Erste Staatsprüfung für das Lehramt an Gymnasien
Wissenschaftliche Hausarbeit im Fach Informatik

Subreguläre Sprachklassen und ihre Verwendung zur Steuerung in externen kontextualen Grammatiken

vorgelegt von MarvinKödding
Straße
Stadt
Gutachterin: *Gutachterin*

12. Dezember 2023

Inhaltsverzeichnis

Kapitel 1

Einleitung

In der Welt der formalen Sprachen und Grammatiken spielen subreguläre Sprachen eine entscheidende Rolle bei der Untersuchung und Modellierung von Strukturen, die in vielen natürlichen und künstlichen Systemen vorkommen. Diese spezielle Klasse von Sprachen, die sich unterhalb der Ebene regulärer Sprachen befindet, ist Gegenstand intensiver Forschung und bietet ein tiefgehendes Verständnis für die komplexen Interaktionen in sprachlichen und nicht-sprachlichen Systemen. Es hat sich die Anwendung von subregulären Sprachen als Steuerungselement für externe kontextuelle Grammatiken als besonders vielversprechend erwiesen.

In seiner Veröffentlichung [10] hat SOLOMON MARCUS kontextuale Grammatiken eingeführt, wobei die externen kontextualen Grammatiken die einfachste Form von kontextualen Grammatiken sind. Hier kann ein Wort x, wenn es in einer Auswahlsprache der Grammatik liegt, mit einem Kontext (u, v) der zugehörigen Menge von Kontexten versehen und damit zu uxv abgeleitet werden.

Ein Forschungsfeld widmet sich der Verwendung von subregulären Sprachen zur Steuerung externer kontextueller Grammatiken. Dieser Ansatz erweitert die Möglichkeiten der Grammatikgestaltung und -verarbeitung erheblich und eröffnet neue Horizonte für die Realisierung von effizienteren, adaptiveren und kreativeren Systemen. In dieser Arbeit wird die Bedeutung von subregulären Sprachen für die Steuerung externer kontextueller Grammatiken ausführlich untersucht. Wir werden, aufbauend auf den Arbeiten von TRUTHE [4,8,16,17], BORDIHN [1], HOLZER [1, 8], KUTRIB [1], DASSOW [2, 4] und weiteren, die bisherige Hierarchie der subregulären Sprachen und fünf andere Sprachfamilien erweitern und deren Verwendung zum Steuern externer kontextualer Grammatiken untersuchen.

Kapitel 2

Grundlagen

2.1 Sprachen, Grammatiken, Automaten

Definition 2.1 – Mathematisches
Wir bezeichnen mit $\mathbb{Z}$ die Menge der ganzen Zahlen und mit $\mathbb{N}$ die Menge der positiven ganzen Zahlen. Mit $\mathbb{N}_0$ bezeichnen wir $\mathbb{N} \cup \{0\}$. Die Potenzmenge einer Menge A bezeichnen wir mit

$$\mathcal{P}(A) = \{B \mid B \subseteq A\} .$$

Definition 2.2 – Buchstabe, Alphabet, Wort, Sprache
Ein *Alphabet* ist eine endliche, nichtleere Menge von Symbolen (genannt *Buchstaben*). Ein *Wort* ist dabei eine endliche Folge von Buchstaben. Die Länge eines Wortes w bezeichnen wir als $|w|$. Das leere Wort λ enthält keine Buchstaben und hat die Länge 0. Es sei $n \in \mathbb{N}$ und $x_1, x_2, \ldots, x_n$ Buchstaben, dann nennen wir $w = x_1 x_2 \ldots x_n$ Wort. Wir bezeichnen die Spiegelung von w mit

$$w^R = x_n x_{n-1} \ldots x_1.$$

Eine Menge an Wörtern über einem Alphabet V nennen wir *Sprache* über dem Alphabet V. Es sei V ein Alphabet, dann verwenden wir die folgenden Bezeichnungen und Notationen:

- V^* bezeichnet die Menge aller Wörter über dem Alphabet V,
- V^+ bezeichnet die Menge aller nichtleeren Wörter über dem Alphabet V (es gilt also $V^+ = V^* \setminus \{\lambda\}$),
- V^k bezeichnet für ein $k \in \mathbb{N}_0$ die Menge aller Wörter über dem Alphabet V mit Länge k.

Für ein Wort $w \in V^*$ und eine Menge $A \subseteq V$ bezeichnen wir mit $|w|_A$, wie oft ein Buchstabe $a \in A$ in w vorkommt. Falls A nur aus einem Buchstaben a besteht, so schreiben wir vereinfacht $|w|_a$. Die Kardinalität einer Menge A bezeichnen wir mit $|A|$.

Definition 2.3 – Konkatenation und Iteration
Die *Konkatenation* von zwei Sprachen U und V ist die Menge aller durch Konkatenation erhaltenen Wörter aus einem Wort der Sprache U mit einem Wort der Sprache V:

$$U \cdot V = \{uv \mid u \in U \text{ und } v \in V\}.$$

Wir schreiben für die Konkatenation von zwei Sprachen $U \cdot V$ von zwei Sprachen U und V auch UV.

Für eine Sprache L und eine natürliche Zahl $i \in \mathbb{N}$ mit $n > 1$ bezeichnen wir mit L^i die Konkatenation der Sprache L^{i-1} mit der Sprache L ($L^1 = L$). Weiterhin ist $L^0 = \{\lambda\}$. Für eine Sprache L verwenden wir folgende Notationen

$$L^* = \bigcup_{i \geq 0} L^i \qquad \text{und} \qquad L^+ = \bigcup_{i \geq 1} L^i$$

analog zur Notation von Alphabeten.

Definition 2.4 – Grammatik und Ableitung
Eine *erzeugende Grammatik* ist ein Tupel

$$G = (N, T, P, S)$$

wobei N und T zwei disjunkte Alphabete sind, P eine nichtleere und endliche Teilmenge von $(N \cup T)^* \setminus T^* \times (N \cup T)^*$ und das Element $S \in N$ das Startsymbol ist. Die Elemente der Menge N nennen wir Nichtterminale, die Elemente aus der Menge T Terminale, die Elemente aus P Produktionen und schreiben $\alpha \to \beta$ statt (α, β). Eine erzeugende Grammatik heißt *rechtslinear*, wenn

$$P \subset N \times (T^* N \cup T^*),$$

und *regulär*, wenn

$$P \subset N \times (TN \cup T)$$

gilt.

Es seien $G = (N, T, P, S)$ eine erzeugende Grammatik und die Menge $V = N \cup T$. Zwei Wörter $w = w_1 \alpha w_2$ und $w' = w_1 \beta w_2$ mit $w_1, w_2, \alpha, \beta \in V^*$ stehen in Relation zueinander, wenn es eine Regel $(\alpha, \beta) \in P$ gibt. Diese Relation heißt Ableitungsrelation und wir sagen, dass wir w zu w' *ableiten* können. Wir schreiben dafür

$$w \underset{G}{\Longrightarrow} w'$$

oder

$$w \Longrightarrow w',$$

wenn durch den Kontext eindeutig ist, welche Grammatik angewandt wird. Wir bezeichnen mit

$$w \underset{G}{\Longrightarrow}^k w'$$

die Anwendung von k Ableitungen. Den reflexiven und transitiven Abschluss der Relation $\underset{G}{\Longrightarrow}$ bezeichnen wir mit

$$\underset{G}{\Longrightarrow}^*.$$

Definition 2.5 – Sprache einer Grammatik
Die *von einer Grammatik G erzeugte Sprache* $L(G)$ ist die Menge aller Wörter, die durch die Grammatik G mit dem Axiom S abgeleitet werden kann:

$$L(G) = \left\{ w \mid S \underset{G}{\Longrightarrow}^* w \right\}.$$

Wir bezeichnen die Menge der Sprachen, die durch eine reguläre Grammatik erzeugt werden kann, als reguläre Sprachen und die Menge der Sprachen, die durch eine rechtslineare Grammatik erzeugt werden kann als rechtslineare Sprachen.

Es ist bereits bekannt, dass die Familie der rechtslinearen Sprachen mit der Familie der Sprachen, die von regulären Grammatiken erzeugt werden, übereinstimmt.

Definition 2.6 – Reguläre Ausdrücke
Reguläre Sprachen können auch durch *reguläre Ausdrücke* beschrieben werden. Es sei V ein Alphabet. Ein regulärer Ausdruck wird folgendermaßen definiert:

1. $\emptyset$ ist ein regulärer Ausdruck;
2. für jedes $x \in V$ ist x ein regulärer Ausdruck;
3. falls $\mathcal{R}$ und $\mathcal{S}$ reguläre Ausdrücke sind, so ist auch die Konkatenation $\mathcal{R} \cdot \mathcal{S}$, die Vereinigung $\mathcal{R} \cup \mathcal{S}$ und der Kleene-Abschluss $\mathcal{R}^*$ ein regulärer Ausdruck;
4. für jeden regulären Ausdruck existiert ein $n \in \mathbb{N}$, sodass wir diesen mit n Operationen aus den atomaren Elementen $\emptyset$ und $x \in V$ erhalten können

Die durch reguläre Ausdrücke beschriebenen Sprachen sind wie folgt definiert:

1. $L(\emptyset) = \emptyset$;

2. für jedes $x \in V$ ist $L(x) = \{x\}$;
3. falls $\mathcal{R}$ und $\mathcal{S}$ reguläre Ausdrücke sind, so gilt

$$\begin{aligned} L(\mathcal{R} \cdot \mathcal{S}) &= L(\mathcal{R}) \cdot L(\mathcal{S}), \\ L(\mathcal{R} \cup \mathcal{S}) &= L(\mathcal{R}) \cup L(\mathcal{S}), \\ L(\mathcal{R}^*) &= (L(\mathcal{R}))^*. \end{aligned}$$

Der Operator $\cdot$ wird oft weggelassen; statt des Vereinigungssymbols $\cup$ wird auch oft $+$ verwendet.

Definition 2.7 – Endlicher Automat
Ein nichtdeterministischer *endlicher Automat* (NFA) ist ein 5-Tupel

$$\mathcal{A} = (V, Z, z_0, F, \delta)$$

wobei V als Eingabealphabet bezeichnet wird, Z eine endliche, nichtleere Menge aller Zustände, $z_0 \in Z$ der Startzustand, $F \subseteq Z$ die Menge der akzeptierenden Zustände und die Abbildung $\delta : Z \times V \to \mathcal{P}(Z)$ die Übergangsfunktion ist. Ein endlicher Automat heißt deterministisch (DFA), wenn jede Menge der Übergangsfunktion $\delta(z, a)$ mit $z \in Z$ und $a \in V$ einelementig ist.

Die Abbildung δ kann zu einer Funktion $\delta^* : Z \times V^* \to \mathcal{P}(Z)$ erweitert werden mit

$$\delta^*(z, va) = \bigcup_{z' \in \delta^*(z,v)} \delta\left(z', a\right) \text{ und } \delta^*(z, \lambda) = \{z\}$$

für $z \in Z$, $a \in V$, und $v \in V^*$.

Wir werden δ für beide Versionen, also die originale und die erweiterte Version, verwenden.

Definition 2.8 – Akzeptierte Sprache eines Automaten
Es sei $\mathcal{A} = (V, Z, z_0, F, \delta)$ ein endlicher Automat. Ein beliebiges Wort $w \in V^*$ wird genau dann von diesem endlichen Automaten $\mathcal{A}$ akzeptiert, wenn der Automat nach dem Einlesen

des Wortes w einen akzeptierenden Zustand erreicht:

$$\delta(z_0, w) \cap F \neq \emptyset.$$

Die *von einem endlichen Automaten akzeptierte Sprache* $L(\mathcal{A})$ bezeichnet die Menge aller vom Automaten $\mathcal{A}$ akzeptierten Wörter:

$$L(\mathcal{A}) = \{w \mid w \in V^* \text{ und } \delta(z_0, w) \cap F \neq \emptyset\}$$

Die Sprache, die von einem endlichen Automaten akzeptiert wird, ist immer regulär. Auch andersrum existiert für jede reguläre Sprache ein Automat, der diese akzeptiert.

Definition 2.9 – Myhill–Nerode-Relation
Es sei V ein Alphabet und $L \subseteq V^*$ eine Sprache über diesem Alphabet. Mit D_xL für ein Wort $x \in V^*$ bezeichnen wir die Menge

$$D_xL = \{w \mid xw \in L\}$$

Wir definieren eine binäre Relation $\equiv_L \subseteq V^* \times V^*$ mit

$$x \equiv_L y \Leftrightarrow D_xL = D_yL$$

für zwei Wörter $x \in V^*$ und $y \in V^*$. Die Relation $\equiv_L$ ist eine Äquivalenzrelation und heißt *Myhill–Nerode-Relation* der Sprache L. Die Anzahl der Äquivalenzklassen heißt Index der Relation. Die Ergebnisse von J. R. Myhill und A. Nerode sind in [9] zu finden. Eine Sprache heißt regulär, falls der Index der Relation $\equiv_L$ endlich ist. Die minimale Anzahl an Zuständen, die ein deterministischer endlicher Automat braucht, um die Sprache L zu akzeptieren ist gleich dem Index der Myhill–Nerode-Relation $\equiv_L$. Wir nennen solch einen deterministischen endlichen Automaten dann minimal.

Definition 2.10 – Teilmengen und Unvergleichbarkeit
Es seien X und Y zwei Sprachen.

Wir sagen $X \subseteq Y$, wenn jedes Wort aus X auch in Y liegt.

Wir sagen $X \subset Y$, wenn $X \subseteq Y$, aber $X \neq Y$.

Wir sagen X und Y sind unvergleichbar, wenn weder $X \subseteq Y$, noch $Y \subseteq X$ gilt oder anders, wenn $X \setminus Y \neq \emptyset \neq Y \setminus X$.

Die Definition von Teilmengen und Unvergleichbarkeit von Sprachfamilien ist analog.

2.2 Subreguläre Sprachfamilien

Wir definieren nun die subregulären Sprachfamilien basierend auf strukturellen Eigenschaften.

Definition 2.11 – Subreguläre Sprachfamilien
Für eine Sprache L über einem Alphabet V definieren wir

$$\begin{aligned} Comm(L) = \{a_{i_1} \ldots a_{i_n} \mid a_1 \ldots a_n \in L, n \geq 1, \\ \{i_1, i_2, \ldots, i_n\} = \{1, 2, \ldots, n\}\} \end{aligned}$$

als ihren kommutativen Abschluss.

Wir bezeichnen mit

$$Circ(L) = \{vu \mid uv \in L,\ u, v \in V^*\}$$

ihren zirkulären Abschluss, mit

$$Suf(L) = \{v \mid uv \in L,\ u, v \in V^*\}$$

ihren Suffixabschluss, mit

$$Pre(L) = \{v \mid vu \in L,\ u, v \in V^*\}$$

ihren Präfixabschluss und mit

$$Inf(L) = \{v \mid uvw \in L,\ u, v, w \in V^*\}$$

ihren Infixabschluss.

Wir definieren die folgenden Einschränkungen für reguläre Sprachen. In Klammern geben wir an, wie wir die jeweilige Familie aller dieser Sprachen bezeichnen.

Es sei L eine Sprache über einem Alphabet V. Wir nennen die Sprache L

- **kombinational** (*COMB*), wenn sie die Form
  $$L = V^*A$$
  hat mit einer Teilmenge $A \subseteq V$,
- **definit** (*DEF*), wenn sie in der Form
  $$L = A \cup V^*B$$
  dargestellt werden kann, wobei A und B endliche Teilmengen von V^* sind,
- **nilpotent** (*NIL*), wenn L endlich ist oder ihr Komplement $V^* \setminus L$ endlich ist,
- **kommutativ** (*COMM*), wenn zu jedem Wort alle seine Permutationen auch in der Sprache sind, also
  $$L = Comm(L)$$
  gilt,
- **zirkulär** (*CIRC*), wenn zu jedem Wort auch alle seine zirkulären Verschiebungen zur Sprache gehöhren, also
  $$L = Circ(L)$$
  gilt,
- **abgeschlossen unter Suffixbildung** (*SUF*), wenn für je zwei Wörter $x \in V^*$ und $y \in V^*$ gilt: $xy \in L$ impliziert $y \in L$ (oder $L = Suf(L)$),

- **abgeschlossen unter Präfixbildung** (*PRE*), wenn je zwei Wörter $x \in V^*$ und $y \in V^*$ gilt: $xy \in L$ impliziert $x \in L$ (oder $L = Pre(L)$),
- **abgeschlossen unter Infixbildung** (*INF*), wenn für je drei Wörter $x \in V^*, y \in V^*$ und $z \in V^*$ gilt: $xyz \in L$ impliziert $y \in L$ (oder $L = Inf(L)$),
- **nicht-zählend** (*NC*), wenn es ein $k \in \mathbb{N}$ gibt, sodass für je drei Wörter $x \in V^*, y \in V^*$ und $z \in V^*$ gilt, dass $xy^kz \in L$ genau dann, wenn $xy^{k+1}z \in L$ gilt,
- **kleenesternfrei** (*SF*), wenn L durch einen regulären Ausdruck dargestellt werden kann, der ohne den Kleene-Stern geschrieben ist,
- **vereinigungsfrei** (*UF*), wenn L durch einen regulären Ausdruck dargestellt werden kann, der ohne den Vereinigungsoperator geschrieben ist,
- **endlich** (*FIN*), wenn $|L|$ eine natürliche Zahl ist,
- **monoidal** (*MON*), falls $L = V^*$ gilt,
- **potenzseparierend** (*PS*), falls es ein $m \in \mathbb{N}$ gibt, sodass für jedes $x \in V^*$ entweder $J_x^m \cap L = \emptyset$ gilt oder $J_x^m \subseteq L$ mit $J_x^m = \{x^n \mid n \geq m\}$,
- **geordnet** (*ORD*), falls es einen DFA $\mathcal{A} = (V, Z, z_0, F, \delta)$ gibt, der die Sprache akzeptiert und eine totale Ordnung $(Z, \prec)$ existiert, sodass für alle Buchstaben $a \in V$, und Zustände $z, z' \in Z$ die Relation $z \preceq z'$ impliziert, dass $\delta(z, a) \preceq \delta(z', a)$ gilt,
- **Stern** (*STAR*), falls $L = H^*$, wobei H eine reguläre Sprache ist mit $H \subseteq V^*$,
- **Komet** (*COM*), falls $L = G^*H$, wobei $G^* \in STAR$ und H eine reguläre Sprache mit $H \subseteq V^*$ und $G \notin \{\emptyset, \{\lambda\}\}$,

- **zweiseitiger Komet** ($2COM$), falls $L = EG^*H$, wobei $G^* \in STAR$ und E, H reguläre Sprachen mit $E, H \subseteq V^*$ und $G \notin \{\emptyset, \{\lambda\}\}$.

In manchen dieser Sprachfamilien liegen auch nicht-reguläre Sprachen. Wir betrachten hier aber nur solche, die regulär sind.

Als Menge der Sprachfamilien, die wir untersuchen werden, setzen wir

$$\begin{aligned}\mathcal{F} = \{&FIN, MON, NIL, COMB, DEF, SUF, PRE, INF, ORD,\\ &COMM, CIRC, NC, SF, PS, UF, STAR, COM, 2COM\}.\end{aligned}$$

2.3 Kontextuale Grammatiken

Nun werden wir kontextuale Grammatiken mit Auswahl aus einer subregulären Sprachfamilie und die von einer kontextualen Grammatik erzeugte Sprache definieren.

Definition 2.12 – Kontextuale Grammatik, Ableitung
Es sei $\mathcal{F}$ eine Familie von Sprachen. Eine kontextuale Grammatik mit einer Auswahl aus $\mathcal{F}$ ist ein Tripel $G = (V, \mathcal{S}, A)$ mit

1. V ist ein Alphabet,

2. $\mathcal{S}$ ist eine endliche Menge an Paaren (S, C) mit einer Auswahlsprache S über einer Teilmenge $U \subseteq V$, die zu $\mathcal{F}$ gehört und einer endlichen Menge $C \subset V^* \times V^*$ von Kontexten, in der für jedes $(u, v) \in C$ mindestens eine Komponente nicht leer ist: $uv \neq \lambda$,

3. A ist eine endliche Teilmenge von V^* (deren Elemente heißen Axiome).

Es sei $G = (V, \mathcal{S}, A)$ eine kontextuale Grammatik mit einer Auswahl aus $\mathcal{F}$.

Für ein $(S, C) \in \mathcal{S}$ schreiben wir $S \to (u, v)$ für $(u, v) \in C$.

Ein direkter externer Ableitungsschritt in G wird wie folgt definiert: Ein Wort x wird in ein Wort y abgeleitet ($x \Longrightarrow y$), falls es ein Paar $(S, C) \in \mathcal{S}$ gibt, sodass $x \in S$ und $y = uxv$ mit einem Kontext $(u, v) \in C$.

Ein direkter interner Ableitungsschritt in G wird wie folgt definiert: Ein Wort x wird in ein Wort y abgeleitet, falls es beliebige Wörter $x_1, x_2, x_3 \in V^*$ und ein Paar $(S, C) \in \mathcal{S}$ mit einem Kontext $(u, v) \in C$ so gibt, dass das Wort $x = x_1x_2x_3$ aufgeteilt werden kann, x_2 in S liegt und $y = x_1ux_2vx_3$ gilt.

Wir bezeichnen mit $\Longrightarrow^*$ den reflexiven und transitiven Abschluss der Relation $\Longrightarrow$. Die Sprache, die von G erzeugt wird, ist

$$L = \{z \mid x \Longrightarrow^* z \text{ für } x \in A\}.$$

Wir bezeichnen mit $\mathcal{EC}(\mathcal{F})$ die Familie aller Sprachen, die extern durch kontextuale Grammatiken mit der Auswahl aus $\mathcal{F}$ erzeugt werden.

Wenn eine kontextuale Grammatik extern arbeitet, dann nennen wir sie externe kontextuale Grammatik. Arbeitet eine kontextuale Grammatik intern, so nennen wir sie interne kontextuale Grammatik.

Wir sehen uns nun ein Beispiel an, um die Arbeitsweise einer kontextualen Grammatik zu verstehen.

Beispiel 2.13
Wir wollen nun eine kontextuale Grammatik G definieren, die durch externe Ableitungsschritte die Sprache

$$L = \{a^n b^n \mid n \geq 0\}$$

erzeugt.

Dafür können wir eine Axiomenmenge A definieren, sodass das kürzeste Wort oder die kürzesten Wörter der Sprache L in A liegen. Das ist hier das Leerwort λ. Weiterhin fügen wir auch noch das Wort ab hinzu. Also gilt $A = \{\lambda, ab\}$.

Nun brauchen wir eine Sprache S_1 und einen dazugehörigen Kontext $C_1 = (u, v)$, sodass ab in S_1 liegt und $uabv$ wieder in der Sprache L liegt. Am besten wäre es, wenn wir eine Sprache und einen Kontext so finden würden, dass wir immer Wörter der Form $a^k b^k$ zu $a^{k+1} b^{k+1}$ für $k \geq 1$ ableiten könnten. Wählen wir nun $S_1 = \{a\}^*\{b\}^*$, so sehen wir, dass jedes Wort der Form $a^k b^k$ in der Auswahlsprache liegt und wir können durch den dazugehörigen Kontext $C_1 = (a, b)$ dafür sorgen, dass wir die Wörter der Form $a^k b^k$ zu $a^{k+1} b^{k+1}$ ableiten können.

Also erzeugt die Grammatik

$$G = (\{a, b\}, \{(\{a\}^*\{b\}^*, \{(a, b)\})\}, \{\lambda, ab\})$$

die obige Sprache L. $\Diamond$

Kapitel 3

Normalformen

Für die Arbeit mit subregulären Sprachen ist es sinnvoll, sich Gedanken über Normalformen der jeweiligen Sprachen zu machen. Wir werden später noch zeigen, dass externe kontextuale Grammatiken mit $2COM$-Auswahlsprachen und solche mit beliebigen regulären Auswahlsprachen die gleichen Sprachfamilien erzeugen. Dafür werden wir die Normalform aus Theorem 3.4 für $2COM$-Sprachen verwenden. Um zu dieser Normalform zu gelangen, stellen wir zunächst einige Beobachtungen zusammen.

Lemma 3.1
Jede $2COM$-Sprache $L = EG^*H$ lässt sich als endliche Vereinigung

$$\bigcup_{i=1}^{n} E_i G^* H$$

für ein $n \geq 1$ und mit jeweils vereinigungsfreien Sprachen E_i für alle $1 \leq i \leq n$ darstellen.

Beweis. Jede reguläre Sprache ist die Vereinigung von endlich vielen vereinigungsfreien Sprachen [11]. Es sei E eine Menge mit $E = E_1 \cup E_2 \cup \cdots \cup E_n$ mit vereinigungsfreien Sprachen E_i für alle $1 \leq i \leq n$. Dann folgt

$$EG^*H = E_1G^*H \cup E_2G^*H \cup \cdots \cup E_nG^*H. \qquad \square$$

Um später zu zeigen, dass wir $2COM$-Sprachen in die erwähnte Normalform umwandeln können, zeigen wir nun, dass man eine unendliche vereinigungsfreie Sprache in eine spezielle $2COM$-Form überführen kann.

Lemma 3.2
Zu einer unendlichen regulären Sprache L über einem Alphabet V mit $L \in UF$ existieren Mengen L_l, L_i und L_r so, dass $L = L_l L_i^* L_r$ mit $|L_l| < \infty$ und $L_i \notin \{\emptyset, \{\lambda\}\}$.

Beweis. Wir beweisen die Behauptung induktiv über die Anzahl der Aufbaustufen, die benötigt werden, um einen regulären Ausdruck $\mathcal{R}$ zu erstellen, sodass $L = L(\mathcal{R})$ gilt. Für den Induktionsanfang wählen wir $n = 1$, da in der Aufbaustufe 0 nur endliche Sprachen entstehen können.

Induktionsanfang $n = 1$: Da L unendlich ist, gilt $\mathcal{R} = \{x\}^*$ für einen Buchstaben $x \in V$. Eine gewünschte Form für die Sprache L ist dann $\{\lambda\}\{x\}^*\{\lambda\}$.

Induktionsvoraussetzung: Zu jedem regulären Ausdruck $\mathcal{R}$ ohne Vereinigungsoperator, der eine unendliche Sprache beschreibt und höchstens von der Aufbaustufe n ist, lässt sich die Sprache $L(\mathcal{R})$ als $L(\mathcal{R}) = L_l L_i^* L_r$ darstellen mit $|L_l| < \infty$ und $L_i \notin \{\emptyset, \{\lambda\}\}$.

Induktionsschritt $n \to n + 1$:

Fall 1 – Wenn $\mathcal{R} = \mathcal{S}\mathcal{T}$, dann gilt $L(\mathcal{R}) = L(\mathcal{S})L(\mathcal{T})$ und nach der Induktionsvoraussetzung $L(\mathcal{S}) = S_l S_i^* S_r$ und $L(\mathcal{T}) = T_l T_i^* T_r$ für geeignete Mengen S_l,S_i,S_r,T_l,T_i und T_r. Es gilt also für $R_l = S_l$, $R_i = S_i$ und $R_r = S_r L(\mathcal{T})$, dass $L(\mathcal{R}) = R_l R_i^* R_r$ ist mit $|R_l| < \infty$ und $R_i \notin \{\emptyset, \{\lambda\}\}$.
Fall 2 – Wenn $\mathcal{R} = \mathcal{S}^*$, dann gilt $L(\mathcal{R}) = (L(\mathcal{S}))^*$. Es gilt also $L(\mathcal{R}) = R_l R_i^* R_r$ für $R_l = \{\lambda\}$, $R_i = L(\mathcal{S})$ und $R_r = \{\lambda\}$ mit $|R_l| < \infty$ und $R_i \notin \{\emptyset, \{\lambda\}\}$. □

Wir haben mit Lemma 3.1 gezeigt, dass wir eine $2COM$-Sprache in eine endliche Vereinigung von $2COM$-Sprachen überführen können, bei denen jeweils der erste Bestandteil vereinigungsfrei ist. Das verwenden wir nun und folgern daraus, dass wir eine Darstellung einer $2COM$-Sprache erhalten, bei der die erste Menge endlich ist.

Lemma 3.3
Zu jeder $2COM$-Sprache $L = EG^*H$ über einem Alphabet V mit $E \in UF$ existieren Sprachen E', G' und H' so, dass $EG^*H = E'(G')^*H'$ ist sowie $|E'| < \infty$ und $G' \notin \{\emptyset, \{\lambda\}\}$ gelten.

Beweis. Wir zeigen die Behauptung über die Anzahl der Aufbaustufen, die benötigt werden, um einen regulären Ausdruck $\mathcal{R}$ so zu erstellen, dass $E = L(\mathcal{R})$

Induktionsanfang $n = 0$: Dann gilt $\mathcal{R} = \emptyset$ oder $\mathcal{R} = x$ mit $x \in V$. Dann gilt für $E' = E, G' = G$ und $H' = H$ die Behauptung.

Induktionsvoraussetzung: Zu jedem regulären Ausdruck $\mathcal{R}$ ohne Vereinigungsoperator, sodass $L = L(\mathcal{R})G^*H$ eine $2COM$ Sprache und $\mathcal{R}$ dabei höchstens von der Aufbaustufe n ist, lässt sich die Sprache L als $L = (E')(G')^*(H')$ darstellen mit $E' = L(\mathcal{R})$, $|E'| < \infty$ und $G' \notin \{\emptyset, \{\lambda\}\}$.

Induktionsschritt $n \to n+1$:

Fall 1 – Es gelte $\mathcal{R} = \mathcal{S}\mathcal{T}$. Wenn $|E| < \infty$, dann folgt die Behauptung für $E' = E, G' = G$ und $H' = H$.

Wenn $|E| = \infty$ und $|L(\mathcal{S})| < \infty$, dann folgt $|L(\mathcal{T})| = \infty$. Dann gilt $L(\mathcal{T}) = T_l T_i^* T_r$ mit $|T_l| < \infty$ und $T_i \notin \{\emptyset, \{\lambda\}\}$ nach Lemma 3.2 für Sprachen T_l, T_i, T_r. Es folgt die Behauptung mit $E' = L(\mathcal{S})T_l$, $G' = T_i$ und $H' = T_r G^* H$.

Wenn $|E| = \infty$ und $|L(\mathcal{S})| = \infty$. Dann folgt $L(\mathcal{S}) = S_l S_i^* S_r$ mit $|S_l| < \infty$ und $S_i \notin \{\emptyset, \{\lambda\}\}$ nach Lemma 3.2 für Sprachen S_l, S_i, S_r. Dann folgt die Behauptung für $E' = S_l$, $G' = S_i$ und $H' = S_r L(\mathcal{T}) G^* H$.

Fall 2 – Es gelte $\mathcal{R} = \mathcal{S}^*$.

Wenn $|L(\mathcal{S})| = 0$, dann ist $\mathcal{S} = \mathcal{R} = \emptyset$ und mit $E' = E$, $G' = G$ und $H' = H$ folgt die Behauptung.

Wenn $\mathcal{S} = \lambda$, dann folgt mit $\mathcal{R} = \lambda^* = \lambda$ und $E' = E$, $G' = G$ und $H' = H$ die Behauptung,

Wenn $|L(\mathcal{S})| \geq 1$ und $L(\mathcal{S}) \neq \lambda$, dann gilt $|L(\mathcal{R})| = \infty$. Dann folgt die Behauptung für $E' = \{\lambda\}$, $G' = L(\mathcal{S})$ und $H' = G^*H$. $\square$

Jetzt verbinden wir die vorhergehenden Lemmata und folgern, dass es zu jeder $2COM$-Sprache eine solche Darstellung gibt, bei der die erste Menge der Sprache endlich ist.

Theorem 3.4 – Normalform für $2COM$-Sprachen
Zu jeder Sprache $L \in 2COM$ existiert eine Darstellung $L = EG^*H$ derart, dass $|E| < \infty$ und $G \notin \{\emptyset, \{\lambda\}\}$.

Beweis. Nach Lemma 3.1 lässt sich eine Sprache $L \in 2COM$ als endliche Vereinigung von E_iG^*H darstellen, sodass alle $E_i \in UF$ sind. Nach Lemma 3.3 gilt, dass sich diese E_iG^*H wiederum als $(E_i')(G')^*H'$ darstellen lassen, sodass E_i' endlich ist. Nun lässt sich diese Vereinigung

$$\bigcup_{i=1}^{n} E_i'(G')^*H' = \left(\bigcup_{i=1}^{n} E_i'\right)(G')^*H'$$

wieder als $E'(G')^*H'$ darstellen, sodass E' eine endliche Vereinigung an endlichen Sprachen ist. Damit ist auch E' wieder endlich. $\square$

Wir bezeichnen diese Darstellung als linksendliche Normalform. Auf analoge Weise lässt sich auch eine rechtsendliche Normalform herleiten.

Theorem 3.5 – Rechtsendliche Normalform
Zu jeder Sprache $L \in 2COM$ existiert eine Darstellung $L = EG^*H$, sodass $|H| < \infty$ und $G \notin \{\emptyset, \{\lambda\}\}$.

Beweis. Analog zu Theorem 3.4 □

Im weiteren Verlauf werden wir nur die linksendliche Normalform benötigen.

Kapitel 4

Hierarchie der subregulären Sprachfamilien

In diesem Kapitel untersuchen wir die Teilmengenbeziehungen zwischen den verschiedenen Sprachfamilien. Dafür werfen wir zunächst einen Blick auf die bisherige Forschung und fügen im Anschluss die beiden Hierarchien zusammenfügen, um später zu untersuchen, wie die Hierarchie der dazugehörigen $\mathcal{EC}$-Grammatiken aussieht.

Theorem 4.1 – Bisherige Ergebnisse
Die in Abbildung 4.1 und Abbildung 4.2 dargestellten Inklusionen gelten. Ein Pfeil von einem Knoten X zu einem Knoten Y bedeutet die echte Inklusion $X \subset Y$. Falls zwei Familien nicht durch einen Pfad verbunden sind, so sind diese unvergleichbar.

Die Beschriftung an einer Kante in Abbildung 4.1 verweist auf eine Arbeit, in der die betreffende Relation bewiesen wurde. In manchen Fällen ist das sicherlich nicht die erste Erwähnung, da manche Ergebnisse so offensichtlich sind, dass sie nicht gesondert veröffentlicht werden.

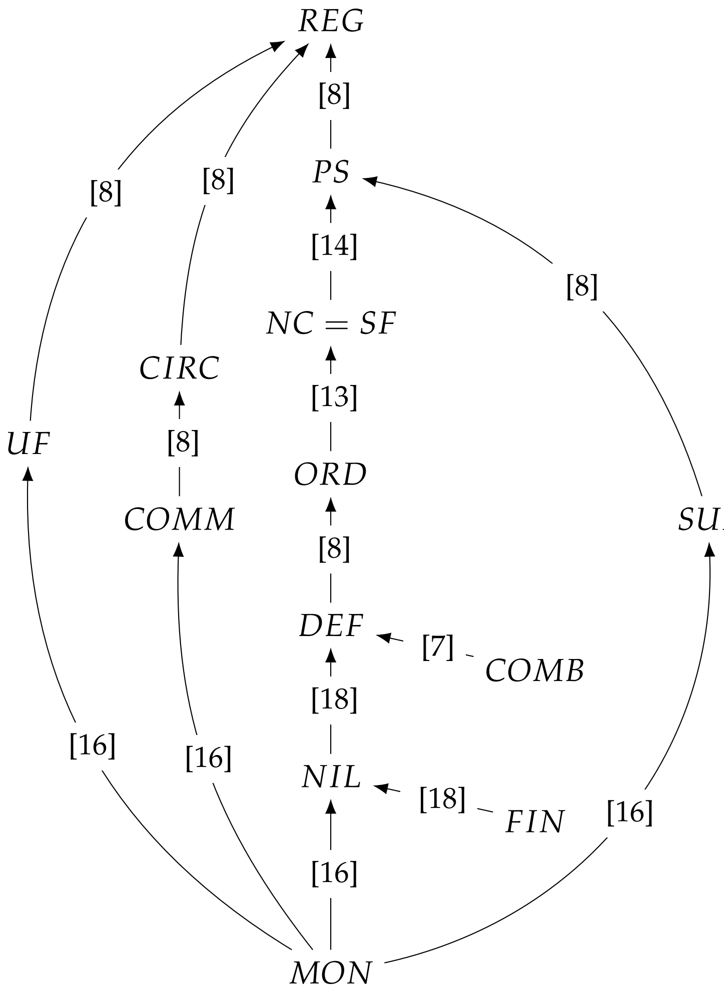

Abbildung 4.1: Hierarchie von subregulären Sprachfamilien

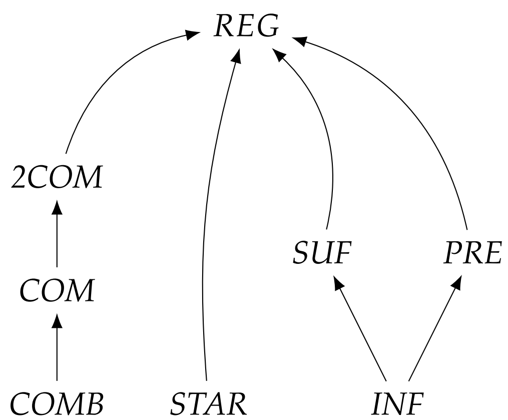

Abbildung 4.2: Hierarchie von subregulären Sprachfamilien (siehe [1])

4.1 *STAR*, *COM* und *2COM*

In der Literatur gilt im Allgemeinen die Inklusion

$$STAR \subset COM$$

mit der Ausnahme, dass $\{\lambda\}$ in der Sprachfamilie *STAR* liegt, aber nicht in *COM* oder *2COM*. Wir wollen aber jede Sprache berücksichtigen, weswegen für uns diese Teilmengenrelation nicht gelten wird. Zunächst werden wir die Unvergleichbarkeit von *STAR* und den Sprachfamilien *COM* und *2COM* zeigen, bevor wir mit den anderen Sprachfamilien fortfahren.

Lemma 4.2
Für die Sprache

$$L = \{b^n a \mid n \geq 0\}$$

gilt

$$L \in COM \setminus STAR.$$

Beweis. Es sei $L = \{b^n a \mid n \geq 0\}$.

i) Dann gilt für die Mengen $G = \{b\}$, $H = \{a\}$ die Gleichheit $L = G^* H$. Damit liegt die Sprache L also in *COM*.

ii) Angenommen L wäre eine Sprache, die sich mit einer Menge H' als $L = (H')^*$ darstellen lässt. Dann gäbe es ein Wort mit einem a in der Sprache H'. Dann läge aber auch ein Wort mit mehr als einem a in der Sprache L. Also gehört die Sprache L nicht zu *STAR*. □

Lemma 4.3
Für die Sprache

$$L = \{\lambda\}$$

gilt

$$L \in STAR \setminus 2COM.$$

Beweis. i) Setze $H = \{\lambda\}$. Dann folgt $L = H^*$; also liegt L in der Familie *STAR*.

ii) Wenn $L = EG^*H$ mit $G \notin \{\{\lambda\}, \emptyset\}$, so hat L entweder unendlich viele Wörter oder ist die leere Sprache, wenn E oder H leer sind. Es gibt also keine Möglichkeit, dass L ausschließlich das leere Wort enthält. Also kann L nicht in *2COM* liegen. □

Lemma 4.4
Die Sprachfamilie *STAR* ist unvergleichbar zu den Sprachfamilien *COM* und *2COM*.

Beweis. Nach Lemma 4.2 gibt es eine Sprache

$$L_1 = \{b^n a \mid n \geq 0\},$$

für die $L_1 \in COM \setminus STAR$ gilt. Da *COM* eine Teilmenge der Sprachfamilie *2COM* ist, folgt, $L_1 \in 2COM \setminus STAR$.

Weiterhin gilt für die Sprache $L_2 = \{\lambda\}$ die Beziehung $L_2 \in STAR \setminus 2COM$ nach Lemma 4.3. Da *COM* eine Teilmenge von *2COM* ist, liegt die Sprache L_2 auch in der Menge $STAR \setminus COM$.

Damit folgt die Behauptung. □

4.2 Zusammenfügen der beiden Hierarchien

Im Folgenden werden wir die beiden Hierarchien zusammenfügen.

4.2.1 Vorbereitungen

Wir tragen einige Sätze über die Struktur von einigen subregulären Sprachfamilien zusammen, die uns später bei verschiedenen Beweisen helfen werden.

Lemma 4.5 [11]
Eine Sprache L mit $L \in UF$ ist entweder unendlich oder enthält höchstens ein Wort.

Lemma 4.6 [11]
Für jede reguläre Sprache L ist die Sprache L^* regulär vereinigungsfrei.

Lemma 4.7 [11]
Das kürzeste Wort einer vereinigungsfreien Sprache ist eindeutig.

Lemma 4.8 [4]
Jede reguläre Sprache kann durch eine endliche Vereinigung von vereinigungsfreien Sprachen dargestellt werden.

Man kann an der Struktur von *2COM*-Sprachen ($L = EG^*H$) sehen, dass durch den Kleene-Abschluss der Menge G jede Sprache unendlich viele Wörter haben muss, wenn keine der drei Mengen E,G und H die leere Menge ist. Es kann aber auch eine der Mengen E oder H die leere sein. Dadurch können wir $L = \emptyset$ erzeugen.

Lemma 4.9
Für jede Sprache $L \in 2COM$ gilt: L ist entweder unendlich oder leer.

Beweis. Wenn $E = \emptyset$ oder $H = \emptyset$, dann gilt auch $EG^*H = \emptyset$ für jede Sprache G.

Falls $E \neq \emptyset$ und $H \neq \emptyset$, so folgt aus der Bedingung $G \notin \{\emptyset, \{\lambda\}\}$, dass EG^*H unendlich ist. □

Nun wollen wir noch etwas Ähnliches für *STAR*-Sprachen zeigen.

Lemma 4.10
Für jede Sprache $L \in STAR$ gilt: L ist entweder unendlich oder beinhaltet ausschließlich das leere Wort λ.

Beweis. Es sei L eine Sprache aus *STAR* über einem Alphabet V. Dann lässt sich L als H^* schreiben für eine Teilmenge $H \subseteq V^*$.

Setzen wir $H = \emptyset$, so folgt durch $L = H^* = \{\lambda\}$, dass die Sprache L aus dem Leerwort λ besteht.

Setzen wir nun $H = \{\lambda\}$, so folgt auch $L = \{\lambda\}^* = \{\lambda\}$. Die Sprache L enthält auch nur das leere Wort.

Sobald H ein anderes Element als das leere Wort λ enthält, wird L durch den Kleene-Abschluss von H unendlich viele Wörter enthalten. □

4.2.2 Einige Sprachen

Wir werden in diesem Abschnitt einige Sprachen in die jeweiligen Sprachfamilien einordnen. Diese werden wir in den Abschnitten zu den Inklusionen und Unvergleichbarkeiten verwenden, um zu zeigen, dass einige Inklusionen echt oder zwei Sprachfamilien unvergleichbar sind.

Lemma 4.11
Für die Sprache

$$L = \{a^{2n} \mid n \geq 0\}$$

gilt

$$L \in STAR \setminus PS.$$

Beweis. i) Setze $H = \{aa\}$, so folgt $L = H^*$. Das heißt

$$\{aa\}^* = \{a^{2n} \mid n \geq 0\} \in STAR.$$

ii) Für jedes $m \in \mathbb{N}$ gilt

$$J_a^m \cap L \neq \emptyset \quad \text{und} \quad J_a^m \not\subseteq L$$

für

$$J_a^m = \{a^n \mid n \geq m\}.$$

Es gilt nämlich für jedes $m \geq 0$, dass a^{2m} zu L gehört, aber a^{2m+1} nicht. Das heißt $\{a^{2n} \mid n \geq 0\} \notin PS$.

□

Lemma 4.12
Für die Sprache
$$L = \{a\}$$
gilt
$$L \in (FIN \cap COMM \cap UF) \setminus (STAR \cup PRE \cup SUF).$$

Beweis. i) Das einzige Wort, das in L enthalten ist, ist das Wort a. Das heißt $|L| = 1$ und damit gilt $L \in FIN$.

ii) Für das einzige Wort a in der Sprache L sind auch alle Permutationen in der Sprache enthalten. Das heißt $L \in COMM$.

iii) Die Sprache wird durch den regulären Ausdruck a beschrieben, welcher ohne Vereinigung beschrieben werden kann. Das heißt $L \in UF$.

iv) Nach Lemma 4.10 sind Sprachen in $STAR$ unendlich oder enthalten nur das leere Wort. Die Sprache L enthält aber das Wort a. Deswegen gilt $L \notin STAR$.

v) Vom Wort $a \in L$ liegt das Präfix und Suffix λ nicht in der Sprache L. Daraus folgt $L \notin PRE \cup SUF$. □

Lemma 4.13
Für die Sprache
$$L = \{a, b\}^*\{a, b\}$$
gilt
$$L \in COM \setminus UF.$$

Beweis. i) Setze $G = \{a, b\}$ und $H = \{a, b\}$, dann folgt $L = G^*H$. Dabei ist $G \notin \{\emptyset, \{\lambda\}\}$; also folgt $L \in COM$.

ii) Die Sprache enthält mit a und b mehr als ein kürzestes Wort. Daher gilt nach Lemma 4.7 die Beziehung $L \notin UF$. □

Lemma 4.14
Für die Sprache
$$L = \{\lambda\} \cup \{a\}$$

gilt

$$L \in INF \setminus (2COM \cup STAR).$$

Beweis. i) Zu dem Wort a gehören die Infixe a und λ zu der Sprache. Zu dem Wort λ gehört das Infix λ zu der Sprache. Damit gilt für alle Wörter, dass alle ihre Infixe in der Sprache liegen, also $L \in INF$.

ii) Jede nichtleere Sprache in $2COM$ ist nach Lemma 4.9 unendlich. Da L nichtleer aber endlich ist, gilt $L \notin 2COM$. □

Lemma 4.15
Für die Sprache

$$L = \{\lambda, a, b, ab\}$$

gilt

$$L \in INF \setminus (CIRC \cup UF).$$

Beweis. i) Für das Wort ab sind alle Infixe ab, a, b und λ in der Sprache. Für das Wort b sind auch alle Infixe b und λ, für das Wort a sind alle Infixe a und λ und für λ ist das Infix λ in der Sprache. Es gilt also, dass für jedes Wort alle Infixe in der Sprache L sind, also gilt $L \in INF$.

ii) Für das Wort ab ist die zirkuläre Permutation ba nicht in der Sprache. Also gilt $L \notin CIRC$.

iii) Nach Lemma 4.5 hat eine Sprache unendlich viele Wörter oder höchstens ein Wort, wenn sie vereinigungsfrei ist. Diese Sprache hat aber vier Wörter und kann demnach nicht vereinigungsfrei sein. □

Lemma 4.16
Für die Sprache

$$L = \{ab\}^*$$

gilt

$$L \in STAR \setminus (CIRC \cup PRE \cup SUF).$$

Beweis. i) Für $H = \{ab\}$ gilt $L = H^*$. Also gilt $L \in STAR$.

ii) Für das Wort ab ist die zirkuläre Permutation ba nicht in der Sprache. Also gilt $L \notin CIRC$.

iii) Für das Wort ab gehören das Präfix a und das Suffix b nicht zu der Sprache. Also gilt $L \notin PRE \cup SUF$. □

Lemma 4.17
Für die Sprache
$$L = \{a, b\}^*\{aa\}$$
gilt
$$L \in COM \setminus COMB.$$

Beweis. i) Für $G = \{a, b\}$ und $H = \{aa\}$ gilt $L = G^*H$ mit $G \notin \{\emptyset, \{\lambda\}\}$. Das heißt $L \in COM$.

ii) Angenommen, es gäbe eine Menge $A \subseteq V$, sodass $L = V^*A$ gelte. Da jedes Wort aus der Sprache auf aa endet, gilt $A = \{a\}$. Nun liegt aber auch das Wort ba in der Sprache V^*A, aber nicht in der Sprache L. Es gibt also keine Menge $A \subseteq V$, sodass $L = V^*A$. Also gilt $L \notin COMB$. □

Lemma 4.18
Für die Sprache
$$L = \{b^n a \mid n \geq 0\} \cup \{b^n c \mid n \geq 0\}$$
gilt
$$L \in COM \setminus UF.$$

Beweis. i) Wir setzen $G = \{b\}$ und $H = \{a, c\}$. Dann ist $L = G^*H$ und damit auch $L \in COM$.

ii) Nach Lemma 4.7 ist für jede Sprache in UF das kürzeste Wort eindeutig. Die Sprache L hat aber zwei kürzeste Wörter a und c. Damit gilt $L \notin UF$. □

Lemma 4.19
Für die Sprache
$$L = \{ab\}$$

gilt

$$L \in FIN \setminus (SUF \cup PRE).$$

Beweis. i) Da die Sprache nur ein Wort enthält, ist sie endlich.

ii) Für das Wort ab ist das Suffix und Präfix λ nicht in der Sprache. □

Lemma 4.20
Für die Sprache

$$L = Inf(\{ab^{2n}a \mid n \geq 1\})$$

gilt

$$L \in INF \setminus NC.$$

Beweis. i) Da wir hier einen Infixabschluss einer Sprache betrachen, ist L unter Infixbildung abgeschlossen.

ii) Angenommen, L wäre nicht-zählend. Dann folgt nach der Definition daraus, dass für alle Wörter $x, y, z \in \{a, b\}^*$ und für eine Zahl $k \geq 1$ die Äquivalenz

$$xy^kz \in L \Leftrightarrow xy^{k+1}z \in L$$

gilt. Wir setzen nun $x = z = a$ und $y = b$.

Wenn k gerade ist, dann folgt aus $ab^ka \in L$ aber $ab^{k+1}a \in L$, was ein Widerspruch ist.

Wenn k ungerade ist, dann folgt aus $ab^{k+1}a \in L$ aber $ab^ka \in L$, was ebenfalls ein Widerspruch ist.

Daraus folgt, dass L ist nicht-zählend. □

4.2.3 Inklusionen

Wir werden nun die Teilmengenbeziehungen zwischen den Sprachfamilien untersuchen. Um zu zeigen, dass die Teilmengenbeziehungen strikt sind, werden wir die Sprachen aus dem Abschnitt 4.2.2 verwenden.

Lemma 4.21
Es gilt die echte Inklusion

$$MON \subset INF.$$

Beweis. Im ersten Schritt werden wir die Inklusion $MON \subseteq INF$ zeigen. Danach werden wir zeigen, dass es eine Sprache $L \in INF \setminus MON$ gibt, die Inklusion also echt ist.

i) Es sei $L_1 \in MON$, dann gilt $L_1 = V^*$ für ein Alphabet V. Für jedes Wort $w \in V^*$ ist jedes Infix auch in V^* enthalten.

ii) Nach Lemma 4.20 gilt

$$L_2 = Inf(\{ab^{2n}a \mid n \geq 1\}) \in INF \setminus NC.$$

Die Sprache L_2 liegt auch in der Menge $INF \setminus MON$, da MON eine Teilmenge von NC ist. □

Lemma 4.22
Es gilt die echte Inklusion

$$MON \subset STAR.$$

Beweis. Es sei L eine Sprache aus MON. Dann gilt $L = V^*$ für ein Alphabet V. Da ein Alphabet eine endliche Menge an Symbolen ist, ist es insbesondere eine reguläre Sprache, woraus folgt, dass $L = H^*$ mit $H = V$ und $H \in REG$. □

Lemma 4.23
Es gilt die echte Inklusion

$$STAR \subset UF.$$

Beweis. Im ersten Schritt werden wir die Inklusion $STAR \subseteq UF$ zeigen. Danach werden wir im zweiten Schritt zeigen, dass es eine Sprache $L \in UF \setminus STAR$ gibt, die Inklusion also echt ist.

i) Nach Lemma 4.6 gilt für jede reguläre Sprache L, dass L^* vereinigungsfrei ist. Da für jede Sprache $L \in STAR$ gilt, dass $L = H^*$ mit einer regulären Sprache H über V ist, liegt jede Sprache aus $STAR$ auch in UF.

ii) Nach Lemma 4.12 liegt die Sprache $L = \{a\}$ in UF, aber nicht in $STAR$. Also gilt die echte Inklusion. □

Lemma 4.24
Es gilt die strikte Inklusion $PRE \subset PS$.

Beweis. Im ersten Schritt werden wir zeigen, dass $PRE \subseteq PS$ gilt. Im zweiten Schritt werden wir dann zeigen, dass diese Inklusion echt ist. Der Beweis wird analog zu [8, Theorem 3.4] geführt.

i) Angenommen, es gäbe eine unter Präfixbildung abgeschlossene Sprache L, die aber nicht potenzseparierend ist. Es sei V ein Alphabet so, dass $L \subseteq V^*$. Da die Sprache L nicht potenzseparierend ist, gilt für jede natürliche Zahl $m \geq 1$, dass ein Wort $x \in V^*$ existiert, sodass es eine natürliche Zahl $k_1 \geq m$ mit $x^{k_1} \in L$ und eine natürliche Zahl $k_2 \geq m$ mit $x^{k_2} \notin L$ gibt. Da L unter Präfixbildung abgeschlossen ist, wissen wir, dass $k_1 < k_2$ ist und dass $x^{k_2+i} \notin L$ für alle $i \geq 0$ gilt.

Die Sprache L ist regulär. Es sei also $G = (N, T, P, S)$ eine reguläre Grammatik, die die Sprache L erzeugt. Es sei

$$m = |N| + 1.$$

Weiterhin seien $x \in V^*$ ein Wort und k_1 und k_2 zwei natürliche Zahlen mit

$$m \leq k_1 \leq k_2$$

so, dass

$$x^{k_1} \in L \text{ und } x^{k_2+i} \notin L$$

für alle $i \geq 0$.

Da das Wort x^{k_1} in der Sprache L liegt, gibt es eine Ableitung

$$A_0 \Longrightarrow^* xA_1 \Longrightarrow^* x^2A_2 \Longrightarrow^* \cdots \Longrightarrow^* x^{k_1-1}A_{k_1-1} \Longrightarrow^* x^{k_1},$$

bei der $S = A_0$ ist und $A_i \in N$ für $0 \leq i \leq k_1 - 1$ gilt. Da $k_1 \geq m > |N|$ gilt, gibt es zwei identische Nichtterminale unter $A_0, A_1, \ldots, A_{k_1-1}$. Es seien e und f zwei Indizes mit $0 \leq e < f \leq k_1 - 1$ und $A_e = A_f$. Dann existiert auch eine Ableitung

$$\begin{aligned}
A_0 &\Longrightarrow^* x^e A_e \\
&\Longrightarrow^* x^{e+f-e}A_f = x^{e+f-e}A_e \\
&\Longrightarrow^* x^{e+2(f-e)}A_f = x^{e+2(f-e)}A_e \\
&\Longrightarrow^* x^{e+k_2(f-e)}A_f = x^{k_2+e+k_2(f-e-1)}A_f \\
&\Longrightarrow^* x^{k_2+e+k_2(f-e-1)+(k_1-f)}.
\end{aligned}$$

Da wir das Wort mit der Grammatik ableiten können, gehört es offensichtlich zur Sprache L, aber

$$e + k_2(f - e - 1) + (k_1 - f) \geq 0$$

und wir haben gezeigt, dass kein Wort x^{k_2+i} mit $i \geq 0$ zur Sprache L gehört. Dieser Widerspruch zeigt, dass es keine unter Präfixbildung abgeschlossene reguläre Sprache geben kann, die dazu nicht potenzseparierend ist. Daraus folgt die Behauptung $PRE \subseteq PS$.

ii) Die Inklusion ist echt, da nach Lemma 4.19 die Sprache

$$L = \{ab\}$$

endlich ist, also in *FIN* liegt, aber nicht in *PRE*. Da diese Sprache in *FIN* liegt und *FIN* eine echte Teilmenge von *PS* ist, liegt sie auch in *PS*. □

4.2.4 Unvergleichbarkeiten

Nachdem wir die Teilmengenbeziehungen untersucht haben, werden wir für die übrigen Sprachfamilien zeigen, dass diese jeweils paarweise unvergleichbar sind.

Lemma 4.25
Die Sprachfamilie *STAR* ist unvergleichbar zu den Sprachfamilien *CIRC* und *COMM*.

Beweis. i) Es ist bereits bekannt, dass *COMM* und *UF* unvergleichbar sind [8, Lemma 4.9]. Also existiert eine Sprache L_1 in $COMM \setminus UF$. Da *STAR* eine echte Teilmenge von *UF* ist, gilt für die Sprache auch, dass L_1 in *COMM* liegt, aber nicht in *STAR*. Da *COMM* eine echte Teilmenge von *CIRC* ist, ist die Sprache L_1 auch in *CIRC* enthalten.

ii) Für die Sprache

$$L_2 = \{ab\}^*$$

gilt, dass sie in *STAR* liegt, aber nicht in *CIRC*. Aus der Inklusion $COMM \subset CIRC$ folgt, dass L_2 auch nicht in *COMM* liegt.

Folglich ist *STAR* ist unvergleichbar zu den Sprachfamilien *COMM* und *CIRC*. □

Lemma 4.26
Die Sprachfamilie *STAR* ist unvergleichbar mit allen Sprachfamilien der Menge $\mathcal{F}$ mit

$$\mathcal{F} = \{FIN, NIL, DEF, ORD, NC, SF, PS\}.$$

Beweis. i) Nach Theorem 4.1 sind die Sprachen *FIN* und *UF* unvergleichbar. Also existiert eine Sprache

$$L_1 \in FIN \setminus UF,$$

die aber nicht vereinigungsfrei ist. Da *STAR* eine Teilmenge von *UF* und *FIN* eine Teilmenge aller Sprachfamilien aus $\mathcal{F}_1$

ist, gilt für diese Sprache L_1, dass sie in $F \setminus STAR$ liegt für alle $F \in \mathcal{F}_1$.

ii) Wir setzen

$$L_2 = \{a^{2n} \mid n \geq 0\}.$$

Nach Lemma 4.11 liegt diese Sprache in $STAR \setminus PS$. Aus obiger Teilmengenrelation folgt, dass $L_2 \in STAR \setminus F$ für alle $F \in \mathcal{F}_1$ gilt.

Damit folgt die Behauptung. □

Lemma 4.27
Jede der Sprachfamilien *STAR* und *UF* ist unvergleichbar mit den Sprachfamilien *PRE*, *SUF* und *INF*.

Beweis. Da *INF* eine Teilmenge von *SUF* und *PRE* ist und diese jeweils eine Teilmenge von *PS*, reicht es zu zeigen, dass es eine Sprache $L_1 \in INF \setminus UF$ gibt, da diese Sprache auch in $PRE \setminus STAR$ und in $SUF \setminus STAR$ liegt und dass es eine Sprache $L_2 \in STAR \setminus PS$ gibt. Diese Sprache liegt dann auch in der Menge $STAR \setminus (PRE \cup SUF \cup INF)$.

i) Wir setzen

$$L_1 = \{\lambda, a, b, ab\}.$$

Nach Lemma 4.15 liegt diese Sprache in $INF \setminus UF$.

ii) Weiterhin setzen wir

$$L_2 = \{a^{2n} \mid n \geq 0\}.$$

Nach Lemma 4.11 liegt diese Sprache in $STAR \setminus PS$.

Damit folgt die Behauptung. □

Lemma 4.28
Die Sprachfamilien *MON*, *STAR* und *UF* sind unvergleichbar mit den Sprachfamilien *COM* und *2COM*.

Beweis. i) Die Sprache

$$L_1 = \{\lambda\}$$

liegt in *MON*, und nach Lemma 4.3 nicht in *2COM*.

Da *MON* eine Teilmenge von *STAR* und *STAR* eine Teilmenge von *UF* ist, liegt die Sprache L_1 auch in $STAR \setminus 2COM$ und in $UF \setminus 2COM$ und da *COM* eine Teilmenge von *2COM* ist, liegt L_1 auch in $STAR \setminus COM$ und in $UF \setminus COM$.

iii) Nach Lemma 4.13 liegt die Sprache

$$L_2 = \{a, b\}^*\{a, b\}$$

in *COM* und damit auch in *2COM*, aber nicht in *UF* und damit nicht in *STAR* und nicht in *MON*.

Damit folgt die Behauptung. □

Lemma 4.29
Die Sprachfamilien *INF* und *PRE* sind unvergleichbar zu den Sprachfamilien der Menge $\mathcal{F}_2$ mit

$$\mathcal{F}_2 = \{FIN, NIL, DEF, ORD, NC \text{ und } SF\}.$$

Beweis. Da *INF* eine echte Teilmenge von *PRE* ist und jede Familie $F \in \mathcal{F}_2$ eine Teilmenge von *NC* und Obermenge von *FIN* ist, müssen wir nur zeigen, dass es zwei Sprachen $L_1 \in FIN \setminus PRE$ und $L_2 \in INF \setminus NC$ gibt.

i) Nach Lemma 4.19 liegt die Sprache

$$L_1 = \{ab\}$$

in *FIN*, aber nicht in *PRE*.

ii) Nach Lemma 4.20 liegt die Sprache

$$L_2 = Inf(\{ab^{2n}a \mid n \geq 1\})$$

in *INF*, aber nicht in *NC*.

Damit folgt die Behauptung. □

4.3 Ergebnisse

Aus den gezeigten Beziehungen folgt die in Abbildung 4.3 dargestellte Gesamthierarchie.

Theorem 4.30 – Neue Hierarchie
Die in Abbildung 4.3 dargestellten Inklusionen gelten.

Ein Pfeil von einem Knoten X zu einem Knoten Y bedeutet die echte Inklusion $X \subset Y$. Falls zwei Familien nicht durch einen Pfad verbunden sind, so sind diese unvergleichbar.

Beweis. Die Verweise zu den Beweisen für die Inklusionen sind auf den jeweiligen Pfeilen angegeben.

Verweise zu bereits bekannten Teilmengenbeziehungen wurden nicht erneut aufgeschrieben. Diese sind in Abbildung 4.1 und Abbildung 4.2 zu finden.

Die Unvergleichbarkeiten wurden mit den Lemmata aus dem Abschnitt 4.2.4 bewiesen. $\square$

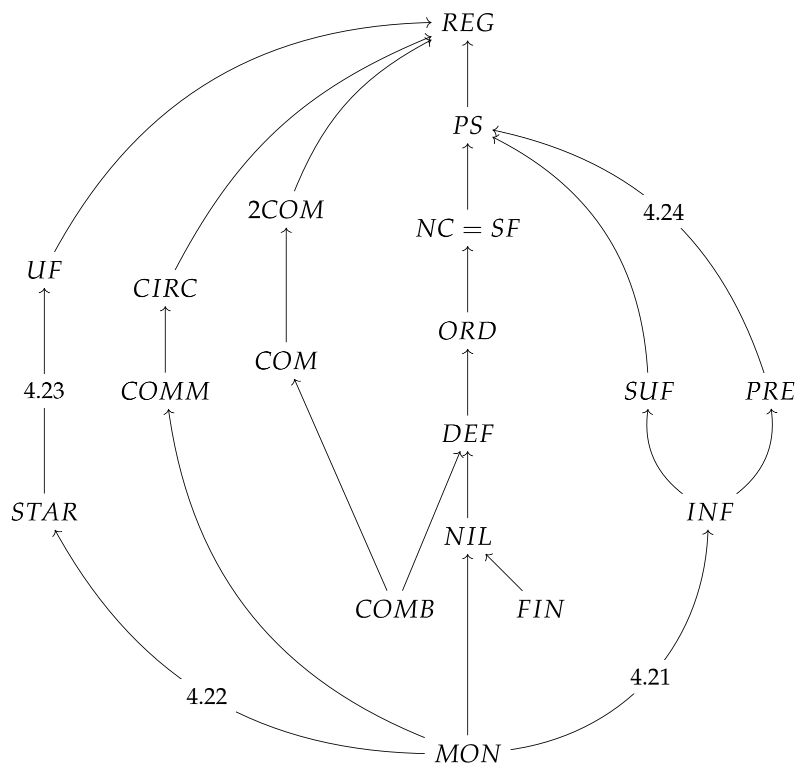

Abbildung 4.3: Hierarchie von subregulären Sprachfamilien

Kapitel 5

Hierarchie der $\mathcal{EC}$-Sprachfamilien

Im nächsten Abschnitt wollen wir die Hierarchie von Sprachfamilien, die mit externen kontextualen Grammatiken mit subregulären Steuersprachen erzeugt werden, untersuchen.

Wenn bei einer kontextualen Grammatik alle Auswahlsprachen zu einer Sprachfamilie X gehören, so liegen sie auch in jeder Obermenge Y von X. Daher wird jede Sprache in $\mathcal{EC}(X)$ auch von einer kontextualen Grammatik mit Auswahlsprachen aus Y erzeugt und es gilt die folgende Monotonie.

Lemma 5.1
Für zwei Sprachfamilien X und Y mit $X \subseteq Y$ gilt

$$\mathcal{EC}(X) \subseteq \mathcal{EC}(Y).$$

In Abbildung 5.1 geben wir den aktuellen Stand der Forschung an. Ein Pfeil von einer Sprachfamilie X zu einer Sprachfamilie Y zeigt, dass die Sprachfamilie X eine echte Teilmenge der Sprachfamilie Y ist. Sind zwei Sprachfamilien nicht verbunden, so sind sie unvergleichbar.

Theorem 5.2 – Aktueller Forschungsstand
Es gelten die Beziehungen in Abbildung 5.1.

Beweis. Auf den Pfeilen ist jeweils ein Verweis auf die Veröffentlichung, in der die jeweilige Inklusion bewiesen wurde. Die Unvergleichbarkeiten wurden zusammengenommen ebenfalls in diesen Arbeiten bewiesen. □

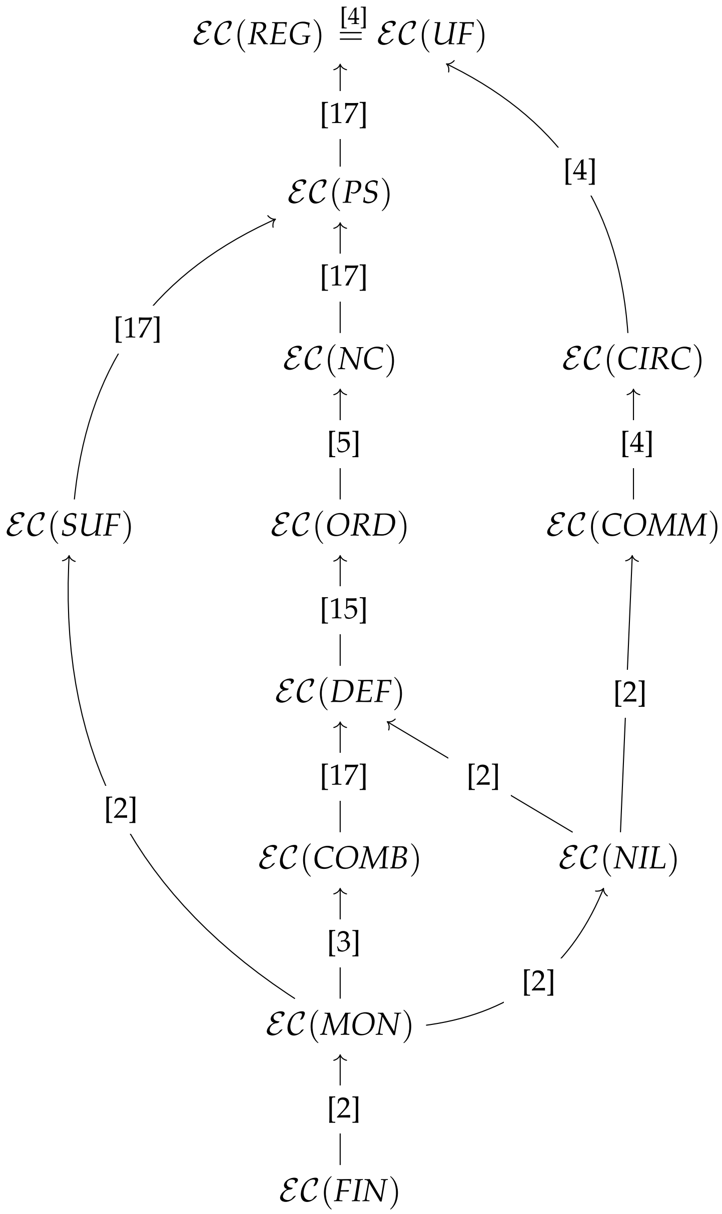

Abbildung 5.1: Bisherige Ergebnisse

5.1 Einige Sprachen

Wir untersuchen zunächst ein paar Sprachen hinsichtlich ihrer Zugehörigkeit zu Sprachfamilien. Damit zeigen wir im Anschluss echte Teilmengenbeziehungen und Unvergleichbarkeiten.

Den folgenden Beweisen liegt die gleiche Methode zugrunde wie in [2].

Lemma 5.3
Für die Sprache

$$L = \{a^n b^n \mid n \geq 1\} \cup \{b^n \mid n \geq 1\}$$

gilt

$$L \in (\mathcal{EC}(COMM) \cap \mathcal{EC}(PRE)) \setminus (\mathcal{EC}(DEF) \cup \mathcal{EC}(SUF)).$$

Beweis. In [2, Lemma 3.3] wurde bereits gezeigt, dass

$$L \in \mathcal{EC}(COMM) \setminus (\mathcal{EC}(DEF) \cup \mathcal{EC}(SUF))$$

gilt. Wir zeigen noch, dass $L \in \mathcal{EC}(PRE)$.

Die kontextuale Grammatik $G = (\{a,b\}, \mathcal{S}, \{ab, b\})$, wobei $\mathcal{S}$ aus den Regeln

$$\begin{array}{rcl} Pre(\{a^n b^m \mid n,m \geq 1\}) & \to & (a,b) \\ \{b\}^+ & \to & (\lambda, b) \end{array}$$

besteht, erzeugt die Sprache L, wobei jede der Auswahlsprachen von $\mathcal{S}$ unter Präfixbildung abgeschlossen ist.

Auf das erste Element der Axiomenmenge ab können wir die erste Regel und nur die anwenden. Wir haben also nur die Möglichkeit, die folgende Ableitung auszuführen:

$$ab \underset{ex}{\Longrightarrow} a^2 b^2.$$

Diese Ableitungsregel können wir nun immer wieder anwenden und damit $a^k b^k$ zu $a^{k+1}b^{k+1}$ ableiten. Damit können wir die Sprache $\{a^n b^n \mid n \geq 1\}$ erzeugen.

Nehmen wir stattdessen das zweite Element der Axiomenmenge b. Damit können wir die erste Regel nicht anwenden, da b nicht in der Sprache $Pre(\{a^n b^m \mid n, m \geq 1\})$ liegt. Die zweite Regel können wir aber verwenden und damit b zu b^2 ableiten. Auf alle anderen Wörter b^i mit $i \geq 2$ können wir auch immer die zweite Regel anwenden, aber die erste nicht, sodass wir damit $\{b^n \mid n \geq 1\}$ erzeugen.

Wir können also entweder $\{a^n b^n \mid n \geq 1\}$ oder $\{b^n \mid n \geq 1\}$, also genau L erzeugen. Folglich gilt $L = L(G)$. $\square$

Für die folgende Sprache wandeln wir die Sprache aus Lemma 5.3 etwas ab.

Lemma 5.4
Für die Sprache

$$L = \{a^n b^n \mid n \geq 1\} \cup \{a^n \mid n \geq 1\}$$

gilt $L \in (\mathcal{EC}(COMM) \cap \mathcal{EC}(SUF)) \setminus \mathcal{EC}(PRE)$.

Beweis. i) Die kontextuale Grammatik $G = (\{a, b\}, \mathcal{S}, \{ab, a\})$, wobei $\mathcal{S}$ aus den Regeln

$$\begin{aligned} \{w \mid w \in \{a,b\}^+, |w|_b > 0\} \quad &\rightarrow \quad (a, b) \\ \{a\}^+ \quad &\rightarrow \quad (\lambda, a) \end{aligned}$$

besteht, erzeugt die Sprache L, wobei jede der Auswahlsprachen von $\mathcal{S}$ kommutativ ist. Also gilt $L \in \mathcal{EC}(COMM)$.

ii) Die kontextuale Grammatik $G = (\{a, b\}, \mathcal{S}, \{ab, a\})$, wobei $\mathcal{S}$ aus den Regeln

$$\begin{aligned} Suf(\{a^n b^m \mid n, m \geq 1\}) \quad &\rightarrow \quad (a, b) \\ \{a\}^+ \quad &\rightarrow \quad (\lambda, a) \end{aligned}$$

besteht, erzeugt die Sprache L, wobei jede der Auswahlsprachen von $\mathcal{S}$ unter Suffixbildung abgeschlossen ist.

Der Beweis kann so ähnlich wie bei Lemma 5.3 geführt werden.

iii) Wir nehmen an, dass $L \in \mathcal{EC}(PRE)$. Dann gilt $L = L(G)$ für eine kontextuale Grammatik

$$G = (\{a, b\}, \{(S_1, C_1), (S_2, C_2), \dots, (S_n, C_n)\}, A),$$

bei der alle Auswahlsprachen S_i mit $(1 \leq i \leq n)$ unter Präfixbildung abgeschlossen sind. Wir definieren

$$\begin{aligned} t_1 &= \max\{|z| \mid z \in A\}, \\ t_2 &= \max\{|u| + |v| \mid (u, v) \in C_i,\ 1 \leq i \leq n\} \end{aligned}$$

und wählen $a^n b^n \in L$ für ein $n \geq t_1 + t_2 + 1$. Es gibt also ein Wort $z \in L(G)$, sodass wir z mit einem externen Ableitungsschritt zu $a^n b^n$ ableiten können. Das heißt, es gibt einen Index i mit $1 \leq i \leq n$, sodass $(u, v) \in C_i$ mit $a^n b^n = uzv$. Es gilt $z = a^m b^m$ mit einem $m \geq 1$, $u = a^l$ und $v = b^l$ für ein $l \geq 1$. Also gilt $a^m b^m \in S_i$. Nach unserer Voraussetzung ist S_i unter Präfixbildung abgeschlossen, also ist auch $a^m \in S_i$. Wir können also $a^m \in L = L(G)$ um einen Kontext (u, v) erweitern, sodass

$$L \ni a \underset{ex}{\Longrightarrow} a^{m+l} b^l \notin L$$

gilt. Aus diesem Widerspruch folgt die Behauptung. □

Lemma 5.5
Für die Sprache

$$L = \{bbba^n \mid n \geq 1\} \cup \{bb\}$$

gilt $L \in \mathcal{EC}(NIL) \setminus \mathcal{EC}(PRE)$.

Beweis. i) Die kontextuale Grammatik

$$G = (\{a,b\}, \mathcal{S}, \{bbba, bb\}),$$

wobei $\mathcal{S}$ aus der Regel

$$\{a,b\}^4\{a,b\}^* \quad \to \quad (\lambda, a)$$

besteht, erzeugt die Sprache L, wobei die Auswahlsprache von $\mathcal{S}$ nilpotent ist.

Aus dem Wort bb entstehen keine Wörter, aus dem Wort $bbba$ entstehen sukzessive alle anderen Wörter der Sprache.

ii) Angenommen die Sprache L würde von einer Grammatik G erzeugt werden, die nur unter Präfixbildung abgeschlossene Auswahlsprachen enthält. Da die Sprache unendlich ist, gibt es eine Auswahlsprache, die beim Ableiten verwendet wird und das Wort $bbba^n$ für ein $n \geq 1$ enthält. Der dazugehörige Kontext ist (λ, a^m) für ein $m \geq 1$. Da die Auswahlsprache unter Präfixbildung abgeschlossen ist, enthält sie auch das Wort bb. Wir können also den Kontext an das Wort bb anhängen, wobei aber bba^m für kein $m \geq 1$ zu L gehört. Aus diesem Widerspruch folgt die Behauptung. □

Lemma 5.6
Für die Sprache

$$L = \{bbba^n \mid n \geq 1\} \cup \{\lambda\}$$

gilt $L \in \mathcal{EC}(NIL) \setminus \mathcal{EC}(STAR)$.

Beweis. i) Die kontextuale Grammatik

$$G = (\{a,b\}, \mathcal{S}, \{bbba, \lambda\}),$$

wobei $\mathcal{S}$ aus den Regeln

$$\{a,b\}^4\{a,b\}^* \quad \to \quad (\lambda, a)$$

besteht, erzeugt die Sprache L, wobei die Auswahlsprache von $\mathcal{S}$ nilpotent ist.

ii) Angenommen die Sprache L würde von G erzeugt werden, wobei alle Auswahlsprachen von G in $STAR$ liegen. Da die Sprache L unendlich ist, gibt es eine Auswahlsprache, die das Wort $bbba^n$ für ein $n \geq 1$ enthält. Der dazugehörige Kontext ist (λ, a^m) für ein $m \geq 1$. Da die Auswahlsprache in $STAR$ liegt, enthält sie auch das Leerwort λ. Wir können also den Kontext um das Wort λ bauen, wobei aber a^m für kein $m \geq 1$ zu L gehört. Aus diesem Widerspruch folgt die Behauptung. □

Lemma 5.7
Für die Sprache

$$L = \{b^n a \mid n \geq 0\} \cup \{\lambda\}$$

gilt $L \in \mathcal{EC}(COMB) \setminus (\mathcal{EC}(PRE) \cup \mathcal{EC}(STAR))$.

Beweis. i) Die kontextuale Grammatik $G = (\{a,b\}, \mathcal{S}, \{\lambda, a\})$, wobei $\mathcal{S}$ aus der Regel

$$\{a,b\}^*\{a\} \quad \rightarrow \quad (b, \lambda)$$

besteht, erzeugt die Sprache L, wobei die Auswahlsprache von $\mathcal{S}$ kombinational ist.

Aus dem leeren Wort λ entstehen keine neuen Wörter. Es ist einfach ein Element der Sprache. Das zweite Axiom a liegt aber in der Auswahlsprache. Wir können also a mit dem Kontext (b, λ) versehen, sodass a zu ba abgeleitet wird. Sukzessive erhalten wir den ersten Teil der Sprache. Da wir keine weiteren Axiome und keine weiteren Kontexte haben, können wir nur Wörter der Form $b^n a$ für $n \geq 1$ oder eben das Leerwort λ erzeugen.

ii) Angenommen die Sprache L würde von einer Grammatik $G = (\{a,b\}, \mathcal{S}, A)$ erzeugt werden, wobei $\mathcal{S}$ nur unter Präfixbildung abgeschlossene Auswahlsprachen enthält. Da die

Sprache L unendlich ist, gibt es eine Auswahlsprache, die das Wort $b^n a$ für ein $n \geq 1$ enthält. Der dazugehörige Kontext ist (b^m, λ) für ein $m \geq 1$. Da die Auswahlsprache unter Präfixbildung abgeschlossen ist, enthält sie auch das Leerwort λ. Wir können also den Kontext um λ bauen, wobei aber b^m für kein $m \geq 1$ zu L gehört. Aus diesem Widerspruch folgt die Behauptung.

iii) Angenommen die Sprache L würde von einer Grammatik $G = (\{a, b\}, \mathcal{S}, A)$ erzeugt werden, wobei $\mathcal{S}$ nur Auswahlsprachen aus *STAR* enthält. Da die Sprache unendlich ist, gibt es eine Auswahlsprache, die das Wort $b^n a$ für ein $n \geq 1$ enthält. Der dazugehörige Kontext ist (b^m, λ) für ein $m \geq 1$. Da die Auswahlsprache in *STAR* liegt, enthält sie auch das Leerwort λ. Wir können also den Kontext um λ bauen, wobei aber b^m für kein $m \geq 1$ zu L gehört. Aus diesem Widerspruch folgt die Behauptung. □

Lemma 5.8
Für die Sprache

$$L = \{a^n b^n \mid n \geq 1\} \cup \{b^n a^n \mid n \geq 1\}$$

gilt $L \in (\mathcal{EC}(INF) \cap \mathcal{EC}(STAR)) \setminus \mathcal{EC}(CIRC)$.

Beweis. i) Die kontextuale Grammatik

$$G = (\{a, b\}, \mathcal{S}, \{ab, ba\}),$$

wobei $\mathcal{S}$ aus den Regeln

$$\begin{array}{lcl} Inf(\{a\}^*\{b\}^*) & \to & (a, b) \\ Inf(\{b\}^*\{a\}^*) & \to & (b, a) \end{array}$$

besteht, erzeugt die Sprache L, wobei jede der Auswahlsprachen von $\mathcal{S}$ unter Infixbildung abgeschlossen ist.

Wir sehen uns zunächst die möglichen Ableitungen mit dem ersten Axiom an. Wir können das Axiom ab mit dem ersten

Kontext erweitern, da es in der ersten Auswahlsprache liegt. Also gilt

$$ab \underset{ex}{\Longrightarrow} a^2b^2.$$

Auf die gleiche Weise entstehen aus den Wörtern a^nb^n für $n \geq 1$ die Wörter $a^{n+1}b^{n+1}$ und keine anderen. Das zweite Axiom funktioniert analog. Hier erhalten wir die Wörter der Form b^na^n für $n \geq 1$. Daraus folgt, dass $L = L(G)$ gilt.

ii) Die kontextuale Grammatik $G = (\{a,b\}, \mathcal{S}, \{ab, ba\})$, wobei $\mathcal{S}$ aus den Regeln

$$\begin{array}{lcl} \{\{a\}^+\{b\}^+\}^* & \rightarrow & (a,b) \\ \{\{b\}^+\{a\}^+\}^* & \rightarrow & (b,a) \end{array}$$

besteht, erzeugt die Sprache L, wobei jede der Auswahlsprachen von $\mathcal{S}$ in *STAR* liegt.

Wie im ersten Teil entstehen auch hier die Wörter a^nb^n für $n \geq 1$ aus ab und die anderen aus ba.

iii) Wir nehmen an, dass L in $\mathcal{EC}(CIRC)$ liegt. Dann gilt $L = L(G)$ für eine kontextuale Grammatik

$$G = (\{a,b\}, \{(S_1, C_1), (S_2, C_2), \ldots, (S_n, C_n)\}, A),$$

bei der alle Auswahlsprachen S_i mit $(1 \leq i \leq n)$ unter zirkulärer Permutation abgeschlossen sind.

Wir definieren

$$t = \max\{|z| \mid z \in A\}$$

und wählen $a^nb^n \in L$ für ein $n \geq t$. Es gibt also ein $z \in L(G)$, sodass wir z mit einem externen Ableitungsschritt zu a^nb^n ableiten können. Das heißt, es gibt ein i mit $1 \leq i \leq n$, sodass $(u,v) \in C_i$ mit $a^nb^n = uzv$ existiert. Es gilt $z = a^mb^m$ für eine Zahl $m \geq 1$, $u = a^l$ und $v = b^l$ für ein $l \geq 1$. Also gilt $a^mb^m \in S_i$. Nach unserer Voraussetzung ist S_i unter zirkulärer Permutation abgeschlossen, also ist auch $b^ma^m \in S_i$. Wir können also $b^m a^m \in L = L(G)$ um den Kontext (u, v) erweitern, sodass

$$L \ni b^m a^m \underset{ex}{\Longrightarrow} a^l b^m a^m b^l \notin L$$

gilt. Aus diesem Widerspruch folgt die Behauptung. □

Lemma 5.9
Es seien

$$\begin{aligned} L_1 &= \{a^m bc^{2n} ba^m \mid n \geq 1, m \geq 0\}, \\ L_2 &= \{c^n \mid n \geq 2\}, \\ L_3 &= \{bc^n b \mid n \geq 2\}, \\ L_4 &= \{ac^n a \mid n \geq 2\}. \end{aligned}$$

Für die Sprache

$$L = L_1 \cup L_2 \cup L_3 \cup L_4$$

gilt

$$L \in \mathcal{EC}(INF) \setminus \mathcal{EC}(NC).$$

Beweis. i) Die kontextuale Grammatik $G = (\{a, b, c\}, \mathcal{S}, \{cc\})$, wobei $\mathcal{S}$ aus den Regeln

$$\begin{aligned} \{c\}^* &\rightarrow (\lambda, c) \mid (b, b) \\ Inf(\{a^m bc^{2n} ba^m \mid n \geq 1, m \geq 0\}) &\rightarrow (a, a) \end{aligned}$$

besteht, erzeugt die Sprache L, wobei jede der Auswahlsprachen von $\mathcal{S}$ unter Infixbildung abgeschlossen ist.

Das Axiom cc liegt in beiden Auswahl sprachen. Wenden wir die erste Regel beliebig oft an, so erhalten wir genau die Wörter der Form c^n für $n \geq 2$. Wenden wir den zweiten Kontext der ersten Auswahlsprache an, so erhalten wir die Wörter $bc^n b$ für $n \geq 2$. Sobald wir diesen Kontext verwenden, liegt unser Wort nicht mehr in der ersten Auswahlsprache. Wenn die Anzahl der c in unserem Wort gerade ist, so können wir die Ersetzungsregel der zweiten Auswahlsprache wiederholt verwenden, um das Wort zu einem beliebigen Wort der Form $a^m bc^{2n} ba^m$ für $m \geq 1$ abzuleiten. Andere Wörter entstehen dabei nicht.

Wenden wir auf die Wörter c^n mit $n \geq 2$ die Ersetzung der zweiten Auswahlsprache an, erhalten wir genau die Wörter $ac^n a$ für $n \geq 2$. Diese Wörter liegen in keiner der beiden Auswahlsprachen.

Wir sehen also, dass die Grammatik die Sprache L erzeugt.

ii) Wir nehmen an, dass $L \in \mathcal{EC}(NC)$ gilt. Dann gilt $L = L(G)$ für eine kontextuale Grammatik

$$G = (\{a, b, c\}, \{(S_1, C_1), (S_2, C_2), \ldots, (S_n, C_n)\}, A),$$

bei der alle Auswahlsprachen S_i mit einem Index i mit $1 \leq i \leq n$ nicht-zählend sind.

Da alle Auswahlsprachen S_i nicht-zählend sind, gibt es zu jeder Sprache eine natürliche Zahl k so, dass für alle Wörter $x, y, z \in V^*$ gilt $xy^k z \in S_i \Leftrightarrow xy^{k+1} z \in S_i$. Wir bezeichnen unsere Auswahlsprachen S_i als $S_i^{(k)}$, wobei k jeweils die kleinste natürliche Zahl im Sinne der Definition der nicht-zählenden Sprachen ist.

Des Weiteren definieren wir

$$p = \max\{k \mid (S^{(k)}, C) \in \mathcal{S}\}.$$

Damit gilt für jede Auswahlsprache S_i die Aussage: Für alle Wörter $x, y, z \in V^*$ gilt

$$xy^p z = xy^k y^{p-k} z \in S_i \Leftrightarrow xy^{k+1} y^{p-k} z = xy^{p+1} z \in S_i. \quad (5.1)$$

Wir definieren außerdem

$$t = \max\{|z| \mid z \in A\} + 1.$$

Da die Sprache L_1 Wörter mit beliebiger Anzahl an Buchstaben enthält, gibt es eine Ableitung

$$w_0 \Rightarrow^* w_1 \Rightarrow uw_1v$$

mit $w_0 \in A$, $|w_1|_c > p + t$, $|w_1|_a = 0$ und $|uv|_a > 0$.

Wir unterscheiden nun zwei Fälle. Im ersten Fall beginnt das Wort w_1 mit einem b, im zweiten Fall beginnt das Wort w_1 mit einem c.

Fall 1 – w_1 beginnt mit b:

Daraus folgt, dass $w_1 = bc^kb$ mit $k > p + t$ gilt. Da uw_1v in der Sprache L_1 liegt (es liegt in L und enthält sowohl ein a als auch ein b), ist k gerade.

Die verwendete Auswahlsprache in dem Ableitungsschritt $w_1 \Rightarrow uw_1v$ sei S_i. Da w_1 in S_i liegt, erhalten wir wegen der Beziehung (5.1)

$$w_1 = bc^pc^{k-p}b \in S_i \Leftrightarrow bc^{p+1}c^{k-p}b = bc^{k+1}b \in S_i.$$

Da $bc^{k+1}b$ in L_3 und damit auch in L liegt, wird auch das Wort $ubc^{k+1}bv$ abgeleitet. Da aber $k + 1$ ungerade und $|uv|_a > 0$ ist, liegt dieses Wort nicht in L, was ein Widerspruch ist zu $L = L(G)$ ist.

Fall 2 – w_1 beginnt mit c:

Dann folgt $w_1 = c^k$ mit $k > p + t$. Da $uw_1v \in L_1$ gilt, folgt daraus $|u|_b = |v|_b = 1$ und nach unserer Bedingung auch $|uv|_a > 0$.

Die verwendete Auswahlsprache in dem Ableitungsschritt $w_1 \Rightarrow uw_1v$ sei S_i. Da w_1 in S_i liegt, erhalten wir wegen der Beziehung (5.1)

$$w_1 = c^pc^{k-p} \in S_i \Leftrightarrow c^{p+1}c^{k-p} = c^{k+1} \in S_i.$$

Da c^{k+1} in L_2 und damit in L liegt, wird auch das Wort

$uc^{k+1}v$ abgeleitet. Da $|uv|_a > 0$ und $|uv|_b > 0$ gilt, liegt das Wort $uc^{k+1}v$ höchstens in L_1. Da aber uc^kv in L_1 liegt, ist $|uc^kv|_c$ eine gerade Zahl und $|uc^{k+1}v|_c$ eine ungerade Zahl. Folglich liegt $uc^{k+1}v$ nicht in L_1 und auch nicht in L, was ein Widerspruch zu $L = L(G)$ ist.

Da alle Fälle zu einem Widerspruch führen, ist die Annahme, dass L in $\mathcal{EC}(NC)$ liegt, falsch. □

Lemma 5.10
Es seien

$$L_1 = \{ba^{2n}b \mid n \geq 1\},$$
$$L_2 = \{a^n \mid n \geq 2\}.$$

Die Sprache

$$L = L_1 \cup L_2$$

liegt in

$$\mathcal{EC}(STAR) \setminus \mathcal{EC}(PS).$$

Beweis. i) Die kontextuale Grammatik

$$G = (\{a, b\}, \mathcal{S}, \{aa\}),$$

wobei $\mathcal{S}$ aus den Regeln

$$\begin{aligned} \{a^n \mid n \geq 0\} &\rightarrow (\lambda, a) \\ \{a^{2n} \mid n \geq 0\} &\rightarrow (b, b) \end{aligned}$$

besteht, erzeugt die Sprache L. Die Auswahlsprache

$$S_1 = \{a^n \mid n \geq 0\}$$

lässt sich mit $H = \{a\}$ als $S_1 = H^*$ und die Auswahlsprache

$$S_2 = \{a^{2n} \mid n \geq 0\}$$

lässt sich mit $H' = \{aa\}$ durch $S_2 = (H')^*$ beschreiben. Damit genügen beide Sprachen den Voraussetzungen einer *STAR*-Sprache.

ii) Wir nehmen an, dass $L \in \mathcal{EC}(PS)$ gilt. Dann gilt $L = L(G)$ für eine kontextuale Grammatik

$$G = (\{a, b, c\}, \{(S_1, C_1), (S_2, C_2), \ldots, (S_n, C_n)\}, A),$$

bei der alle Auswahlsprachen S_i mit einem Index i mit $1 \leq i \leq n$ potenzseparierend sind.

Da alle Auswahlsprachen potenzseparierend sind, gibt es zu jeder Sprache ein $m \in \mathbb{N}$, sodass für jedes $x \in V^*$ entweder $J_x^m \cap S_i = \emptyset$ gilt oder $J_x^m \subseteq S_i$ mit $J_x^m = \{x^n \mid n \geq m\}$.

Wir bezeichnen unsere Auswahlsprachen S_i als $S_i^{(m)}$, wobei m jeweils die kleinste natürliche Zahl im Sinne der Definition der potenzseparierenden Sprachen ist.

Des Weiteren definieren wir

$$p = \max\{m \mid (S^{(m)}, C) \in \mathcal{S}\}.$$

Damit gilt für jede Auswahlsprache S_i die folgende Aussage: Für alle Wörter $x \in V^*$ gilt: Für $J_x^m = \{x^n \mid n \geq m\}$ gilt

$$\text{entweder } J_x^m \cap S_i = \emptyset \text{ oder } J_x^m \subseteq S_i \tag{5.2}$$

Wir definieren außerdem

$$t = \max\{|z| \mid z \in A\} + 1.$$

Da die Sprache L_1 Wörter mit beliebiger Anzahl an Buchstaben enthält, gibt es eine Ableitung

$$w_0 \Rightarrow^* w_1 \Rightarrow uw_1v$$

mit $w_0 \in A$, $|w_1|_a > p + t$, $|w_1|_b = 0$ und $|uv|_b > 0$. Daraus folgt $w_1 = a^k$ mit $k > p + t$.

Die verwendete Auswahlsprache sei S_i. Da w_1 in S_i liegt, erhalten wir mit der Beziehung (5.2)

$$w_1 = a^k \in S_i \Leftrightarrow a^{k+1} \in S_i$$

Da a^{k+1} in L_2 und damit in L liegt, wird auch das Wort $ua^{k+1}v$ abgeleitet. Da $|uv|_b > 0$ gilt, liegt das Wort $ua^{k+1}v$ höchstens in L_1. Da aber ua^kv in L_1 liegt, ist $|ua^kv|_a$ eine gerade Zahl und $|ua^{k+1}v|_a$ eine ungerade Zahl. Folglich liegt $ua^{k+1}v$ nicht in L_1 und auch nicht in L, was ein Widerspruch zu $L = L(G)$ ist. □

Lemma 5.11
Die Sprache

$$L = (a^*b^*)^*(a^+b^+) \cup (ca^+b^+c)$$

liegt in

$$\mathcal{EC}(INF) \setminus \mathcal{EC}(STAR).$$

Beweis. i) Die kontextuale Grammatik

$$G = (\{a, b\}, \mathcal{S}, \{ab\}),$$

wobei $\mathcal{S}$ aus den Regeln

$$\begin{aligned} Inf(a^+b^+) &\rightarrow (c, c) \\ Inf((a \cup b)^*) &\rightarrow (a, \lambda) \mid (b, \lambda) \mid (\lambda, b) \end{aligned}$$

besteht, erzeugt die Sprache L.

ii) Angenommen die Sprache L würde von einer Grammatik $G = (\{a, b\}, \mathcal{S}, A)$ erzeugt werden, wobei $\mathcal{S}$ nur Auswahlsprachen aus $STAR$ enthält.

Wir definieren

$$t = \max\{|z| \mid z \in A\} + 1.$$

Da die Sprache L Wörter mit beliebiger Anzahl an Buchstaben und genau zwei c enthält, gibt es eine Ableitung

$$w_0 \Rightarrow^* w_1 \Rightarrow uw_1v$$

mit $w_0 \in A$, $|w_1| > t$, $|w_1|_c = 0$ und $|uv|_c = 2$. Daraus folgt $w_1 = a^k b^l$ für zwei Zahlen $k > 0$, $l > 0$ und $k + l > t$.

Die verwendete Auswahlsprache für dieses Ableitungsschritt sei S_i und es gilt $w_1 \in S_i$. Da S_i in $STAR$ liegt, gilt $S_i = H^*$ für eine Menge H. Da $(H^*)^* = H^*$, gilt auch $S_i^* = S_i$ und da w_1 in S_i liegt, liegt auch $w_1^n = (a^k b^l)^n$ in S_1 für jede natürliche Zahl n. Also können wir $w_1^2 = a^k b^l a^k b^l$ um den Kontext $C_i = (u, v)$ erweitern, um so

$$L \ni w_1^2 \underset{ex}{\Longrightarrow} uw_1^2v \notin L$$

mit $|uw_1^2v|_c > 0$ abzuleiten.

Aus diesem Widerspruch folgt die Behauptung. □

5.2 Inklusionen

Wir werden nun die Teilmengenbeziehungen zwischen den Sprachfamilien untersuchen. Um zu zeigen, dass die Teilmengenbeziehungen strikt sind, werden wir die Sprachen aus dem Abschnitt 5.1 verwenden.

Lemma 5.12
Es gilt $\mathcal{EC}(REG) = \mathcal{EC}(2COM) = \mathcal{EC}(COM)$.

Beweis. i) Da die Sprachfamilien $2COM$ und COM jeweils Teilmenge von REG sind, gilt nach Lemma 5.1 auch, dass $\mathcal{EC}(2COM) \subseteq \mathcal{EC}(REG)$ und $\mathcal{EC}(COM) \subseteq \mathcal{EC}(REG)$.

ii) Es sei $G = (V, \mathcal{S}, A)$ eine Grammatik, deren Auswahlsprachen beliebig regulär sind und X ein neues Symbol ($X \notin V$).

Die Grammatik

$$G' = (V \cup \{X\}, \{(S'_1, C_1), \ldots, (S'_n, C_n)\}, A)$$

mit $S'_i = \{X\}^* S_i$ für Zahlen $1 \leq i \leq n$ erzeugt die gleiche Sprache wie G. Die Auswahlsprachen von G' liegen alle in der Familie *COM*. Der Buchstabe X kommt weder in den Axiomen noch den Kontexten vor. Also hat der Bestandteil $\{X\}^*$ der Auswahlsprachen keinen Einfluss auf die möglichen Ableitungen. Daher gilt $\mathcal{EC}(REG) = \mathcal{EC}(COM)$

iii) Da die Familie *2COM* eine Obermenge von der Familie *COM* ist, ist auch die Familie $\mathcal{EC}(2COM)$ eine Obermenge von $\mathcal{EC}(COM)$. Da aber $\mathcal{EC}(REG)$ eine Teilmenge von $\mathcal{EC}(COM)$ ist, gilt auch $\mathcal{EC}(REG) = \mathcal{EC}(2COM)$.

Damit folgt die Behauptung. □

Lemma 5.13
Die Inklusionen

$$\mathcal{EC}(INF) \subset \mathcal{EC}(SUF)$$

und

$$\mathcal{EC}(INF) \subset \mathcal{EC}(PRE)$$

sind strikt.

Beweis. i) Nach Lemma 5.1 folgen aus den beiden Inklusionen $INF \subseteq SUF$ und $INF \subseteq PRE$ auch die Inklusionen $\mathcal{EC}(INF) \subseteq \mathcal{EC}(SUF)$ und $\mathcal{EC}(INF) \subseteq \mathcal{EC}(PRE)$.

ii) Nach Lemma 5.4 liegt die Sprache

$$L_1 = \{a^n b^n \mid n \geq 1\} \cup \{a^n \mid n \geq 1\}$$

in der Familie $\mathcal{EC}(SUF)$, aber nicht in $\mathcal{EC}(PRE)$ und damit auch nicht in $\mathcal{EC}(INF)$.

iii) Nach Lemma 5.3 liegt die Sprache

$$L_2 = \{a^n b^n \mid n \geq 1\} \cup \{b^n \mid n \geq 1\}$$

in der Familie $\mathcal{EC}(PRE)$, aber nicht in $\mathcal{EC}(SUF)$ und damit auch nicht in $\mathcal{EC}(INF)$.

Damit folgt die Behauptung. □

Lemma 5.14
Es sei
$$\mathcal{F} = \{INF, PRE, STAR\}.$$

Es gilt die strikte Inklusion

$$\mathcal{EC}(MON) \subset \mathcal{EC}(F)$$

für jede Sprachfamilie F aus $\mathcal{F}$.

Beweis. i) Da INF und $STAR$ Obermengen von MON sind, folgt mit Lemma 5.1, dass $\mathcal{EC}(INF)$ und $\mathcal{EC}(STAR)$ Obermengen von $\mathcal{EC}(MON)$ sind.

Da PRE eine Obermenge von INF ist, folgt auch, dass $\mathcal{EC}(PRE)$ eine Obermenge von $\mathcal{EC}(INF)$ und damit von $\mathcal{EC}(MON)$ ist.

ii) Es seien

$$\begin{aligned} L_1 &= \{a^m b c^{2n} b a^m \mid n \geq 1, m \geq 0\}, \\ L_2 &= \{c^n \mid n \geq 2\}, \\ L_3 &= \{b c^n b \mid n \geq 2\}, \\ L_4 &= \{a c^n a \mid n \geq 2\}. \end{aligned}$$

Nach Lemma 5.9 liegt die Sprache

$$L = L_1 \cup L_2 \cup L_3 \cup L_4$$

in $\mathcal{EC}(INF)$ und damit auch in $\mathcal{EC}(PRE)$, aber nicht in $\mathcal{EC}(NC)$ und damit auch nicht in $\mathcal{EC}(MON)$.

iii) Es seien

$$L_5 = \{ba^{2n}b \mid n \geq 1\},$$
$$L_6 = \{a^n \mid n \geq 2\}.$$

Nach Lemma 5.10 liegt die Sprache

$$L' = L_5 \cup L_6$$

in der Familie $\mathcal{EC}(STAR)$, aber nicht in $\mathcal{EC}(PS)$ und damit nicht in $\mathcal{EC}(MON)$.

Damit gilt die echte Inklusion. □

Lemma 5.15
Es sei

$$\mathcal{F} = \{INF, PRE, STAR\}.$$

Es gilt die strikte Inklusion

$$\mathcal{EC}(FIN) \subset \mathcal{EC}(F)$$

für jede Sprachfamilie F aus $\mathcal{F}$.

Beweis. Nach Lemma 5.14 ist $\mathcal{EC}(MON)$ eine echte Teilmenge einer jeden Sprachfamilie $\mathcal{EC}(F)$ mit F aus $\mathcal{F}$. Da $\mathcal{EC}(FIN)$ eine Teilmenge von $\mathcal{EC}(MON)$ ist, folgt die Behauptung. □

Lemma 5.16
Die Inklusion $\mathcal{EC}(STAR) \subset \mathcal{EC}(COM)$ ist strikt.

Beweis. i) Wie im Abschnitt 4.1 erwähnt, gilt die Inklusion $STAR \setminus \{\{\lambda\}\} \subset COM$. Der einzige Fall, bei dem ein Auswahlpaar $(\{\lambda\}, (u, v)) \in \mathcal{S}$ einer kontextualen Grammatik $G = (V, \mathcal{S}, A)$ eine Rolle spielt ist, wenn λ ein Element der Axiomenmenge A ist. In diesem Fall können wir eine äquivalente Grammatik G' so definieren, dass wir für alle Paare $(\{\lambda\}, C) \in \mathcal{S}$ alle Wörter uv mit $(u, v) \in C$ der Axiomenmenge hinzufügen, auf diese Paare verzichten und alles andere von G übernehmen. Damit erzeugt die Grammatik G' die gleiche Sprache wie G und alle ihre Auswahlsprachen liegen in COM.

ii) Nach Lemma 5.7 liegt die Sprache

$$L = \{b^n a \mid n \geq 0\} \cup \{\lambda\}$$

in $\mathcal{EC}(COMB)$, aber nicht in $\mathcal{EC}(STAR)$. Da $\mathcal{EC}(COMB)$ eine Teilmenge von $\mathcal{EC}(COM)$ ist, liegt die Sprache L auch in $\mathcal{EC}(COM)$, aber nicht in $\mathcal{EC}(STAR)$.

Damit gilt die echte Inklusion. □

Lemma 5.17
Die Inklusionen

$$\mathcal{EC}(PRE) \subset \mathcal{EC}(PS)$$

und

$$\mathcal{EC}(INF) \subset \mathcal{EC}(PS)$$

ist strikt.

Beweis. i) Nach Lemma 5.1 folgen aus der Inklusion $PRE \subseteq PS$ die Inklusion $\mathcal{EC}(PRE) \subseteq \mathcal{EC}(PS)$

ii) Nach Lemma 5.5 liegt die Sprache

$$L = \{bbba^n \mid n \geq 1\} \cup \{bb\}$$

in $\mathcal{EC}(NIL)$ und damit auch in $\mathcal{EC}(PS)$, aber nicht in $\mathcal{EC}(PRE)$.

iii) Da die Sprachfamilie $\mathcal{EC}(PRE)$ nach i) und ii) eine echte Teilmenge der Sprachfamilie $\mathcal{EC}(PS)$ ist, folgt mit Lemma 5.13, dass auch $\mathcal{EC}(INF)$ eine echte Teilmenge von $\mathcal{EC}(PS)$ ist.

Damit folgt die Behauptung. □

5.3 Unvergleichbarkeiten

Nachdem wir die Teilmengenbeziehungen untersucht haben, werden wir für die übrigen Sprachfamilien zeigen, dass diese jeweils paarweise unvergleichbar sind.

Lemma 5.18
Die Sprachfamilien $\mathcal{EC}(SUF)$ und $\mathcal{EC}(PRE)$ sind unvergleichbar.

Beweis. i) Nach Lemma 5.3 liegt die Sprache

$$L_1 = \{a^n b^n \mid n \geq 1\} \cup \{b^n \mid n \geq 1\}$$

in $\mathcal{EC}(PRE)$, aber nicht in $\mathcal{EC}(SUF)$.

ii) Nach Lemma 5.4 liegt die Sprache

$$L_2 = \{a^n b^n \mid n \geq 1\} \cup \{a^n \mid n \geq 1\}$$

in $\mathcal{EC}(SUF)$, aber nicht in $\mathcal{EC}(PRE)$.

Damit gilt die Unvergleichbarkeit. □

Lemma 5.19
Es sei

$$\mathcal{F} = \{COMB, DEF, ORD, NC\}.$$

Die Sprachfamilien $\mathcal{EC}(INF)$ und $\mathcal{EC}(PRE)$ sind unvergleichbar zu allen Sprachfamilien $\mathcal{EC}(F)$ mit $F \in \mathcal{F}$.

Beweis. i) Nach Lemma 5.7 liegt die Sprache

$$L_1 = L = \{b^n a \mid n \geq 0\} \cup \{\lambda\}$$

in $\mathcal{EC}(COMB)$, aber nicht in $\mathcal{EC}(PRE)$, und damit auch in $\mathcal{EC}(F) \setminus \mathcal{EC}(PRE)$ und $\mathcal{EC}(F) \setminus \mathcal{EC}(INF)$ für alle $F \in \mathcal{F}$.

ii) Es seien

$$L_1 = \{a^m bc^{2n} ba^m \mid n \geq 1, m \geq 0\},$$
$$L_2 = \{c^n \mid n \geq 2\},$$
$$L_3 = \{bc^n b \mid n \geq 2\},$$
$$L_4 = \{ac^n a \mid n \geq 2\}.$$

Nach Lemma 5.9 liegt die Sprache

$$L = L_1 \cup L_2 \cup L_3 \cup L_4$$

in $\mathcal{EC}(INF)$, aber nicht in $\mathcal{EC}(NC)$, und damit auch in $\mathcal{EC}(INF) \setminus \mathcal{EC}(F)$ und $\mathcal{EC}(PRE) \setminus \mathcal{EC}(F)$ für alle $F \in \mathcal{F}$.

Damit folgt die Unvergleichbarkeit. □

Lemma 5.20
Es sei
$$\mathcal{F} = \{NIL, COMM, CIRC\}.$$

Die Sprachfamilien $\mathcal{EC}(INF)$ und $\mathcal{EC}(PRE)$ sind unvergleichbar zu allen Sprachfamilien $\mathcal{EC}(F)$ mit $F \in \mathcal{F}$.

Beweis. i) Nach Lemma 5.8 liegt die Sprache

$$L_1 = \{a^n b^n \mid n \geq 1\} \cup \{b^n a^n \mid n \geq 1\}$$

in $\mathcal{EC}(INF)$, aber nicht in $\mathcal{EC}(CIRC)$, und damit auch in $\mathcal{EC}(INF) \setminus \mathcal{EC}(F)$ und $\mathcal{EC}(PRE) \setminus \mathcal{EC}(F)$ für alle $F \in \mathcal{F}$.

ii) Nach Lemma 5.5 liegt die Sprache

$$L_2 = \{bbba^n \mid n \geq 1\} \cup \{bb\}$$

in $\mathcal{EC}(NIL)$, aber nicht in $\mathcal{EC}(PRE)$, und damit auch in $\mathcal{EC}(F) \setminus \mathcal{EC}(PRE)$ und $\mathcal{EC}(F) \setminus \mathcal{EC}(INF)$ für alle $F \in \mathcal{F}$.

Damit folgt die Unvergleichbarkeit. □

Lemma 5.21
Es sei
$$\mathcal{F} = \{NIL, COMM, CIRC\}.$$

Die Sprachfamilie $\mathcal{EC}(STAR)$ ist unvergleichbar zu allen Sprachfamilien $\mathcal{EC}(F)$ mit $F \in \mathcal{F}$.

Beweis. i) Nach Lemma 5.8 liegt die Sprache

$$L_1 = \{a^n b^n \mid n \geq 1\} \cup \{b^n a^n \mid n \geq 1\}$$

in $\mathcal{EC}(STAR)$, aber nicht in $\mathcal{EC}(CIRC)$ und damit auch nicht in den Familien $\mathcal{EC}(NIL)$ und $\mathcal{EC}(COMM)$.

ii) Nach Lemma 5.6 liegt die Sprache

$$L = \{bbba^n \mid n \geq 1\} \cup \{\lambda\}$$

in $\mathcal{EC}(NIL)$ und damit auch in $\mathcal{EC}(COMM)$ und $\mathcal{EC}(CIRC)$, aber nicht in $\mathcal{EC}(STAR)$.

Damit folgt die Unvergleichbarkeit. □

Lemma 5.22
Es sei
$$\mathcal{F} = \{COMB, DEF, ORD, NC, PS\}.$$

Die Sprachfamilie $\mathcal{EC}(STAR)$ ist unvergleichbar zu allen Sprachfamilien $\mathcal{EC}(F)$ mit $F \in \mathcal{F}$.

Beweis. i) Nach Lemma 5.7 liegt die Sprache

$$L_1 = \{b^n a \mid n \geq 0\} \cup \{\lambda\}$$

in $\mathcal{EC}(COMB)$ und damit auch in $\mathcal{EC}(F)$ für alle $F \in \mathcal{F}$, aber nicht in $\mathcal{EC}(STAR)$.

ii) Nach Lemma 5.10 liegt die Sprache

$$L = \{ba^{2n}b \mid n \geq 1\} \cup \{a^n \mid n \geq 2\}$$

in $\mathcal{EC}(STAR)$, aber nicht in $\mathcal{EC}(PS)$ und damit auch nicht in $\mathcal{EC}(F)$ für alle $F \in \mathcal{F}$.

Damit gilt die echte Inklusion. □

Lemma 5.23
Es sei

$$\mathcal{F} = \{PRE, SUF, INF\}.$$

Die Sprachfamilie $\mathcal{EC}(STAR)$ ist unvergleichbar zu allen Sprachfamilien $\mathcal{EC}(F)$ mit $F \in \mathcal{F}$.

Beweis. i) Nach Lemma 5.10 liegt die Sprache

$$L = \{ba^{2n}b \mid n \geq 1\} \cup \{a^n \mid n \geq 2\}$$

in $\mathcal{EC}(STAR)$, aber nicht in $\mathcal{EC}(PS)$ und damit auch nicht in $\mathcal{EC}(F)$ für alle $F \in \mathcal{F}$.

ii) Nach Lemma 5.11 liegt die Sprache

$$L = (a^*b^*)^*(a^+b^+) \cup (ca^+b^+c)$$

in $\mathcal{EC}(INF)$ und damit auch in allen Sprachfamilien $\mathcal{EC}(F)$ für alle $F \in \mathcal{F}$, aber nicht in $\mathcal{EC}(STAR)$. □

5.4 Ergebnisse

Wir wollen nun unsere Ergebnisse in Tabelle 5.1 festhalten. Die Zeilen und Spalten stehen jeweils für eine Sprachfamilie. Bei den Sprachfamilien handelt es sich jeweils um $\mathcal{EC}(F)$, wobei aus Platzgründen jeweils nur F angegeben ist.

Ein grüner Verweis zeigt zu einem Lemma, in dem bewiesen wurde, dass die Sprachfamilie in dieser Zeile eine Teilmenge der Sprachfamilie dieser Spalte ist, ein blauer Verweis zeigt, dass es sich um eine Obermenge handelt. Ein roter Verweis zeigt an, dass dort die Unvergleichbarkeit der beiden Sprachfamilien bewiesen wurde. Ein **schwarzer** Verweis zeigt an, dass dort die Gleichheit bewiesen wurde.

Daraus ergibt sich folgendes Resultat:

Theorem 5.24 – Hierarchie der $\mathcal{EC}$-Sprachfamilien
Es gelten die Beziehungen in Abbildung 5.2.

Beweis. In Tabelle 5.1 sind alle Verweise zu den Beweisen von Teilmengenbeziehungen oder Unvergleichbarkeiten zu finden. Deswegen wird auf eine Kennzeichnung in der Abbildung verzichtet.

Ein Pfeil von einer Sprachfamilie X zu einer Sprachfamilie Y sagt, dass Sprachfamilie X eine echte Teilmenge der Sprachfamilie Y ist. Sind zwei Sprachfamilien nicht verbunden, so geht die Unvergleichbarkeit aus einem Lemma des Abschnitts 5.3 hervor. $\square$

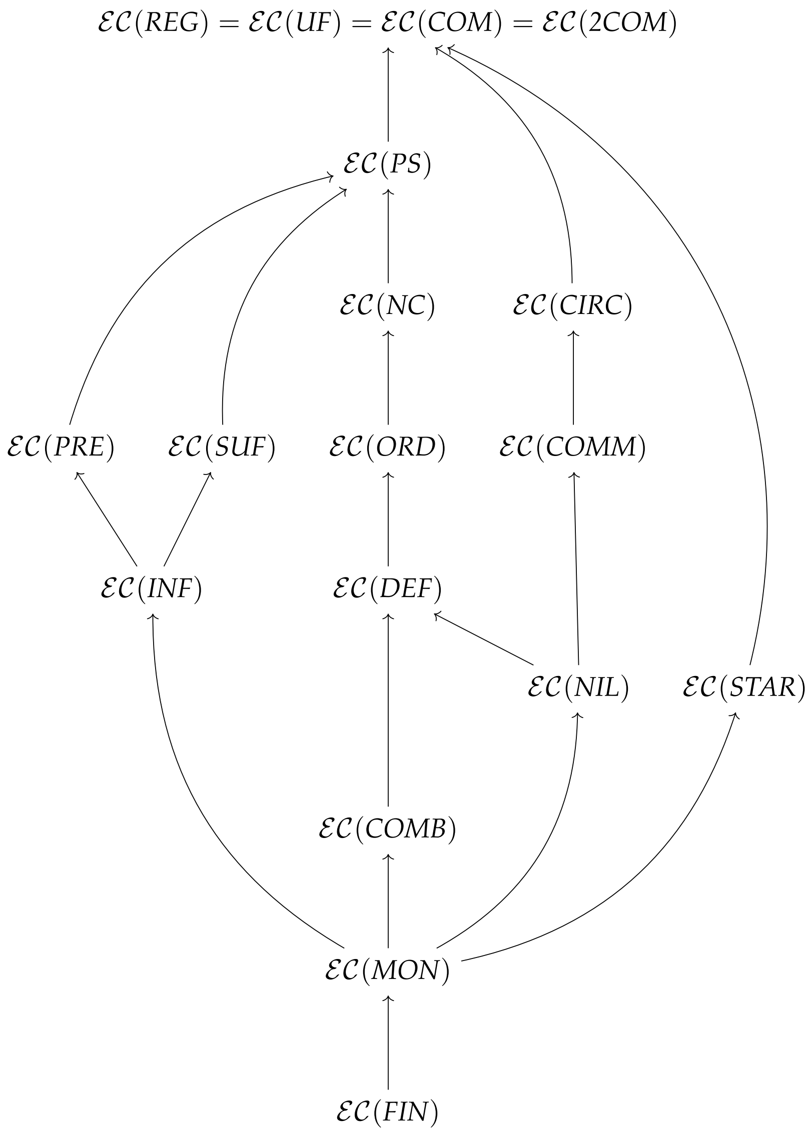

Abbildung 5.2: Ergebnisse

	MON	COMB	FIN	NIL	DEF	ORD	NC	PS	INF	SUF	PRE	COM	2COM	COMM	CIRC	STAR	UF
MON	=	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.14	5.2	5.14	5.12	5.12	5.2	5.2	5.14	5.2
COMB	5.2	=	5.2	5.2	5.2	5.2	5.2	5.2	5.19	5.2	5.19	5.12	5.12	5.2	5.2	5.22	5.2
FIN	5.2	5.2	=	5.2	5.2	5.2	5.2	5.2	5.15	5.2	5.15	5.12	5.12	5.2	5.2	5.15	5.2
NIL	5.2	5.2	5.2	=	5.2	5.2	5.2	5.2	5.20	5.2	5.20	5.12	5.12	5.2	5.2	5.21	5.2
DEF	5.2	5.2	5.2	5.2	=	5.2	5.2	5.2	5.19	5.2	5.19	5.12	5.12	5.2	5.2	5.22	5.2
ORD	5.2	5.2	5.2	5.2	5.2	=	5.2	5.2	5.19	5.2	5.19	5.12	5.12	5.2	5.2	5.22	5.2
NC	5.2	5.2	5.2	5.2	5.2	5.2	=	5.2	5.19	5.2	5.19	5.12	5.12	5.2	5.2	5.22	5.2
PS	5.2	5.2	5.2	5.2	5.2	5.2	5.2	=	5.17	5.2	5.17	5.12	5.12	5.2	5.2	5.22	5.2
INF	5.14	5.19	5.15	5.20	5.19	5.19	5.19	5.17	=	5.13	5.13	5.12	5.12	5.20	5.20	5.23	5.2
SUF	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.13	=	5.18	5.12	5.12	5.2	5.2	5.23	5.2
PRE	5.14	5.19	5.15	5.20	5.19	5.19	5.19	5.17	5.13	5.18	=	5.12	5.12	5.20	5.20	5.23	5.2
COM	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	=	5.12	5.12	5.12	5.12	5.12
2COM	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12	=	5.12	5.12	5.12	5.12
COMM	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.20	5.2	5.20	5.12	5.12	=	5.2	5.21	5.2
CIRC	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.20	5.2	5.20	5.12	5.12	5.2	=	5.21	5.2
STAR	5.14	5.22	5.15	5.21	5.22	5.22	5.22	5.22	5.23	5.23	5.23	5.12	5.12	5.21	5.21	=	5.2
UF	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.2	5.12	5.12	5.2	5.2	5.2	=

Tabelle 5.1: Ergebnisse – Ein grüner Verweis steht für eine Teilmengenbeziehung, ein blauer Verweis für eine Obermengenbeziehung, ein roter Verweis für eine Unvergleichbarkeit und ein schwarzer Verweis für Gleichheit.

Kapitel 6

Schluss und Ausblick

Im Rahmen dieser Arbeit wurde die derzeitige Hierarchie an subregulären Sprachen, sowie von externen kontextualen Grammatiken mit subregulären Steuersprachen erweitert.

Gegenstand zukünftiger Forschung wird sein, die Beziehungen von internen kontextualen Grammatiken oder baumgesteuerten Grammatiken mit den hier verwendeten subregulären Steuersprachen zu untersuchen.

Es wurden auch in anderen Arbeiten verschiedene subreguläre Sprachfamilien untersucht [1,6,12]. Ein weiterer Forschungsgegenstand folgender Arbeiten wird sein, die aktuelle Hierarchie von subregulären Sprachfamilien um weitere Familien zu erweitern und Sprachen daraus als Steuersprachen für externe kontextuale Grammatiken zu verwenden. Explizit seien hier die präfix-, suffix- und infixfreien Sprachen genannt, deren Untersuchung bereits begonnen hat. Weiterhin ist auch geplant, die Hierarchie der subregulären Sprachfamilien, erweitert um die Familien der präfix-, suffix- und infixfreien Sprachen, mit den Hierarchien aus den Sprachfamilien, die durch ihre beschränkten Ressourcen definiert sind, zu vereinigen.

Literaturverzeichnis

[1] Henning Bordihn, Markus Holzer, and Martin Kutrib. Determination of finite automata accepting subregular languages. *Theoretical Computer Science*, 410(35):3209–3222, 2009.

[2] Jürgen Dassow. Contextual grammars with subregular choice. *Fundamenta Informaticae*, 64(1–4):109–118, 2005.

[3] Jürgen Dassow. Contextual languages with strictly locally testable and star free selection languages. *Analele Universitatii Bucuresti*, (62):25–36, 2015.

[4] Jürgen Dassow, Florin Manea, and Bianca Truthe. On external contextual grammars with subregular selection languages. *Theoretical Computer Science*, 449:64–73, 2012.

[5] Jürgen Dassow and Bianca Truthe. Relations of contextual grammars with strictly locally testable selection languages. *RAIRO Theoretical Informatics and Applications*, 57:#10, 2023.

[6] Yo-Sub Han and Kai Salomaa. State complexity of basic operations on suffix-free regular languages. *Theoretical Computer Science*, 410(27):2537–2548, 2009.

[7] Ivan M. Havel. The theory of regular events. II. *Kybernetika*, 5(6):520–544, 1969.

[8] Markus Holzer and Bianca Truthe. On relations between some subregular language families. In Rudolf Freund, Markus Holzer, Nelma Moreira, and Rogério Reis, editors, *Seventh Workshop on Non-Classical Models of Automata and Applications – NCMA 2015, Porto, Portugal, August 31 – September 1, 2015. Proceedings*, volume 318 of *books@ocg.at*, pages 109–124. Österreichische Computer Gesellschaft, 2015.

[9] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, 1979.

[10] Solomon Marcus. Contextual grammars. *Revue Roumaine de Mathématique Pures et Appliquées*, 14:1525–1534, 1969.

[11] Benedek Nagy. Union-freeness, deterministic union-freeness and union-complexity. In Michal Hospodár, Galina Jirásková, and Stavros Konstantinidis, editors, *Descriptional Complexity of Formal Systems, 21st IFIP WG 1.02 International Conference, DCFS 2019, Košice, Slovakia, July 17–19, 2019, Proceedings*, pages 46–56. Springer, Cham, 2019.

[12] Viktor Olejár and Alexander Szabari. Closure properties of subregular languages under operations. In Jérôme Durand-Lose and György Vaszil, editors, *Machines, Computations, and Universality*, pages 126–142, Cham, 2022. Springer International Publishing.

[13] H. J. Shyr and G. Thierrin. Ordered automata and associated languages. *Tamkang Journal of Mathematics*, 5(1):9–20, 1974.

[14] H. J. Shyr and G. Thierrin. Power-separating regular languages. *Mathematical Systems Theory*, 8(1):90–95, 1974.

[15] Bianca Truthe. A relation between definite and ordered finite automata. In Suna Bensch, Rudolf Freund, and Friedrich Otto, editors, *Sixth Workshop on Non-Classical Models for Automata and Applications – NCMA 2014, Kassel, Germany, July 28–29, 2014. Proceedings*, volume 304 of *books@ocg.at*, pages 235–247. Österreichische Computer Gesellschaft, 2014.

[16] Bianca Truthe. Hierarchy of subregular language families. Technical report, Justus-Liebig-Universität Giessen, Institut für Informatik, IFIG Research Report 1801, 2018.

[17] Bianca Truthe. Generative capacity of contextual grammars with subregular selection languages. *Fundamenta Informaticae*, 180(1–2):123–150, 2021.

[18] B. Wiedemann. *Vergleich der Leistungsfähigkeit endlicher determinierter Automaten*. Diplomarbeit, Universität Rostock, 1978.

Eigenständigkeitserklärung

Ich versichere hiermit, dass ich die Arbeit selbstständig verfasst, keine anderen, als die angegebenen Hilfsmittel verwandt und die Stellen, die anderen benutzten Druck- und digitalisierten Werken im Wortlaut oder dem Sinn nach entnommen sind, mit Quellenangaben kenntlich gemacht habe.

Marvin Ködding