sprache.

Lemma 3.38 – Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Schnitt mit regulärer Sprache

Unter den Voraussetzungen $n \in \mathbb{N}_0$ und einer beliebigen regulären Sprache R gelten

i. $\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(n)$ für $R = \emptyset$,

ii. $\{0\} = \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(0)$,

iii. $\{0, 1\} = \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(n)$ für $n \geq 1$ und $|R| = 1$,

iv. $\{0, 1\} = \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(1)$ für $R \neq \emptyset$,

Beweis: Zu i. Es sei L eine beliebige kontextfreie Sprache. Offenbar gilt für $R = \emptyset$ die Aussage

$$\cap_\mathrm{R}(L) = \cap_\emptyset(L) = \emptyset.$$

Nach Lemma 3.1 ergibt sich

$$\mathrm{Prod}(\cap_\mathrm{R}(L)) = \mathrm{Prod}(\cap_\emptyset(L)) = \mathrm{Prod}(\emptyset) = 0.$$

Zu ii. Für eine kontextfreie Sprache L mit $\mathrm{Prod}(L) = 0$ kommt nach Lemma 3.1 nur $L = \emptyset$ in Betracht. Es folgt

$$\cap_\mathrm{R}(L) = \cap_\mathrm{R}(\emptyset) = \emptyset$$

für eine beliebige reguläre Sprache R. Nach Lemma 3.1 erhalten wir

$$\mathrm{Prod}(\cap_\mathrm{R}(L)) = \mathrm{Prod}(\cap_\mathrm{R}(\emptyset)) = \mathrm{Prod}(\emptyset) = 0.$$

Zu iii. Es sei L eine beliebige kontextfreie Sprache und $R = \{w\}$, $w \in T^*$ für ein Alphabet T. Dann gelten
$$\cap_{\mathrm{R}}(L) = \cap_{\{\mathrm{w}\}}(L) = \{w\}$$
im Fall von $w \in L$ und
$$\cap_{\mathrm{R}}(L) = \cap_{\{\mathrm{w}\}}(L) = \emptyset$$
im Fall $w \notin L$. Nach Lemma 3.1 ergeben sich

$$\mathrm{Prod}(\cap_{\mathrm{R}}(L)) = \mathrm{Prod}\Big(\cap_{\{\mathrm{w}\}}(L)\Big) = \mathrm{Prod}(\{w\}) = 1$$

bzw.
$$\mathrm{Prod}(\cap_{\mathrm{R}}(L)) = \mathrm{Prod}\Big(\cap_{\{\mathrm{w}\}}(L)\Big) = \mathrm{Prod}(\emptyset) = 0.$$

Zu iv. Es soll $\mathrm{Prod}(L) = 1$ für eine kontextfreie Sprache L gelten. Nach Lemma 3.1 kommt dann nur $L = \{w\}$, $w \in T^*$, für ein Alphabet T infrage. Wegen $R \neq \emptyset$ existiert ein $w' \in R$. Setzen wir $L = \{w'\}$, erhalten wir

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{w'\}) = \{w'\},$$

was nach Lemma 3.1 die Tatsache

$$\mathrm{Prod}(\cap_{\mathrm{R}}(L)) = \mathrm{Prod}(\cap_{\mathrm{R}}(\{w'\})) = \mathrm{Prod}(\{w'\}) = 1$$

liefert. Setzen wir nun $L = \{a\}$ für ein $a \in T$, $a \notin \mathrm{Alph}(R)$, gilt

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{a\}) = \emptyset.$$

Lemma 3.1 liefert

$$\mathrm{Prod}(\cap_{\mathrm{R}}(L)) = \mathrm{Prod}(\cap_{\mathrm{R}}(\{a\})) = \mathrm{Prod}(\emptyset) = 0.$$

Der Schnitt von $L = \{w\}$ mit $R \neq \emptyset$ ist stets entweder

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{w\}) = \emptyset$$

oder
$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{w\}) = \{w\}$$
je nachdem, ob $w \notin R$ bzw. $w \in R$ gilt. Aufgrund von Lemma 3.1 gilt also immer $\mathrm{Prod}(\cap_{\mathrm{R}}(L)) = 0$ bzw. $\mathrm{Prod}(\cap_{\mathrm{R}}(L)) = 1$. □

Lemma 3.39 – Es gibt ein R, sodass $0, \ldots, n \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n)$ gilt
Für ein Alphabet T und $a \in T$ sowie $n \in \mathbb{N}_0$ und die reguläre Sprache

$$R = \{a\}^+$$

gilt $\{0, \ldots, n\} = \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n)$.

Beweis: Es seien T ein Alphabet mit mindestens 2 Buchstaben, paarweise verschiedene

$a, b \in T$ und $0 \leq k \leq n$ sowie

$$L_{n,k} = \left\{ a^{2^i} \,\middle|\, 1 \leq i \leq k \right\} \cup \left\{ b^{2^i} \,\middle|\, 1 \leq i \leq n-k \right\}.$$

Die Sprache $L_{n,k}$ lässt sich mittels

$$\begin{aligned} A_k &= \left\{ a^{2^i} \,\middle|\, 1 \leq i \leq k \right\}, \\ B_{n,k} &= \left\{ b^{2^i} \,\middle|\, 1 \leq i \leq n-k \right\} \end{aligned}$$

als $A_k \cup B_{n,k}$ ausdrücken. Nach den Lemmata 3.2 und 3.1 (falls $A_k = \emptyset$ oder $B_{n,k} = \emptyset$) gelten $\mathrm{Prod}(A_k) = k$ und $\mathrm{Prod}(B_{n,k}) = n - k$. Da A_k und $B_{n,k}$ disjunkt sind, dürfen wir Satz 2.22 anwenden und erhalten

$$\begin{aligned} \mathrm{Prod}(L_{n,k}) &= \mathrm{Prod}(A_k \cup B_{n,k}) \\ &= \mathrm{Prod}(A_k) + \mathrm{Prod}(B_{n,k}) \\ &= k + (n-k) = n. \end{aligned}$$

Es gilt $\cap_{\mathrm{R}}(L_{n,k}) = \{a^{2^i} \mid 1 \leq i \leq k\} = A_k$. Wie zuvor folgt

$$\mathrm{Prod}(\cap_{\mathrm{R}}(L_{n,k})) = \mathrm{Prod}(A_k) = k.$$

Es sei L eine kontextfreie Sprache und $G_L = (N_L, T_L, P_L, S_L)$ eine regel-minimale kontextfreie Grammatik, die L erzeugt. Wir zeigen nun $\mathrm{Symb}(\cap_{\mathrm{R}}(L)) \leq n$. Die Sprache $\cap_{\mathrm{R}}(L) = \cap_{\{\mathrm{a}\}^+}(L)$ enthält nur Wörter, die gänzlich aus dem Buchstaben a bestehen und in L enthalten sind. Aufgrund dieser Tatsache können wir leicht eine kontextfreie Grammatik G konstruieren, welche die Sprache $\cap_{\mathrm{R}}(L)$ erzeugt, indem wir alle Regeln p, die in Ableitungen der Art $S_L \Longrightarrow^*_{G_L} w_a \in \{a\}^+$ angewendet werden, in G einfügen. Die Regeln p bestehen neben dem Pfeil nur aus Symbolen aus $N_L \cup \{a, \lambda\}$. Somit gilt stets

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = \mathrm{Symb}(G) \leq \mathrm{Symb}(G_L) = \mathrm{Symb}(L) = n.$$

□

Lemma 3.40 – Es gibt ein R_m, sodass $0, \ldots, m \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_m}}}(n)$ gilt
Unter den Voraussetzungen $n \in \mathbb{N}$, $n \geq 2$, und $m \in \mathbb{N}$, $m \geq 2$, sowie einem Alphabet T, $a \in T$ und der regulären Sprache $R_m = \{a^{2^i} \mid 1 \leq i \leq m\}$ gelten

i. $0, 1 \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_m}}}(n)$,

ii. $2, \ldots, m \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_m}}}(2)$,

iii. $2, \ldots, m \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_m}}}(n)$ für $n \geq 3$.

Beweis: Es seien T ein Alphabet mit mindestens 2 Buchstaben und paarweise verschiedene $a, b \in T$.

Zu i. Für $k \in \{0, 1\}$ sei

$$L_{n,k} = \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq k \right\} \cup \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq n - k \right\}$$

eine kontextfreie Sprache. Die Sprache $L_{n,k}$ lässt sich mittels

$$A_k = \left\{ a^{2^i} \;\middle|\; 1 \leq i \leq k \right\},$$
$$B_{n,k} = \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq n - k \right\}$$

als $A_k \cup B_{n,k}$ ausdrücken. Nach den Lemmata 3.2 und 3.1 (falls $A_k = \emptyset$) gelten $\mathrm{Prod}(A_k) = k$ und $\mathrm{Prod}(B_{n,k}) = n - k$. Da die Alphabete von A_k und $B_{n,k}$ disjunkt sind, dürfen wir Satz 2.22 anwenden und erhalten

$$\begin{aligned} \mathrm{Prod}(L_{n,k}) &= \mathrm{Prod}(A_k \cup B_{n,k}) \\ &= \mathrm{Prod}(A_k) + \mathrm{Prod}(B_{n,k}) \\ &= k + (n - k) = n. \end{aligned}$$

Es gilt $\cap_{\mathrm{R_m}}(L_{n,k}) = \{a^{2^i} \mid 1 \leq i \leq k\}$. Wegen Lemma 3.2 folgt daraus

$$\mathrm{Prod}(\cap_{\mathrm{R_m}}(L_{n,k})) = \mathrm{Prod}\left(\left\{ a^{2^i} \;\middle|\; 1 \leq i \leq k \right\}\right) = k.$$

Zu ii. Für $2 \leq k \leq m$ sei

$$L_{m,k} = \left\{ a^i \;\middle|\; i \geq 2^{m-k+1} \right\}$$

eine kontextfreie Sprache. Nach Lemma 3.3 gilt

$$\mathrm{Prod}(L_{m,k}) = \mathrm{Prod}\left(\left\{ a^i \;\middle|\; i \geq 2^{m-k+1} \right\}\right) = 2.$$

Es gilt $\cap_{\mathrm{R_m}}(L_{m,k}) = \{a^{2^i} \mid m - k + 1 \leq i \leq m\}$. Lemma 3.2 liefert

$$\mathrm{Prod}(\cap_{\mathrm{R_m}}(L_{m,k})) = \mathrm{Prod}\left(\left\{ a^{2^i} \;\middle|\; m - k + 1 \leq i \leq m \right\}\right) = k.$$

Zu iii. Für $n \geq 3$ und $2 \leq k \leq m$ sei

$$L_{n,k} = \left\{ a^i \;\middle|\; 1 \leq i \leq 2^k \right\} \cup \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq n - 3 \right\}$$

eine kontextfreie Sprache. Die Sprache $L_{n,k}$ lässt sich mittels

$$A_k = \left\{ a^i \;\middle|\; 1 \leq i \leq 2^k \right\},$$
$$B_n = \left\{ b^{2^i} \;\middle|\; 1 \leq i \leq n - 3 \right\}$$

als $A_k \cup B_n$ ausdrücken. Nach Lemma 3.5 gilt $\mathrm{Prod}(A_k) = 3$ und nach Lemma 3.2 sowie 3.1 (falls $B_n = \emptyset$) $\mathrm{Prod}(B_n) = n - 3$. Die Alphabete von A_k und B_n sind disjunkt. Daher dürfen wir Satz 2.22 anwenden und erhalten

$$\mathrm{Prod}\,(L_{n,k}) = \mathrm{Prod}\,(A_k \cup B_n) = \mathrm{Prod}\,(A_k) + \mathrm{Prod}\,(B_n) = 3 + (n-3) = n.$$

Es gilt $\cap_{\mathrm{R_m}}(L_{n,k}) = \{a^{2^i} \mid 1 \le i \le k\}$. Aufgrund von Lemma 3.2 folgt

$$\mathrm{Prod}\,(\cap_{\mathrm{R_m}}(L_{n,k})) = \mathrm{Prod}\left(\left\{a^{2^i} \,\middle|\, 1 \le i \le k\right\}\right) = k.$$

□

Satz 3.41 – Teilbestimmung des Raums der Regelkomplexität unter Schnitt mit regulärer Sprache

Es sei $n \in \mathbb{N}_0$.

i. Für eine beliebige reguläre Sprache R gelten

$$\begin{aligned} \{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(0), \\ \{0,1\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(1), \text{ falls } R \neq \emptyset, \\ \{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(n), \text{ falls } R = \emptyset, \\ \{0,1\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(n) \text{ für } n \geq 1, \text{ falls } |R| = 1. \end{aligned}$$

ii. Es existiert eine reguläre Sprache R, für welche die folgende Aussage korrekt sind

$$\{0,\dots,n\} = \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(n).$$

Es existieren ein $m \in \mathbb{N}_0$ und eine reguläre Sprache R_m, für welche die folgenden Aussagen korrekt sind:

$$\begin{aligned} \{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R_m}}(0), \\ \{0,1\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R_m}}(1), \\ 0,\dots,\max(\{n,m\}) &\in \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R_m}}(n) \text{ für } n \geq 2. \end{aligned}$$

iii. Für die Menge aller Schnitte mit regulärer Sprache Reg *gelten*

$$\begin{aligned} \{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(0), \\ \{0,1\} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(1), \\ \mathbb{N}_0 &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(n) \text{ für } n \geq 2. \end{aligned}$$

Beweis: Anstrich i und ii sind eine Zusammenfassung der Lemmata 3.38 und 3.39 und 3.40. Anstrich iii ergibt sich aus Anstrich i und Anstrich ii. □

Bemerkung: Es seien $n, k \in \mathbb{N}_0$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 3.41 vollständig dargelegt:

- Gegeben sei eine beliebige reguläre Sprache R mit $|R| \geq 2$. Gilt $k \in \mathrm{C}_{\cap_{\mathrm{R}}}^{\mathrm{Prod}}(n)$ für $n \geq 2$?

$\triangle$

4. Symbolkomplexität

Zu Beginn bestimmen wir im Abschnitt 4.1 die Symbolkomplexität einiger Sprachen, die wir in den folgenden Abschnitten verwenden werden. In diesen bestimmen wir (in Teilen) die Räume der Symbolkomplexität unter den Operationen Spiegelbild (in 4.2), Vereinigung (in 4.3), Konkatenation (in 4.4), Kleene-Abschluss (in 4.5), Homomorphismus (in 4.6), inverser Homomorphismus (in 4.7) und Schnitt mit regulärer Sprache (in 4.8) analog wie wir es in Kapitel 3 bezüglich der Regelkomplexität getan haben.

4.1. Allgemeines

Wir geben die Symbolkomplexitäten einiger Sprachen, die wir später benötigen werden, an.

Lemma 4.1 – Kontextfreie Sprachen mit der Symbolkomplexität höchstens 8
Die folgenden Aussagen sind für kontextfreie Sprachen L wahr, wobei $a, b, c, d, e \in T$, $T = \mathrm{Alph}(L)$, nicht notwendig verschieden seien:

i. Es gilt $\mathrm{Symb}(L) = 0$ genau dann, wenn $L = \emptyset$ ist.

ii. Es gilt nie $\mathrm{Symb}(L) = 1$.

iii. Es gilt $\mathrm{Symb}(L) = 2$ genau dann, wenn $L = \{\lambda\}$ ist.

iv. Es gilt $\mathrm{Symb}(L) = 3$ genau dann, wenn $L = \{a\}$ ist.

v. Es gilt $\mathrm{Symb}(L) = 4$ genau dann, wenn $L = \{ab\}$ ist.

vi. Es gilt $\mathrm{Symb}(L) = 5$ genau dann, wenn $L = \{abc\}$ oder $L = \{\lambda, a\}$ ist.

vii. Es gilt $\mathrm{Symb}(L) = 6$ genau dann, wenn $L = \{abcd\}$, $L = \{\lambda, ab\}$, $L = \{a, b\}$, $a \neq b$, oder $L = \{a\}^*$ ist.

viii. Wenn $L = \{a\}^+$, $L = \{ab\}^*$, $L = \{abcde\}$ oder $L = \{a, bc\}$ ist, gilt $\mathrm{Symb}(L) = 7$.

ix. Wenn $L = \{abc\}^*$ ist, gilt $\mathrm{Symb}(L) = 8$.

Beweis: Es sei $G = (N, T, P, S)$ mit $\mathrm{L}(G) = L$ eine symbol-minimale kontextfreie Grammatik.

Zu i. Es sei

$$\mathrm{Symb}(L) = \mathrm{Symb}(G) = 0.$$

Es existiert also keine Regel in G. Damit gibt es kein $v \in T^*$, für das $S \Longrightarrow^* v$ gilt, woraus unmittelbar $L = \mathrm{L}(G) = \emptyset$ folgt. Wir geben nun direkt eine symbol-minimale kontextfreie Grammatik an, welche die Sprache $\emptyset$ erzeugt:

$$(\emptyset, \emptyset, \emptyset, \lambda)\,.$$

(Eine Grammatik mit weniger als 0 Symbolen existiert nicht.)

Zu ii. Es sei $G = (N,\ T,\ P,\ S)$ eine kontextfreie Grammatik. Aus $|P| = 0$ folgt $\mathrm{Symb}(G) = 0$ und aus $|P| \geq 1$ folgt $\mathrm{Symb}(G) \geq 2$ (eine kürzeste Regel hat immer die Form $A \to \lambda,\ A \in N$). Mithin existiert $\mathrm{Symb}(G) = 1$ nicht, und daher auch keine kontextfreie Sprache L mit $\mathrm{Symb}(L) = 1$.

Zu iii. Betrachten wir $\mathrm{Symb}(L) = 2$, so erhalten wir $P = \{S \to \lambda\}$ als einzige Möglichkeit, und damit $L = \{\lambda\}$. Demgegenüber impliziert $L = \{\lambda\}$ direkt $P = \{S \to \lambda\}$, und damit $\mathrm{Symb}(G) = \mathrm{Symb}(L) = 2$. Die Ungleichung $\mathrm{Symb}(L) < 2$ ist wegen der zuvor behandelten Fälle nicht erreichbar.

Zu iv. Aus $\mathrm{Symb}(L) = 3$ folgt $P = \{S \to a\}$ als einzige Option. Mithin steht $L = \{a\}$ fest. Zum anderen kann wegen des Vorherigen für $L = \{a\}$ nicht $\mathrm{Symb}(L) < 3$ gelten.

Zu v. Die Aussage $\mathrm{Symb}(L) = 4$ impliziert $P = \{S \to ab\}$ als erschöpfende Möglichkeit. Es ergibt sich $L = \{ab\}$. Aufgrund des Obigen folgt außerdem, dass für $L = \{ab\}$ die Aussage $\mathrm{Symb}(L) < 4$ nicht wahr ist.

Zu vi. Für $\mathrm{Symb}(L) = 5$ ergeben sich ausschließlich

$$\begin{aligned} P &= \{S \to abc\} \ \text{ oder} \\ P &= \{S \to \lambda, S \to a\}\,. \end{aligned}$$

Andere Kombinationen führen zu keiner symbol-minimalen kontextfreien Grammatik. Somit gilt $L = \{abc\}$ oder $L = \{\lambda, a\}$. Die Sprache $L = \{abc\}$ kann wegen des zuvor Betrachteten nicht mit weniger als 5 Symbolen erzeugt werden, Analoges gilt für $L = \{\lambda,\ a\}$.

Zu vii. Aus $\mathrm{Symb}(L) = 6$ folgen ausschließlich

$$\begin{aligned} P &= \{S \to abcd\}\,, \\ P &= \{S \to \lambda, S \to ab\}\,, \\ P &= \{S \to a, S \to b\}\,, a \neq b, \ \text{ oder} \\ P &= \{S \to aS, S \to \lambda\} \ \text{ (ohne Beschränkung der Allgemeinheit).} \end{aligned}$$

Weitere Kombinationen führen zu keiner symbol-minimalen kontextfreien Grammatik. Die möglichen erzeugbaren Sprachen sind demnach $L = \{abcd\}$, $L = \{\lambda, ab\}$, $L = \{a, b\}$ oder $L = \{a\}^*$. Da diese Sprachen in keinem der zuvor behandelten Fälle auftreten, sind sie nicht mit weniger als 6 Symbolen erzeugbar.

Zu viii. Da in den zuvor gezeigten Fällen $L = \{a\}^+$ nicht vorkommt, muss $\mathrm{Symb}(L) \geq 7$ korrekt sein. Für

$$P = \{S \to aS, S \to a\}$$

erzeugt G die Sprache L mithilfe von 7 Symbolen, also $\mathrm{Symb}(L) \leq 7$. Aus den beiden

Relationen folgt $\text{Symb}(L) = 7$. Analoges gilt für $L = \{ab\}^*$, $L = \{abcde\}$ und $L = \{a, bc\}$ – die entsprechenden Regelmengen sind

$$P = \{S \to abS, S \to \lambda\}\,,$$
$$P = \{S \to abcde\} \text{ bzw.}$$
$$P = \{S \to a, S \to bc\}\,.$$

Zu ix. Setzen wir

$$P = \{S \to abcS, S \to \lambda\}\,,$$

erzeugt G die Sprache L mithilfe von 8 Symbolen, also $\text{Symb}(L) \leq 8$. Wir zeigen nachstehend $\text{Symb}(L) \geq 8$. Es sei $G' = (N', T', P', S')$ mit $T' = \{a, b, c\}$ und $\mathrm{L}(G') = L$ eine symbol-minimale kontextfreie Grammatik, für die $\text{Symb}(G') \leq 7$ gilt. Aus der zuvor erfolgten Betrachtung der Sprachen und Symbolkomplexitäten wissen wir $\text{Symb}(G') \geq 7$ – es folgt $\text{Symb}(G') = 7$. Aufgrund von $\lambda \in L$ muss es eine Regel $A \to \lambda$ für $A \in N'$ geben. Des Weiteren muss G nach Lemma 2.16 eine Schleifenableitung $B \Longrightarrow^* xBy$ für $B \in N'$ und $x, y \in {T'}^*$, $xy \neq \lambda$, erzeugen. Nach Lemma 2.23 existieren in P' mindestens 2 Regeln, die aus insgesamt 5 Symbolen der Art Pfeil und Nichtterminal bestehen. Es kommt mindestens der Buchstabe a und damit 1 weiteres Symbol hinzu. Wir nehmen nun an, G' enthält mindestens 3 Regeln. Eine 3. Regel besteht dann aus mindestens 2 Symbolen. Zusammen erhalten wir mindestens 8 Symbole, was $\text{Symb}(G') = 7$ widerspricht. Mithin enthält G' genau 2 Regeln. Da die Schleifenableitung $B \Longrightarrow^* xBy$ und λ mittels genau 2 Regeln und genau 7 Symbolen zu erzeugen sind, bleiben nur $A \to \lambda$ und $B \to z$ für $z \in (T' \cup N')^*$, $|z| = 3$ und $|z|_B \geq 1$ übrig.

- $A = B \neq S'$: Nicht möglich, da es in G keine Regel mit S' auf der linken Seite gäbe, und so nur $\emptyset$ erzeugt würde.
- $A = S' \neq B$: Nicht möglich, da $S' \to \lambda$ die einzige Regel mit S auf der linken Seite wäre, und damit nur die Sprache $\{\lambda\}$ generiert würde.
- $A \neq B = S'$: Nicht möglich, da jedes Vorkommen von A auf der rechten Seite einer Regel durch λ ersetzt werden könnte, und somit die nutzlose Regel $A \to \lambda$ getilgt werden könnte. Folglich würden weniger Symbole benötigt, und die geforderte Minimalität würde verletzt werden.
- $A = B = S'$: Für $d, e \in T'$ ergeben sich die Fälle:
  - $S' \to \lambda, S' \to S'S'S'$: Nicht möglich, da $\mathrm{L}(G') = \{\lambda\}$ folgte.
  - $S' \to \lambda, S' \to dS'S'$ (ohne Beschränkung der Allgemeinheit): Nicht möglich, da $d \in \mathrm{L}(G')$ folgte.
  - $S' \to \lambda, S' \to deS'$ (ohne Beschränkung der Allgemeinheit): Nicht möglich, da $de \in \mathrm{L}(G')$ folgte.

Es sind alle Möglichkeiten ausgeschöpft, L durch G' mittels höchstens 7 Symbolen zu erzeugen. Mithin gilt $\text{Symb}(L) \geq 8$. Zusammen mit der obigen Aussage $\text{Symb}(L) \leq 8$ erhalten wir schließlich $\text{Symb}(L) = 8$. □

Lemma 4.2 – Die Symbolkomplexität einer kontextfreien Sprachen L ist mindestens $|\mathrm{Alph}(L)| + 2$
Für jede kontextfreie Sprache $L \neq \emptyset$ gilt $\mathrm{Symb}(L) \geq |\mathrm{Alph}(L)| + 2$, und die Regelmenge einer jeden symbol-minimalen kontextfreien Grammatik, die L erzeugt, enthält mindestens alle Buchstaben aus $\mathrm{Alph}(L)$ sowie ein Nichtterminal und einen Pfeil.

Beweis: Es sei $G = (N, T, P, S)$ mit $\mathrm{L}(G) = L$ und $\mathrm{Symb}(G) = \mathrm{Symb}(L)$. Jeder Buchstabe aus $\mathrm{Alph}(L)$ kommt in L mindestens einmal vor. Daher muss jeder Buchstabe aus $\mathrm{Alph}(L)$ auch mindestens einmal auf den rechten Seiten der Regeln aus P vorkommen. Mithin gilt $\mathrm{Symb}(G) \geq |\mathrm{Alph}(L)|$. Des Weiteren muss es in P mindestens eine Regel geben, wodurch 2 zusätzliche Symbole vorhanden sind: ein Nichtterminal und ein Pfeil. Insgesamt erhalten wir $\mathrm{Symb}(L) = \mathrm{Symb}(G) \geq |\mathrm{Alph}(L)| + 2$. □

Lemma 4.3 – Symbolkomplexität von $\{a^{\pi_n}\}$
Es seien T ein Alphabet und $a \in T$. Dann existiert für jedes $n \in \mathbb{N}^{\not\ni 1}$ ein $\pi_n \in \mathbb{N}_0$, sodass folgende Aussagen gelten:

i. $\mathrm{Symb}(\{a^{\pi_n}\}) = n$,

ii. für jedes $m \in \mathbb{N}$ und $m \geq n + 1$, existiert ein π_m, sodass $\mathrm{Symb}(\{a^{\pi_n}\}) = n$, $\mathrm{Symb}(\{a^{\pi_m}\}) = m$ und $\pi_m \geq \pi_n + 1$ gelten,

iii. jede symbol-minimale kontextfreie Grammatik, die $\{a^{\pi_n}\}$ erzeugt, generiert keine Schleifenableitungen.

Beweis: Zu i. Ein Beweis befindet sich in [Gru72] (Beweis zu Theorem 1) in kurzer Form. Wir geben im Folgenden einen darauf aufbauenden, ausführlichen Beweis an, mit dem wir später auch noch die Gültigkeit des Anstrichs ii zeigen können.

Wir beweisen zunächst die Aussage

$$\mathrm{Symb}\left(\left\{a^{\pi_n+1}\right\}\right) \leq \mathrm{Symb}(\{a^{\pi_n}\}) + 1. \tag{1.1}$$

Es sei dazu $G_n = (N_n, \{a\}, P_n, S)$ eine symbol-minimale kontextfrei Grammatik mit $\mathrm{L}(G_n) = \{a^{\pi_n}\}$. Nach Lemma 2.14 ist G_n reduziert. Da G_n genau ein Wort erzeugt, existiert für ein gewisses $p_n \in \mathbb{N}$ eine Ableitung der Form

$$S \Longrightarrow w_{1,n} \Longrightarrow \cdots \Longrightarrow w_{p_n,n} = a^{\pi_n} \text{ für } w_{i,n} \in (N_n \cup \{a\})^*, 1 \leq i \leq p_n.$$

Somit existiert die Regel $S \to w_{1,n} \in P_n$. Wegen Lemma 2.16 erzeugt G_n keine Schleifenableitung, woraus wir folgern, dass in $w_{1,n}$ kein S vorkommt. Wir konstruieren nun die kontextfreie Grammatik

$$G'_n = (N_n, \{a\}, P'_n, S)$$

mit

$$P'_n = (P_n \setminus \{S \to w_{1,n}\}) \cup \{S \to w_{1,n}a\}.$$

Die Grammatik G'_n generiert offenbar nur ein Wort, und es existiert die Ableitung

$$S \Longrightarrow w_{1,n}a \Longrightarrow \cdots \Longrightarrow w_{p_n,n}a = a^{\pi_n}a = a^{\pi_n+1},$$

weswegen $\mathrm{L}(G'_n) = \{a^{\pi_n+1}\}$ gilt. Sie besteht aus genau einem Symbol mehr als G_n, woraus

$$\mathrm{Symb}\left(\left\{a^{\pi_n+1}\right\}\right) \leq \mathrm{Symb}(\mathrm{L}(G'_n)) = \mathrm{Symb}(G_n) + 1 = \mathrm{Symb}(\{a^{\pi_n}\}) + 1$$

folgt (G'_n ist nicht zwingend symbol-minimal).

Mittels vollständiger Induktion werden wir zeigen, dass zu jedem n ein π_n so existiert, dass $\mathrm{Symb}(\{a^{\pi_n}\}) = n$ gilt.

Für $n = 2$ erhalten wir nach Lemma 4.1 als einzige Möglichkeit $2 = \mathrm{Symb}(\{\lambda\}) = \mathrm{Symb}(\{a^0\})$, und wegen $2 = \mathrm{Symb}(\{a^{\pi_2}\}) = \mathrm{Symb}(\{a^0\})$ die Aussage $\pi_2 = 0$.

Es gelte nun $\mathrm{Symb}(\{a^{\pi_k}\}) = k$, $k \in \mathbb{N}^{\not\ni 1}$. Wir legen die Menge

$$\mathscr{G}_{\leq k} = \left\{G \,\middle|\, G \in \mathrm{CF}, \mathrm{L}(G) = \left\{a^i\right\}, i \in \mathbb{N}_0, \mathrm{Symb}(G) \leq k\right\}$$

fest, die alle kontextfreien Grammatiken enthält, die aus höchstens k Symbolen bestehen und eine Sprache der Form $\{a^i\}$, $i \in \mathbb{N}_0$, erzeugen (eine solche Grammatik gibt es stets). Offenbar existieren nur endlich viele kontextfreie Grammatiken, welche aus höchstens k Symbolen bestehen. Mithin ist $\mathscr{G}_{\leq k}$ endlich. Daher ist

$$i_{\max,k} = \max\left(\left\{i \,\middle|\, G \in \mathscr{G}_{\leq k}, \mathrm{L}(G) = \left\{a^i\right\}, i \in \mathbb{N}_0\right\}\right)$$

bestimmbar. Es gilt $\mathrm{Symb}(\{a^{i_{\max,k}}\}) \leq k$, da es eine Grammatik in $\mathscr{G}_{\leq k}$ gibt, die $\{a^{i_{\max,k}}\}$ generiert. Durch Umformung entsteht die äquivalente Aussage

$$\mathrm{Symb}\left(\left\{a^{i_{\max,k}}\right\}\right) + 1 \leq k + 1. \tag{1.2}$$

Aufgrund der Wahl von $i_{\max,k}$ als größter Potenz existiert keine Grammatik in $\mathscr{G}_{\leq k}$, die $\{a^{i_{\max,k}+1}\}$ erzeugt. Folglich gilt

$$\mathrm{Symb}\left(\left\{a^{i_{\max,k}+1}\right\}\right) \geq k + 1. \tag{1.3}$$

Wir substituieren π_n in der Ungleichung (1.1) mit $i_{\max,k}$ und erhalten

$$\mathrm{Symb}\left(\left\{a^{i_{\max,k}+1}\right\}\right) \leq \mathrm{Symb}\left(\left\{a^{i_{\max,k}}\right\}\right) + 1. \tag{1.4}$$

Aus den Ungleichungen (1.2), (1.3) und (1.4) folgt

$$k + 1 \leq \mathrm{Symb}\left(\left\{a^{i_{\max,k}+1}\right\}\right) \leq \mathrm{Symb}\left(\left\{a^{i_{\max,k}}\right\}\right) + 1 \leq k + 1,$$

woraus wir $\mathrm{Symb}(\{a^{i_{\max,k}+1}\}) = k + 1$ ablesen. Somit haben wir zu

$$\mathrm{Symb}(\{a^{\pi_k}\}) = k$$

den Nachfolger
$$\mathrm{Symb}(\{a^{\pi_{k+1}}\}) = k+1 \text{ für } \pi_{k+1} = i_{\max,k}+1$$
gefunden.

Zu ii. Aus dem Beweis zu Anstrich i lesen wir ab, dass $i_{\max,k} \geq \pi_k$ gilt, da $i_{\max,k}$ die größte Potenz ist und eine Grammatik in $\mathscr{G}_{\leq k}$ existiert, die $\{a^{\pi_k}\}$ erzeugt ($\mathrm{Symb}(\{a^{\pi_k}\}) = k$). Hieraus folgt sicher $i_{\max,k}+1 \geq \pi_k+1$. Dem Beweis zu Anstrich i folgend, ergibt sich

$$\mathrm{Symb}(\{a^{\pi_{k+1}}\}) = k+1 \text{ für } \pi_{k+1} = i_{\max,k}+1.$$

Zusammen ergeben sich

$$\begin{aligned} \mathrm{Symb}(\{a^{\pi_k}\}) &= k, \\ \mathrm{Symb}(\{a^{\pi_{k+1}}\}) &= k+1, \\ \pi_{k+1} &\geq \pi_k + 1. \end{aligned}$$

Für $m = n+1$ ersetzen wir k durch n und erhalten aus Obigem

$$\pi_m = \pi_{n+1} \geq \pi_n + 1.$$

Für $m \geq n+2$ folgt aus

$$\pi_m \geq \pi_{m-1} + 1 \geq \cdots \geq \pi_n + 1$$

die Aussage $\pi_m \geq \pi_n + 1$.

Zu iii. Da $\{a^{\pi_n}\}$ eine endliche Sprache ist, folgt die Aussage direkt aus Lemma 2.16. □

Bemerkung: Aufgrund von Lemma 4.1 wissen wir bereits

- $\pi_2 = 0$, folglich $\{a^{\pi_2}\} = \{a^0\} = \{\lambda\}$,
- $\pi_3 = 1$, folglich $\{a^{\pi_3}\} = \{a^1\}$,
- $\pi_4 = 2$, folglich $\{a^{\pi_4}\} = \{a^2\}$,
- $\pi_5 = 3$, folglich $\{a^{\pi_5}\} = \{a^3\}$,
- $\pi_6 = 4$, folglich $\{a^{\pi_6}\} = \{a^4\}$.

△

Lemma 4.4 – Symbolkomplexität von $\{w\} \cup \{\lambda\}$
Für ein Alphabet T und $w \in T^+$ ist $\mathrm{Symb}(\{w\} \cup \{\lambda\}) = \mathrm{Symb}(\{w\}) + 2$ wahr.

Beweis: Es sei $L = \{w\} \cup \{\lambda\}$. Weiterhin sei $G = (N, T, P, S)$ eine symbol-minimale kontextfreie Grammatik, die L generiert. Nach Lemma 2.14 ist G reduziert. Um $\lambda \in L$ zu erzeugen, muss eine Regel $A \to \lambda \in P$, $A \in N$, existieren. Wegen Lemma 2.17 muss es neben $A \Longrightarrow^* \lambda$ noch eine Ableitung $A \Longrightarrow^* w'$, $w' \in T^+$, geben, und aufgrund der Reduziertheit von G besteht eine Ableitung $S \Longrightarrow^* xAy$, $x, y \in T^*$. Wäre $xy \neq \lambda$ möglich, gäbe es die Ableitungen $S \Longrightarrow^* xy \in T^+$ und $S \Longrightarrow^* xw'y \in T^+$, mithin zwei unterschiedliche von λ abweichende Terminalwörter. Die Sprache L enthält aber nur ein Wort, dass nicht λ ist – ein Widerspruch. Daher ist ausschließlich $S \Longrightarrow^* A$ für A möglich und korrekt. Es ergeben sich insgesamt $S \Longrightarrow^* A \Longrightarrow^* w' \in L$ und $S \Longrightarrow^* A \Longrightarrow^* \lambda \in L$, ergo $w' = w$. Dies und die von G verlangte Minimalität lassen nur noch $A = S$ zu – ansonsten gäbe es „überflüssige“ Regeln mit S auf der linken Seite.

Da aufgrund von Lemma 2.16, Anstrich i in G keine Schleifenableitung der Form $S \Longrightarrow^* xSy$ für $x, y \in T^*$, $xy \neq \lambda$ existiert, wird die Regel $S \to \lambda$ während der Generierung von w nicht verwendet. Das Streichen von $S \to \lambda$ in G lieferte folglich eine symbol-minimale kontextfreie Grammatik G', die $L \setminus \{\lambda\} = \{w\}$ erzeugte. Infolgedessen erhalten wir genau $\mathrm{Symb}(G') + 2 = \mathrm{Symb}(\{w\}) + 2$ Symbole in G. □

Lemma 4.5 – Initialschleifenableitung für $\{a\}^+$
Für ein Alphabet T, $a \in T$ und $L = \{a\}^+$ generiert jede symbol-minimale kontextfreie Grammatik, die L erzeugt, eine Initialschleifenableitung.

Beweis: Es sei $G = (N, \{a\}, P, S)$ eine kontextfreie Grammatik mit $\mathrm{L}(G) = L$ und $\mathrm{Symb}(G) = \mathrm{Symb}(L)$. Nach Lemma 4.1 gilt $\mathrm{Symb}(L) = 7 = \mathrm{Symb}(G)$. Die Grammatik G enthält höchstens 2 Regeln, da bereits ab 3 Regeln, aufgrund der Beschränkung von 7-Symbolen, auf den rechten Seiten der Regeln insgesamt nur noch 1 Symbol vorkommen dürfte, und L damit sicher nicht erzeugt werden könnte.

Nach Lemma 2.16 erzeugt G eine Schleifenableitung der Form $A \Longrightarrow^* xAy$ für $A \in N$ und $x, y \in \{a\}^*$, $xy \neq \lambda$. Wir können diese mit den höchstens 2 Regeln von G nur erzeugen, wenn $A = S$ ist – durch Aufstellen aller möglichen Grammatiken mit 7 Symbolen und höchstens 2 Regeln sieht man, dass kein von S verschiedenes Nichtterminal existieren kann. □

Lemma 4.6 – Erweiterung des Satzes 2.22 im Fall von $\{a\}^+$
Es seien T ein Alphabet, $a \in T$ und $L_1 = \{a\}^+$ sowie L_2 eine kontextfreie Sprache, für die $a \notin \mathrm{Alph}(L_2)$ und $\lambda \notin L_2$ gelten. Außerdem sei $G_2 = (N_2, T_2, P_2, S_2)$ eine symbol-minimale kontextfreie Grammatik mit $\mathrm{L}(G_2) = L_2$ und $\mathrm{Symb}(G_2) = \mathrm{Symb}(L_2)$, die keine Initialschleifenableitung erzeugt.

Unter diesen Voraussetzungen gibt es eine symbol-minimale kontextfreie Grammatik G mit $\mathrm{L}(G) = L_1 \cup L_2$:

- $$G = (\{S\}, \{a\}, \{S \to aS, S \to a\}, S),$$
  $$\mathrm{Symb}(G) = \mathrm{Symb}(G_1) = 7,$$

  falls $\mathrm{L}(G_2) = \emptyset$ ist.

- $$G = ((\{S, A\} \cup N_2) \setminus S_2, \{a\} \cup T_2, P, S) \text{ mit}$$

$$P = \{S \to A, A \to aA, A \to a\} \cup \{S \to w \mid S_2 \to w \in Q_2\} \cup (P_2 \setminus Q_2),$$
$$\mathrm{Symb}(G) = \mathrm{Symb}(G_2) + 10$$

für $Q_2 = \{S_2 \to w \mid S_2 \to w \in P_2\}$ und $S, A \notin N_2$ (ohne Beschränkung der Allgemeinheit), falls $\mathrm{L}(G_2) = \emptyset$ nicht gilt. In diesem Fall erzeugt G keine Initialschleifenableitung.

Beweis: Nach Lemma 4.1 gelten $\mathrm{Symb}(\emptyset) = 0$ und $\mathrm{Symb}(L_1) = 7$. Im Fall von $L_2 = \emptyset$ folgt $L_1 \cup \emptyset = L_1$, und damit $\mathrm{Symb}(G) = \mathrm{Symb}(L_1) = 7$. Die Konstruktion von G entnehmen wir dem Beweis zu Lemma 4.1.

Es sei $L_2 \neq \emptyset$. Wir konstruieren eine kontextfreie Grammatik $G' = (N', \{a\} \cup T_2, P', S')$, die $L_1 \cup L_2$ mit genau $\mathrm{Symb}(L_1 \cup L_2)$ Symbolen erzeugt. Wir verwenden die Erkenntnis aus dem Beweis zu Satz 2.22, dass wir dazu eine symbol-minimale kontextfreie Grammatik $G'_1 = (N'_1, \{a\}, P'_1, S'_1)$ für L_1 finden müssen, die keine Initialschleifenableitung erzeugt (ohne Beschränkung der Allgemeinheit gelte $N'_1 \cap N_2 = \emptyset$). Offenbar muss es nach Lemma 2.16 eine Schleifenableitung $A \Longrightarrow^*_{G'_1} xAy$ für $A \in N'_1$ und $x, y \in \{a\}^*$, $xy \neq \lambda$, in G'_1 geben. Darüber hinaus muss die Schleifenableitung $A \Longrightarrow^*_{G'_1} xAy$ mit einer Ableitung $A \Longrightarrow^*_{G'_1} w$ für $w \in \{a\}^*$ beendet werden können (Begründung analog zum Beweis von Lemma 2.14). Im Übrigen gilt $A \neq S'_1$. Eine Schleifenableitung $A \Longrightarrow^*_{G'_1} xAy$ muss für ein $n \in \mathbb{N}$ mittels der Regeln

$$A_1 \to u_1 A_2 v_1, A_2 \to u_2 A_3 v_2, \ldots, A_{n-1} \to u_{n-1} A_n v_{n-1}, A_n \to u_n A_1 v_n \tag{1.5}$$

für $A = A_1$, $A_i \in N'_1$, $A_i \neq A_j$, sowie $u_i, v_i \in (N'_1 \cup \{a\})^*$, wobei $1 \leq i \leq n$, $1 \leq j \leq n$ und $i \neq j$ gelten, realisiert werden. Weiterhin benötigen wir Regeln für $A \Longrightarrow^*_{G'_1} w$. Dies können nicht nur Regeln aus (1.5) sein, da nach jedem Ableitungsschritt stets ein Nichtterminal im abgeleiteten Wort stünde – es gibt also mindestens eine weitere. Offenbar sind dann die Regeln $A \to uAv$ für $uv = a$ sowie $A \to \lambda$, die bezüglich der Anzahl der Symbole kleinsten Regeln, die eine Schleifenableitung $A \Longrightarrow^*_{G'_1} xAy$ und eine Ableitung $A \Longrightarrow^*_{G'_1} w$ ermöglichen. Wir benötigen zusätzlich eine Regel, welche die notwendige Ableitung $S'_1 \Longrightarrow^*_{G'_1} u'Av'$ für $u', v' \in (N'_1 \cup \{a\})^*$ ermöglicht. Die hinsichtlich der Anzahl der Symbole kleinste ist dann $S'_1 \to A$.

Die Regeln $S'_1 \to A, A \to uAv$ $(uv = a)$ und $A \to \lambda$ erzeugen die Sprache $L_1 = \{a\}^+$ nicht. Da L_1 mit nicht weniger als in den zuvor genannten Regeln verwendeten Symbolen generiert werden kann, müssen wir davon ausgehen, dass ein weiteres Symbol benötigt wird. Tatsächlich erzeugt die kontextfreie Grammatik

$$G'_1 = (\{S'_1, A\}, \{a\}, \{S'_1 \to A, A \to aA, A \to a\}, S'_1)$$

die Sprache L_1 mit einem weiteren Symbol. Wir erhalten $\mathrm{Symb}(G'_1) = 10$.

An dieser Stelle können wir den Beweis zum Satz 2.22 (Fall iii) folgen und konstruieren die kontextfreie Grammatik

$$G' = ((N'_1 \cup N_2) \setminus S_2, \{a\} \cup T_2, P', S'_1)$$

mit

$$\begin{aligned} P' &= \{S_1' \to A, A \to aA, A \to a\} \cup \{S_1' \to w \mid S_2 \to w \in Q_2\} \cup (P_2 \setminus Q_2)\,, \\ Q_2 &= \{S_2 \to w \mid S_2 \to w \in P_2\}\,, \end{aligned}$$

die $L_1 \cup L_2$ symbol-minimal erzeugt. Wir zählen $\mathrm{Symb}(G') = \mathrm{Symb}(G_2) + 10$. □

Lemma 4.7 – Symbolkomplexität von $\{a_1, \dots, a_n\}$
Es seien $n \in \mathbb{N}$, T_n ein Alphabet mit mindestens n Buchstaben und $a_i \in T_n$, $1 \leq i \leq n$, paarweise verschieden. Dann gilt für die kontextfreie Sprache $L_n = \{a_i \,|\, 1 \leq i \leq n\}$ die Aussage $\mathrm{Symb}(L_n) = 3n$.

Beweis: Wir wenden Lemma 2.10 an, indem wir das dort definierte X als $X = \{a_i \,|\, 1 \leq i \leq n\} = L_n$ feststellen. Mithin existiert in jeder kontextfreien Grammatik, die L_n erzeugt, mindestens eine separate Regel für jedes a_i, $1 \leq i \leq n$, mit mindestens einem a_i auf der rechten Seite – diese Regeln bestehen folglich jeweils aus mindestens 3 Symbolen. Insgesamt erhalten wir $\mathrm{Symb}(L_n) \geq 3n$.

Die kontextfreie Grammatik

$$G_n = (\{S\}, \{a_i \mid 1 \leq i \leq n\}, P_n, S)$$

mit

$$P_n = \{S \to a_i \mid 1 \leq i \leq n\}$$

erzeugt die Sprache L_n mit $\mathrm{Symb}(G_n) = 3n$ Symbolen. Mithin gilt $\mathrm{Symb}(L_n) \leq 3n$, und zusammen mit dem Obigen erhalten wir $\mathrm{Symb}(L_n) = 3n$. □

Lemma 4.8 – Symbolkomplexität von $\{a_1, \dots, a_n\}^*$
Es seien $n \in \mathbb{N}$, T_n ein Alphabet mit mindestens n Buchstaben und $a_i \in T_n$, $1 \leq i \leq n$, paarweise verschieden. Dann gilt für die kontextfreie Sprache $L_n = \{a_i \,|\, 1 \leq i \leq n\}^*$ die Aussage

$$\mathrm{Symb}(L_n) = \begin{cases} 4n + 2 & \text{für } n \leq 3 \\ 3n + 6 & \text{sonst.} \end{cases}$$

Beweis: Fall 1: $n \leq 3$. Die kontextfreie Grammatik

$$G_n' = (\{S\}, \{a_i \mid 1 \leq i \leq n\}, P_n', S)$$

mit

$$P_n' = \{S \to a_i S \mid 1 \leq i \leq n\} \cup \{S \to \lambda\}$$

erzeugt die Sprache L_n mit $\mathrm{Symb}(G_n') = 4n + 2$ Symbolen. Mithin gilt $\mathrm{Symb}(L_n) \leq 4n + 2$.

Es sei nun $G_n = (N_n,\ T_n,\ P_n,\ S)$ mit $T_n = \{a_i \,|\, 1 \leq i \leq n\}$ eine (reduzierte) kontextfreie Grammatik, für die $\mathrm{L}(G_n) = L_n$ und $\mathrm{Symb}(G_n) = \mathrm{Symb}(L_n)$ gilt. Für jedes a_i, $1 \leq i \leq n$, muss es eine Ableitung $S \Longrightarrow^* a_i \in L_n$ in G_n geben. Das erzwingt die Existenz einer Regel $A_i \to w_i \in P_n$ mit $A_i \in N_n$, $w_i \in (\{a_i\} \cup N_n)^*$ und $|w_i|_{a_i} = 1$. Es gibt also für jedes a_i, $1 \leq i \leq n$, eine separate Regel mit a_i auf der rechten Seite (Lemma 2.10). Des Weiteren existiert die Ableitung $S \Longrightarrow^* \lambda$. Folglich muss es eine Regel $A \to \lambda$, $A \in N_n$, geben. Insgesamt erhalten wir mindestens $3n + 2$ Symbole ($|w_i| \geq 1$).

Offenbar ist für jedes i, $1 \leq i \leq n$, die Sprache $\{a_i\}^*$ eine Teilsprache von L_n. Daher muss es eine Schleifenableitung $B_i \Longrightarrow^* x_i B_i y_i$ für $B_i \in N_n$ und $x_i, y_i \in \{a_i\}^*$, $x_i y_i \neq \lambda$, geben. Für $|w_i| = 1$ ermöglichen die zuvor ermittelten Regeln $A_i \to w_i$ – genauer $A_i \to a_i$ – und $A \to \lambda$ solche Schleifenableitungen nicht allein. Wegen $\mathrm{Symb}(L_n) \leq 4n+2$ und $\mathrm{Symb}(G_n) \geq 3n + 2$ kann es höchstens n weitere Symbole in G_n geben, die zur Realisierung der Schleifenableitungen beitragen. Eine zusätzliche Regel $C \to w_C$ für $C \in N_n$ und $w_C \in (N_n \cup T_n)^*$ wäre der Beschränkung $|w_c| \leq n - 2$ unterlegen ($n \leq 3$, daher könnte es nicht mehr als eine zusätzliche Regel geben). Die rechte Seite w_C dieser Regel müsste aber mindestens die Länge 2 haben, damit eine Schleifenableitung entsteht. Folglich kann es keine zusätzliche Regel geben. Somit verbleibt nur die Möglichkeit $|w_i| \geq 2$. Dies muss für jedes i, $1 \leq i \leq n$, gelten, da jede Teilsprache $\{a_i\}^*$ erzeugt werden muss und dort ausschließlich der Buchstabe a_i vorkommt. Mithin erhalten wir $\mathrm{Symb}(G_n) \geq 4n+2$ und mit obiger Aussage $\mathrm{Symb}(L_n) \leq 4n+2$ insgesamt $\mathrm{Symb}(L_n) = 4n + 2$.

Fall 2: $n \geq 4$. Die kontextfreie Grammatik

$$G'_n = (\{S\}, \{a_i \mid 1 \leq i \leq n\}, P'_n, S)$$

mit

$$\begin{aligned} P'_n = \{&S \to a_i \mid 1 \leq i \leq n\} \cup \\ \{&S \to SS, S \to \lambda\} \end{aligned}$$

erzeugt die Sprache L_n mit $\mathrm{Symb}(G'_n) = 3n + 6$ Symbolen. Mithin gilt $\mathrm{Symb}(L_n) \leq 3n + 6$.

Es sei $G_n = (N_n, T_n, P_n, S)$ mit $T_n = \{a_i \mid 1 \leq i \leq n\}$ eine (reduzierte) kontextfreie Grammatik, für die $\mathrm{L}(G_n) = L_n$ und $\mathrm{Symb}(G_n) = \mathrm{Symb}(L_n)$ gilt. Wie im zuvor betrachteten Fall $n \leq 3$ begründen wir die Existenz der Regeln $A_i \to w_i \in P_n$ mit $A_i \in N_n$, $w_i \in (\{a_i\} \cup N_n)^*$ und $|w_i|_{a_i} = 1$ für $1 \leq i \leq n$ und die Existenz der Regel $A \to \lambda$, $A \in N_n$. Insgesamt erhalten wir folgerichtig ebenso mindestens $3n + 2$ Symbole ($|w_i| \geq 1$).

Wie im vorherigen Fall ist für jedes i, $1 \leq i \leq n$, die Sprache $\{a_i\}^*$ eine Teilsprache von L_n. Daher muss es eine Schleifenableitung $B_i \Longrightarrow^* x_i B_i y_i$ für $B_i \in N_n$ und $x_i, y_i \in \{a_i\}^*$, $x_i y_i \neq \lambda$, geben. Für $|w_i| = 1$ ermöglichen die zuvor ermittelten Regeln $A_i \to w_i$ – genauer $A_i \to a_i$ – und $A \to \lambda$ solche Schleifenableitungen nicht allein. Im Gegensatz zum obigen Fall muss hier für $n \geq 5$ eine zusätzliche Regel existieren, da höchstens $(3n+6)-(3n+2) = 4$ Symbole hinzukommen und für $|w_i| \geq 2$ mindestens n Symbole statt nur 4 Symbole hinzukämen. Eine zusätzliche Regel, welche die Schleifenableitungen realisieren, muss hier aus mindestens 4 Symbolen bestehen. Für $n = 4$ existiert entweder eine zusätzliche Regel oder es werden 4 weitere Symbole in den bereits ermittelten Regeln verwendet (wie oben) – die Symbolanzahl ist gleich. Mithin erhalten wir $\mathrm{Symb}(G_n) \geq 3n+2+4 = 3n+6$

und mit obiger Aussage $\mathrm{Symb}(L_n) \leq 3n + 6$ insgesamt $\mathrm{Symb}(L_n) = 3n + 6$. □

Lemma 4.9 – Symbolkomplexität von $\{a_1 \cdots a_n\}$
Für $n \in \mathbb{N}$, ein Alphabet T_n mit mindestens n Buchstaben und paarweise verschiedene $a_i \in T_n$, $1 \leq i \leq n$, sowie $L_n = \{a_1 \cdots a_n\}$ gilt $\mathrm{Symb}(L_n) = n + 2$.

Beweis: Aufgrund von Lemma 4.2 folgt $\mathrm{Symb}(L_n) \geq |\mathrm{Alph}(L_n)| + 2 = n + 2$.
Die kontextfreie Grammatik

$$(\{S\}, \{a_1, \ldots, a_n\}, \{S \to a_1 \cdots a_n\}, S)$$

erzeugt L_n mit $n + 2$ Symbolen. Mit Obigem folgt $\mathrm{Symb}(L_n) = n + 2$. □

Lemma 4.10 – Symbolkomplexität von $\{a_1 \cdots a_n\}^*$
Für $n \in \mathbb{N}$, ein Alphabet T_n mit mindestens n Buchstaben und paarweise verschiedene $a_i \in T_n$, $1 \leq i \leq n$, sowie $L_n = \{a_1 \cdots a_n\}^*$ gilt $\mathrm{Symb}(L_n) = n + 5$.

Beweis: Es sei $G_n = (N_n, T_n, P_n, S)$ eine symbol-minimale kontextfreie Grammatik mit $\mathrm{L}(G_n) = L_n$. Da L_n eine unendliche Sprache ist, enthält G_n nach Lemma 2.23 mindestens 5 Symbole der Art Pfeil und Nichtterminal. Offenbar müssen die n Buchstaben $a_1, \ldots, a_n$ in den Regeln von G_n vorkommen. Zusammen folgt $\mathrm{Symb}(L_n) = \mathrm{Symb}(G_n) \geq n + 5$.
Die kontextfreie Grammatik

$$(\{S\}, \{a_1, \ldots, a_n\}, \{S \to a_1 \cdots a_n S, S \to \lambda\}, S)$$

generiert L_n mit $n + 5$ Symbolen. Mittels der obigen Aussage erhalten wir daher $\mathrm{Symb}(L_n) = n + 5$. □

Lemma 4.11 – Symbolkomplexität von $\{a_1 \cdots a_n\}^*$ ohne Initialschleifenableitung
Es seien $n \in \mathbb{N}$, T_n ein Alphabet mit mindestens n Buchstaben und $a_i \in T_n$, $1 \leq i \leq n$, paarweise verschieden sowie $L_n = \{a_1 \cdots a_n\}^*$. Für eine kontextfreie Grammatik G_n, die L_n mit möglichst wenigen Symbolen erzeugt und keine Initialschleifenableitung generiert, gilt $\mathrm{Symb}(G_n) = n + 8$. Außerdem enthält G_n nicht die Regel $S \to \lambda$.

Beweis: Es sei $G_n = (N_n, T_n, P_n, S)$ mit $\mathrm{L}(G_n) = L_n$ eine symbol-minimale kontextfreie Grammatik ohne Initialschleifenableitung. Es kann analog zum Beweis von Lemma 2.14 gezeigt werden, dass G_n reduziert ist. Offenbar müssen die n Buchstaben $a_1, \ldots, a_n$ in den Regeln von G_n enthalten sein (Lemma 4.2). Es folgt $\mathrm{Symb}(G_n) \geq n$.

Nach Voraussetzung und Lemma 2.16 muss es in G_n eine Schleifenableitung der Form $A \Longrightarrow^* xAy$ für ein $A \in N_n \setminus \{S\}$ und $x, y \in {T_n}^*$, $xy \neq \lambda$, und eine Ableitung der Form $A \Longrightarrow^* w$, $w \in {T_n}^*$ geben. Nach Lemma 2.23 können diese Ableitungen nur mithilfe von mindestens 2 Regeln, die aus mindestens 5 Symbolen der Art Pfeil und Nichtterminal bestehen, erzeugt werden. Außerdem darf bei diesen Regeln kein S auf der linken oder rechten Seite stehen. Zusammen mit der obigen Aussage erhalten wir daher $\mathrm{Symb}(G_n) \geq n + 5$.

Wegen der Reduziertheit von G_n existiert ferner eine Ableitung $S \Longrightarrow^* uAv$ für $u, v \in T_n{}^*$. Dies impliziert eine Regel $S \to u'A'v'$ für $u', v' \in ((N_n \setminus \{S\}) \cup T_n)^*$ und $A' \in N_n \setminus \{S\}$. Eine der kürzesten Regeln hat dann die Form $S \to A'$ und besteht daher aus 3 Symbolen. Gemeinsam mit der vorherigen Betrachtung erhalten wir $\mathrm{Symb}(G_n) \geq n + 5 + 3 = n + 8$.

Wir konstruieren nun die kontextfreie Grammatik

$$G'_n = (\{S, A\}, \{a_1, \ldots, a_n\}, \{S \to A, A \to a_1 \cdots a_n A, A \to \lambda\}, S),$$

die L_n mit $\mathrm{Symb}(G'_n) = n + 8$ Symbolen erzeugt. Dies impliziert zusammen mit der vorherigen Aussage, dass G'_n die Sprache L_n ohne S-Schleifenableitungen mit minimaler Symbolanzahl generiert.

Es gibt keine symbol-minimale kontextfreie Grammatik $G''_n = (N''_n, T''_n, P''_n, S)$, die L_n ohne S-Schleifenableitungen erzeugt und eine Regel $S \to \lambda$ enthält. Dies begründen wir mit folgender Tatsache: Die Grammatik G'_n wurde so konstruiert, dass sie nur solche Regeln enthält, die im ersten Teil des Beweises verlangt wurden – nämlich 2 Regeln, bei denen kein S auf der linken oder rechten Seite vorkommen darf und eine Regel der Form $S \to u'A'v'$ für $u', v' \in ((N_n \setminus \{S\}) \cup T_n)^*$. □

Lemma 4.12 – Symbolkomplexität von $\{a^m c_1 \cdots c_n b^m \mid m \geq 0\}$
Für $n \in \mathbb{N}$, ein Alphabet T_n mit mindestens $n+2$ Buchstaben und paarweise verschiedene $a, b, c_i \in T_n$, $1 \leq i \leq n$, sowie $L_n = \{a^m c_1 \cdots c_n b^m \mid m \geq 0\}$ gilt $\mathrm{Symb}(L_n) = n + 7$.

Beweis: Es sei $G_n = (N_n, T_n, P_n, S)$ mit $T_n = \{a, b, c_1, \ldots, c_n\}$ eine symbol-minimale kontextfreie Grammatik, die L_n erzeugt. Nach Lemma 4.2 müssen alle Buchstaben aus $\mathrm{Alph}(L_n) = T_n$ in P_n vorkommen. Somit enthält P_n mindestens $n+2$ Symbole. Da L_n eine unendliche Sprache ist, existieren nach Lemma 2.23 in P_n weiterhin mindestens 5 Symbole der Art Pfeil und Nichtterminal. Zusammen gilt folglich $\mathrm{Symb}(L_n) = \mathrm{Symb}(G_n) \geq n+7$.

Die kontextfreie Grammatik

$$(\{S\}, T_n, \{S \to aSb, S \to c_1 \cdots c_n\}, S)$$

generiert L_n mittels $n + 7$ Symbolen – also $\mathrm{Symb}(L_n) \leq n + 7$. Mit Obigem folgt $\mathrm{Symb}(L_n) = n + 7$. □

Lemma 4.13 – Symbolkomplexität von $\{a^m c_1 \cdots c_n b^m \mid m \geq 0\}^*$
Für $n \in \mathbb{N}$, ein Alphabet T_n mit mindestens $n+2$ Buchstaben und paarweise verschiedene $a, b, c_i \in T_n$, $1 \leq i \leq n$, sowie $L_n = \{a^m c_1 \cdots c_n b^m \mid m \geq 0\}$ gilt $\mathrm{Symb}(L_n{}^*) = n + 13$.

Beweis: Die kontextfreie Grammatik

$$G'_n = (\{S, S'\}, \{a, b\} \cup \{c_1, \ldots, c_n\}, P'_n, S)$$

mit

$$P'_n = \{S \to SS', S \to \lambda, S' \to aS'b, S' \to c_1 \cdots c_n\}$$

erzeugt ${L_n}^*$ mittels $n + 13$ Symbole.

Es sei $G_n = (N_n, T_n, P_n, S)$, $T_n = \{a, b\} \cup \{c_1, \ldots, c_n\}$, mit $\mathrm{L}\,(G_n) = {L_n}^*$ eine symbolminimale kontextfreie Grammatik. Lemma 2.14 begründet die Reduziertheit von G_n. Aufgrund von G'_n gilt $\mathrm{Symb}\,(G_n) \leq \mathrm{Symb}\,(G'_n) = n + 13$. Nach Lemma 4.2 kommen alle Buchstaben aus T_n in Regeln von P_n vor, woraus $\mathrm{Symb}\,(G_n) \geq n + 2$ folgt.

Die Existenz der Ableitung $S \Longrightarrow^* \lambda \in {L_n}^*$ bedingt eine Regel $B \to \lambda \in P_n$ für $B \in N_n$.

Aus $S \Longrightarrow^* c_1 \cdots c_n \in {L_n}^*$ folgt das Vorhandensein mindestens einer Regel $C \to w_C \in P_n$ für ein $C \in N_n$ und $w_C \in (N_n \cup \{c_1, \ldots, c_n\})^+$, wobei für mindestens ein i, $1 \leq i \leq n$, die Aussage $|w_C|_{c_i} \geq 1$ gilt – es kommen also weder a noch b vor.

Aufgrund von $S \Longrightarrow^* ac_1 \cdots c_n b \in {L_n}^*$ muss es eine Regel geben, in der ein Buchstabe a vorkommt. In den Regeln $B \to \lambda$ und $C \to w_C$ kommt kein a vor. Folglich existiert eine weitere Regel $A \to w_A \in P_n$ für $A \in N_n$ und $w_A \in (N_n \cup T_n)^+$ mit der Bedingung $|w_A|_a \geq 1$. Außerdem folgt aus dem nur einmaligen Vorkommen von a in $S \Longrightarrow^* ac_1 \cdots c_n b$ die Tatsache $|w_A|_a \leq 1$, zusammen also $|w_A|_a = 1$. Insgesamt enthält P_n somit mindestens 3 Regeln.

Wir werden nun zeigen, dass G_n, enthielte sie nur 3 Regeln, ${L_n}^*$ nicht erzeugte. Angenommen, in G_n gibt es genau 3 Regeln. Die Ableitung $S \Longrightarrow^* c_1 \cdots c_n \in {L_n}^*$ impliziert $|w_C|_{c_i} = 1$. Da w_A ein a enthält, kann die Regel $A \to w_A$ nicht bei $S \Longrightarrow^* c_1 \cdots c_n$ angewendet worden sein. Deshalb bleibt als einzige Möglichkeit für C nur S übrig. Analog leiten wir $B = S$, also $S \to \lambda$, her. Wenn $A \neq S$ und $|w_C|_A \geq 1$ gälten, wäre $S \Longrightarrow^* c_1 \cdots c_n$ nicht realisierbar, da stets auch ein a erzeugt werden würde. Gälten $A \neq S$ und $|w_C|_A = 0$, fände die Regel $A \to w_A$ nie Anwendung. Wir schlussfolgern hieraus $A = S$, also $S \to w_A$. Für alle $m \in \mathbb{N}_0$ existiert die Ableitung $S \Longrightarrow^* a^m c_1 \cdots c_n b^m$. Die Buchstaben a^m können nicht mithilfe von $S \to w_C$ erzeugt werden, da sonst $|w_C|_S \geq 1$ gelten müsste und mit jeder erneuten Anwendung von $S \to w_C$ ein weiteres c_i zur Ableitung hinzukäme. Daher bleibt nur $|w_A|_S \geq 1$ übrig. Das Terminalwort w'_A in der Ableitung $S \Longrightarrow w_A \Longrightarrow^* w'_A \in {T_n}^*$ enthält, wenn $C \to w_C = S \to w_C$ nicht angewendet wird, sondern nur $S \to \lambda$, kein c_i, und ist somit kein Wort von ${L_n}^*$. Folglich gilt $|w_A|_S = 0$. Dies widerspricht dem zuvor Festgestellten $|w_A|_S \geq 1$. Somit müssen wir die obige Annahme, dass G_n nur 3 Regeln enthält, fallen lassen – es gilt $|P_n| \geq 4$.

Angenommen, es gilt $|P_n| \geq 5$. Dann enthält P_n mindestens 5 Pfeile und mindestens 5 Nichtterminale auf den linken Seiten von Regeln. Außerdem kommen die $n + 2$ Buchstaben aus T_n hinzu. Da ${L_n}^*$ eine unendliche Sprache ist, muss G_n nach Lemma 2.16 eine Schleifenableitung generieren. Mit den bereits gezählten Symbolen in P_n ist dies nicht möglich, weil es dann kein Nichtterminal auf der rechten Seite einer Regel gäbe, was aber eine Voraussetzung zur Erzeugung einer Schleifenableitung ist. Somit kommt mindestens 1 weiteres Nichtterminal dazu. Insgesamt erhalten wir $5 + 5 + (n + 2) + 1 = n + 13$ Symbole. Da die Grammatik G'_n von oben ${L_n}^*$ mittels 4 Regeln sowie $n + 13$ Symbolen erzeugt, wir zuvor $|P_n| \geq 4$ festgestellt haben und $|P_n| \geq 5$ die Existenz von mindestens $n + 13$ Symbolen in P_n impliziert, dürfen wir $|P_n| = 4$ schließen.

Insgesamt bestehen die 4 Regeln in P_n aus mindestens folgenden Symbolen:

- 4 Nichtterminale von den linken Seiten der Regeln und 4 Pfeile (8 Symbole),
- die Buchstaben $c_1, \ldots, c_n$ sowie a und b ($n + 2$ Symbole),

- 2 Nichtterminale auf der rechten Seite einer Regel und 1 weiteres Nichtterminal auf der rechten Seite einer der anderen Regeln $A \to w_A$, $B \to \lambda$ oder $C \to w_C$, damit die beiden verschiedenen Schleifenableitungen $X \Longrightarrow^* c_1 \cdots c_n X$, $X \in N_n$, (ohne Beschränkung der Allgemeinheit) und $Y \Longrightarrow^* aYb$, $Y \in N_n$, (a und b können nicht anders mit gleicher Anzahl generiert werden) gewährleistet werden können (3 Symbole).

Somit gilt $\mathrm{Symb}(G_n) \geq n + 13$, und mit der zum Anfang festgestellten Eigenschaft $\mathrm{Symb}(G_n) \leq n + 13$ folgt das Resultat $\mathrm{Symb}(G_n) = n + 13$. □

Lemma 4.14 – Symbolkomplexität von $\{a\} \cup \{a^{\pi_n}\}$
Für $n \in \mathbb{N}$, $n \geq 4$, ein Alphabet T, $a \in T$ und $L_n = \{a\} \cup \{a^{\pi_n}\}$ gilt $\mathrm{Symb}(L_n) = n + 3$.

Beweis: Es sei $G_n = (N_n, \{a\}, P_n, S)$ eine symbol-minimale kontextfreie Grammatik mit $\mathrm{L}(G_n) = L_n$. Nach Lemma 2.14 ist G_n reduziert. Wir betrachten die möglichen Ableitungen für $a \in L_n$ in G_n, also

$$S \Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r \Longrightarrow^* a \tag{1.6}$$

für $r \in \mathbb{N}_0$ und $u, u_i \in \{a\}^*$ sowie $A_i \in N_n$, $A_i \neq S$, für $1 \leq i \leq r$. Die Bedingung $A_i \neq S$ gilt, da L_n eine endliche Sprache ist und somit keine Schleifenableitungen existieren (Lemma 2.16).

Es sei nun $r \geq 2$. Ohne Beschränkung der Allgemeinheit ergeben sich nur die Möglichkeiten

$$u = a, u_i = \lambda, A_i \Longrightarrow^* \lambda \tag{1.7}$$

und

$$u = u_i = \lambda, A_1 \Longrightarrow^* a, A_j \Longrightarrow^* \lambda \text{ für } j \in \mathbb{N}, 2 \leq j \leq r, \tag{1.8}$$

um die Ableitung (1.6) zu realisieren. Im Fall (1.7) gelten nach Lemma 2.17 auch $A_1 \Longrightarrow^* a^{k_1}$ und $A_2 \Longrightarrow^* a^{k_2}$ für $k_1, k_2 \in \mathbb{N}$. Daher können wir die folgenden Wörter ableiten:

$$\begin{aligned}
S &\Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r = aA_1 \cdots A_r \Longrightarrow^* a\lambda\lambda,\\
S &\Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r = aA_1 \cdots A_r \Longrightarrow^* aa^{k_1}\lambda,\\
S &\Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r = aA_1 \cdots A_r \Longrightarrow^* aa^{k_1}a^{k_2}.
\end{aligned}$$

Wir können demnach 3 unterschiedliche Wörter erzeugen. Die Sprache L_n enthält allerdings nur 2 Wörter – ein Widerspruch. Für den Fall (1.8) ergibt sich in ähnlicher Weise ein Widerspruch. Nach Lemma 2.17 existieren weiterhin die Ableitungen $A_1 \Longrightarrow^* a^{k_1}$ für $k_1 \in \mathbb{N}_0^{\neq 1}$ (es gilt $a^{k_1} \neq a$) und $A_2 \Longrightarrow^* a^{k_2}$ für $k_2 \in \mathbb{N}$. Somit können wir mindestens 3 unterschiedliche Wörter ableiten:

$$\begin{aligned}
S &\Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r = A_1 \cdots A_r \Longrightarrow^* a\lambda,\\
S &\Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r = A_1 \cdots A_r \Longrightarrow^* aa^{k_2},\\
S &\Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r = A_1 \cdots A_r \Longrightarrow^* a^{k_1}a^{k_2},
\end{aligned}$$

$$S \Longrightarrow uA_1u_1 \cdots u_{r-1}A_ru_r = A_1 \cdots A_r \Longrightarrow^* a^{k_1}.$$

Insgesamt ist daher der Fall $r \geq 2$ auszuschließen.

Wir prüfen den Fall $r = 1$. Ohne Beschränkung der Allgemeinheit ergeben sich nur

$$u = a, u_1 = \lambda, A_1 \Longrightarrow^* \lambda \tag{1.9}$$

und

$$u = u_1 = \lambda, A_1 \Longrightarrow^* a, \tag{1.10}$$

um die Ableitung (1.6) zu realisieren. Für den Fall (1.9) gilt nach Lemma 2.17 die Aussage $A_1 \Longrightarrow^* a^k$ für $k \in \mathbb{N}$. Da a^{π_n} neben a das einzige weitere Wort in L_n ist, folgt aus $S \Longrightarrow aA_1 \Longrightarrow^* aa^k = a^{\pi_n}$ der Wert $k = \pi_n - 1$. Für die Ableitung $S \Longrightarrow aA_1$ kommt als kürzeste Regel nur $S \to aA_1$ infrage – und wie oben gezeigt, gilt $A_1 \neq S$. Da es in L_n genau 2 Wörter gibt und $A_1 \Longrightarrow^* \lambda$ sowie $A_1 \Longrightarrow^* a^{\pi_n - 1}$ gelten, ergibt sich $\mathrm{L}(G_n, A_1) = \{\lambda, a^{\pi_n-1}\}$. Nach Lemma 4.4 stimmt die Aussage

$$\mathrm{Symb}(\mathrm{L}(G_n, A_1)) = \mathrm{Symb}\left(\left\{a^{\pi_n-1}\right\}\right) + 2.$$

Wir bestimmen als Nächstes eine untere Schranke für $\mathrm{Symb}(\{a^{\pi_n-1}\})$. Dazu verwenden wir die Aussage (1.1) aus dem Beweis von Lemma 4.3:

$$\mathrm{Symb}(\{a^{\pi_n}\}) \leq \mathrm{Symb}\left(\left\{a^{\pi_n-1}\right\}\right) + 1.$$

Nach Lemma 4.3 führt dies zu

$$\mathrm{Symb}\left(\left\{a^{\pi_n-1}\right\}\right) \geq \mathrm{Symb}(\{a^{\pi_n}\}) - 1 = n - 1.$$

Es folgt

$$\mathrm{Symb}(\mathrm{L}(G_n, A_1)) = \mathrm{Symb}\left(\left\{a^{\pi_n-1}\right\}\right) + 2 \geq (n-1) + 2 = n + 1.$$

Es sei nun $G_{A_1,n} = (N_n, \{a\}, P_{A_1,n}, A_1)$ eine symbol-minimale kontextfreie Grammatik, die $\mathrm{L}(G_n,\ A_1)$ erzeugt. Nach Vorherigem ist $\mathrm{Symb}(G_{A_1,n}) \geq n + 1$. Wir zeigen nun folgende Aussage: Die Regelmenge P_n von G_n muss mindestens so viele Regeln enthalten wie $P_{A_1,n}$. Wäre das nicht so, könnten wir eine kontextfreie Grammatik mit Regeln aus P_n zur Generierung von $\mathrm{L}(G_n,\ A_1)$ konstruieren, indem wir nur die Regeln wählten, die bei den Ableitungen $A_1 \Longrightarrow^* \lambda$ und $A_1 \Longrightarrow^* a^{\pi_n-1}$ verwendet werden und sicher nur $\mathrm{L}(G_n,\ A_1) = \{\lambda,\ a^{\pi_n-1}\}$ erzeugen. Diese Grammatik bestünde dann aus weniger als $n + 1$ Symbolen – ein Widerspruch zur angenommenen Minimalität. Damit gilt $\mathrm{Symb}(G_n) \geq n + 1$. Da die oben festgestellte Regel $S \to aA_1$ nicht in den Ableitungen $A_1 \Longrightarrow^* \lambda$ und $A_1 \Longrightarrow^* a^{\pi_n-1}$ benutzt wird, kommen noch 4 Symbole dazu: $\mathrm{Symb}(G_n) \geq (n + 1) + 4 = n + 5$. Für den Fall (1.10) gilt neben $A_1 \Longrightarrow^* a$ nach Lemma 2.17 die Aussage $A_1 \Longrightarrow^* a^j$ für $j \in \mathbb{N}_0^{\neq 1}$. Da L_n nur die beiden Wörter a und a^{π_n} enthält, folgen sofort $S \Longrightarrow A_1 \Longrightarrow^* a$ und $S \Longrightarrow A_1 \Longrightarrow^* a^j = a^{\pi_n}$. Damit ist A_1 ein kontextfreies Nichtterminal – dies widerspricht Lemma 2.19. Mithin tritt der Fall (1.10) nicht ein.

Es sei $r = 0$. Dann ist $u = a$, und es muss die Regel $S \to a$ in P_n geben. Da es keine Schleifenableitungen in G_n gibt, kommt S auf keiner rechten Seite einer Regel aus P_n vor. Daher findet die Regel $S \to a$ bei der Ableitung von a^{π_n} keine Anwendung. Somit können wir aus G_n eine kontextfreie Grammatik G'_n konstruieren, die alle Regeln enthält, die für die Erzeugung von a^{π_n} notwendig sind. Die Grammatik G'_n erzeugt dann a^{π_n} mit mindestens $\mathrm{Symb}(a^{\pi_n}) = n$ Symbolen. Mithin gilt für G_n die Aussage $\mathrm{Symb}(L_n) = \mathrm{Symb}(G_n) \geq n + 3$. Folglich ist G_n minimal, wenn $r = 0$ gilt.

Es sei $G'_n = (N'_n,\ \{a\},\ P'_n,\ S)$ eine symbol-minimale kontextfreie Grammatik mit $\mathrm{L}(G'_n) = \{a^{\pi_n}\}$. Nach Lemma 4.3 ist $\mathrm{Symb}(G'_n) = n$. Die Grammatik G'_n enthält keine Initialschleifenableitungen (Lemma 2.16). Unter Zuhilfenahme von G'_n erzeugt die kontextfreie Grammatik

$$(N'_n, \{a\}, P'_n \cup \{S \to a\}, S)$$

die Sprache L_n mit $\mathrm{Symb}(G'_n) + 3 = n + 3$ Symbolen zum Schluss minimal. □

Lemma 4.15 – Symbolkomplexität von $\{ab^{\pi_n}\}$ und $\{acb^{\pi_n}\}$

Es seien T ein Alphabet, $a, b, c \in T$ paarweise verschieden und $n \in \mathbb{N}^{\not\ni 1}$.

i. Für $L_n = \{ab^{\pi_n}\}$ gilt $\mathrm{Symb}(L_n) = n + 1$.

ii. Für $L_n = \{acb^{\pi_n}\}$ gilt $\mathrm{Symb}(L_n) = n + 2$.

Beweis: Zu i. Es seien $L'_n = \{b^{\pi_n}\}$ sowie $G'_n = (N'_n,\ T',\ P'_n,\ S')$, $T' = \{b\}$, mit $\mathrm{L}(G'_n) = L'_n$ und $\mathrm{Symb}(G'_n) = \mathrm{Symb}(L'_n)$. Nach Lemma 2.14 ist G'_n reduziert. Es existiert eine Regel $S' \to w \in P'_n$ für $w \in (N'_n \cup T')^*$. Weiterhin gilt $S' \Longrightarrow w \Longrightarrow^* b^{\pi_n}$, da b^{π_n} das einzige Wort in L'_n ist. Außerdem existieren aufgrund von Lemma 2.16, Anstrich i keine Schleifenableitungen in G'. Änderten wir daher $S' \to w$ zu $S' \to aw$, erzeugten wir ausschließlich das Wort ab^{π_n}, also L_n. Mithin gilt

$$\mathrm{Symb}(L_n) \leq \mathrm{Symb}(L'_n) + 1 = n + 1$$

(die Addition von 1 rührt vom zusätzlichen a her).

Es sei $G_n = (N_n, T, P_n, S)$, $T = \{a, b\}$, mit $\mathrm{L}(G_n) = L_n$ und $\mathrm{Symb}(G_n) = \mathrm{Symb}(L_n)$. Aufgrund von Lemma 2.14 ist G_n reduziert. Da L_n außerdem nur ein Wort beinhaltet, kommt a genau einmal in genau einer Regel aus P_n vor – anderenfalls hätte das einzige Wort in L_n mehr als nur ein Vorkommen von a oder es gäbe zwei Ableitungen, die zu ab^{π_n} führten. Würden wir das a aus der entsprechenden Regel in P_n streichen, erzeugten wir die Sprache $L'_n = \{b^{\pi_n}\}$. Wie wir aus Lemma 4.3 wissen, ergäbe sich $\mathrm{Symb}(L'_n) = n$. Es muss ferner $\mathrm{Symb}(L'_n) \leq \mathrm{Symb}(L_n) - 1$ (die Subtraktion von 1 rührt vom gestrichenen a her), also

$$\mathrm{Symb}(L_n) \geq \mathrm{Symb}(L'_n) + 1 = n + 1$$

gelten. Mit dem vorhergehenden Ergebnis – $\mathrm{Symb}(L_n) \leq n + 1$ – resultiert $\mathrm{Symb}(L_n) = n + 1$.

Zu ii. Der Beweis kann analog zum Beweis von Anstrich i geführt werden. □

Lemma 4.16 – Raum der Symbolkomplexität für Symbolkomplexität 1
Es seien $n \in \mathbb{N}$ und $x_i \in \mathbb{N}_0$, $1 \leq i \leq n$. Für eine n-stellige Operation $\circ_n$, unter der die kontextfreien Sprachen abgeschlossen sind, gelten

- $\emptyset = \mathrm{C}^{\mathrm{Symb}}_{\circ_n}(x_1, \ldots, x_n)$, wenn für mindestens ein $x \in \{x_1, \ldots, x_n\}$ die Aussage $x = 1$ stimmt,
- $1 \notin \mathrm{C}^{\mathrm{Symb}}_{\circ_n}(x_1, \ldots, x_n)$.

Beweis: Die Aussage von Lemma 4.1 Anstrich ii, dass $\mathrm{Symb}(L) = 1$ für eine kontextfreie Sprache L nie gilt, führt zu den beiden zu beweisenden Aussagen. □

4.2. Spiegelbild

Wir untersuchen den Raum der Symbolkomplexität unter Spiegelbild.

Satz 4.17 – Raum der Symbolkomplexität unter Spiegelbild
Für $n \in \mathbb{N}_0$ gelten

$$\begin{aligned}\{0\} &= \mathrm{C}^{\mathrm{Symb}}_{()^{\mathrm{R}}}(0),\\ \emptyset &= \mathrm{C}^{\mathrm{Symb}}_{()^{\mathrm{R}}}(1),\\ \{n\} &= \mathrm{C}^{\mathrm{Symb}}_{()^{\mathrm{R}}}(n) \text{ für } n \geq 2.\end{aligned}$$

Beweis: Die Aussage $\emptyset = \mathrm{C}^{\mathrm{Symb}}_{()^{\mathrm{R}}}(1)$ entnehmen wir Lemma 4.16. Ansonsten kann der Beweis analog zu Satz 3.13 geführt werden. □

4.3. Vereinigung

Wir untersuchen den Raum der Symbolkomplexität unter Vereinigung.

Lemma 4.18 – Kommutativität von $\mathrm{C}^{\mathrm{Symb}}_{\cup}$
Für $n, m \in \mathbb{N}_0$ gilt

$$\mathrm{C}^{\mathrm{Symb}}_{\cup}(n, m) = \mathrm{C}^{\mathrm{Symb}}_{\cup}(m, n).$$

Beweis: Die Fälle $n = 1$ oder $m = 1$ sind in Lemma 4.16 behandelt worden. Für die übrigen Fälle kann analog zum Beweis von Lemma 3.14 verfahren werden. □

Lemma 4.19 – Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Vereinigung
Unter den Voraussetzungen $n, m \in \mathbb{N}_0^{\neq 1}$ gelten

i. $0 \notin \mathrm{C}^{\mathrm{Symb}}_{\cup}(n, m)$ für $n \geq 2$,

ii. $\{n\} = \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, 0)$,

iii. $n \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, n)$,

iv. $n + m + 4, \ldots \notin \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 2$ und $2 \leq m \leq 5$

v. $n + m + 7, \ldots \notin \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$.

Beweis: Es seien L und J zwei kontextfreie Sprachen, für die $\mathrm{Symb}(L) = n$, $n \in \mathbb{N}_0^{\not\ni 1}$, und $\mathrm{Symb}(J) = m$, $m \in \mathbb{N}_0^{\not\ni 1}$, gelten.

Zu i. Es sei $\mathrm{Symb}(L \cup J) = 0$. Dann muss nach Lemma 4.1 $L \cup J = \emptyset$ gelten. Dies impliziert $L = \emptyset$ und $J = \emptyset$. Ebenfalls wegen Lemma 4.1 folgen sofort $\mathrm{Symb}(L) = 0$ und $\mathrm{Symb}(J) = 0$, und damit $n = m = 0$.

Zu ii. Es sei $\mathrm{Symb}(J) = m = 0$. Aufgrund von Lemma 4.1 gilt $J = \emptyset$. Infolgedessen erhalten wir $L \cup J = L \cup \emptyset = L$, und damit $\mathrm{Symb}(L \cup J) = \mathrm{Symb}(L) = n$.

Zu iii. Wir setzen $L = J$ und erhalten $\mathrm{Symb}(L) = \mathrm{Symb}(J) = n$. Es gilt $L \cup J = L = J$, und damit $\mathrm{Symb}(L \cup J) = n$.

Zu iv. Es seien $n \geq 2$ und $2 \leq m \leq 5$. Weiterhin seien J_m eine kontextfreie Sprache sowie $G = (N, T, P, S)$ mit $\mathrm{L}(G) = L$, $\mathrm{Symb}(G) = \mathrm{Symb}(L) = n$ und $G_m = (N_m, T_m, P_m, S')$ mit $\mathrm{L}(G_m) = J_m$, $\mathrm{Symb}(G_m) = \mathrm{Symb}(J_m) = m$, kontextfreie Grammatiken, für die wir ohne Beschränkung der Allgemeinheit $S'' \notin N$ und $S'' \notin N_m$ voraussetzen. Für gewisse $a, b, c \in T_m$ gelten nach Lemma 4.1 und dessen Beweis

$$\begin{aligned}
P_2 &= \{S' \to \lambda\}, \\
P_3 &= \{S' \to a\}, \\
P_4 &= \{S' \to ab\}, \\
P_5 &= \{S' \to abc\} \text{ oder } P_5 = \{S' \to \lambda, S' \to a\}.
\end{aligned}$$

Wir können nun stets die kontextfreie Grammatik

$$(N \cup \{S''\}, T \cup T_m, P''_m, S'')$$

mit

$$P''_m = P \cup \{S'' \to S\} \cup \{S'' \to w \mid S' \to w \in P_m, w \in {T_m}^*\}$$

angeben, die $L \cup J_m$ mit genau $\mathrm{Symb}(G) + \mathrm{Symb}(G_m) + 3$ Symbolen erzeugt. Mithin erhalten wir $\mathrm{Symb}(L \cup J_m) \leq \mathrm{Symb}(G) + \mathrm{Symb}(G_m) + 3 = n + m + 3$.

Zu v. Wählen wir die Grammatiken G mit $\mathrm{L}(G) = L$ und G' mit $\mathrm{L}(G') = J$ aus dem Beweis des Lemmas 2.8 so, dass $\mathrm{Symb}(G) = \mathrm{Symb}(L) = n$ und $\mathrm{Symb}(G') = \mathrm{Symb}(J) = m$ gelten, dann zählen wir in $G_{\cup}$ mit $\mathrm{L}(G_{\cup}) = L \cup J$ höchstens $n + m + 6$ Symbole (Beweis zu Lemma 2.8). Mithin gilt $\mathrm{Symb}(L \cup J) \leq \mathrm{Symb}(G_{\cup}) = n + m + 6$. □

Lemma 4.20 – Spezielle Ergebnisse zum Raum der Symbolkomplexität unter Vereinigung

Unter den Voraussetzungen $n, m \in \mathbb{N}_0^{\not\ni 1}$ gelten

i.
$$\{2\} = \mathrm{C}_{\cup}^{\mathrm{Symb}}(2,2),$$

ii.
$$\begin{aligned} \{5\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(3,2), \\ \{3,6\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(3,3), \end{aligned}$$

iii.
$$\begin{aligned} &\{6\} = \mathrm{C}_{\cup}^{\mathrm{Symb}}(4,2), \\ &\{7\} = \mathrm{C}_{\cup}^{\mathrm{Symb}}(4,3), \\ &4,8 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(4,4) \not\ni 2,3,5,6, \end{aligned}$$

iv.
$$\begin{aligned} \{5,7\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,2), \\ 5,8 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,3) \not\ni 2,3,4,6, \\ 9 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,4) \not\ni 2,\dots,6, \\ 5,10 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,5) \not\ni 2,3,4,6, \end{aligned}$$

v.
$$6 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,m) \text{ für } 2 \le m \le 6,$$

vi.
$$\begin{aligned} 6 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(7,2), \\ 7, m+10 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(7,m) \text{ für } 3 \le m \le 7. \end{aligned}$$

Beweis: Es seien L und J zwei kontextfreie Sprachen sowie T ein beliebiges Alphabet und nicht notwendig verschiedene Buchstaben $a,b,c,d \in T$. Weiterhin seien $n,m \in \mathbb{N}^{\not\ni 1}$.

Zu i. Es sei $\mathrm{Symb}(L) = \mathrm{Symb}(J) = 2$. Dann gilt nach Lemma 4.1 ausschließlich $L = J = \{\lambda\}$. Es folgt $L \cup J = \{\lambda\} \cup \{\lambda\} = \{\lambda\}$, und nach Lemma 4.1 $\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{\lambda\}) = 2$.

Zu ii. Es seien $\mathrm{Symb}(L) = 3$ und $\mathrm{Symb}(J) = 2$. Aufgrund von Lemma 4.1 gelten ausschließlich $L = \{a\}$ und $J = \{\lambda\}$. Wir erhalten $L \cup J = \{a\} \cup \{\lambda\} = \{\lambda, a\}$. Lemma 4.1 liefert $\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{\lambda, a\}) = 5$.

Für den Fall $\mathrm{Symb}(L) = 3$ und $\mathrm{Symb}(J) = 3$ gelten nach Lemma 4.1 ausschließlich $L = \{a\}$ und $J = \{b\}$. Ist $a = b$, folgt $L \cup J = \{a\} \cup \{a\} = \{a\}$, und damit wegen Lemma 4.1 $\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{a\}) = 3$. Ist $a \neq b$, folgt $L \cup J = \{a\} \cup \{b\} = \{a,b\}$, und aufgrund von Satz 2.22 und Lemma 4.1

$$\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{a\} \cup \{b\}) = \mathrm{Symb}(\{a\}) + \mathrm{Symb}(\{b\}) = 3 + 3 = 6.$$

Zu iii. Es seien $\mathrm{Symb}(L) = 4$ und $\mathrm{Symb}(J) = 2$. Dann gelten nach Lemma 4.1 ausschließlich $L = \{ab\}$ und $J = \{\lambda\}$. Es folgt $L \cup J = \{ab\} \cup \{\lambda\} = \{\lambda, ab\}$, und nach Lemma 4.4 sowie Lemma 4.1

$$\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{ab\} \cup \{\lambda\}) = \mathrm{Symb}(\{ab\}) + \mathrm{Symb}(\{\lambda\}) = 4 + 2 = 6.$$

Es seien $\mathrm{Symb}(L) = 4$ und $\mathrm{Symb}(J) = 3$. Aufgrund von Lemma 4.1 gelten ausschließlich $L = \{ab\}$ und $J = \{c\}$. Wir erhalten $L \cup J = \{ab\} \cup \{c\} = \{ab, c\}$. Nach Lemma 4.1 gilt dann $\mathrm{Symb}(L \cup J) = 7$.

Für den Fall $\mathrm{Symb}(L) = 4$ und $\mathrm{Symb}(J) = 4$ gelten nach Lemma 4.1 ausschließlich $L = \{ab\}$ und $J = \{cd\}$. Ist $ab = cd$, folgt $L \cup J = \{ab\} \cup \{ab\} = \{ab\}$, und damit wegen Lemma 4.1 $\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{ab\}) = 4$. Haben ab und cd *keine* gemeinsamen Buchstaben, folgt $L \cup J = \{ab\} \cup \{cd\} = \{ab, cd\}$, und aufgrund von Satz 2.22 und Lemma 4.1

$$\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{ab\} \cup \{cd\}) = \mathrm{Symb}(\{ab\}) + \mathrm{Symb}(\{cd\}) = 4 + 4 = 8.$$

Gilt $ab \neq cd$, folgt $L \cup J = \{ab, cd\}$. Da keine kontextfreie Sprache aus Lemma 4.1 diese Form hat, ergibt sich $\mathrm{Symb}(L \cup J) \geq 7$.

Zu iv. Der Beweis kann analog zu den Beweisen der zuvor betrachteten Aussagen erfolgen.

Zu v. Es seien $L = \{a\}^*$ und $J_m = \{a^{\pi_m}\}$, $2 \leq m \leq 6$. Nach den Lemmata 4.1 und 4.3 gelten $\mathrm{Symb}(L) = 6$ und $\mathrm{Symb}(J_m) = m$. Vereinigen wir die beiden Sprachen, erhalten wir $L \cup J_m = L$. Somit gilt $\mathrm{Symb}(L \cup J_m) = \mathrm{Symb}(L) = 6$.

Zu vi. Nach Lemma 4.1 gilt $\mathrm{Symb}(L) = 7$ für $L = \{a\}^+$. Dasselbe Lemma impliziert $J = \{\lambda\}$, wenn $\mathrm{Symb}(J) = 2$ gelten soll. Es ergibt sich

$$L \cup J = \{a\}^+ \cup \{\lambda\} = \{a\}^*,$$

und nach Lemma 4.1

$$\mathrm{Symb}(L \cup J) = \mathrm{Symb}(\{a\}^*) = 6.$$

Wir setzen $J_m = \{a^{\pi_m}\}$ für $2 \leq m \leq 7$. Nach Lemma 4.3 gilt $\mathrm{Symb}(J_m) = m$. Wir erhalten

$$L \cup J_m = \{a\}^+ \cup \{a^{\pi_m}\} = \{a\}^+ = L,$$

und damit

$$\mathrm{Symb}(L \cup J_m) = \mathrm{Symb}(L) = 7.$$

Wählen wir $J_m = \{b^{\pi_m}\}$, $b \neq a$, so erhalten wir $L \cup J_m = \{a\}^+ \cup \{b^{\pi_m}\}$, und wegen Lemma 4.6 die Aussage

$$\mathrm{Symb}(L \cup J_m) = \mathrm{Symb}\left(\{a\}^+ \cup \{b^{\pi_m}\}\right) = m + 10.$$

□

Lemma 4.21 – Ergebnisse zum Raum der Symbolkomplexität unter Vereinigung
Unter den Voraussetzungen $n, m \in \mathbb{N}_0^{\neq 1}$ gelten

i. $2, \ldots, 5 \notin \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 6$ und $2 \leq m \leq n$,

ii. $m \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 13$ und $13 \leq m \leq n$,

iii. $n \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 5$ und $2 \leq m \leq n - 3$,

iv. $n \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 13$ und $3 \leq m \leq n$,

v. $n + m \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 3$ und $2 \leq m \leq n$,

vi. $23, \ldots, n \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 23$ und $13 \leq m \leq n$,

vii. $n + 3, \ldots, n + m - 2 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m)$ für $n \geq 13$ und $5 \leq m \leq n$.

Beweis: Es seien T ein beliebiges Alphabet und nicht notwendig verschiedene Buchstaben $a, b, c, d \in T$.

Zu i. Für $n \geq 6$ und $2 \leq m \leq n$ seien L_n und J_m kontextfreie Sprachen, für die $\mathrm{Symb}(L_n) = n$ und $\mathrm{Symb}(J_m) = m$ gelten. Weiterhin sei K_p eine kontextfreie Sprache mit $\mathrm{Symb}(K_p) = p$ für $p \in \mathbb{N}$, $2 \leq p \leq 5$. Wir nehmen nun an, es gilt $L_n \subseteq K_p$. Für K_p kommen nur die in Lemma 4.1 angegebenen Sprachen infrage. Wir erhalten nun die folgenden Aussagen:

- Aus $L_n \subseteq K_2 = \{\lambda\}$ folgt $L_n = \emptyset$ oder $L_n = \{\lambda\}$,
- aus $L_n \subseteq K_3 = \{a\}$ folgt $L_n = \emptyset$ oder $L_n = \{a\}$,
- aus $L_n \subseteq K_4 = \{ab\}$ folgt $L_n = \emptyset$ oder $L_n = \{ab\}$,
- aus $L_n \subseteq K_5 = \{abc\}$ folgt $L_n = \emptyset$ oder $L_n = \{abc\}$,
- aus $L_n \subseteq K_5 = \{\lambda, a\}$ folgt $L_n = \emptyset$, $L_n = \{\lambda, a\}$, $L_n = \{\lambda\}$ oder $L_n = \{a\}$.

Nach Lemma 4.1 gilt dann $0 \leq \mathrm{Symb}(L_n) = n \leq 5$. Das widerspricht der Voraussetzung $n \geq 6$. Mithin gilt $L_n \subseteq K_p$ nicht. Bilden wir die Vereinigung von L_n und J_m, erhalten wir $L_n \cup J_m \neq K_p$ für alle möglichen p, $2 \leq p \leq 5$. Somit gilt stets $\mathrm{Symb}(L_n \cup J_m) \neq 2, \ldots, 5$.

Zu ii. Für $n \geq 13$ und $13 \leq m \leq n$ seien

$$L_n = \{a^{\pi_n}\},$$
$$J_m = \{a\}^+ \cup \{b^{\pi_{m-10}}\}.$$

Nach Lemma 4.3 gilt $\mathrm{Symb}(L_n) = n$, und nach Lemma 4.3 sowie Lemma 4.6 gilt

$$\mathrm{Symb}(J_m) = 10 + (m - 10) = m.$$

Aus $L_n \cup J_m = J_m$ folgt $\mathrm{Symb}(L_n \cup J_m) = \mathrm{Symb}(J_m) = m$.

Zu iii. Für $n \geq 5$ und $2 \leq m \leq n - 3$ setzen wir

$$L_{n,m} = \{a^{\pi_m}\} \cup \{b^{\pi_{n-m}}\},$$
$$J_m = \{a^{\pi_m}\}.$$

Im Fall $m = 2$ wenden wir die Lemmata 4.3 und 4.4 an, wodurch wir

$$\mathrm{Symb}(L_{n,m}) = \mathrm{Symb}(\{\lambda\} \cup \{b^{\pi_{n-2}}\}) = 2 + (n - 2) = n$$

sowie
$$\mathrm{Symb}(J_2) = \mathrm{Symb}(\{\lambda\}) = 2$$
erhalten. Die Vereinigung der beiden Sprachen liefert $L_{n,2} \cup J_2 = L_{n,2}$. Es folgt
$$\mathrm{Symb}(L_{n,2} \cup J_2) = \mathrm{Symb}(L_{n,2}) = n.$$
Im Fall $m \geq 3$ gilt $\mathrm{Alph}(\{a^{\pi_m}\}) \cap \mathrm{Alph}(\{b^{\pi_{n-m}}\}) = \emptyset$. Daher dürfen wir Satz 2.22 anwenden und erhalten mittels Lemma 4.3
$$\mathrm{Symb}(L_{n,m}) = \mathrm{Symb}(\{a^{\pi_m}\}) + \mathrm{Symb}(\{b^{\pi_{n-m}}\}) = m + (n-m) = n.$$
Nach Lemma 4.3 gilt
$$\mathrm{Symb}(J_m) = \mathrm{Symb}(\{a^{\pi_m}\}) = m.$$
Die Vereinigung der beiden Sprachen ist $L_{n,m} \cup J_m = L_{n,m}$. Mithin gilt
$$\mathrm{Symb}(L_{n,m} \cup J_m) = \mathrm{Symb}(L_{n,m}) = n.$$
Zu iv. Für $n \geq 13$ und $3 \leq m \leq n$ setzen wir
$$\begin{aligned} L_n &= \{a\}^+ \cup \{b^{\pi_{n-10}}\}, \\ J_m &= \{a^{\pi_m}\}. \end{aligned}$$
Nach den Lemmata 4.3 und 4.6 erhalten wir
$$\mathrm{Symb}(L_n) = 10 + \mathrm{Symb}(b^{\pi_{n-10}}) = 10 + (n-10) = n$$
sowie
$$\mathrm{Symb}(J_m) = \mathrm{Symb}(\{a^{\pi_m}\}) = m$$
aufgrund von Lemma 4.3. Die Vereinigung der beiden Sprachen ergibt $L_n \cup J_m = L_n$. Dies impliziert
$$\mathrm{Symb}(L_n \cup J_m) = \mathrm{Symb}(L_n) = n.$$
Zu v. Für $n \geq 3$ und $2 \leq m \leq n$ seien
$$\begin{aligned} L_n &= \{a^{\pi_n}\}, \\ J_m &= \{b^{\pi_m}\}. \end{aligned}$$
Dann gelten $\mathrm{Symb}(L_n) = n$ und $\mathrm{Symb}(J_m) = m$ wegen Lemma 4.3. Die Aussage $\mathrm{Symb}(L_n \cup J_m) = n + m$ ist im Fall von $m \geq 3$ aufgrund von $\mathrm{Alph}(L_n) \cap \mathrm{Alph}(J_m) = \emptyset$ sowie Satz 2.22 korrekt, und im Fall $m = 2$ wegen Lemma 4.4.

Zu vi. Für $n \geq 23$, $13 \leq m \leq n$ und $23 \leq k \leq n$ seien
$$\begin{aligned} L_n &= \{a^{\pi_{n-(k-10)}}\} \cup \{b^{\pi_{k-20}}\} \cup \{c\}^+, \\ J_m &= \{a\}^+ \cup \{c^{\pi_{m-10}}\}. \end{aligned}$$
Wir legen $A_{n,k} = \{a^{\pi_{n-(k-10)}}\}$, $B_k = \{b^{\pi_{k-20}}\}$ und $C = \{c\}^+$ fest, wobei $L_n = A_{n,k} \cup$

$B_k \cup C$ gilt. Die Aussagen $\mathrm{Symb}(A_{n,k}) = n - (k-10)$, $\mathrm{Symb}(B_k) = k - 20$ und $\mathrm{Symb}(C) = 7$ lesen wir aus Lemma 4.3 bzw. Lemma 4.1 ab. Folglich ergibt sich wegen der Disjunktheit der Alphabete von $A_{n,k}$, B_k und C mittels Satz 2.22 und Lemma 4.6 die Symbolkomplexität

$$\begin{aligned}\mathrm{Symb}(L_n) &= \mathrm{Symb}(A_{n,k} \cup B_k \cup C)\\ &= \mathrm{Symb}(A_{n,k} \cup B_k) + \mathrm{Symb}(C) + 3\\ &= \mathrm{Symb}(A_{n,k}) + \mathrm{Symb}(B_k) + \mathrm{Symb}(C) + 3\\ &= (n-(k-10)) + (k-20) + 7 + 3 = n.\end{aligned}$$

In analoger Weise zeigen wir

$$\begin{aligned}\mathrm{Symb}(J_m) &= \mathrm{Symb}(\{a\}^+ \cup \{c^{\pi_{m-10}}\})\\ &= \mathrm{Symb}(\{a\}^+) + 3 + \mathrm{Symb}(\{c^{\pi_{m-10}}\})\\ &= 7 + 3 + (m-10) = m.\end{aligned}$$

Die Vereinigung von L_n und J_m ist

$$\begin{aligned}L_n \cup J_m &= \{a^{\pi_{n-(k-10)}}\} \cup \{b^{\pi_{k-20}}\} \cup \{c\}^+ \cup \{a\}^+ \cup \{c^{\pi_{m-10}}\}\\ &= \{a\}^+ \cup \{b^{\pi_{k-20}}\} \cup \{c\}^+.\end{aligned}$$

Die Alphabete der Sprachen $\{a\}^+$, $\{b^{\pi_{k-20}}\}$ und $\{c\}^+$ sind disjunkt. Demzufolge erhalten wir durch neuerliche Anwendung der oben benannten Sätze und Lemmata die Aussage

$$\begin{aligned}\mathrm{Symb}(L_n \cup J_m) &= \mathrm{Symb}(\{a\}^+ \cup \{b^{\pi_{k-20}}\} \cup \{c\}^+)\\ &= \mathrm{Symb}(\{a\}^+ \cup \{b^{\pi_{k-20}}\}) + \mathrm{Symb}(\{c\}^+) + 3\\ &= \mathrm{Symb}(\{a\}^+) + 3 + \mathrm{Symb}(\{b^{\pi_{k-20}}\}) + \mathrm{Symb}(\{c\}^+) + 3\\ &= 7 + 3 + (k-20) + 7 + 3 = k.\end{aligned}$$

Zu vii. Für $n \geq 13$, $5 \leq m \leq n$ und $n + 3 \leq k \leq n + m - 2$ legen wir

$$\begin{aligned}L_n &= \{a^{\pi_{n-10}}\} \cup \{b\}^+,\\ J_m &= \{c^{\pi_{k-n}}\} \cup \{b^{\pi_{m-(k-n)}}\}\end{aligned}$$

fest. Wir setzen $A_n = \{a^{\pi_{n-10}}\}$ und $B = \{b\}^+$, wobei $L_n = A_n \cup B$ gilt. Die Symbolkomplexitäten $\mathrm{Symb}(A_n) = n - 10$ und $\mathrm{Symb}(B) = 7$ lesen wir aus Lemma 4.3 bzw. Lemma 4.1 ab. Infolgedessen ergibt sich wegen der Disjunktheit der Alphabete von A_n und B mittels Lemma 4.6 die Aussage

$$\begin{aligned}\mathrm{Symb}(L_n) &= \mathrm{Symb}(A_n \cup B)\\ &= \mathrm{Symb}(A_n) + \mathrm{Symb}(B) + 3\\ &= (n-10) + 7 + 3 = n.\end{aligned}$$

Analog erhalten wir aufgrund des Lemmas 4.3, des Lemmas 4.4 und des Satzes 2.22

$$\begin{aligned}\operatorname{Symb}(J_m) &= \operatorname{Symb}\left(\{c^{\pi_{k-n}}\} \cup \{b^{\pi_{m-(k-n)}}\}\right)\\ &= \operatorname{Symb}\left(\{c^{\pi_{k-n}}\}\right) + \operatorname{Symb}\left(\{b^{\pi_{m-(k-n)}}\}\right)\\ &= (k-n) + (m-(k-n)) = m.\end{aligned}$$

Die Vereinigung von L_n und J_m ist

$$\begin{aligned}L_n \cup J_m &= \{a^{\pi_{n-10}}\} \cup \{b\}^+ \cup \{c^{\pi_{k-n}}\} \cup \{b^{\pi_{m-(k-n)}}\}\\ &= \{a^{\pi_{n-10}}\} \cup \{b\}^+ \cup \{c^{\pi_{k-n}}\}\\ &= L_n \cup \{c^{\pi_{k-n}}\}.\end{aligned}$$

Die Alphabete der Sprachen L_n und $\{c^{\pi_{k-n}}\}$ sind disjunkt. Daher erhalten wir durch abermalige Anwendung der zuvor genannten Lemmata die Symbolkomplexität

$$\begin{aligned}\operatorname{Symb}(L_n \cup J_m) &= \operatorname{Symb}\left(L_n \cup \{c^{\pi_{k-n}}\}\right)\\ &= \operatorname{Symb}(L_n) + \operatorname{Symb}\left(\{c^{\pi_{k-n}}\}\right)\\ &= n + (k-n) = k.\end{aligned}$$

□

Lemma 4.22 – Es gelten $n+m+3, n+m+6 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m)$
Unter den Voraussetzungen $n, m \in \mathbb{N}_0^{\not\exists 1}$ gelten

i. $n+m+6 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m)$ für $n \geq 6$ und $6 \leq m \leq n$,

ii. $n+m+3 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m)$ für $n \geq 6$ und $3 \leq m \leq n$.

Beweis: Zu i. Es seien $T_{n,m}$ ein beliebiges Alphabet mit mindestens $n+m-10$ Buchstaben und paarweise verschiedene $a_i, b_j \in T_{n,m}$ für $1 \leq i \leq n-5$, $1 \leq j \leq m-5$ sowie $n \geq 6$, $6 \leq m \leq n$ und

$$\begin{aligned}L_n &= \{a_1 \cdots a_{n-5}\}^*,\\ J_m &= \{b_1 \cdots b_{m-5}\}^*.\end{aligned}$$

Für diese Sprachen gelten $\operatorname{Symb}(L_n) = n$ bzw. $\operatorname{Symb}(J_m) = m$, und für die Vereinigung dieser $\operatorname{Symb}(L_n \cup J_m) = n+m+6$. Die Aussagen $\operatorname{Symb}(L_n) = n$ und $\operatorname{Symb}(J_m) = m$ ergeben sich aus Lemma 4.10.

Es sei $G_{n,m} = (N_{n,m}, T_{n,m}, P_{n,m}, S)$ eine symbol-minimale kontextfreie Grammatik mit $\mathrm{L}(G_{n,m}) = L_n \cup J_m$, $T_n^a = \{a_1, \ldots, a_{n-5}\}$, $T_m^b = \{b_1, \ldots, b_{m-5}\}$ und $T_{n,m} = T_n^a \cup T_m^b$. Wir dürfen $G_{n,m}$ als reduziert annehmen (Lemma 2.14).

Wir zeigen zunächst, dass es in $G_{n,m}$ keine Initialschleifenableitungen gibt: Gäbe es eine Schleifenableitung $S \Longrightarrow^* x_S S y_S$ für $x_S, y_S \in {T_{n,m}}^*$, $x_S y_S \neq \lambda$, wäre für $S \Longrightarrow^* a_1 \cdots a_n \in L_n$ und $S \Longrightarrow^* b_1 \cdots b_m \in J_m$ eines der dann ableitbaren Terminalwörter $x_S a_1 \cdots a_n y_S$ oder $x_S b_1 \cdots b_m y_S$ nicht in $L_n \cup J_m$ enthalten, da Buchstaben aus T_n^a und Buchstaben aus T_m^b in einem Wort vorkämen.

Nach Lemma 2.19 enthält $G_{n,m}$ keine kontextfreien Nichtterminale. Folglich gibt es für jedes $A \in N_{n,m} \setminus \{S\}$ eine Ableitung $S \Longrightarrow^* x_A A y_A$ für $x_A, y_A \in (N_{n,m} \cup T_{n,m})^*$,

$x_A y_A \neq \lambda$. Wegen der Reduziertheit von $G_{n,m}$ dürfen wir sogar $x_A, y_A \in T_{n,m}{}^*$ annehmen. Ferner existiert wegen der Reduziertheit und Lemma 2.17 eine Ableitung $A \Longrightarrow^* w_A$ für $w_A \in T_{n,m}{}^+$. Zusammen erhalten wir die Ableitung

$$S \Longrightarrow^* x_A A y_A \Longrightarrow^* x_A w_A y_A \in T_{n,m}{}^+.$$

Es gilt $x_A w_A y_A \in L_n \cup J_m$, und damit entweder $x_A w_A y_A \in L_n$ oder $x_A w_A y_A \in J_m$ (wegen $\mathrm{Alph}(L_n) \cap \mathrm{Alph}(J_m) = \emptyset$). Mithin erhalten wir entweder $w_A \in T_n^{a+}$ oder $w_A \in {T_m^b}^+$. Aus Ableitungen der Form $S \Longrightarrow^* x'_A A y'_A$ für $x'_A, y'_A \in T_{n,m}{}^*$, $x'_A y'_A \neq \lambda$, folgt stets $x'_A, y'_A \in T_n^{a*}$, wenn $w_A \in T_n^{a+}$ ist und $x'_A, y'_A \in {T_m^b}^*$, wenn $w_A \in {T_m^b}^+$ ist – ansonsten wäre ein Wort ableitbar, das Buchstaben aus T_n^a und Buchstaben aus T_m^b enthielte. Somit ist jedes von S verschiedene Nichtterminal von $G_{n,m}$ entweder an der Generierung von Wörtern aus L_n oder von denen aus J_m beteiligt – also nicht an einem Wort aus L_n und einem aus J_m zugleich. Mithin können wir für jede Regel mit A auf der linken Seite sagen, ob sie entweder an Ableitungen für Wörter aus L_n oder J_m beteiligt ist.

Allgemein hat eine Regel mit S auf der linken Seite die Form

$$S \to x X_1 x_1 \cdots x_{r-1} X_r x_r$$

für $x, x_i \in T_{n,m}{}^*$ und $X_i \in N_{n,m}$ für $1 \leq i \leq r$ und $r \in \mathbb{N}_0$. Da kein Wort aus $L_n \cup J_m$ Buchstaben aus T_n^a und Buchstaben aus T_m^b enthält, gilt sogar entweder $x x_1 \cdots x_r \in T_n^{a*}$ oder $x x_1 \cdots x_r \in {T_m^b}^*$. Aufgrund der Tatsache, dass für jedes Nichtterminal $A \neq S$ aus $G_{n,m}$ determiniert ist, ob es entweder an der Ableitung für Wörter aus L_n oder J_m beteiligt ist, ist dies damit – bis auf $S \to \lambda$ – auch für Regeln $S \to x X_1 x_1 \cdots x_{r-1} X_r x_r$ gegeben.

Wir konstruieren nun aus $G_{n,m}$ eine kontextfreie Grammatik G_{L_n}, die L_n erzeugt, indem wir in G_{L_n} alle Regeln aus $G_{n,m}$ aufnehmen, die an Ableitungen zur Erzeugung der Wörter aus L_n beteiligt sind. Wie wir oben gesehen haben, sind – bis auf eventuell $S \to \lambda$ – diese Regeln nicht an der Generierung von Wörtern aus J_m beteiligt. Außerdem existieren keine Initialschleifenableitungen in G_{L_n}. Folglich ist G_{L_n} eine kontextfreie Grammatik, die L_n ohne Initialschleifenableitungen erzeugt. Nach Lemma 4.11 gilt dann $\mathrm{Symb}(G_{L_n}) \geq (n-5)+8 = n+3$. Im Fall von $\mathrm{Symb}(G_{L_n}) = n+3$ folgt nach demselben Lemma, dass in G_{L_n} die Regel $S \to \lambda$ nicht enthalten ist. Andererseits kann die Regel $S \to \lambda$ enthalten sein, wenn $\mathrm{Symb}(G_{L_n}) \geq n+4$ ist.

Aus $G_{n,m}$ können wir analog zu G_{L_n} eine kontextfreie Grammatik G_{J_m} konstruieren, die J_m generiert. Es gelten die zu G_{L_n} analogen Aussagen hinsichtlich der Symbolanzahl für G_{J_m}.

Wir unterscheiden 2 mögliche Fälle:

- Ist die Regel $S \to \lambda$ in G_{L_n} und G_{J_m} enthalten, gelten $\mathrm{Symb}(G_{L_n}) \geq n+4$ und $\mathrm{Symb}(G_{J_m}) \geq m+4$. Es folgt

$$\begin{aligned}\mathrm{Symb}(G_{n,m}) &\geq \mathrm{Symb}(G_{L_n}) + \mathrm{Symb}(G_{J_m}) - 2\\ &\geq (n+4)+(m+4)-2 = n+m+6.\end{aligned}$$

Es müssen 2 Symbole abgezogen werden, da wir sonst die Symbole von $S \to \lambda$ doppelt zählen würde.

- In den anderen Fällen gilt $\operatorname{Symb}(G_{L_n}) \geq n+3$ und $\operatorname{Symb}(G_{J_m}) \geq m+3$. Es folgt

$$\begin{aligned}\operatorname{Symb}(G_{n,m}) &\geq \operatorname{Symb}(G_{L_n}) + \operatorname{Symb}(G_{J_m})\\ &\geq (n+3)+(m+3) = n+m+6.\end{aligned}$$

(In diesem Fall kommt $S \to \lambda$ höchstens einmal vor.)

Insgesamt erhalten wir $\operatorname{Symb}(G_{n,m}) \geq n+m+6$.

Die kontextfreie Grammatik

$$\left(\{S, S_a, S_b\}, T_{n,m}, P_n^a \cup P_m^b, S\right)$$

mit

$$\begin{aligned}P_n^a &= \{S \to S_a, S_a \to a_1 \cdots a_{n-5} S_a, S_a \to \lambda\},\\ P_m^b &= \{S \to S_b, S_b \to b_1 \cdots b_{m-5} S_b, S_b \to \lambda\}\end{aligned}$$

erzeugt $L_{n,m}$ mittels $n+m+6$ Symbole. Zusammen mit der obigen Aussage folgt $\operatorname{Symb}(G_{n,m}) = \operatorname{Symb}(L_{n,m}) = n+m+6$.

Zu ii. Für $n \geq 6$ und $3 \leq m \leq n$ legen wir

$$\begin{aligned}L_n &= \{a_1 \cdots a_{n-5}\}^*,\\ J_m &= \{b_1 \cdots b_{m-2}\}\end{aligned}$$

fest. Aus Lemma 4.10 und Lemma 4.9 folgen die Aussagen $\operatorname{Symb}(L_n) = n$ bzw. $\operatorname{Symb}(J_m) = m$.

Den Rest des Beweises führen wir analog zum Beweis von Anstrich i durch. Offenbar enthält jede kontextfreie Grammatik, die J_m erzeugt, nicht die Regel $S \to \lambda$ (S sei das Startsymbol), da das leere Wort nicht in J_m vorkommt. Wir geben eine symbol-minimale kontextfreie Grammatik an, die $L_n \cup J_m$ generiert:

$$\left(\{S, S_a, S_b\}, \{a_1, \ldots, a_{n-5}, b_1, \ldots, b_{m-2}\}, P_n^a \cup P_m^b, S\right)$$

mit

$$\begin{aligned}P_n^a &= \{S \to S_a, S_a \to a_1 \cdots a_{n-5} S_a, S_a \to \lambda\},\\ P_m^b &= \{S \to b_1 \cdots b_{m-2}\}.\end{aligned}$$

Schlussendlich gilt $\operatorname{Symb}(L_n \cup J_m) = n+m+3$. □

Satz 4.23 – Teilbestimmung des Raums der Symbolkomplexität unter Vereinigung
Für $n, m \in \mathbb{N}_0$ und $m \leq n$ gelten

- $$\begin{aligned}\{n\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,0) \text{ \textit{für} } n \neq 1,\\ \emptyset &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(1,0),\end{aligned}$$

- $$\emptyset = \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,1),$$

- $$\begin{aligned}\{2\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(2,2),\\ \{5\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(3,2),\\ \{6\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(4,2),\\ \{5,7\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,2),\\ 6,8 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,2) \not\ni 0,\dots,5,12,\dots,\\ 6,7,9 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(7,2) \not\ni 0,\dots,5,13,\dots,\\ n,n+2 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,2) \not\ni 0,\dots,5,n+6,\dots \text{ \textit{für} } n \geq 8,\end{aligned}$$

- $$\begin{aligned}\{3,6\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(3,3),\\ \{7\} &= \mathrm{C}_{\cup}^{\mathrm{Symb}}(4,3),\\ 5,8 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,3) \not\ni 0,\dots,4,6,12,\dots,\\ 6,9,12 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,3) \not\ni 0,\dots,5,13,\dots,\\ n,n+3,n+6 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,3) \not\ni 0,\dots,5,n+7,\dots \text{ \textit{für} } n \geq 7,\end{aligned}$$

- $$\begin{aligned}4,8 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(4,4) \not\ni 0,\dots,3,5,6,12,\dots,\\ 9 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,4) \not\ni 0,\dots,6,13,\dots,\\ 6,10,13 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,4) \not\ni 0,\dots,5,14,\dots,\\ n,n+4,n+7 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,4) \not\ni 0,\dots,5,n+8,\dots \text{ \textit{für} } n \geq 7,\end{aligned}$$

- $$\begin{aligned}5,10 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,5) \not\ni 0,\dots,4,6,14,\dots,\\ 6,11,14 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,5) \not\ni 0,\dots,5,15,\dots,\\ n,n+5,n+8 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,5) \not\ni 0,\dots,5,n+9,\dots \text{ \textit{für} } n \geq 7,\\ n,n+3,n+5,n+8 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,5) \not\ni 0,\dots,5,n+9,\dots \text{ \textit{für} } n \geq 13,\end{aligned}$$

- $$\begin{aligned}6,12,15,18 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,6) \not\ni 0,\dots,5,19,\dots,\\ 7,13,16,19 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(7,6) \not\ni 0,\dots,5,20,\dots,\\ 14,17,20 &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(8,6) \not\ni 0,\dots,5,21,\dots,\end{aligned}$$

$$n, n+6, n+9, n+12 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,6) \not\ni 0,\dots,5, n+13,\dots \quad \textit{für } n \geq 9,$$
$$n, n+3, n+4, n+6, n+9, n+12 \in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,6) \not\ni 0,\dots,5, n+13,\dots \quad \textit{für } n \geq 13,$$

- $$\begin{aligned} &n+m, n+m+3, n+m+6 \\ &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m) \not\ni 0,\dots,5, n+m+7,\dots \quad \textit{für } n \geq 7, m \geq 7; \\ &n, n+m, n+m+3, n+m+6 \\ &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m) \not\ni 0,\dots,5, n+m+7,\dots \quad \textit{für } n \geq 10, 7 \leq m \leq n-3; \\ &n, n+3,\dots, n+m-2, n+m, n+m+3, n+m+6 \\ &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m) \not\ni 0,\dots,5, n+m+7,\dots \quad \textit{für } n \geq 13, m \geq 7; \\ &m, n, n+3,\dots, n+m-2, n+m, n+m+3, n+m+6 \\ &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m) \not\ni 0,\dots,5, n+m+7,\dots \quad \textit{für } n \geq 13, m \geq 13; \\ &m, 23,\dots, n, n+3,\dots, n+m-2, n+m, n+m+3, n+m+6 \\ &\in \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m) \not\ni 0,\dots,5, n+m+7,\dots \quad \textit{für } n \geq 23, m \geq 13; \end{aligned}$$
- $$\mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m) = \mathrm{C}_{\cup}^{\mathrm{Symb}}(m,n).$$

Beweis: Der Satz ist eine Zusammenfassung der Lemmata 4.16, 4.19, 4.20, 4.21 und 4.22 unter Zuhilfenahme des Lemmas 4.18. □

Bemerkung: Es seien $n, m \in \mathbb{N}_0$ und $m \leq n$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 4.23 vollständig dargestellt:

- $$\begin{aligned} 7, 9, 10, 11 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,2), \\ 8, 10, 11, 12 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(7,2), \\ 6,\dots, n-1, n+1, n+3,\dots, n+5 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,2) \text{ für } n \geq 8, \end{aligned}$$
- $$\begin{aligned} 7, 9, 10, 11 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,3), \\ 7, 8, 10, 11 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,3), \\ 6,\dots, n-1, n+1, n+2, n+4, n+5 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,3) \text{ für } n \geq 7, \end{aligned}$$
- $$\begin{aligned} 7, 9, 10, 11 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(4,4), \\ 7, 8, 10, 11, 12 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(5,4), \\ 7, 8, 9, 11, 12 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(6,4), \\ 6,\dots, n-1, n+1, n+2, n+3, n+5, n+6 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n,4) \text{ für } n \geq 7, \end{aligned}$$

- $$\begin{aligned} 7, 8, 9, 11, 12, 13 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(5, 5), \\ 7, \ldots, 10, 12, 13 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(6, 5), \\ 6, \ldots, n-1, n+1, \ldots, n+4, n+6, n+7 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, 5) \text{ für } 7 \leq n \leq 12, \\ 6, \ldots, n-1, n+1, n+2, n+4, n+6, n+7 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, 5) \text{ für } n \geq 13, \end{aligned}$$

- $$\begin{aligned} 7, \ldots, 11, 13, 14, 16, 17 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(6, 6), \\ 6, 8, \ldots, 12, 14, 15, 17, 18 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(7, 6), \\ 6, \ldots, 13, 15, 16, 18, 19 &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(8, 6), \\ 7, \ldots, n-1, n+1, \ldots, n+5, &n+7, n+8, n+10, n+11 \\ &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, 6) \text{ für } 9 \leq n \leq 12, \\ 7, \ldots, n-1, n+1, n+2, n+5, &n+7, n+8, n+10, n+11 \\ &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, 6) \text{ für } n \geq 13, \end{aligned}$$

- $$\begin{aligned} &6, \ldots, n+m-1, n+m+1, n+m+2, n+m+4, n+m+5 \\ &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m) \text{ für } 7 \leq n \leq 9, m \geq 7 \text{ od. } 7 \leq n \leq 12, m \geq n-3; \\ &6, \ldots, n-1, n+1, \ldots, n+m-1, n+m+1, n+m+2, n+m+4, n+m+5 \\ &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m) \text{ für } 10 \leq n \leq 12, 7 \leq m \leq n-3; \\ &6, \ldots, n-1, n+1, n+2, \\ &n+m-1, n+m+1, n+m+2, n+m+4, n+m+5 \\ &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m) \text{ für } n \geq 13, 7 \leq m \leq 12; \\ &6, \ldots, m-1, m+1, \ldots, n-1, n+1, n+2, \\ &n+m-1, n+m+1, n+m+2, n+m+4, n+m+5 \\ &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m) \text{ für } 13 \leq n \leq 22, m \geq 13; \\ &x, n+1, n+2, \\ &n+m-1, n+m+1, n+m+2, n+m+4, n+m+5 \\ &\in^? \mathrm{C}_{\cup}^{\mathrm{Symb}}(n, m) \text{ für } m \neq x \in \mathbb{N}, 6 \leq x \leq 22, \text{ und } n \geq 23, m \geq 13. \end{aligned}$$

$\triangle$

4.4. Konkatenation

Wir untersuchen den Raum der Symbolkomplexität unter Konkatenation.

Lemma 4.24 – Kommutativität von $\mathrm{C}_{\cdot}^{\mathrm{Symb}}$
Für $n, m \in \mathbb{N}_0$ gilt
$$\mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, m) = \mathrm{C}_{\cdot}^{\mathrm{Symb}}(m, n).$$

Beweis: Der Beweis kann analog zum Beweis von Lemma 3.18 geführt werden. □

Lemma 4.25 – Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Konkatenation

Unter den Voraussetzungen $n, m \in \mathbb{N}_0^{\not\ni 1}$ gelten

i.
$$\{0\} = \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, 0),$$
$$\{n\} = \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, 2),$$

ii.
$$0, 2, 3 \notin \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, m) \text{ für } n \geq 3 \text{ und } m \geq 3,$$
$$4 \notin \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, m) \text{ für } n \geq 4 \text{ und } m \geq 3,$$
$$5 \notin \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, m) \text{ für } n \geq 5 \text{ und } m \geq 3,$$

iii.
$$6 \notin \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, 3) \text{ für } n \geq 6,$$
$$6 \notin \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, 4) \text{ für } n \geq 6,$$

iv.
$$n + m + 5, \ldots \notin \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n, m).$$

Beweis: Es seien L und J zwei kontextfreie Sprachen, für die $\mathrm{Symb}(L) = n$, $n \in \mathbb{N}_0^{\not\ni 1}$, und $\mathrm{Symb}(J) = m$, $m \in \mathbb{N}_0^{\not\ni 1}$, gelten sowie T ein beliebiges Alphabet und nicht notwendig verschiedene Buchstaben $a, b, c, d, e \in T$.

Zu i. Es sei $\mathrm{Symb}(J) = m = 0$. Aufgrund von Lemma 4.1 gilt $J = \emptyset$. Infolgedessen erhalten wir $L \cdot J = L \cdot \emptyset = \emptyset$, und damit $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\emptyset) = 0$ (Lemma 4.1). Im Fall von $\mathrm{Symb}(J) = m = 2$ ergibt sich mithilfe von Lemma 4.1 $L \cdot J = L \cdot \{\lambda\} = L$, und somit $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(L) = n$.

Zu ii. Wegen Anstrich i dürfen wir im Folgenden $m \geq 3$, also $\mathrm{Symb}(J) = m \geq 3$, annehmen.

Wir betrachten den Fall $\mathrm{Symb}(L \cdot J) = 0$, $n \geq 3$. Lemma 4.1 impliziert dann $L \cdot J = \emptyset$. Da $\mathrm{Symb}(L) = n \geq 3$ gelten soll, ist die Gleichung $L \cdot J = \emptyset$ nach Lemma 4.1 nicht lösbar – sie wäre es nur für $L = \emptyset$ oder $J = \emptyset$, dann aber gälten $\mathrm{Symb}(L) = 0$ oder $\mathrm{Symb}(J) = 0$. Für den Fall $\mathrm{Symb}(L \cdot J) = 2$, $n \geq 3$ erhalten wir $L \cdot J = \{\lambda\}$. Auch diese Gleichung ist nicht unter den gegebenen Bedingungen lösbar. Analoges gilt für den Fall $\mathrm{Symb}(L \cdot J) = 3$, $n \geq 3$ und der resultierenden Gleichung $L \cdot J = \{a\}$. Auch die Ergebnisse der letzten beiden Fälle $\mathrm{Symb}(L \cdot J) = 4$, $n \geq 4$ und $\mathrm{Symb}(L \cdot J) = 5$, $n \geq 5$ lassen sich mithilfe von Lemma 4.1 auf gleiche Weise zeigen.

Zu iii. Es sei $\mathrm{Symb}(J) = 3$. Nach Lemma 4.1 erhalten wir $J = \{a\}$. Für den Fall $n \geq 6$, $\mathrm{Symb}(L \cdot J) = 6$ entstehen unter Beachtung desselben Lemmas die möglichen Gleichungen

$$L \cdot J = L \cdot \{a\} = \{bcde\},$$
$$L \cdot J = L \cdot \{a\} = \{\lambda, bc\},$$

$$L \cdot J = L \cdot \{a\} = \{b, c\}\ (b \neq c) \text{ oder}$$
$$L \cdot J = L \cdot \{a\} = \{b\}^*.$$

Offenbar existiert kein L für $\mathrm{Symb}(L) = n \geq 6$, das auch nur eine dieser Gleichungen löst. Analoges lässt sich im Fall von $\mathrm{Symb}(J) = 4$, also $J = \{ab\}$, zeigen.

Zu iv. Wählen wir die Grammatiken G mit $\mathrm{L}(G) = L$ und G' mit $\mathrm{L}(G') = J$ aus dem Beweis des Lemmas 2.8 so, dass $\mathrm{Symb}(G) = \mathrm{Symb}(L) = n$ und $\mathrm{Symb}(G') = \mathrm{Symb}(J) = m$ gelten, dann zählen wir in $G_\cdot$ mit $\mathrm{L}(G_\cdot) = L \cdot J$ höchstens $n + m + 4$ Symbole (Beweis zu Lemma 2.8). Mithin gilt $\mathrm{Symb}(L \cdot J) \leq \mathrm{Symb}(G_\cdot) = n + m + 4$. □

Lemma 4.26 – Ergebnisse zum Raum der Symbolkomplexität unter Konkatenation
Es gelten

i.
$$\{4\} = \mathrm{C}_\cdot^{\mathrm{Symb}}(3, 3),$$

ii.
$$\{5\} = \mathrm{C}_\cdot^{\mathrm{Symb}}(4, 3),$$
$$\{6\} = \mathrm{C}_\cdot^{\mathrm{Symb}}(4, 4),$$

iii.
$$\{6, 7\} = \mathrm{C}_\cdot^{\mathrm{Symb}}(5, 3),$$
$$7 \in \mathrm{C}_\cdot^{\mathrm{Symb}}(5, 4) \not\ni 0, \ldots, 6,$$
$$\mathrm{C}_\cdot^{\mathrm{Symb}}(5, 5) \not\ni 0, \ldots, 6,$$

iv.
$$7 \in \mathrm{C}_\cdot^{\mathrm{Symb}}(6, 3) \not\ni 0, \ldots, 6,$$
$$\mathrm{C}_\cdot^{\mathrm{Symb}}(6, 4) \not\ni 0, \ldots, 6,$$
$$6 \in \mathrm{C}_\cdot^{\mathrm{Symb}}(6, 5) \not\ni 0, \ldots, 5,$$
$$6 \in \mathrm{C}_\cdot^{\mathrm{Symb}}(6, 6) \not\ni 0, \ldots, 5.$$

Beweis: Es seien L und J zwei kontextfreie Sprachen sowie T ein beliebiges Alphabet und nicht notwendig verschiedene Buchstaben $a, b, c, d, e \in T$.

Zu i. Es sei $\mathrm{Symb}(L) = \mathrm{Symb}(J) = 3$. Dann gelten nach Lemma 4.1 ausschließlich $L = \{a\}$ und $J = \{b\}$. Es folgt $L \cdot J = \{a\} \cdot \{b\} = \{ab\}$, und nach Lemma 4.1 $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\{ab\}) = 4$.

Zu ii. Es seien $\mathrm{Symb}(L) = 4$ und $\mathrm{Symb}(J) = 3$. Aufgrund von Lemma 4.1 gelten ausschließlich $L = \{ab\}$ und $J = \{c\}$. Wir erhalten $L \cdot J = \{ab\} \cdot \{c\} = \{abc\}$. Lemma 4.1 liefert $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\{abc\}) = 5$.

Aus Lemma 4.1 folgen für den Fall $\mathrm{Symb}(L) = 4$ und $\mathrm{Symb}(J) = 4$ die Aussagen $L = \{ab\}$ und $J = \{cd\}$. Mithin gilt $L \cdot J = \{ab\} \cdot \{cd\} = \{abcd\}$, und schließlich $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\{abcd\}) = 6$.

Zu iii. Es seien $\mathrm{Symb}(L) = 5$ und $\mathrm{Symb}(J) = 3$. Nach Lemma 4.1 gelten dann allein $L = \{abc\}$ oder $L = \{\lambda, a\}$ und $J = \{d\}$. Die möglichen Konkatenationen ergeben

$$\{abc\} \cdot \{d\} = \{abcd\}\,,$$
$$\{\lambda, a\} \cdot \{d\} = \{d, ad\}\,.$$

Lemma 4.1 liefert für diese Sprachen $\mathrm{Symb}(\{abcd\}) = 6$ bzw. $\mathrm{Symb}(\{d, ad\}) = 7$.

Für den Fall $\mathrm{Symb}(L) = 5$ und $\mathrm{Symb}(J) = 4$ gelten aufgrund von Lemma 4.1 die Aussagen $L = \{abc\}$ und $J = \{de\}$. Die Konkatenation ergibt

$$L \cdot J = \{abc\} \cdot \{de\} = \{abcde\}\,,$$

was nach demselben Lemma $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\{abcde\}) = 7$ ergibt. Es kann darüber hinaus nie $\mathrm{Symb}(L \cdot J) \in \{0, \ldots, 6\}$ gelten, da die einzige weitere Möglichkeit $L = \{\lambda, a\}$ konkateniert mit $J = \{de\}$ – also $\{de, ade\}$ – keine Sprache aus Lemma 4.1 ist.

Die Aussage für $\mathrm{Symb}(L) = 5$ und $\mathrm{Symb}(J) = 5$ ergibt sich analog zu den vorherigen Aussagen.

Zu iv. Die Aussagen können wie zuvor mit dem Lemma 4.1 bewiesen werden. Es seien

- $\mathrm{Symb}(L) = 6$ und $\mathrm{Symb}(J) = 3$: Aus $L = \{abcd\}$ und $J = \{e\}$ folgt $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\{abcde\}) = 7$.
- $\mathrm{Symb}(L) = 6$ und $\mathrm{Symb}(J) = 5$: Aus $L = \{a\}^*$ und $J = \{\lambda, a\}$ folgt $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\{a\}^*) = 6$.
- $\mathrm{Symb}(L) = 6$ und $\mathrm{Symb}(J) = 6$: Aus $L = \{a\}^*$ und $J = \{\lambda, aa\}$ folgt $\mathrm{Symb}(L \cdot J) = \mathrm{Symb}(\{a\}^*) = 6$.

Die Aussagen $\mathrm{Symb}(L \cdot J) \notin \{0, \ldots, 6\}$ oder $\mathrm{Symb}(L \cdot J) \notin \{0, \ldots, 5\}$ ergeben sich analog zum Anstrich iii durch Lemma 4.1 und durch das Bilden der Konkatenationen aller möglichen Sprachen für L und J. □

Lemma 4.27 – Es gilt $n + 4, \ldots, n + m - 2 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n, m)$

Für $n \in \mathbb{N}_0$, $n \geq 6$, und $m \in \mathbb{N}_0$, $6 \leq m \leq n$, gilt $n + 4, \ldots, n + m - 2 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n, m)$.

Beweis: Es seien $T_{n,m}$ ein beliebiges Alphabet mit mindestens $n + m - 4$ Buchstaben und paarweise verschiedene $a_i, b_j, c_p \in T_{n,m,k}$ für $1 \leq i \leq n - x - 2$, $1 \leq j \leq m - x - 2$, $1 \leq p \leq x$, $p \in \mathbb{N}$ sowie

$$\begin{aligned} x &= n + m - k + 2, \\ L_{n,m,k} &= \{a_1 \cdots a_{n-x-2} \cdot c_1 \cdots c_x\}\,, \\ J_{n,m,k} &= \{b_1 \cdots b_{m-x-2} \cdot c_1 \cdots c_x\} \end{aligned}$$

und $k \in \mathbb{N}_0, n{+}4 \leq k \leq n{+}m{-}2$. Wir werden $\mathrm{Symb}(L_{n,m,k}) = n$, $\mathrm{Symb}(J_{n,m,k}) = m$ und $\mathrm{Symb}(L_{n,m,k} \cdot J_{n,m,k}) = k$ zeigen. Aufgrund von Lemma 4.2 wissen wir $\mathrm{Symb}(L_{n,m,k}) \geq n$. Die kontextfreie Grammatik

$$(\{S\}\,, \mathrm{Alph}(L_{n,m,k}), \{S \to a_1 \cdots a_{n-x-2} c_1 \cdots c_x\}\,, S)$$

erzeugt offenbar $L_{n,m,k}$ mit $2{+}n{-}x{-}2{+}x = n$ Symbolen. Mithin gilt $\mathrm{Symb}(L_{n,m,k}) = n$. Analog beweisen wir $\mathrm{Symb}(J_{n,m,k}) = m$.

Es sei

$$w = a_1 \cdots a_{n-x-2} \cdot c_1 \cdots c_x \cdot b_1 \cdots b_{m-x-2} \cdot c_1 \cdots c_x.$$

Für die Konkatenation ergibt sich $L_{n,m,k} \cdot J_{n,m,k} = \{w\}$.

Wir bestimmen nun die Symbolkomplexität der Sprache $\{w\}$. Dazu sei $G = (N, T, P, S)$ eine symbol-minimale kontextfreie Grammatik mit $\mathrm{L}(G) = \{w\}$. Wenn G nur eine Regel hat, muss diese $S \to w$ sein, und es folgt $\mathrm{Symb}(G) = n + m - 2$ (Lemma 4.2).

Wir nehmen nun an, dass G genau zwei verschiedene Regeln p und q hat ($p \neq q$). Gäbe es nur ein Nichtterminal, hätten die beiden Regeln die Form

$$\begin{aligned} p &= S \to uX_1u_1 \cdots u_{r-1}X_ru_r, \\ q &= S \to vY_1v_1 \cdots v_{s-1}Y_sv_s \end{aligned}$$

für $r, s \in \mathbb{N}_0$ und $u, u_i, v, v_j \in T^*$ sowie $X_i = S, Y_j = S$ für $1 \leq i \leq r$ und $1 \leq j \leq s$. Offenbar kommen $S \to \lambda$ oder $S \to S$ nicht infrage. Weiterhin darf nicht $r = s = 0$ gelten, da G sonst die zwei verschiedenen Wörter u und v erzeugte. Gelten hingegen $r \geq 1$ oder $s \geq 1$, befindet sich in der Regel p oder q mindestens ein S auf der rechten Seite, und da $S \to S$ nicht möglich ist, muss die rechte Seite aus mindestens zwei Buchstaben bestehen. Damit entstehen allerdings Schleifenableitungen für S und folglich unendlich viele Wörter.

Angenommen, in G finden zwei verschiedene Nichtterminale S und A Anwendung ($S \neq A$). Dann haben die beiden Regeln die Form

$$\begin{aligned} p &= S \to uX_1u_1 \cdots u_{r-1}X_ru_r, \\ q &= A \to vY_1v_1 \cdots v_{s-1}Y_sv_s \end{aligned}$$

für $r, s \in \mathbb{N}_0$ und $u, u_i, v, v_j \in T^*$ sowie $X_i, Y_j \in \{S, A\}$ für $1 \leq i \leq r$ und $1 \leq j \leq s$. Die Möglichkeiten $S \to \lambda$, $A \to \lambda$, $S \to S$ und $A \to A$ können wir ausschließen. Außerdem gilt $X_i \neq S$, da sonst Schleifenableitungen entstünden. Des Weiteren ist $s = 0$, anderenfalls gäbe es für A keine terminierende Regel (es wäre kein Terminal-Wort ableitbar). Offenbar wäre q nutzlos, wenn $r = 0$ wäre. Gälte $r = 1$, so könnten wir das einzige Nichtterminal A auf der rechten Seite von p durch v – die rechte Seite von q – ersetzen und erhielten weniger Symbole. Wäre $r \geq 3$, gäbe es im einzigen Wort w aus $L_{n,m,k} \cdot J_{n,m,k}$ drei Vorkommen von v, was nicht der Fall ist, da jeder Buchstabe aus T höchstens zweimal auftritt. Mithin gilt $r = 2$. Da nur das Wort $c_1 \cdots c_x$ genau zweimal in w vorkommt, kann v ausschließlich ein Teilwort von $c_1 \cdots c_x$ sein. Daher müssen die übrigen in w vorkommenden Buchstaben in p enthalten sein. Wir erhalten

$$\begin{aligned} \mathrm{Symb}(G) &= (2 + |w| + 2 - 2 \cdot |v|) + (2 + |v|) \\ &= |w| - |v| + 6. \end{aligned}$$

Der Term $|w| - |v| + 6$ wird möglichst klein, wenn $v = c_1 \cdots c_x$ gilt, also $|v|$ möglichst groß ist. Folglich erhalten wir

$$\begin{aligned}\operatorname{Symb}(G) &= |w| - x + 6\\ &= (n - x - 2 + x) + (m - x - 2 + x) - x + 6\\ &= n + m - x + 2\\ &= n + m - (n + m - k + 2) + 2\\ &= k.\end{aligned}$$

Indem wir

$$\begin{aligned}p &= S \to a_1 \cdots a_{n-x-2} A b_1 \cdots b_{m-x-2} A,\\ q &= A \to c_1 \cdots c_x\end{aligned}$$

setzen, vergewissern wir uns, dass die vorherige Betrachtung eine korrekte kontextfreie Grammatik liefert, die $L_{n,m,k} \cdot J_{n,m,k}$ mit k Symbolen erzeugt.

Zuletzt nehmen wir an, dass in G mindestens drei Regeln enthalten sind. Mit den insgesamt

$$(n - x - 2) + (m - x - 2) + x = n + m - x - 4$$

verschiedenen in w vorkommenden Buchstaben und den linken Seiten und Pfeilen der Regeln zählen wir mindestens

$$(n + m - x - 4) + 6 = n + m - x + 2 = k$$

Symbole. Mithin können wir mit drei und mehr Regeln $L_{n,m,k} \cdot J_{n,m,k}$ mit nicht weniger als k Symbolen erzeugen. Es ergibt sich $\operatorname{Symb}(L_{n,m,k} \cdot J_{n,m,k}) = k$. $\square$

Lemma 4.28 – Es gilt $n + m + 4 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n, m)$

Für $n \in \mathbb{N}_0$, $n \geq 8$, und $m \in \mathbb{N}_0$, $8 \leq m \leq n$, gilt $n + m + 4 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n, m)$.

Beweis: Es seien $T_{n,m}$ ein Alphabet mit mindestens $n + m - 10$ Buchstaben, a, b, d, e, c_i, $f_j \in T_{n,m}$ für $1 \leq i \leq n - 7$ und $1 \leq j \leq m - 7$ paarweise verschieden sowie

$$\begin{aligned}L_n &= \left\{ a^i c_1 \cdots c_{n-7} b^i \;\middle|\; i \geq 0 \right\},\\ J_m &= \left\{ d^i f_1 \cdots f_{m-7} e^i \;\middle|\; i \geq 0 \right\}.\end{aligned}$$

Aufgrund von Lemma 4.12 ergeben sich $\operatorname{Symb}(L_n) = n$ und $\operatorname{Symb}(J_m) = m$.

Wir zeigen im Folgenden, dass die Symbolkomplexität für

$$L_n \cdot J_m = \left\{ a^i c_1 \cdots c_{n-7} b^i d^j f_1 \cdots f_{m-7} e^j \;\middle|\; i \geq 0, j \geq 0 \right\}$$

$n + m + 4$ beträgt.

Es sei $G_{n,m} = (N_{n,m}, T_{n,m}, P_{n,m}, S)$ eine symbol-minimale kontextfreie Grammatik, die $L_n \cdot J_m$ erzeugt (Lemma 2.14). Wir können nun so wie im Beweis zum Lemma 3.22 argumentieren, indem wir dort $C_n = \{c_1 \cdots c_{n-7}\}$ und $F_m = \{f_1 \cdots f_{m-7}\}$ festlegen. Zwar wird in dem Beweis eine regel-minimale kontextfreie Grammatik verwendet, allerdings ist die für uns jetzt wichtige Aussage von der Regelminimalität unabhängig: $P_{n,m}$ enthält mindestens 5 Regeln p_S, p_1, p_2, p_c und p_f derart, dass p_S, p_1, p_2 keine Buchstaben aus Wörtern aus $C_n \cup F_m$ enthalten, dass hingegen Buchstaben aus Wörtern aus $C_n \cup F_m$ in p_f, p_c vorkommen, dass p_1, p_2 an der Realisierung je einer der Schleifenableitungen $S \neq A_k \Longrightarrow^* u_k A_k v_k$, $k \in \{1,2\}$, mit $A_k \in N_{n,m}$ und $u_1, v_1 \in \{a,b\}^+$ sowie $u_2, v_2 \in \{d,e\}^+$, wobei $|u_k| = |v_k|$ gilt, beteiligt sind und dass in p_S das Nichtterminal S auf der linken Seite steht. Weiterhin wissen wir aufgrund von Lemma 4.2, dass alle Buchstaben aus

$$\mathrm{Alph}\,(L_n \cdot J_m) = \{a, b, d, e, c_1, \ldots, c_{n-7}, f_1, \ldots, f_{m-7}\}$$

mindestens einmal in den Regeln aus $P_{n,m}$ vorkommen müssen.

Wir geben die kontextfreie Grammatik

$$G'_{n,m} = \Big(\{S, S_1, S_2\}, T_{n,m}, P'_{n,m}, S\Big)$$

mit

$$\begin{aligned} P'_{n,m} = \{&S \to S_1S_2, S_1 \to aS_1b, S_2 \to dS_2e\} \cup \\ &\{S_1 \to c_1 \cdots c_{n-7}\} \cup \{S_2 \to f_1 \cdots f_{m-7}\} \end{aligned}$$

an, die $L_n \cdot J_m$ erzeugt und aufgrund der vorherigen Aussagen nicht mit weniger Symbolen konstruiert werden kann (die beiden Regeln für die beiden Schleifenableitungen können nicht kürzer sein und die S-Regel auch nicht). Es folgt

$$\mathrm{Symb}\,(L_n \cdot J_m) = \mathrm{Symb}\Big(G'_{n,m}\Big) = 18 + (n-7) + (m-7) = n + m + 4.$$

□

Satz 4.29 – Teilbestimmung des Raums der Symbolkomplexität unter Konkatenation

Für $n, m \in \mathbb{N}_0$ und $m \leq n$ gelten

- $$\begin{aligned} \{0\} &= \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n, 0) \textit{ für } n \neq 1, \\ \emptyset &= \mathrm{C}^{\mathrm{Symb}}_{\cdot}(1, 0), \end{aligned}$$

- $$\emptyset = \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n, 1),$$

- $$\{n\} = \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n, 2) \textit{ für } n \geq 2,$$

- $$\begin{aligned} \{4\} &= \mathrm{C}^{\mathrm{Symb}}_{\cdot}(3, 3), \\ \{5\} &= \mathrm{C}^{\mathrm{Symb}}_{\cdot}(4, 3), \\ \{6, 7\} &= \mathrm{C}^{\mathrm{Symb}}_{\cdot}(5, 3), \\ 7 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}&(6, 3) \not\ni 0, \ldots, 6, 14, \ldots, \end{aligned}$$

$$\mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,3) \not\ni 0,\dots,6,n+8,\dots \text{ für } n \geq 7,$$

- $$\begin{aligned} \{6\} &= \mathrm{C}^{\mathrm{Symb}}_{\cdot}(4,4), \\ 7 &\in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(5,4) \not\ni 0,\dots,6,14,\dots, \\ & \mathrm{C}^{\mathrm{Symb}}_{\cdot}(6,4) \not\ni 0,\dots,6,15,\dots, \\ & \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,4) \not\ni 0,\dots,6,n+9,\dots \text{ für } n \geq 7, \end{aligned}$$

- $$\begin{aligned} &\mathrm{C}^{\mathrm{Symb}}_{\cdot}(5,5) \not\ni 0,\dots,6,15,\dots, \\ 6 \in\ &\mathrm{C}^{\mathrm{Symb}}_{\cdot}(6,5) \not\ni 0,\dots,5,16,\dots, \\ &\mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,5) \not\ni 0,\dots,6,n+10,\dots \text{ für } n \geq 7, \end{aligned}$$

- $$\begin{aligned} 6,10 \in\ &\mathrm{C}^{\mathrm{Symb}}_{\cdot}(6,6) \not\ni 0,\dots,5,17,\dots, \\ &\mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,6) \not\ni 0,\dots,5,n+11,\dots \text{ für } n \geq 7, \end{aligned}$$

- $$n+4, n+5 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,7) \not\ni 0,\dots,5,n+12,\dots \text{ für } n \geq 7,$$

- $$\begin{aligned} &n+4,\dots,n+m-2,n+m+4 \\ &\in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,m) \not\ni 0,\dots,5,n+m+5,\dots \text{ für } n \geq 8, m \geq 8, \end{aligned}$$

- $$\mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,m) = \mathrm{C}^{\mathrm{Symb}}_{\cdot}(m,n).$$

Beweis: Der Satz ist eine Zusammenfassung der Lemmata 4.16, 4.25, 4.26, 4.27 und 4.28 unter Zuhilfenahme des Lemmas 4.24. □

Bemerkung: Es seien $n, m \in \mathbb{N}_0$ und $m \leq n$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 4.29 vollständig dargestellt:

- $$\begin{aligned} 8,\dots,13 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(6,3), \\ 7,\dots,n+7 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,3) \text{ für } n \geq 7, \end{aligned}$$

- $$\begin{aligned} 8,\dots,13 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(5,4), \\ 7,\dots,14 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(6,4), \\ 7,\dots,n+8 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,4) \text{ für } n \geq 7, \end{aligned}$$

- $$\begin{aligned} 7,\dots,14 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(5,5), \\ 7,\dots,15 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(6,5), \end{aligned}$$

$$7, \ldots, n+9 \in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,5) \text{ für } n \geq 7,$$

- $$7,8,9,11,\ldots,16 \in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(6,6),$$
  $$6,\ldots,n+10 \in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,6) \text{ für } n \geq 7,$$

- $$6,\ldots,n+3,n+6,\ldots,n+11 \in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,7) \text{ für } n \geq 7,$$

- $$6,\ldots,n+3,n+m-1,\ldots,n+m+3 \in^? \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,m) \text{ für } n \geq 8, m \geq 8.$$

Wie in [DS08] bezüglich der Variablenkomplexität bei der Konkatenation ist also auch hier im Wesentlichen unklar, ob die Zahlen $2,\ldots,n$ im Raum der Symbolkomplexität enthalten sind oder nicht. Analoges hatten wir in der Bemerkung zu Satz 3.28 festgestellt. $\triangle$

4.5. Kleene-Abschluss

Wir untersuchen den Raum der Symbolkomplexität unter Kleene-Abschlus.

Lemma 4.30 – Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Kleene-Abschluss

Unter der Voraussetzung $n \in \mathbb{N}_0^{\neq 1}$ gelten

i. $0 \notin \mathrm{C}^{\mathrm{Symb}}_{()^*}(n)$,

ii. $2,\ldots,5 \notin \mathrm{C}^{\mathrm{Symb}}_{()^*}(n)$ für $n \geq 3$,

iii. $n+7,\ldots \notin \mathrm{C}^{\mathrm{Symb}}_{()^*}(n)$.

Beweis: Es sei L eine kontextfreie Sprache mit $\mathrm{Symb}(L) = n$.

Zu i. Nach Lemma 2.7 gilt stets $\lambda \in L^*$ und nach Lemma 4.1 kommt für $\mathrm{Symb}(L^*) = 0$ nur $L^* = \emptyset$ in Frage – ein Widerspruch.

Zu ii. Nach Lemma 4.1 gilt $L \neq \emptyset$ und $L \neq \{\lambda\}$, wenn $n \geq 3$ wahr ist. Somit existiert in L immer ein Wort $w \neq \lambda$. Aus Lemma 2.7 lesen wir dann $|L^*| = |\mathbb{N}|$ ab. Für jede kontextfreie Sprache J in Lemma 4.1 mit $\mathrm{Symb}(J) \leq 5$ gilt $|J| \in \mathbb{N}_0$. Folglich kann $L^* = J$ nicht gelten, und wir erhalten abermals aufgrund von Lemma 4.1 $\mathrm{Symb}(L^*) \geq 6$.

Zu iii. Wählen wir die Grammatik G mit $\mathrm{L}(G) = L$ aus dem Beweis des Lemmas 2.8 so, dass $\mathrm{Symb}(G) = \mathrm{Symb}(L) = n$ gilt, sind in $G_{()^*}$ mit $\mathrm{L}(G_{()^*}) = L^*$ höchstens $n+6$ Symbole enthalten (Beweis zu Lemma 2.8). Folglich gilt $\mathrm{Symb}(L^*) \leq \mathrm{Symb}(G_{()^*}) = n+6$. $\square$

Lemma 4.31 – Spezielle Ergebnisse zum Raum der Symbolkomplexität unter Kleene-Abschluss

Es gelten

$$\begin{aligned}\{2\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(0),\\ \{2\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(2),\\ \{6\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(3),\\ \{7\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(4),\\ \{6,8\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(5),\\ 6,7 &\in \mathrm{C}^{\mathrm{Symb}}_{()^*}(6) \not\ni 11,\ldots\end{aligned}$$

Beweis: Es seien L eine kontextfreie Sprache, T ein Alphabet und nicht notwendig verschiedene Buchstaben $a, b, c, d \in T$.

Nach Lemma 4.1 kommt für $\mathrm{Symb}(L) = 0$ nur $L = \emptyset$ infrage. Der Kleene-Abschluss liefert $L^* = \emptyset^* = \{\lambda\}$. Dasselbe Lemma legt $\mathrm{Symb}(\{\lambda\}) = 2$ dar.

Für $\mathrm{Symb}(L) = 2$ gilt nach Lemma 4.1 ausschließlich $L = \{\lambda\}$. Es folgt $L^* = \{\lambda\}^* = \{\lambda\}$. Lemma 4.1 liefert $\mathrm{Symb}(\{\lambda\}) = 2$.

Aus $\mathrm{Symb}(L) = 3$ folgt aufgrund von Lemma 4.1 $L = \{a\}$. Demselben Lemma entnehmen wir $\mathrm{Symb}(L^*) = \mathrm{Symb}(\{a\}^*) = 6$.

Nach Lemma 4.1 gilt $\mathrm{Symb}(L) = 4$ nur, wenn $L = \{ab\}$ ist. Nach demselben Lemma gilt für $L^* = \{ab\}^*$ die Symbolanzahl $\mathrm{Symb}(L^*) = \mathrm{Symb}(\{ab\}^*) = 7$.

Für $\mathrm{Symb}(L) = 5$ gelten nach Lemma 4.1 einzig $L = \{abc\}$ oder $L = \{\lambda, a\}$. Dasselbe Lemma liefert $\mathrm{Symb}(\{abc\}^*) = 8$ bzw. für $\{\lambda, a\}^* = \{a\}^*$ die Aussage $\mathrm{Symb}(\{\lambda, a\}^*) = \mathrm{Symb}(\{a\}^*) = 6$.

Im Fall von $\mathrm{Symb}(L) = 6$ kommen nach Lemma 4.1 ausschließlich $L = \{abcd\}$, $L = \{\lambda, ab\}$, $L = \{a, b\}$, $a \neq b$, oder $L = \{a\}^*$ infrage. Dasselbe Lemma zeigt $\mathrm{Symb}(\{\lambda, ab\}^*) = 7$, da $\{\lambda, ab\}^* = \{ab\}^*$, und für $(\{a\}^*)^* = \{a\}^*$ die Aussage

$$\mathrm{Symb}((\{a\}^*)^*) = \mathrm{Symb}(\{a\}^*) = 6.$$

Die Sprache $\{abcd\}^*$ können wir mit

$$(\{S\}, \{a, b, c, d\}, \{S \to abcdS, S \to \lambda\}, S)$$

erzeugen und erhalten daher $\mathrm{Symb}(\{abcd\}^*) \leq 9$. Analog erhalten wir für $\{a, b\}^*$ mittels der kontextfreien Grammatik

$$(\{S\}, \{a, b\}, \{S \to aS, S \to bS, S \to \lambda\}, S)$$

die Aussage $\mathrm{Symb}(\{a, b\}^*) \leq 10$. □

Lemma 4.32 – Ergebnisse zum Raum der Symbolkomplexität unter Kleene-Abschluss

Unter der Voraussetzung $n \in \mathbb{N}_0^{\not\ni 1}$ gelten

i. $6 \in \mathrm{C}^{\mathrm{Symb}}_{()^*}(n)$ für $n \geq 7$,

ii. $10, \ldots, n \in \mathrm{C}^{\mathrm{Symb}}_{()^*}(n)$ für $n \geq 10$,

iii. $n + 3 \in \mathrm{C}^{\mathrm{Symb}}_{()^*}(n)$ für $n \geq 3$,

iv. $n + 6 \in \mathrm{C}^{\mathrm{Symb}}_{()^*}(n)$ für $n \geq 8$.

Beweis: Zu i. Wir setzen $L_n = \{a\} \cup \{a^{\pi_{n-3}}\}$ und erhalten aufgrund von Lemma 4.14 die Aussage $\mathrm{Symb}(L_n) = n$. Die Anwendung des Kleene-Abschlusses ergibt ${L_n}^* = \{a\}^*$, woraus nach Lemma 4.1 $\mathrm{Symb}({L_n}^*) = 6$ folgt.

Zu ii. Es seien $k \in \mathbb{N}$, $10 \leq k \leq n$, T_k ein beliebiges Alphabet mit mindestens $k' + 1$ Buchstaben und paarweise verschiedene $a_i, b \in T_k$ für $1 \leq i \leq k'$ sowie

$$\begin{aligned} n' &= n - 5, k' = k - 9, \\ L_{n,k} &= \{a_1 \cdots a_{k'}\} \cup \{b\} \cup \{b^{\pi_{n'-k'}}\}. \end{aligned}$$

Wir können $L_{n,k}$ in $A_{k'} = \{a_1 \cdots a_{k'}\}$ und $B_{n,k} = \{b\} \cup \{b^{\pi_{n'-k'}}\}$ zerlegen: $L_{n,k} = A_{k'} \cup B_{n,k}$. Aus Lemma 4.9 lesen wir $\mathrm{Symb}(A_{k'}) = k' + 2$ ab und aus Lemma 4.14 die Aussage $\mathrm{Symb}(B_{n,k}) = (n' - k') + 3$. Satz 2.22 liefert

$$\begin{aligned} \mathrm{Symb}(L_{n,k}) &= \mathrm{Symb}(A_{k'}) + \mathrm{Symb}(B_{n,k}) \\ &= (k' + 2) + (n' - k' + 3) \\ &= n' + 5 = (n - 5) + 5 = n. \end{aligned}$$

Der Kleene-Abschluss von $L_{n,k}$ ist ${L_{n,k}}^* = (\{a_1 \cdots a_{k'}\} \cup \{b\})^*$ und wird zum Beispiel von der kontextfreien Grammatik

$$G_k = (\{S\}, \{a_i \mid 1 \leq i \leq k'\} \cup \{b\}, P_k, S)$$

mit

$$P_k = \{S \to a_1 \cdots a_{k'} S, S \to bS, S \to \lambda\}$$

erzeugt. Daher gilt $\mathrm{Symb}({L_{n,k}}^*) \leq k' + 9 = k$.

Es sei $G_{n,k} = (N_{n,k},\ T_k,\ P_{n,k},\ S)$ eine symbol-minimale kontextfreie Grammatik, die ${L_{n,k}}^*$ erzeugt. Aus $\lambda \in {L_{n,k}}^*$ folgt die Existenz einer Regel $A \to \lambda$ für $A \in N_{n,k}$. Die mögliche Ableitung $S \Longrightarrow^* b$ impliziert $B \to w_b \in P_{n,k}$ für $B \in N_{n,k}$ und $w_b \in (\{b\} \cup N_{n,k})^*$ unter der Bedingung $|w_b|_b = 1$. Aufgrund der Ableitung $S \Longrightarrow^* a_1 \cdots a_{k'}$ muss es Regeln geben, in denen die Buchstaben aus $\{a_i \mid 1 \leq i \leq k'\}$ enthalten sind. Da keine Buchstaben aus $\{a_i \mid 1 \leq i \leq k'\}$ in $A \to \lambda$ und $B \to w_b$ vorkommen, muss es mindestens eine weitere Regel $C \to w \in P_{n,k'}$ für $C \in N_{n,k}$ und $w \in (T_k \cup N_{n,k})^*$ geben, in der Buchstaben aus $\{a_i \mid 1 \leq i \leq k'\}$ vorkommen. Insgesamt ergibt sich aus $A \to \lambda$, $B \to w_b$, $C \to w$ und den Buchstaben $a_1, \ldots, a_{k'}$ die Aussage $\mathrm{Symb}({L_{n,k}}^*) = \mathrm{Symb}(G_{n,k}) \geq k' + 7 = k - 2$. Mit Obigem erhalten wir $k - 2 \leq \mathrm{Symb}({L_{n,k}}^*) \leq k$.

Im Falle von $\mathrm{Symb}(G_{n,k}) = k - 2$ sind wegen $w_b = b$ und $w = a_1 \cdots a_{k'}$ keine Schleifenableitungen realisierbar, was Lemma 2.16 widerspräche. Die Annahme

$\mathrm{Symb}(G_{n,k}) = k-1$ bedingt folgende mögliche Fälle:

- $|w_b| = 2$, also $|w_b|_X = 1$ für ein $X \in N_{n,k}$ und $w = a_1 \cdots a_{k'}$,
- $|w_b| = 1$, also $w_b = b$, und $|w| = k'+1$, $|w|_{a_i} = 1$, $1 \le i \le k'$, $|w|_x = 1$ für ein $x \in \{b\} \cup N_{n,k}$

Es ist egal, wie wir X oder x wählen: Entweder könnten nie die Wörter b und b^2 gemeinsam generiert werden oder nie die Wörter $a_1 \cdots a_{k'}$ und $(a_1 \cdots a_{k'})^2$.

Zum Schluss bleibt der Fall $\mathrm{Symb}(G_{n,k}) = k$ übrig. Setzen wir etwa $G_{n,k} = G_k$, so erhalten wir $\mathrm{Symb}({L_{n,k}}^*) = k$.

Zu iii. Es seien T_n ein beliebiges Alphabet mit mindestens $n-2$ Buchstaben und paarweise verschiedene $a_i \in T_n$ für $1 \le i \le n-2$. Wählen wir die kontextfreie Sprache $L_n = \{a_1 \cdots a_{n-2}\}$, erhalten wir nach Lemma 4.9 die Aussage $\mathrm{Symb}(L_n) = n$, und nach Lemma 4.10 die Aussage $\mathrm{Symb}({L_n}^*) = n+3$.

Zu iv. Es seien T_n ein beliebiges Alphabet mit mindestens $n-5$ Buchstaben und paarweise verschiedene $a, b, c_i \in T_n$ für $1 \le i \le n-7$. Wir setzen $L_n = \{a^m c_1 \cdots c_{n-7} b^m \mid m \ge 0\}$ und erhalten nach Lemma 4.12 die Aussage $\mathrm{Symb}(L_n) = n$, und nach Lemma 4.13 die Aussage $\mathrm{Symb}({L_n}^*) = n+6$. □

Satz 4.33 – Teilbestimmung des Raums der Symbolkomplexität unter Kleene-Abschluss

Für $n \in \mathbb{N}_0$ gelten

$$\begin{aligned}
\{2\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(0),\\
\emptyset &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(1),\\
\{2\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(2),\\
\{6\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(3),\\
\{7\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(4),\\
\{6,8\} &= \mathrm{C}^{\mathrm{Symb}}_{()^*}(5),\\
6,7 &\in \mathrm{C}^{\mathrm{Symb}}_{()^*}(6) \not\ni 0,\dots,5,11,\dots,\\
6,10 &\in \mathrm{C}^{\mathrm{Symb}}_{()^*}(7) \not\ni 0,\dots,5,14,\dots,\\
6,11,14 &\in \mathrm{C}^{\mathrm{Symb}}_{()^*}(8) \not\ni 0,\dots,5,15,\dots,\\
6,12,15 &\in \mathrm{C}^{\mathrm{Symb}}_{()^*}(9) \not\ni 0,\dots,5,16,\dots,\\
6,10,\dots,n,n+3,n+6 &\in \mathrm{C}^{\mathrm{Symb}}_{()^*}(n) \not\ni 0,\dots,5,n+7,\dots \quad \textit{für } n \ge 10.
\end{aligned}$$

Beweis: Der Satz ist eine Zusammenfassung der Lemmata 4.16, 4.30, 4.31 und 4.32. □

Bemerkung: Es sei $n \in \mathbb{N}_0$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 4.33 vollständig dargestellt:

$$8,9,10 \in^? \mathrm{C}^{\mathrm{Symb}}_{()^*}(6),$$

$$7, 8, 9, 11, 12, 13 \in^? \mathrm{C}_{()^*}^{\mathrm{Symb}}(7),$$
$$7, \ldots, 10, 12, 13 \in^? \mathrm{C}_{()^*}^{\mathrm{Symb}}(8),$$
$$7, \ldots, 11, 13, 14 \in^? \mathrm{C}_{()^*}^{\mathrm{Symb}}(9),$$
$$7, 8, 9, n+1, n+2, n+4, n+5 \in^? \mathrm{C}_{()^*}^{\mathrm{Symb}}(n) \text{ für } n \geq 10.$$

△

4.6. Homomorphismus

Wir untersuchen den Raum der Symbolkomplexität unter Homomorphismus.

Lemma 4.34 – Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Homomorphismus

Für zwei Alphabete X, Y, einen Homomorphismus $\mathrm{h}\colon X^* \to Y^*$ und $n \in \mathbb{N}_0^{\not\ni 1}$ gelten

i. $0 \notin \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$ für $n \geq 2$,

ii. $\{0\} = \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(0)$,

iii. $\{2\} = \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(2)$,

iv. $3 + (n-2) \cdot \max(\{|\mathrm{h}(a)| \mid a \in X\}), \ldots \notin \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$ für $n \geq 3$.

Beweis: Es sei L eine kontextfreie Sprache, für die $\mathrm{Symb}(L) = n$, $n \in \mathbb{N}_0^{\not\ni 1}$, gilt.

Zu i. Es sei $\mathrm{Symb}(\mathrm{h}(L)) = 0$. Dann erhalten wir nach Lemma 4.1 die Sprache $\mathrm{h}(L) = \emptyset$, woraus wir $L = \emptyset$ folgern. Aufgrund von Lemma 4.1 ergibt sich $\mathrm{Symb}(L) = \mathrm{Symb}(\emptyset) = 0$. Für $n \geq 2$ gilt nach Lemma 4.1 die Aussage $L \neq \emptyset$, und damit $\mathrm{Symb}(L) \neq 0$.

Zu ii. Ist $\mathrm{Symb}(L) = 0$, so folgt $L = \emptyset$ (Lemma 4.1), und aufgrund dessen $\mathrm{h}(L) = \mathrm{h}(\emptyset) = \emptyset$. Nach Lemma 4.1 benötigen wir für $\emptyset$ genau 0 Symbole, mithin gilt die Aussage $\mathrm{Symb}(\mathrm{h}(L)) = \mathrm{Symb}(\emptyset) = 0$.

Zu iii. Lemma 4.1 liefert für $\mathrm{Symb}(L) = 2$ die Sprache $L = \{\lambda\}$. Die Anwendung eines Homomorphismus ergibt $\mathrm{h}(L) = \mathrm{h}(\{\lambda\}) = \{\lambda\}$. Lemma 4.1 begründet dann $\mathrm{Symb}(\mathrm{h}(L)) = \mathrm{Symb}(\mathrm{h}(\{\lambda\})) = \mathrm{Symb}(\{\lambda\}) = 2$.

Zu iv. Der folgende Beweis verwendet die Konstruktion aus dem Beweis von Lemma 2.8 bezüglich des Homomorphismus. Wir wählen zunächst einen Buchstaben

$$a_{\max} \in \{a \mid a \in X, |\mathrm{h}(a)| = \max(\{|\mathrm{h}(a')| \mid a' \in X\})\},$$

dessen Bild nach Anwendung des Homomorphismus h die größte Wortlänge hat. Mithilfe von $a_{\max}$ und $n \in \mathbb{N}$, $n \geq 3$, konstruieren wir die kontextfreie Grammatik

$$G_{a_{\max},n} = \left(\{S\}, \{a_{\max}\}, \left\{S \to a_{\max}^{n-2}\right\}, S\right),$$

für die

$$\mathrm{Symb}(\mathrm{L}(G_{a_{\max},n})) \leq \mathrm{Symb}(G_{a_{\max},n}) = n$$

gilt. Anschließend erstellen wir die kontextfreie Grammatik

$$G_{\mathrm{h},a_{\max},n} = \left(\{S\}, Y, \left\{S \to \mathrm{h}\left(a_{\max}^{n-2}\right)\right\}, S\right)$$

wie im Beweis zu Lemma 2.8, die sicher die Sprache $\mathrm{h}(\mathrm{L}(G_{a_{\max},n}))$ generiert. Wir errechnen

$$\mathrm{Symb}(\mathrm{h}(\mathrm{L}(G_{a_{\max},n}))) \leq \mathrm{Symb}(G_{\mathrm{h},a_{\max},n}) = 2 + (n-2) \cdot |\mathrm{h}(a_{\max})|.$$

Wir haben $a_{\max}$, $G_{a_{\max},n}$ und $G_{\mathrm{h},a_{\max},n}$ so gewählt, dass es keine symbol-minimale kontextfreie Grammatik G_n mit $\mathrm{Symb}(G_n) = n$, $L_n = \mathrm{L}(G_n)$ und keine symbol-minimale kontextfreie Grammatik $G_{\mathrm{h},n}$ mit $\mathrm{L}(G_{\mathrm{h},n}) = \mathrm{h}(L_n)$ derart gibt, dass $\mathrm{Symb}(G_{\mathrm{h},n}) \geq \mathrm{Symb}(G_{\mathrm{h},a_{\max},n}) + 1$ gelten kann. Wir haben die Anzahl der zu ersetzenden Buchstaben maximiert und ersetzen Buchstaben durch Buchstaben mit maximaler Länge. Darüber hinaus ist im Beweis zu Lemma 2.8 eine Konstruktion, die auch für $G_{\mathrm{h},a_{\max},n}$ verwendet wurde, dargelegt, mit der stets eine kontextfreie Grammatik für $\mathrm{h}(L_n)$ angegeben werden kann und die folglich aus nicht weniger als $\mathrm{Symb}(G_{\mathrm{h},n}) = \mathrm{Symb}(\mathrm{h}(L_n))$ besteht. Somit haben wir die obere Schranke

$$\begin{aligned}\mathrm{Symb}(\mathrm{h}(L_n)) &\leq 2 + (n-2) \cdot |\mathrm{h}(a_{\max})| \\ &= 2 + (n-2) \cdot \max(\{|\mathrm{h}(a')| \mid a' \in X\})\end{aligned}$$

gefunden. □

Lemma 4.35 – Es gibt ein $\mathbf{h}_k$, sodass $k \in \mathbf{C}_{\mathbf{h}_k}^{\mathbf{Symb}}(n)$ gilt

Für ein Alphabet T mit mindestens 2 Buchstaben, paarweise verschiedene $a, b \in T$, natürliche Zahlen $n \in \mathbb{N}$, $n \geq 3$, und $k \in \mathbb{N}$, $k \geq 2$, sowie den Homomorphismus

$$\mathrm{h}_k \colon \{a, b\}^* \to \{a\}^*$$

mit

$$\begin{aligned}\mathrm{h}_k(a) &= a^{\pi_k}, \\ \mathrm{h}_k(b) &= \lambda\end{aligned}$$

gilt $k \in \mathrm{C}_{\mathrm{h}_k}^{\mathrm{Symb}}(n)$.

Beweis: Es sei

$$L_n = \{ab^{\pi_{n-1}}\}$$

eine kontextfreie Sprache. Nach Lemma 4.15 gilt $\mathrm{Symb}(L_n) = n$. Die Applikation des Homomorphismus h_k ergibt $\mathrm{h}_k(L_n) = \{a^{\pi_k}\}$. Aus Lemma 4.3 entnehmen wir $\mathrm{Symb}(\mathrm{h}_k(L_n)) = \mathrm{Symb}(\{a^{\pi_k}\}) = k$. □

Lemma 4.36 – Es gilt $\{2\} \in \mathrm{C}_{\mathbf{h}}^{\mathbf{Symb}}(n)$, wenn h total-löschend ist
Für $n \in \mathbb{N}^{\not\ni 1}$, zwei Alphabete X, Y und einen (total-löschenden) Homomorphismus

$$\mathrm{h}\colon X^* \to Y^* \text{ mit } \mathrm{h}(x) = \lambda, x \in X,$$

gilt $\{2\} \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$.

Beweis: Es sei $L \subseteq X^*$ eine kontextfreie Sprache mit $\mathrm{Symb}(L) = n$. Die Anwendung des Homomorphismus h ergibt $\mathrm{h}(L) = \{\lambda\}$. Aus Lemma 4.1 lesen wir $\mathrm{Symb}(\mathrm{h}(L)) = \mathrm{Symb}(\{\lambda\}) = 2$ ab. □

Lemma 4.37 – Ergebnisse zum Raum der Symbolkomplexität unter einem speziellen Homomorphismus
Für ein Alphabet T mit mindestens 3 Buchstaben, paarweise verschiedene $a, b, c \in T$ und $n \in \mathbb{N}_0^{\not\ni 1}$ sowie den Homomorphismus

$$\mathrm{h}\colon \{a, b, c\}^* \to \{a, c\}^* \text{ mit } \mathrm{h}(a) = a, \mathrm{h}(b) = \lambda, \mathrm{h}(c) = c$$

gelten

i. $2 \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$ für $n \geq 2$,

ii. $3 \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$ für $n \geq 3$,

iii. $4 \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$ für $n \geq 4$,

iv. $5, \ldots, n-1 \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$ für $n \geq 6$,

v. $n \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)$ für $n \geq 2$.

Beweis: Zu i. Für $n \geq 2$ legen wir

$$L_n = \{b^{\pi_n}\}$$

fest. Nach Lemma 4.3 gilt $\mathrm{Symb}(L_n) = n$. Aus $\mathrm{h}(L_n) = \{\lambda\}$ folgt aufgrund von Lemma 4.1 die Aussage $\mathrm{Symb}(\mathrm{h}(L_n)) = \mathrm{Symb}(\{\lambda\}) = 2$.

Zu ii. Für $n \geq 3$ sei

$$L_n = \{ab^{\pi_{n-1}}\}.$$

Nach Lemma 4.15 gilt $\mathrm{Symb}(L_n) = n$. Die Anwendung des Homomorphismus h ergibt $\mathrm{h}(L_n) = \{a\}$. In Bezug auf Lemma 4.1 erhalten wir $\mathrm{Symb}(\mathrm{h}(L_n)) = \mathrm{Symb}(\{a\}) = 3$.

Zu iii. Für $n \geq 4$ erhalten wir mit der Sprache

$$L_n = \{acb^{\pi_{n-2}}\}$$

analog zum Beweis von Anstrich ii die Aussage $\mathrm{Symb}(L_n) = n$ und $\mathrm{Symb}(\mathrm{h}(L_n)) = \mathrm{Symb}(\{ac\}) = 4$.

Zu iv. Für $n \geq 6$ und $5 \leq k \leq n-1$ setzen wir

$$\begin{aligned} k' &= k-2, \\ L_{n,k} &= \{a^{\pi_{k'}}\} \cup \{b^{\pi_{n-k'}}\}. \end{aligned}$$

Da $L_{n,k}$ aus der Vereinigung der beiden Mengen $\{a^{\pi_{k'}}\}$ und $\{b^{\pi_{n-k'}}\}$ besteht und beide Mengen keine Buchstaben gemeinsam haben, dürfen wir Satz 2.22 anwenden, woraus sich unter Berücksichtigung von Lemma 4.3

$$\mathrm{Symb}(L_{n,k}) = \mathrm{Symb}(\{a^{\pi_{k'}}\}) + \mathrm{Symb}(\{b^{\pi_{n-k'}}\}) = k' + (n-k') = n$$

ergibt. Die Anwendung des Homomorphismus h auf $L_{n,k}$ liefert die Sprache

$$\mathrm{h}(L_{n,k}) = \{a^{\pi_{k'}}\} \cup \{\lambda\}.$$

Aus Lemma 4.3 und Lemma 4.4 lesen wir

$$\mathrm{Symb}(\mathrm{h}(L_{n,k})) = \mathrm{Symb}(\{a^{\pi_{k'}}\}) + 2 = k' + 2 = (k-2) + 2 = k$$

ab.

Zu v. Für $n \geq 2$ legen wir

$$L_n = \{a^{\pi_n}\}$$

fest. Nach Lemma 4.3 gilt $\mathrm{Symb}(L_n) = n$. Die Anwendung des Homomorphismus h ergibt $\mathrm{h}(L_n) = \{a^{\pi_n}\} = L_n$. Es folgt $\mathrm{Symb}(\mathrm{h}(L_n)) = \mathrm{Symb}(L_n) = n$. □

Satz 4.38 – Teilbestimmung des Raums der Symbolkomplexität unter Homomorphismus

Es sei $n \in \mathbb{N}_0$.

i. Für zwei Alphabete X, Y und einen beliebigen Homomorphismus $\mathrm{h}\colon X^* \to Y^*$ gelten

$$\begin{aligned} \{0\} &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(0), \\ \emptyset &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(1), \\ \{2\} &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(2), \\ &\mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n) \not\ni 0, 1, 3 + (n-2) \cdot \max(\{|\mathrm{h}(x)| \mid x \in X\}), \ldots \text{ für } n \geq 3. \end{aligned}$$

ii. Es existiert ein Homomorphismus h, für den die folgenden Aussagen korrekt sind:

$$\begin{aligned} \{0\} &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(0), \\ \emptyset &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(1), \\ \{2, \ldots, n\} &= \mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n) \text{ für } n \geq 2. \end{aligned}$$

iii. Für die Menge aller Homomorphismen Hom *gelten*

$$\begin{aligned} \{0\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}}(0), \\ \emptyset &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}}(1), \\ \{2\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}}(2), \\ \mathbb{N}^{\not\ni 1} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}}(n) \textit{ für } n \geq 3. \end{aligned}$$

Beweis: Anstrich i und ii sind eine Zusammenfassung der Lemmata 4.16, 4.34 und 4.37. Insbesondere verwenden wir in Anstrich ii konkret den Homomorphismus $\mathrm{h}\colon \{a,b\}^* \to \{b\}^*$ mit $\mathrm{h}(a) = a$, $\mathrm{h}(b) = \lambda$, für den unter den Voraussetzungen $n, k \in \mathbb{N}^{\not\ni 1}$ und $n \geq 3$ die Aussage

$$\begin{aligned} k \notin \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}}(n) \text{ für } k &\geq 3 + (n-2) \cdot \max(\{|\mathrm{h}(x)| \mid x \in X\}) \\ &= 3 + (n-2) \cdot \max(\{\mathrm{h}(a), \mathrm{h}(b)\}) \\ &= 3 + (n-2) \cdot \max(\{1, 0\}) \\ &= 3 + (n-2) \cdot 1 \\ &= n + 1 \end{aligned}$$

gilt. Der Anstrich iii ergibt sich aus dem Anstrich i und dem Lemma 4.35. □

Bemerkung: Es seien $n \in \mathbb{N}_0$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 4.38 vollständig dargelegt:

- Es seien zwei Alphabete X, Y, ein beliebiger Homomorphismus $\mathrm{h}\colon X^* \to Y^*$ und $m = \max(\{|\mathrm{h}(x)| \mid x \in X\})$ gegeben.

$$\left.\begin{array}{ll} 3, \ldots, n & \text{für } m = 1 \\ 3, \ldots, 2 + (n-2) \cdot m & \text{für } m \geq 2 \end{array}\right\} \in^? \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}}(n) \text{ für } n \geq 3.$$

  Für $m = 0$ ist h ein total-löschender Homomorphismus wie in Lemma 4.36.

△

4.7. Inverser Homomorphismus

Wir untersuchen den Raum der Symbolkomplexität unter inversem Homomorphismu.

Lemma 4.39 – Es gilt $\{0\} = \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{-1}}(0)$
Für einen inversen Homomorphismus h^{-1} gilt $\{0\} = \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{-1}}(0)$.

Beweis: Es sei L eine kontextfreie Sprache mit $\mathrm{Symb}(L) = 0$. Dann folgt $L = \emptyset$ nach Lemma 4.1, und aufgrund dessen $\mathrm{h}^{-1}(L) = \mathrm{h}^{-1}(\emptyset) = \emptyset$. Nach Lemma 4.1 benötigen wir für $\emptyset$ genau 0 Symbole – mithin gilt $\mathrm{Symb}(\mathrm{h}^{-1}(L)) = \mathrm{Symb}(\emptyset) = 0$. □

Lemma 4.40 – Es gibt kein h, sodass $0 \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{-1}}(2)$ gilt
Für einen inversen Homomorphismus h^{-1} gilt $0 \notin \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{-1}}(2)$.

Beweis: Es gibt nach Lemma 4.1 nur eine kontextfreie Sprache L, für die $\mathrm{Symb}(L) = 2$ gilt – nämlich $L = \{\lambda\}$. Offenbar ist $\lambda \in \mathrm{h}^{-1}(L) = \mathrm{h}^{-1}(\{\lambda\})$ korrekt, da $\mathrm{h}(\lambda) = \lambda$ stimmt. Soll für eine kontextfreie Sprache J allerdings $\mathrm{Symb}(J) = 0$ gelten, so folgt nach Lemma 4.1 unmittelbar $J = \emptyset$ als einzige Möglichkeit. Mithin kann $\mathrm{Symb}(\mathrm{h}^{-1}(L)) = 0$ nicht gelten. $\square$

Lemma 4.41 – Es gibt ein $\mathrm{h}_{k,n}$, sodass $3k \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}_{k,n}^{-1}}(n)$ gilt
Für $n \in \mathbb{N}^{\not\ni 1}$, $n \geq 3$, $k \in \mathbb{N}$, zwei Alphabete $X_k = \{a_i \,|\, 1 \leq i \leq k\}$, $Y = \{a\}$ und den Homomorphismus

$$\mathrm{h}_{k,n}\colon {X_k}^* \to Y^* \text{ mit } \mathrm{h}_{k,n}(a_i) = a^{\pi_n} \text{ für } 1 \leq i \leq k$$

gilt $3k \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}_{k,n}^{-1}}(n)$.

Beweis: Es sei $L_n = \{a^{\pi_n}\}$. Nach Lemma 4.3 gilt $\mathrm{Symb}(L_n) = n$. Die Anwendung des inversen Homomorphismus ergibt $\mathrm{h}_{k,n}^{-1}(L_n) = \{a_i \,|\, 1 \leq i \leq k\}$. Lemma 4.7 liefert

$$\mathrm{Symb}(\mathrm{h}_{k,n}^{-1}(L_n)) = \mathrm{Symb}(\{a_i \mid 1 \leq i \leq k\}) = 3k.$$

$\square$

Lemma 4.42 – Es gibt ein h_k, sodass $3k + 6 \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}_k^{-1}}(2)$ gilt
Für $k \in \mathbb{N}$, zwei Alphabete $X_k = \{a_i \,|\, 1 \leq i \leq k\}$, $Y = \{a\}$ und den Homomorphismus

$$\mathrm{h}_k\colon {X_k}^* \to Y^* \text{ mit } \mathrm{h}_k(a_i) = \lambda \text{ für } 1 \leq i \leq k$$

gilt

$$\left.\begin{matrix} 4k+2 & \text{für } k \leq 3 \\ 3k+6 & \text{sonst} \end{matrix}\right\} \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}_k^{-1}}(2).$$

Beweis: Nach Lemma 4.1 kommt nur $L = \{\lambda\}$ als eine Sprache mit der Symbolkomplexität 2 infrage. Die Anwendung des inversen Homomorphismus ergibt $\mathrm{h}_k^{-1}(L) = \mathrm{h}_k^{-1}(\{\lambda\}) = \{a_i \,|\, 1 \leq i \leq k\}^*$. Lemma 4.8 liefert

$$\mathrm{Symb}(\mathrm{h}_k^{-1}(L)) = \mathrm{Symb}(\{a_i \mid 1 \leq i \leq k\}^*) = \begin{cases} 4k+2 & \text{für } k \leq 3 \\ 3k+6 & \text{sonst.} \end{cases}$$

$\square$

Lemma 4.43 – Ergebnisse zum Raum der Symbolkomplexität unter einem speziellen inversen Homomorphismus
Es seien $n \in \mathbb{N}_0^{\not\ni 1}$ und $X = \{a, b\}$, $Y = \{a, b, c\}$ zwei Alphabete sowie

$$\mathrm{h}\colon X^* \to Y^* \text{ mit } \mathrm{h}(a) = c, \mathrm{h}(b) = b$$

ein Homomorphismus. Dann gelten

i. $0 \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-1}}}(n)$ für $n \neq 2$,

ii. $\{2\} = \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-1}}}(2)$,

iii. $\{0, 3\} = \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-1}}}(3)$,

iv. $\{0, 4\} = \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-1}}}(4)$,

v. $2, \ldots, n-3, n \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-1}}}(n)$ für $n \geq 5$.

Beweis: Zu i. Für $n \in \mathbb{N}_0^{\neq 1}$, $n \neq 2$ sei

$$L_n = \begin{cases} \emptyset & \text{für } n = 0 \\ \{a^{\pi_n}\} & \text{sonst} \end{cases}$$

eine kontextfreie Sprache. Für $n = 0$ gilt aufgrund des Lemmas 4.1 die Aussage $\mathrm{Symb}(L_n) = \mathrm{Symb}(\emptyset) = 0$ und für $n \geq 3$ gilt wegen Lemma 4.3 die Aussage $\mathrm{Symb}(L_n) = \mathrm{Symb}(\{a^{\pi_n}\}) = n$.

Durch die Anwendung des inversen Homomorphismus $\mathrm{h}^{\text{-1}}$ ergibt sich $\mathrm{h}^{\text{-1}}(L_n) = \emptyset$. Lemma 4.1 begründet $\mathrm{Symb}(\mathrm{h}^{\text{-1}}(L_n)) = \mathrm{Symb}(\emptyset) = 0$.

Zu ii. Es sei L eine kontextfreie Sprache mit $\mathrm{Symb}(L) = 2$. Aus Lemma 4.1 lesen wir dann $L = \{\lambda\}$ ab. Die Anwendung des inversen Homomorphismus $\mathrm{h}^{\text{-1}}$ ergibt $\mathrm{h}^{\text{-1}}(L) = \mathrm{h}^{\text{-1}}(\{\lambda\}) = \{\lambda\}$. Lemma 4.1 erneut anwendend, erhalten wir $\mathrm{Symb}(\mathrm{h}^{\text{-1}}(L)) = \mathrm{Symb}(\{\lambda\}) = 2$.

Zu iii. Es sei L eine kontextfreie Sprache mit $\mathrm{Symb}(L) = 3$. Nach Lemma 4.1 gilt dann $L = \{x\}$ für ein $x \in Y$. Hieraus ergeben sich die folgenden möglichen Fälle:

- Aus $x = a$ folgt $\mathrm{h}^{\text{-1}}(L) = \mathrm{h}^{\text{-1}}(\{a\}) = \emptyset$,
- aus $x = b$ folgt $\mathrm{h}^{\text{-1}}(L) = \mathrm{h}^{\text{-1}}(\{b\}) = \{b\}$,
- aus $x = c$ folgt $\mathrm{h}^{\text{-1}}(L) = \mathrm{h}^{\text{-1}}(\{c\}) = \{a\}$.

Aufgrund von Lemma 4.1 kommen somit nur $\mathrm{Symb}(\mathrm{h}^{\text{-1}}(L)) = \mathrm{Symb}(\mathrm{h}^{\text{-1}}(\emptyset)) = 0$ bzw. $\mathrm{Symb}(\mathrm{h}^{\text{-1}}(L)) = \mathrm{Symb}(\mathrm{h}^{\text{-1}}(\{b\})) = \mathrm{Symb}(\mathrm{h}^{\text{-1}}(\{a\})) = 3$ infrage.

Zu iv. Die gewünschte Aussage lässt sich mithilfe von Lemma 4.1 analog zum Beweis der Aussage in Anstrich iii zeigen.

Zu v. Für $n \geq 5$ und $2 \leq k \leq n-3$ sei

$$L_{n,k} = \{a^{\pi_{n-k}}\} \cup \{b^{\pi_k}\}$$

eine kontextfreie Sprache. Wir deklarieren $A_{n,k} = \{a^{\pi_{n-k}}\}$ und $B_k = \{b^{\pi_k}\}$ – es gilt $L_{n,k} = A_{n,k} \cup B_k$. Aufgrund von Lemma 4.3 erhalten wir die Aussagen $\mathrm{Symb}(A_{n,k}) = n-k$

und $\mathrm{Symb}(B_k) = k$. Daraus ergibt sich wegen der Disjunktheit der Alphabete von $A_{n,k}$ und B_k mittels Satz 2.22 und Lemma 4.4 die Symbolkomplexität

$$\begin{aligned}\mathrm{Symb}(L_{n,k}) &= \mathrm{Symb}(A_{n,k} \cup B_k) = \mathrm{Symb}(A_{n,k}) + \mathrm{Symb}(B_k) \\ &= (n-k) + k = n.\end{aligned}$$

Es gilt $\mathrm{h}^{\text{-}1}(L_{n,k}) = \{b^{\pi_k}\}$. Aus Lemma 4.3 lesen wir

$$\mathrm{Symb}(\mathrm{h}^{\text{-}1}(L_{n,k})) = \mathrm{Symb}(\{b^{\pi_k}\}) = k$$

ab.

Um $n \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(n)$, $n \geq 5$, zu zeigen, setzen wir

$$L_n = \{b^{\pi_n}\}.$$

Lemma 4.3 liefert $\mathrm{Symb}(L_n) = n$. Die inverse Anwendung von h ergibt $\mathrm{h}^{\text{-}1}(L_n) = \{b^{\pi_n}\} = L_n$. Mithin gilt $\mathrm{Symb}(\mathrm{h}^{\text{-}1}(L_n)) = \mathrm{Symb}(L_n) = n$. □

Satz 4.44 – Teilbestimmung des Raums der Symbolkomplexität unter inversem Homomorphismus

Es seien $n \in \mathbb{N}_0$ und $k \in \mathbb{N}$.

i. Für einen beliebigen inversen Homomorphismus $\mathrm{h}^{\text{-}1}$ *gelten*

$$\begin{aligned}\{0\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(0),\\ \emptyset &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(1),\\ 0,1 &\notin \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(2),\\ 1 &\notin \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(n) \text{ für } n \geq 3.\end{aligned}$$

ii. Es existiert ein inverser Homomorphismus $\mathrm{h}^{\text{-}1}$*, für den die folgenden Aussagen korrekt sind:*

$$\begin{aligned}\{0\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(0),\\ \emptyset &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(1),\\ \{2\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(2),\\ \{0,3\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(3),\\ \{0,4\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(4),\\ 0,2,\ldots,n-3,n &\in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(n) \not\ni 1 \text{ für } n \geq 5.\end{aligned}$$

iii. Für die Menge aller inversen Homomorphismen $\mathrm{Hom}^{\text{-}1}$ *gilt*

$$\{0\} = \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(0),$$

$$\begin{aligned} \emptyset &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(1), \\ 2, 6, 10, 14, 3k+6 &\in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(2) \not\ni 0, 1 \textit{ für } k \geq 4, \\ 0, 3k &\in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(3) \not\ni 1, \\ 0, 4, 3k &\in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(4) \not\ni 1, \\ 0, 2, \ldots, n-3, n, 3k &\in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(n) \not\ni 1 \textit{ für } n \geq 5. \end{aligned}$$

Beweis: Die Anstriche i, ii und iii sind eine Zusammenfassung der Lemmata 4.16, 4.39, 4.40, 4.41, 4.42 und 4.43. □

Bemerkung: Es seien $n, k \in \mathbb{N}_0$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 4.44 vollständig dargelegt:

- Gegeben sei ein beliebiger inverser Homomorphismus $\mathrm{h}^{\text{-}1}$:

$$\begin{aligned} k &\in^? \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(2) \text{ für } k \geq 2, \\ 0, k &\in^? \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(n) \text{ für } k \geq 2, n \geq 3. \end{aligned}$$

- Für den in Anstrich ii verwendeten inversen Homomorphismus:

$$n-2, n-1, n+1, \ldots \in^? \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{\text{-}1}}(n) \text{ für } n \geq 5.$$

- Für die Menge aller inversen Homomorphismen $\mathrm{Hom}^{\text{-}1}$:

$$\begin{aligned} 3, 4, 5, 7, 8, 9, 11, 12, 13, 15, 16, 17, 3k+7, 3k+8 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(2) \text{ für } k \geq 4, \\ 2, 3k+1, 3k+2 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(3), \\ 2, 3k+1, 3k+2 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(4), \\ n-2, n-1, 3k+1, 3k+2 &\in^? \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Hom}^{\text{-}1}}(n) \text{ für } n \geq 5. \end{aligned}$$

△

4.8. Schnitt mit regulärer Sprache

Wir untersuchen den Raum der Symbolkomplexität unter Schnitt mit regulärer Sprache.

Lemma 4.45 – Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Schnitt mit regulärer Sprache

Für eine beliebige reguläre Sprache R und $n, k \in \mathbb{N}^{\not\ni 1}$ gelten

i. $\{0\} = \mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R}}}(0)$,

ii. $0 \in \mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R}}}(n)$ für $n \geq 3$, iii.

$$\left.\begin{array}{ll} \{0\} & \lambda \notin R \\ \{2\} & \text{sonst} \end{array}\right\} = \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(2),$$

iv.

$$\begin{aligned} &\{0\} \cup \\ \{x \mid x \in \{3\}, \mathrm{Alph}(R) \cap R \neq \emptyset\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(3), \end{aligned}$$

v.

$$\begin{aligned} &\{0\} \cup \\ \left\{x \,\middle|\, x \in \{4\}, \mathrm{Alph}(R)^2 \cap R \neq \emptyset\right\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(4), \end{aligned}$$

vi.

$$\begin{aligned} \{0\} \cup \\ \{x \mid x \in \{2\}, \lambda \in R\} \cup \\ \{x \mid x \in \{3\}, \lambda \notin R, \mathrm{Alph}(R) \cap R \neq \emptyset\} \cup \\ \{x \mid x \in \{5\}, \lambda \in R, \mathrm{Alph}(R) \cap R \neq \emptyset\} \cup \\ \left\{x \,\middle|\, x \in \{5\}, \mathrm{Alph}(R)^3 \cap R \neq \emptyset\right\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(5), \end{aligned}$$

vii.

$$\begin{aligned} \lambda \in R \Leftrightarrow 2 \in \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(n) \text{ für } n \geq 6, \\ \mathrm{Alph}(R) \cap R \neq \emptyset \Leftrightarrow 3 \in \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(n) \text{ für } n \geq 6, \\ \mathrm{Alph}(R)^2 \cap R \neq \emptyset \Leftrightarrow 4 \in \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(n) \text{ für } n \geq 7. \end{aligned}$$

Beweis: Es seien L eine kontextfreie Sprache, T ein Alphabet und nicht notwendig verschiedene $a, b, c \in T$. Aufgrund der Endlichkeit von $\mathrm{Alph}(R)$ existiert stets ein Alphabet X, sodass X und $\mathrm{Alph}(R)$ disjunkt sind. Wir wählen einen Buchstaben $x \in X$ – es gilt $x \notin \mathrm{Alph}(R)$.

Zu i. Es sei $\mathrm{Symb}(L) = 0$. Nach Lemma 4.1 gilt dann stets $L = \emptyset$. Somit folgt

$$\cap_\mathrm{R}(L) = \cap_\mathrm{R}(\emptyset) = \emptyset,$$

und daraus mithilfe von Lemma 4.1 die Aussage

$$\mathrm{Symb}(\cap_\mathrm{R}(L)) = \mathrm{Symb}(\emptyset) = 0.$$

Zu ii. Für die kontextfreie Sprache $L_n = \{x^{\pi_n}\}$, $n \geq 3$, erhalten wir nach Lemma 4.3 die Aussage $\mathrm{Symb}(\{x^{\pi_n}\}) = n$. Der Schnitt mit R ergibt

$$\cap_\mathrm{R}(L_n) = \cap_\mathrm{R}(\{x^{\pi_n}\}) = \emptyset,$$

da der Buchstabe x nicht in $\mathrm{Alph}(R)$ enthalten ist. Lemma 4.1 zufolge gilt dann

$$\mathrm{Symb}(\cap_\mathrm{R}(L_n)) = \mathrm{Symb}(\emptyset) = 0.$$

Zu iii. Es sei $\mathrm{Symb}(L) = 2$. Nach Lemma 4.1 kommt dann nur $L = \{\lambda\}$ infrage. Im Fall von $\lambda \in R$ folgt stets

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{\lambda\}) = \{\lambda\},$$

und damit

$$\operatorname{Symb}(\cap_{\mathrm{R}}(L)) = \operatorname{Symb}(\{\lambda\}) = 2$$

mithilfe von Lemma 4.1. Im Fall $\lambda \notin R$ folgen stets

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{\lambda\}) = \{\emptyset\}$$

und

$$\operatorname{Symb}(\cap_{\mathrm{R}}(L)) = \operatorname{Symb}(\{\emptyset\}) = 0$$

unter Beachtung von Lemma 4.1.

Zu iv. Es sei $\operatorname{Symb}(L) = 3$. Nach Lemma 4.1 gilt dann $L = \{a\}$. Ist a nicht in R enthalten, folgt

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{a\}) = \{\emptyset\},$$

und damit

$$\operatorname{Symb}(\cap_{\mathrm{R}}(L)) = \operatorname{Symb}(\{\emptyset\}) = 0$$

unter Zuhilfenahme von Lemma 4.1. Ist a in R enthalten, gelten andererseits

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{a\}) = \{a\}$$

und

$$\operatorname{Symb}(\cap_{\mathrm{R}}(L)) = \operatorname{Symb}(\{a\}) = 3$$

(Lemma 4.1).

Wir konstruieren die kontextfreie Sprache $J = \{x\}$, für die nach Lemma 4.1 immer $\operatorname{Symb}(J) = 3$ gilt. Es folgen

$$\cap_{\mathrm{R}}(J) = \cap_{\mathrm{R}}(\{x\}) = \emptyset$$

und

$$\operatorname{Symb}(\cap_{\mathrm{R}}(J)) = \operatorname{Symb}(\emptyset) = 0$$

(Lemma 4.1). Gilt $\operatorname{Alph}(R) \cap R \neq \emptyset$, setzen wir $J = \{y\}$ für ein $y \in \operatorname{Alph}(R) \cap R$, wodurch nach Lemma 4.1 die Bedingung $\operatorname{Symb}(J) = 3$ nicht verletzt wird. Es folgen

$$\cap_{\mathrm{R}}(J) = \cap_{\mathrm{R}}(\{y\}) = \{y\}$$

und

$$\operatorname{Symb}(\cap_{\mathrm{R}}(J)) = \operatorname{Symb}(\{y\}) = 3$$

(Lemma 4.1).

Zu v. Den Beweis können wir mithilfe von Lemma 4.1 analog zum Beweis von Anstrich iv führen. Statt $L = \{a\}$ verwenden wir $L = \{ab\}$, und anstelle von $J = \{x\}$ die Sprache $J = \{xx\}$ bzw. statt $J = \{y\}$ die Sprache $J = \{yz\}$ für $y, z \in \operatorname{Alph}(R)^2 \cap R$, falls $\operatorname{Alph}(R)^2 \cap R \neq \emptyset$ gilt.

Zu vi. Es sei $\operatorname{Symb}(L) = 5$. Aus Lemma 4.1 lesen wir die Möglichkeiten $L = \{abc\}$

oder $L = \{\lambda, a\}$ ab.

In den Fällen, in denen $abc \notin R$ und $\lambda, a \notin R$ sind, gelten

$$\cap_{\mathrm{R}}(L) = \emptyset$$

und

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = \mathrm{Symb}(\emptyset) = 0$$

(Lemma 4.1).

Im Fall von $\lambda \notin R$ und $a \in R$ folgen

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{\lambda, a\}) = \{a\}$$

und

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = \mathrm{Symb}(\{a\}) = 3$$

durch Setzen von $L = \{\lambda, a\}$ und mithilfe von Lemma 4.1. Analoges gilt für $\lambda \in R, a \in R$, nur dass $\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = 5$ folgt.

Gilt $abc \in R$, erhalten wir, indem wir $L = \{abc\}$ verwenden und Lemma 4.1 anwenden,

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{abc\}) = \{abc\}$$

und

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = \mathrm{Symb}(\{abc\}) = 5.$$

Wir setzen $J = \{xxx\}$. Nach Lemma 4.1 gilt $\mathrm{Symb}(J) = 5$. Es ergeben sich

$$\cap_{\mathrm{R}}(J) = \cap_{\mathrm{R}}(\{xxx\}) = \emptyset$$

und

$$\mathrm{Symb}(\cap_{\mathrm{R}}(J)) = \mathrm{Symb}(\emptyset) = 0$$

(Lemma 4.1).

Für den Fall $\lambda \in R$ gelten unter Verwendung von $J = \{\lambda, x\}$ ($\mathrm{Symb}(J) = 5$ nach Lemma 4.1) die Aussagen

$$\cap_{\mathrm{R}}(J) = \cap_{\mathrm{R}}(\{\lambda, x\}) = \{\lambda\}$$

und

$$\mathrm{Symb}(\cap_{\mathrm{R}}(J)) = \mathrm{Symb}(\{\lambda\}) = 2$$

(Lemma 4.1).

Gelten $\lambda \notin R$ und $\mathrm{Alph}(R) \cap R \neq \emptyset$, können wir $L = \{\lambda, y\}$ für ein $y \in \mathrm{Alph}(R) \cap R$ wählen ($\mathrm{Symb}(L) = 5$ nach Lemma 4.1), und es folgen

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{\lambda, y\}) = \{y\}$$

sowie

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = \mathrm{Symb}(\{y\}) = 3$$

wegen Lemma 4.1. Analoges gilt für $\lambda \in R$ und $\mathrm{Alph}(R) \cap R \neq \emptyset$, nur dass die Aussage $\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = 5$ folgt.

Gilt $\mathrm{Alph}(R)^3 \cap R \neq \emptyset$, setzen wir $L = \{uvy\}$ für ein $uvy \in \mathrm{Alph}(R)^3 \cap R$. Nach Lemma 4.1 erhalten wir $\mathrm{Symb}(L) = 5$. Es folgen

$$\cap_{\mathrm{R}}(L) = \cap_{\mathrm{R}}(\{uvy\}) = \{uvy\}$$

und

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = \mathrm{Symb}(\{uvy\}) = 5,$$

indem wir Lemma 4.1 anwenden.

Zu vii. Gilt $\lambda \in R$, können wir die kontextfreie Sprache $L_n = \{\lambda, x^{\pi_{n-2}}\}$, für die nach den Lemmata 4.3 und 4.4 die Aussage $\mathrm{Symb}(L_n) = 2 + (n-2) = n$ korrekt ist, angeben ($n \geq 6$), und es folgen

$$\cap_{\mathrm{R}}(L_n) = \cap_{\mathrm{R}}(\{\lambda, x^{\pi_{n-2}}\}) = \{\lambda\}$$

sowie

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L_n)) = \mathrm{Symb}(\{\lambda\}) = 2$$

mithilfe von Lemma 4.1. Gilt andererseits $\lambda \notin R$, folgt für eine kontextfreie Sprache L_n mit $\mathrm{Symb}(L_n) = n$ stets $\lambda \notin \cap_{\mathrm{R}}(L_n)$. Lemma 4.1 impliziert dann die Aussage $\mathrm{Symb}(\cap_{\mathrm{R}}(L_n)) \neq 2$.

Gilt $\mathrm{Alph}(R) \cap R \neq \emptyset$, können wir die kontextfreie Sprache $L_n = \{y, x^{\pi_{n-3}}\}$ für ein $y \in \mathrm{Alph}(R) \cap R$ und $n \geq 6$ angeben, für die nach Satz 2.22 und Lemma 4.3 die Aussage $\mathrm{Symb}(L_n) = 3 + (n-3) = n$ wahr ist. Es ergeben sich

$$\cap_{\mathrm{R}}(L_n) = \cap_{\mathrm{R}}(\{y, x^{\pi_{n-3}}\}) = \{y\}$$

und

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L_n)) = \mathrm{Symb}(\{y\}) = 3$$

aufgrund von Lemma 4.1. Gilt andererseits $3 = \mathrm{Symb}(\cap_{\mathrm{R}}(L_n))$ für eine kontextfreie Sprache L_n mit $\mathrm{Symb}(L_n) = n$, folgt nach Lemma 4.1 die Tatsache $\{y\} = \cap_{\mathrm{R}}(L_n)$ für ein $y \in \mathrm{Alph}(\cap_{\mathrm{R}}(L_n))$, und mithin gilt $y \in R$.

Analog zum vorherigen Absatz kann die letzte Aussage bewiesen werden, indem $L_n = \{uv, x^{\pi_{n-4}}\}$ für $uv \in \mathrm{Alph}(R)^2 \cap R$ und $n \geq 7$ gesetzt wird. $\square$

Lemma 4.46 – Es gibt ein R, sodass $2, \ldots, n-3, n \in \mathrm{C}^{\mathbf{Prod}}_{\cap_{\mathbf{R}}}(n)$ gilt

Für ein Alphabet T, $a \in T$ und die reguläre Sprache $R = \{a\}^*$ sowie $n \in \mathbb{N}_0^{\not\ni 1}$ gelten

i. $2, \ldots, n-3 \in \mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R}}}(n) \not\ni n+1, \ldots$ für $n \geq 5$,

ii. $n \in \mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R}}}(n)$ für $n \geq 2$.

Beweis: Es seien T ein beliebiges Alphabet mit mindestens 2 Buchstaben und paarweise verschiedene $a, b \in T$ sowie $k \in \mathbb{N}_0^{\not\ni 1}$.

Zu i. Für $n \geq 5$ und $2 \leq k \leq n-3$ sei

$$L_{n,k} = \{a^{\pi_k}\} \cup \{b^{\pi_{n-k}}\}$$

eine kontextfreie Sprache. Wir setzen $A_k = \{a^{\pi_k}\}$ und $B_{n,k} = \{b^{\pi_{n-k}}\}$ – es gilt $L_{n,k} = A_k \cup B_{n,k}$. Aufgrund von Lemma 4.3 erhalten wir die Symbolkomplexitäten $\mathrm{Symb}(A_k) = k$ und $\mathrm{Symb}(B_{n,k}) = n - k$. Daraus ergibt sich im Fall von $k = 2$, also $A_2 = \{\lambda\}$, wegen Lemma 4.4 sowie im Fall von $k \geq 3$ wegen der Disjunktheit der Alphabete von A_k und $B_{n,k}$ unter Anwendung des Satzes 2.22 die Aussage

$$\begin{aligned}\mathrm{Symb}(L_{n,k}) &= \mathrm{Symb}(A_k \cup B_{n,k}) = \mathrm{Symb}(A_k) + \mathrm{Symb}(B_{n,k}) \\ &= k + (n-k) = n.\end{aligned}$$

Für den Schnitt von $L_{n,k}$ mit R ergibt sich

$$\cap_{\mathrm{R}}(L_{n,k}) = \cap_{\{\mathrm{a}\}^*}(\{a^{\pi_k}\} \cup \{b^{\pi_{n-k}}\}) = \{a^{\pi_k}\} = A_k,$$

woraus

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L_{n,k})) = \mathrm{Symb}(A_k) = k$$

folgt.

Es sei L eine kontextfreie Sprache und $G_L = (N_L, T_L, P_L, S_L)$ eine regel-minimale kontextfreie Grammatik, die L erzeugt. Wir zeigen nun $\mathrm{Symb}(\cap_{\mathrm{R}}(L)) \leq n$. Die Sprache $\cap_{\mathrm{R}}(L) = \cap_{\{\mathrm{a}\}^*}(L)$ enthält nur Wörter, die aus dem Buchstaben a bestehen und in L enthalten sind, oder das leere Wort λ, falls $\lambda \in L$ ist. Aufgrund dieser Tatsache können wir leicht eine kontextfreie Grammatik G konstruieren, die die Sprache $\cap_{\mathrm{R}}(L)$ erzeugt, indem wir alle Regeln p, die in Ableitungen der Art $S_L \Longrightarrow^*_{G_L} w_a \in \{a\}^*$ angewendet werden in G einfügen. Die Regeln p enthalten nur Buchstaben aus $N_L \cup \{a\}$. Somit gilt stets

$$\mathrm{Symb}(\cap_{\mathrm{R}}(L)) = \mathrm{Symb}(G) \leq \mathrm{Symb}(G_L) = \mathrm{Symb}(L) = n.$$

Zu ii. Für $n \geq 2$ legen wir die kontextfreie Sprache

$$L_n = \{a^{\pi_n}\}$$

fest. Lemma 4.3 impliziert $\mathrm{Symb}(L_n) = n$. Wir erhalten

$$\cap_{\mathrm{R}}(L_n) = \cap_{\{\mathrm{a}\}^*}(\{a^{\pi_n}\}) = \{a^{\pi_n}\} = L_n,$$

woraus sofort $\mathrm{Symb}(\cap_{\mathrm{R}}(L_n)) = \mathrm{Symb}(L_n) = n$ folgt. □

Lemma 4.47 – Es gibt ein R_k, sodass $k \in \mathbf{C}^{\mathbf{Symb}}_{\cap_{\mathbf{R_k}}}(n)$ gilt

Unter den Voraussetzungen $n \in \mathbb{N}$, $n \geq 13$, und $k \in \mathbb{N}$, $k \geq 3$, sowie einem Alphabet T, $a \in T$ und der regulären Sprache $R_k = \{b^{\pi_k}\}$ gilt $k \in \mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R_k}}}(n)$.

Beweis: Es seien T ein Alphabet mit mindestens 2 Buchstaben und $a, b \in T$ paarweise verschieden sowie

$$L_n = \{a^{\pi_{n-10}}\} \cup \{b\}^+.$$

Wir legen $A_n = \{a^{\pi_{n-10}}\}$ und $B = \{b\}^+$ fest, wodurch sicher $L_n = A_n \cup B$ gilt. Die Symbolkomplexitäten $\mathrm{Symb}(A_n) = n - 10$ und $\mathrm{Symb}(B) = 7$ lesen wir aus Lemma 4.3

bzw. Lemma 4.1 ab. Infolgedessen ergibt sich hinsichtlich der Disjunktheit der Alphabete von A_n und B mittels Lemma 4.6 die Aussage

$$\begin{aligned}\mathrm{Symb}(L_n) &= \mathrm{Symb}(A_n \cup B) = \mathrm{Symb}(A_n) + \mathrm{Symb}(B) + 3\\ &= (n-10) + 7 + 3 = n.\end{aligned}$$

Die Schnittmenge $\cap_{\mathrm{R_k}}(L_n)$ ist R_k. Lemma 4.3 impliziert $\mathrm{Symb}(R_k) = k$, woraus $\mathrm{Symb}(\cap_{\mathrm{R_k}}(L_n)) = \mathrm{Symb}(R_k) = k$ folgt. □

Satz 4.48 – Teilbestimmung des Raums der Symbolkomplexität unter Schnitt mit regulärer Sprache

Es sei $n \in \mathbb{N}_0$.

i. Für eine reguläre Sprache R und nicht notwendig verschiedene $a, b, c \in \mathrm{Alph}(R)$ gelten

$$\begin{aligned}
\{0\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(0),\\
\emptyset &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(1),\\
\left.\begin{matrix}\{2\} & \lambda \in R\\ \{0\} & \textit{sonst}\end{matrix}\right\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(2),\\
\{0\} \cup{}&\\
\{x \mid x \in \{3\}, \mathrm{Alph}(R) \cap R \neq \emptyset\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(3),\\
\{0\} \cup{}&\\
\left\{x \,\middle|\, x \in \{4\}, \mathrm{Alph}(R)^2 \cap R \neq \emptyset\right\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(4),\\
\{0\} \cup{}&\\
\{x \mid x \in \{2\}, \lambda \in R\} \cup{}&\\
\{x \mid x \in \{3\}, \lambda \notin R, \mathrm{Alph}(R) \cap R \neq \emptyset\} \cup{}&\\
\{x \mid x \in \{5\}, \lambda \in R, \mathrm{Alph}(R) \cap R \neq \emptyset\} \cup{}&\\
\left\{x \,\middle|\, x \in \{5\}, \mathrm{Alph}(R)^3 \cap R \neq \emptyset\right\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(5),\\
0 &\in \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(n) \not\ni 1 \textit{ für } n \geq 6,\\
\lambda \in R \Leftrightarrow 2 &\in \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(n) \textit{ für } n \geq 6,\\
\mathrm{Alph}(R) \cap R \neq \emptyset \Leftrightarrow 3 &\in \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(n) \textit{ für } n \geq 6,\\
\mathrm{Alph}(R)^2 \cap R \neq \emptyset \Leftrightarrow 4 &\in \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(n) \textit{ für } n \geq 7.
\end{aligned}$$

ii. Es existiert eine reguläre Sprache R, für die die folgenden Aussagen korrekt sind:

$$\begin{aligned}
\{0\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(0),\\
\emptyset &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(1),\\
\{2\} &= \mathrm{C}_{\cap_\mathrm{R}}^{\mathrm{Symb}}(2),
\end{aligned}$$

$$\begin{aligned}
\{0,3\} &= \mathrm{C}^{\mathrm{Symb}}_{\cap_\mathrm{R}}(3),\\
\{0,4\} &= \mathrm{C}^{\mathrm{Symb}}_{\cap_\mathrm{R}}(4),\\
\{0,2,5\} &= \mathrm{C}^{\mathrm{Symb}}_{\cap_\mathrm{R}}(5),\\
0,2,\ldots,n-3,n \in \mathrm{C}^{\mathrm{Symb}}_{\cap_\mathrm{R}}&(n) \not\ni 1, n+1, \ldots \quad \textit{für } n \geq 6.
\end{aligned}$$

iii. Für die Menge aller Schnitte mit regulärer Sprache Reg *gelten*

$$\begin{aligned}
\{0\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(0),\\
\emptyset &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(1),\\
\{0,2\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(2),\\
\{0,3\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(3),\\
\{0,4\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(4),\\
\{0,2,3,5\} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(5),\\
0,2,\ldots,n-3,n \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}&(n) \not\ni 1 \quad \textit{für } n \geq 6,\\
\mathbb{N}_0^{\not\ni 1} &= \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(n) \quad \textit{für } n \geq 13.
\end{aligned}$$

Beweis: Anstrich i und ii sind eine Zusammenfassung der Lemmata 4.16, 4.45 und 4.46. Anstrich iii folgt aus Anstrich i und ii sowie Lemma 4.47. □

Bemerkung: Es seien $n, k \in \mathbb{N}_0$. Im Folgenden sind die noch offenen Fälle zu den Ergebnissen aus Satz 4.48 vollständig dargelegt:

- Gegeben sei eine beliebige reguläre Sprache R. Gilt $k \in \mathrm{C}^{\mathrm{Prod}}_{\cap_\mathrm{R}}(n)$ für $k \geq 5$ und $n \geq 6$?
- Existieren reguläre Sprachen, sodass $n-2, n-1, n+1, \ldots \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(n)$ für $6 \leq n \leq 12$ gilt?

△

5. Ergebnisse und Aussicht

Wir fassen die Ergebnisse der Abschnitte 3 und 4 im Folgenden zusammen. Hierbei ignorieren wir einige – für prinzipielle Aussagen nicht notwendige – Ergebnisse, um eine bessere Übersicht zu erhalten. Es seien $n, m \in \mathbb{N}_0$ und stets $m \leq n$. Dann gelten die im Folgenden dargestellten Aussagen über die Räume der Regelkomplexität und Symbolkomplexität sowie über die maximal möglichen Regelkomplexitäten und Symbolkomplexitäten:

Zur Regelkomplexität (aus Abschnitt 3)

- Spiegelbild nach Satz 3.13:

$$\begin{aligned} \mathrm{C}^{\mathrm{Prod}}_{()^{\mathrm{R}}}(n) &= \{n\}, \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{()^{\mathrm{R}}}(n)\right) &= n. \end{aligned}$$

- Vereinigung nach Satz 3.17:

$$\begin{aligned} 6, \ldots, n+m+2 \in \mathrm{C}^{\mathrm{Prod}}_{\cup}(n,m) &= \mathrm{C}^{\mathrm{Prod}}_{\cup}(m,n) \\ &\not\ni n+m+3, \ldots \text{ für } n \geq 2, m \geq 2, \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{\cup}(n,m)\right) &= \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{\cup}(m,n)\right) &= \begin{cases} n & \text{für } m = 0 \\ 2 & \text{für } n = 1, m = 1 \\ n+2 & \text{für } n \geq 2, m = 1 \\ n+m+2 & \text{sonst.} \end{cases} \end{aligned}$$

- Konkatenation nach Satz 3.28:

$$\begin{aligned} n+2, \ldots, n+m+1 \in \mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m) &= \mathrm{C}^{\mathrm{Prod}}_{\cdot}(m,n) \\ &\not\ni n+m+2, \ldots \text{ für } n \geq 5, m \geq 5, \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{\cdot}(n,m)\right) &= \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{\cdot}(m,n)\right) &= \begin{cases} 0 & \text{für } m = 0 \\ 1 & \text{für } n = 1, m = 1 \\ \text{noch offen} & \text{für } n \geq 2, m = 1 \\ n+m+1 & \text{sonst.} \end{cases} \end{aligned}$$

- Kleene-Abschluss nach Satz 3.30:

$$\mathrm{C}_{()^*}^{\mathrm{Prod}}(n) = \begin{cases} \{1\} & \text{für } n = 0 \\ \{1,2\} & \text{für } n = 1 \\ \{2,\dots,n+2\} & \text{sonst,} \end{cases}$$
$$\max\left(\mathrm{C}_{()^*}^{\mathrm{Prod}}(n)\right) = \begin{cases} 1 & \text{für } n = 0 \\ 2 & \text{für } n = 1 \\ n+2 & \text{sonst.} \end{cases}$$

- Homomorphismus nach Satz 3.33:
  Für einen beliebigen Homomorphismus h gelten

  $$\mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n) \not\ni n+1,\dots,$$
  $$\max\left(\mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n)\right) \le n.$$

  Es existiert ein Homomorphismus h, für den gelten

  $$\mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n) = \begin{cases} \{0\} & \text{für } n = 0 \\ \{1,\dots,n\} & \text{sonst,} \end{cases}$$
  $$\max\left(\mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n)\right) = n.$$

  Für die Menge aller Homomorphismen Hom gelten

  $$\mathrm{C}_{\mathrm{Hom}}^{\mathrm{Prod}}(n) = \begin{cases} \{0\} & \text{für } n = 0 \\ \{1,\dots,n\} & \text{sonst,} \end{cases}$$
  $$\max\left(\mathrm{C}_{\mathrm{Hom}}^{\mathrm{Prod}}(n)\right) = n.$$

- Inverser Homomorphismus nach Satz 3.37:
  Für einen beliebigen inversen Homomorphismus $\mathrm{h}^{\text{-}1}$ gelten

  $$\mathrm{C}_{\mathrm{h}^{\text{-}1}}^{\mathrm{Prod}}(n) \text{ und } \max\left(\mathrm{C}_{\mathrm{h}^{\text{-}1}}^{\mathrm{Prod}}(n)\right) \text{ sind noch unbekannt.}$$

  Es existiert ein inverser Homomorphismus $\mathrm{h}^{\text{-}1}$, für den gelten

  $$\{0\} = \mathrm{C}_{\mathrm{h}^{\text{-}1}}^{\mathrm{Prod}}(0),$$
  $$0,\dots,n \in \mathrm{C}_{\mathrm{h}^{\text{-}1}}^{\mathrm{Prod}}(n) \text{ für } n \ge 1,$$
  $$\max\left(\mathrm{C}_{\mathrm{h}^{\text{-}1}}^{\mathrm{Prod}}(n)\right) = \begin{cases} 0 & \text{für } n = 0 \\ \text{noch offen} & \text{sonst.} \end{cases}$$

  Für die Menge aller inversen Homomorphismen $\mathrm{Hom}^{\text{-}1}$ gelten

  $$\{0\} = \mathrm{C}_{\mathrm{Hom}^{\text{-}1}}^{\mathrm{Prod}}(0),$$
  $$\mathbb{N} = \mathrm{C}_{\mathrm{Hom}^{\text{-}1}}^{\mathrm{Prod}}(n) \text{ für } n \ge 1,$$

$$\max\left(\mathrm{C}^{\mathrm{Prod}}_{\mathrm{Hom}^{\text{-}1}}(n)\right) = \begin{cases} 0 & \text{für } n = 0 \\ \text{nicht existent} & \text{sonst.} \end{cases}$$

- Schnitt mit regulärer Sprache nach Satz 3.41:
  Für eine beliebige reguläre Sprache R gelten

  $$\mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n) \text{ und } \max\left(\mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n)\right) \text{ sind noch unbekannt.}$$

  Es existiert eine reguläre Sprache R, für die gelten

  $$\begin{aligned} \{0, \ldots, n\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n), \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n)\right) &= n. \end{aligned}$$

  Es existiert für ein frei wählbares $k \in \mathbb{N}_0$ eine reguläre Sprache R_k, für die gelten

  $$\begin{aligned} \{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_k}}}(0), \\ \{0, 1\} &= \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_k}}}(1), \\ 0, \ldots, \max(\{n, k\}) &\in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_k}}}(n) \text{ für } n \geq 2, \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_k}}}(n)\right) &= \begin{cases} n & \text{für } 0 \leq n \leq 1 \\ \text{noch offen} & \text{sonst.} \end{cases} \end{aligned}$$

  Für die Menge aller Schnitte mit regulärer Sprache Reg gelten

  $$\begin{aligned} \{0\} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(0), \\ \{0, 1\} &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(1), \\ \mathbb{N}_0 &= \mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(n) \text{ für } n \geq 2, \\ \max\left(\mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(n)\right) &= \begin{cases} n & \text{für } 0 \leq n \leq 1 \\ \text{nicht existent} & \text{sonst.} \end{cases} \end{aligned}$$

Zur Symbolkomplexität (aus Abschnitt 4)

- Spiegelbild nach Satz 4.17:

  $$\begin{aligned} \mathrm{C}^{\mathrm{Symb}}_{()^{\mathrm{R}}}(n) &= \begin{cases} \emptyset & \text{für } n = 1 \\ \{n\} & \text{sonst,} \end{cases} \\ \max\left(\mathrm{C}^{\mathrm{Symb}}_{()^{\mathrm{R}}}(n)\right) &= \begin{cases} \text{nicht existent} & \text{für } n = 1 \\ n & \text{sonst.} \end{cases} \end{aligned}$$

- Vereinigung nach Satz 4.23:

  $$\begin{aligned} &23, \ldots, n, n+3, \ldots, n+m-2, \\ &\qquad n+m+6 \in \mathrm{C}^{\mathrm{Symb}}_{\cup}(n, m) = \mathrm{C}^{\mathrm{Symb}}_{\cup}(m, n) \end{aligned}$$

$$\not\ni n+m+7,\dots \text{ für } n \geq 23, m \geq 13,$$

$$\max\left(\mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m)\right) =$$

$$\max\left(\mathrm{C}_{\cup}^{\mathrm{Symb}}(m,n)\right) = \begin{cases} n & \text{für } n \neq 1, m = 0 \\ \text{nicht existent} & \text{für } n = 1 \text{ od. } m = 1 \\ \text{noch offen} & \text{für } n \geq 2, 2 \leq m \leq 3 \\ \text{noch offen} & \text{für } 4 \leq n \leq 5, 4 \leq n \leq 5 \\ n+7 & \text{für } n \geq 6, m = 4 \\ n+8 & \text{für } n \geq 6, m = 5 \\ n+m+6 & \text{sonst,} \end{cases}$$

$$\max\left(\mathrm{C}_{\cup}^{\mathrm{Symb}}(n,m)\right) \leq n+m+6.$$

- Konkatenation nach Satz 4.29:

$$n+4,\dots,n+m-2, n+m+4 \in \mathrm{C}_{\cdot}^{\mathrm{Symb}}(n,m) = \mathrm{C}_{\cdot}^{\mathrm{Symb}}(m,n)$$

$$\not\ni n+m+5,\dots \text{ für } n \geq 8, m \geq 8,$$

$$\max\left(\mathrm{C}_{\cdot}^{\mathrm{Symb}}(n,m)\right) =$$

$$\max\left(\mathrm{C}_{\cdot}^{\mathrm{Symb}}(m,n)\right) = \begin{cases} 0 & \text{für } n \neq 1, m = 0 \\ \text{nicht existent} & \text{für } n = 1 \text{ od. } m = 1 \\ n & \text{für } n \geq 2, m = 2 \\ \text{noch offen} & \text{für } n \geq 3, 3 \leq m \leq 7 \\ n+m+4 & \text{sonst,} \end{cases}$$

$$\max\left(\mathrm{C}_{\cdot}^{\mathrm{Symb}}(n,m)\right) \leq n+m+4.$$

- Kleene-Abschluss nach Satz 4.33:

$$10,\dots,n,n+6 \in \mathrm{C}_{()^*}^{\mathrm{Symb}}(n) \not\ni n+7,\dots \text{ für } n \geq 10,$$

$$\max\left(\mathrm{C}_{()^*}^{\mathrm{Symb}}(n)\right) = \begin{cases} 2 & \text{für } n = 0 \\ \text{nicht existent} & \text{für } n = 1 \\ 2 & \text{für } n = 2 \\ n+3 & \text{für } 3 \leq n \leq 5 \\ \text{noch offen} & \text{für } 6 \leq n \leq 7 \\ n+6 & \text{sonst,} \end{cases}$$

$$\max\left(\mathrm{C}_{()^*}^{\mathrm{Symb}}(n)\right) \leq n+6.$$

- Homomorphismus nach Satz 4.38:
  Für einen beliebigen Homomorphismus h, wobei X der Definitionsbereich von h sei, gelten

$$\mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n) \not\ni 3 + (n-2) \cdot \max(\{|\mathrm{h}(x)| \mid x \in X\}), \ldots \text{ für } n \geq 3,$$
$$\max\left(\mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)\right) \leq 3 + (n-2) \cdot \max(\{|\mathrm{h}(x)| \mid x \in X\}), \ldots \text{ für } n \geq 3.$$

Es existiert ein Homomorphismus h, für den gelten

$$\mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n) = \begin{cases} \{0\} & \text{für } n = 0 \\ \emptyset & \text{für } n = 1 \\ \{2, \ldots, n\} & \text{sonst,} \end{cases}$$
$$\max\left(\mathrm{C}_{\mathrm{h}}^{\mathrm{Symb}}(n)\right) = \begin{cases} \text{nicht existent} & \text{für } n = 1 \\ n & \text{sonst.} \end{cases}$$

Für die Menge aller Homomorphismen Hom gelten

$$\mathrm{C}_{\mathrm{Hom}}^{\mathrm{Symb}}(n) = \begin{cases} \{0\} & \text{für } n = 0 \\ \emptyset & \text{für } n = 1 \\ \{2\} & \text{für } n = 2 \\ \mathbb{N}^{\not\ni 1} & \text{sonst,} \end{cases}$$
$$\max\left(\mathrm{C}_{\mathrm{Hom}}^{\mathrm{Symb}}(n)\right) = \begin{cases} 0 & \text{für } n = 0 \\ \text{nicht existent} & \text{für } n = 1 \\ 2 & \text{für } n = 2 \\ \text{nicht existent} & \text{sonst.} \end{cases}$$

- Inverser Homomorphismus nach Satz 4.44:
  Für einen beliebigen inversen Homomorphismus h^{-1} gelten

  $$\mathrm{C}_{\mathrm{h}^{-1}}^{\mathrm{Symb}}(n) \text{ und } \max\left(\mathrm{C}_{\mathrm{h}^{-1}}^{\mathrm{Symb}}(n)\right) \text{ sind noch unbekannt.}$$

  Es existiert ein inverser Homomorphismus h^{-1}, für den gelten

  $$2, \ldots, n-3, n \in \mathrm{C}_{\mathrm{h}^{-1}}^{\mathrm{Symb}}(n) \text{ für } n \geq 5,$$
  $$\max\left(\mathrm{C}_{\mathrm{h}^{-1}}^{\mathrm{Symb}}(n)\right) \text{ ist noch unbekannt.}$$

  Für die Menge aller inversen Homomorphismen Hom^{-1} gelten

  $$2, \ldots, n-3, n, 3k \in \mathrm{C}_{\mathrm{Hom}^{-1}}^{\mathrm{Symb}}(n) \text{ für alle } k \in \mathbb{N}, n \geq 5,$$
  $$\max\left(\mathrm{C}_{\mathrm{Hom}^{-1}}^{\mathrm{Symb}}(n)\right) = \begin{cases} 0 & \text{für } n = 0 \\ \text{nicht existent} & \text{sonst.} \end{cases}$$

- Schnitt mit regulärer Sprache nach Satz 4.48: Für eine beliebige reguläre Sprache R gelten

  $$\mathrm{C}_{\cap_{\mathrm{R}}}^{\mathrm{Symb}}(n) \text{ und } \max\left(\mathrm{C}_{\cap_{\mathrm{R}}}^{\mathrm{Symb}}(n)\right) \text{ sind noch unbekannt.}$$

Es existiert eine reguläre Sprache R, für die gelten

$$\begin{aligned} 2, \dots, n-3, n \in \mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R}}}(n) \not\ni n+1, \dots \quad \text{für } n \geq 6, \\ \max\left(\mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R}}}(n)\right) = n. \end{aligned}$$

Für die Menge aller Schnitte mit regulärer Sprache Reg gelten

$$\mathrm{C}^{\mathrm{Prod}}_{\mathrm{Reg}}(n) = \begin{cases} \{0\} & \text{für } n = 0 \\ \emptyset & \text{für } n = 1 \\ \{0, n\} & \text{für } 2 \leq n \leq 4 \\ \{0, 2, 3, 5\} & \text{für } n = 5 \\ \text{noch offen} & \text{für } 6 \leq n \leq 12 \\ \mathbb{N}_0^{\neq 1} & \text{sonst,} \end{cases}$$

$$\max\left(\mathrm{C}^{\mathrm{Symb}}_{\mathrm{Reg}}(n)\right) = \begin{cases} 0 & \text{für } n = 0 \\ \text{nicht existent} & \text{für } n = 1 \\ n & \text{für } 2 \leq n \leq 5 \\ \text{noch offen} & \text{für } 6 \leq n \leq 12 \\ \text{nicht existent} & \text{sonst.} \end{cases}$$

Die tatsächlich gefundenen Räume der Regel- und Symbolkomplexität können in den einzelnen Sätzen nachgelesen werden. Darüber hinaus gibt es zu jedem Satz eine Bemerkung, die die noch offenen Fälle aufzeigt.

Wir treffen nun Aussagen über die Vollständigkeit der gefundenen Räume der Regel- und Symbolkomplexität. Vollständig oder im Wesentlichen vollständig sind die Räume der Regel- und Symbolkomplexität unter:

> *Spiegelbild, Vereinigung, Kleene-Abschluss, Homomorphismus (ein konkreter, alle), inverser Homomorphismus (nur alle unter Regelkomplexität), Schnitt mit regulärer Sprache (eine konkrete, alle).*

„Im Wesentlichen vollständig“ soll heißen, dass nur einige Zahlen in $\mathrm{C}^{\mathrm{K}}_{\circ}(n)$ oder $\mathrm{C}^{\mathrm{K}}_{\circ}(n, m)$ für $\mathrm{K} \in \{\mathrm{Prod}, \mathrm{Symb}\}$, eine Operation $\circ$ und $n \geq n'$, $m \geq m'$ für recht kleine Zahlen $n', m' \in \mathbb{N}_0$ nicht bekannt sein dürfen. Unvollständig sind die Räume der Regel- und Symbolkomplexität unter:

> *Konkatenation, Homomorphismus (allgemeiner), inverser Homomorphismus (außer alle unter Regelkomplexität), Schnitt mit regulärer Sprache (allgemeine).*

Für einen beliebigen Homomorphismus kennen wir immerhin eine Abschätzung für die maximale Regel- und Symbolkomplexität, während dies für inversen Homomorphismus und Schnitt mit regulärer Sprache im Allgemeinen noch unbekannt ist.

Beziehen wir uns auf den in Abschnitt 1.1 dargelegten möglichen praktischen Nutzen, genügen bereits die im Wesentlichen vollständigen Räume der Regel- und Symbolkomplexität – denn die Regel- oder Symbolanzahl der praktisch verwendeten Grammatiken dürften eher größer sein. Dennoch wäre die vollständige Bestimmung der Räume mindestens aus theoretischer Sicht wünschenswert. Weiterhin steht die Bestimmung der unvollständigen Räume der Regel- und Symbolkomplexität aus. Nicht unerwähnt soll bleiben, dass in einigen Lemmata (z. B. Lemma 3.22) die Regel- oder Symbolkomplexität nur über die Anzahl verschiedener Buchstaben erhöht wird. Dies widerspricht im gewissen Sinne der Praxis, in der die Anzahl der verschiedenen Buchstaben in aller Regel beschränkt ist. Es wäre folglich gut, wenn die höhere Komplexität über die Struktur einer Sprache und nicht über die Anzahl verschiedener Buchstaben erreicht werden könnte.

A. Anhang

A.1. Literaturverzeichnis

[CGKY12] Cui, Bo ; Gao, Yuan ; Kari, Lila ; Yu, Sheng: State complexity of combined operations with two basic operations. In: *Theoretical Computer Science* 437 (2012), S. 82 – 102

[DS08] Dassow, Jürgen ; Stiebe, Ralf: Nonterminal Complexity of Some Operations on Context-Free Languages. In: *Fundamenta Informaticae* 83 (2008), Nr. 1-2, S. 35 – 49

[Ear83] Earley, Jay: An efficient context-free parsing algorithm. In: *Communications of the ACM* 26 (1983), Nr. 1, S. 57 – 61

[Gao10] Gao, Yuan: *Advanced Topics on State Complexity of Combined Operations*, University of Western Ontario, Diss., 2010. `http://ir.lib.uwo.ca/etd/65`

[GMRY16] Gao, Yuan ; Moreira, Nelma ; Reis, Rogério ; Yu, Sheng: A Survey on Operational State Complexity. In: *Journal of Automata, Languages and Combinatorics* 21 (2016), Nr. 4, S. 251–310

[Gru69] Gruska, Jozef: Some Classifications of Context-Free Languages. In: *Information and Control* 14 (1969), S. 152 – 179

[Gru72] Gruska, Jozef: On the Size of Context free Grammars. In: *Kybernetika* 8 (1972), Nr. 3, S. 213 – 218

[HJS05] Hricko, Marek ; Jirásková, Galina ; Szabari, Alexander: Union and Intersection of Regular Languages and Descriptional Complexity. In: *7th International Workshop on Descriptional Complexity of Formal Systems - DCFS 2005, Como, Italy, June 30 - July 2, 2005. Proceedings*, 2005, S. 170–181

[HMU00] Hopcroft, John ; Motwani, Rajeev ; Ullman, Jeffrey: *Introduction to Automata Theory, Languages, and Computation.* 2. Addison-Wesley, 2000. – ISBN 9780201441246

[IEE04] The IEEE and The Open Group (Hrsg.): *The Open Group Base Specifications.* Issue 6. : The IEEE and The Open Group, 2004. `http://pubs.opengroup.org/onlinepubs/009695399/`

[Jir09] Jirásková, Galina: Concatenation of Regular Languages and Descriptional Complexity. In: Frid, Anna (Hrsg.) ; Morozov, Andrey (Hrsg.) ; Rybalchenko, Andrey (Hrsg.) ; Wagner, KlausW. (Hrsg.): *Computer Science - Theory and Applications* Bd. 5675. Springer Berlin Heidelberg, 2009. – ISBN 978–3–642–03350–6, S. 203–214

[Jir14] Jirásková, Galina: The Ranges of State Complexities for Complement, Star, and Reversal of Regular Languages. In: *International Journal of Foundations of Computer Science* 25 (2014), Nr. 01, S. 101–124

[Rei90] REICHEL, Bernd: Some Classifications of Indian Parallel Languages. In: *Journal of Information Processing and Cybernetics* 26 (1990), Nr. 1, S. 85 – 99

[Sch01] SCHÖNING, Uwe: *Theoretische Informatik – kurzgefasst*. Vierte Auflage. Berlin, Heidelberg : Spektrum Akademischer Verlag, 2001. – ISBN 3–8274–1099–1

[SWY04] SALOMAA, Arto ; WOOD, Derick ; YU, Sheng: On the state complexity of reversals of regular languages. In: *Theoretical Computer Science* 320 (2004), Nr. 2 – 3, S. 315 – 329

[Yu00] YU, Sheng: State Complexity of Regular Languages. In: *Journal of Automata, Languages and Combinatorics* 6 (2000), S. 221 – 234

[YZS94] YU, Sheng ; ZHUANG, Qingyu ; SALOMAA, Kai: The state complexities of some basic operations on regular languages. In: *Theoretical Computer Science* 125 (1994), Nr. 2, S. 315 – 328

A.2. Satz-/Lemma-/Definitions-Verzeichnis

Nr.	S.	Beschreibung
3.6	29	Regelkomplexität von $\{a^{2^i} \mid n \leq i \leq m\}$ mit Erweiterungen.
3.7	30	Regelkomplexität von $\{a_i \mid 1 \leq i \leq n\}^+$ und $\{a_i \mid 1 \leq i \leq n\}^*$.
3.8	32	Regelkomplexität von $\{a_i \mid 1 \leq i \leq n\}$.
3.9	32	Regelkomplexität von $\{a^i x b^i \mid i \geq 0, x \in \{c_1, \ldots, c_{n-1}\}\}$.
3.10	33	Regelkomplexität von $\{a^i x b^i \mid i \geq 0, x \in \{c_1, \ldots, c_{n-2}\}\} \cdot \{\alpha\}$.
3.11	35	Regelkomplexität von $\{a^i x b^i \mid i \geq 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-4}\}\}$.
3.12	38	Regelkomplexität von $\{a^i x b^i \mid i \geq 1, x \in \{c\}^+ \cup \{d_1, \ldots, d_{n-5}\}\} \cdot \{\alpha\}$.
3.13	38	Raum der Regelkomplexität unter Spiegelbild.
3.14	39	Kommutativität von $\mathrm{C}_{\cup}^{\mathrm{Prod}}$.
3.15	39	Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Vereinigung.
3.16	40	Ergebnisse zum Raum der Regelkomplexität unter Vereinigung.
3.17	45	Teilbestimmung des Raums der Regelkomplexität unter Vereinigung.
3.18	46	Kommutativität von $\mathrm{C}_{\cdot}^{\mathrm{Prod}}$.
3.19	47	Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Konkatenation.
3.20	48	Es gilt $2, 3, 4 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(2, 2)$.
3.21	49	Es gilt $n \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, 2)$.
3.22	49	Es gilt $n + m + 1 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, m)$.
3.23	52	Es gilt $n + 1 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, 2)$.
3.24	52	Es gilt $n + m \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, m)$.
3.25	53	Es gilt $n + m - 1 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, m)$.
3.26	54	Es gilt $n + m - 2 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, m)$.
3.27	54	Es gilt $n + 2, \ldots, n + m - 3 \in \mathrm{C}_{\cdot}^{\mathrm{Prod}}(n, m)$.
3.28	55	Teilbestimmung des Raums der Regelkomplexität unter Konkatenation.
3.29	56	Ergebnisse zum Raum der Regelkomplexität unter Kleene-Abschluss.
3.30	59	Raum der Regelkomplexität unter Kleene-Abschluss.
3.31	59	Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Homomorphismus.
3.32	60	Es gibt ein h, sodass $1, \ldots, n \in \mathrm{C}_{\mathrm{h}}^{\mathrm{Prod}}(n)$ gilt.
3.33	62	Teilbestimmung des Raums der Regelkomplexität unter Homomorphismus.
3.34	63	Allgemeine Ergebnisse zum Raum der Regelkomplexität unter inversem Homomorphismus.
3.35	63	Es gibt ein h, sodass $0, \ldots, n \in \mathrm{C}_{\mathrm{h}^{-1}}^{\mathrm{Prod}}(n)$ gilt.

Nr.	S.	Beschreibung
3.36	64	Es gibt ein $\mathrm{h}_{n,k}$, sodass $k \in \mathrm{C}^{\mathrm{Prod}}_{\mathrm{h}_{n,k}^{-1}}(n)$ gilt.
3.37	64	Teilbestimmung des Raums der Regelkomplexität unter inversem Homomorphismus.
3.38	65	Allgemeine Ergebnisse zum Raum der Regelkomplexität unter Schnitt mit regulärer Sprache.
3.39	66	Es gibt ein R, sodass $0, \ldots, n \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n)$ gilt.
3.40	67	Es gibt ein R_m, sodass $0, \ldots, m \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R_m}}}(n)$ gilt.
3.41	69	Teilbestimmung des Raums der Regelkomplexität unter Schnitt mit regulärer Sprache.
4.1	71	Kontextfreie Sprachen mit der Symbolkomplexität höchstens 8.
4.2	74	Die Symbolkomplexität einer kontextfreien Sprachen L ist mindestens $\lvert\mathrm{Alph}(L)\rvert + 2$.
4.3	74	Symbolkomplexität von $\{a^{\pi_n}\}$.
4.4	76	Symbolkomplexität von $\{w\} \cup \{\lambda\}$.
4.5	77	Initialschleifenableitung für $\{a\}^+$.
4.6	77	Erweiterung des Satzes 2.22 im Fall von $\{a\}^+$.
4.7	79	Symbolkomplexität von $\{a_1, \ldots, a_n\}$.
4.8	79	Symbolkomplexität von $\{a_1, \ldots, a_n\}^*$.
4.9	81	Symbolkomplexität von $\{a_1 \cdots a_n\}$.
4.10	81	Symbolkomplexität von $\{a_1 \cdots a_n\}^*$.
4.11	81	Symbolkomplexität von $\{a_1 \cdots a_n\}^*$ ohne Initialschleifenableitung.
4.12	82	Symbolkomplexität von $\{a^m c_1 \cdots c_n b^m \mid m \geq 0\}$.
4.13	82	Symbolkomplexität von $\{a^m c_1 \cdots c_n b^m \mid m \geq 0\}^*$.
4.14	84	Symbolkomplexität von $\{a\} \cup \{a^{\pi_n}\}$.
4.15	86	Symbolkomplexität von $\{ab^{\pi_n}\}$ und $\{acb^{\pi_n}\}$.
4.16	87	Raum der Symbolkomplexität für Symbolkomplexität 1.
4.17	87	Raum der Symbolkomplexität unter Spiegelbild.
4.18	87	Kommutativität von $\mathrm{C}^{\mathrm{Symb}}_{\cup}$.
4.19	87	Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Vereinigung.
4.20	88	Spezielle Ergebnisse zum Raum der Symbolkomplexität unter Vereinigung.
4.21	90	Ergebnisse zum Raum der Symbolkomplexität unter Vereinigung.
4.22	94	Es gelten $n + m + 3, n + m + 6 \in \mathrm{C}^{\mathrm{Symb}}_{\cup}(n, m)$.
4.23	96	Teilbestimmung des Raums der Symbolkomplexität unter Vereinigung.
4.24	99	Kommutativität von $\mathrm{C}^{\mathrm{Symb}}_{\cdot}$.

Nr.	S.	Beschreibung
4.25	100	Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Konkatenation.
4.26	101	Ergebnisse zum Raum der Symbolkomplexität unter Konkatenation.
4.27	102	Es gilt $n+4,\ldots,n+m-2 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,m)$.
4.28	104	Es gilt $n+m+4 \in \mathrm{C}^{\mathrm{Symb}}_{\cdot}(n,m)$.
4.29	105	Teilbestimmung des Raums der Symbolkomplexität unter Konkatenation.
4.30	107	Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Kleene-Abschluss.
4.31	108	Spezielle Ergebnisse zum Raum der Symbolkomplexität unter Kleene-Abschluss.
4.32	108	Ergebnisse zum Raum der Symbolkomplexität unter Kleene-Abschluss.
4.33	110	Teilbestimmung des Raums der Symbolkomplexität unter Kleene-Abschluss.
4.34	111	Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Homomorphismus.
4.35	112	Es gibt ein h_k, sodass $k \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}_k}(n)$ gilt.
4.36	113	Es gilt $\{2\} \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}}(n)$, wenn h total-löschend ist.
4.37	113	Ergebnisse zum Raum der Symbolkomplexität unter einem speziellen Homomorphismus.
4.38	114	Teilbestimmung des Raums der Symbolkomplexität unter Homomorphismus.
4.39	115	Es gilt $\{0\} = \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{-1}}(0)$.
4.40	116	Es gibt kein h, sodass $0 \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}^{-1}}(2)$ gilt.
4.41	116	Es gibt ein $\mathrm{h}_{k,n}$, sodass $3k \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}_{k,n}^{-1}}(n)$ gilt.
4.42	116	Es gibt ein h_k, sodass $3k+6 \in \mathrm{C}^{\mathrm{Symb}}_{\mathrm{h}_k^{-1}}(2)$ gilt.
4.43	116	Ergebnisse zum Raum der Symbolkomplexität unter einem speziellen inversen Homomorphismus.
4.44	118	Teilbestimmung des Raums der Symbolkomplexität unter inversem Homomorphismus.
4.45	119	Allgemeine Ergebnisse zum Raum der Symbolkomplexität unter Schnitt mit regulärer Sprache.
4.46	123	Es gibt ein R, sodass $2,\ldots,n-3,n \in \mathrm{C}^{\mathrm{Prod}}_{\cap_{\mathrm{R}}}(n)$ gilt.
4.47	124	Es gibt ein R_k, sodass $k \in \mathrm{C}^{\mathrm{Symb}}_{\cap_{\mathrm{R_k}}}(n)$ gilt.
4.48	125	Teilbestimmung des Raums der Symbolkomplexität unter Schnitt mit regulärer Sprache.

A.3. Ehrenerklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe einer kommerziellen Promotionsberaterin / eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert,
- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche der Urheberin / des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.[1]

Magdeburg, den 08.11.2019
Dipl.-Inform. Ronny Harbich

[1] Die Ehrenerklärung wurde aus der Promotionsordnung vom 7. Januar 1999 (19. Dezember 2012) i. d. F. v. 2. Mai 2018 der Fakultät für Informatik der Otto-von-Guericke-Universität Magdeburg – gemäß der Pflicht durch diese – zitiert.

9 783945 469088