OULU 2009
A 536

ACTA UNIVERSITATIS OULUENSIS

Tuukka Salmi

VERY SMALL FAMILIES GENERATED BY BOUNDED AND UNBOUNDED CONTEXT-FREE LANGUAGES

*FACULTY OF SCIENCE,
DEPARTMENT OF MATHEMATICAL SCIENCES,
UNIVERSITY OF OULU;
INFOTECH OULU,
UNIVERSITY OF OULU*

ACTA UNIVERSITATIS OULUENSIS
A Scientiae Rerum Naturalium 536

TUUKKA SALMI

VERY SMALL FAMILIES GENERATED BY BOUNDED AND UNBOUNDED CONTEXT-FREE LANGUAGES

Academic dissertation to be presented with the assent of the Faculty of Science of the University of Oulu for public defence in Raahensali (Auditorium L10), Linnanmaa, on 14 November 2009, at 12 noon

OULUN YLIOPISTO, OULU 2009

Acta Univ. Oul. A 536, 2009

Supervised by
Docent Juha Kortelainen
Professor Ismo Hakala

Reviewed by
Professor Jean-Michel Autebert
Professor Michel Latteux

ISBN 978-951-42-9273-6 (Paperback)
ISBN 978-951-42-9274-3 (PDF)
http://herkules.oulu.fi/isbn9789514292743/
ISSN 0355-3191 (Printed)
ISSN 1796-220X (Online)
http://herkules.oulu.fi/issn03553191/

Cover design
Raimo Ahonen

OULU UNIVERSITY PRESS
OULU 2009

Salmi, Tuukka, Very small families generated by bounded and unbounded context-free languages.
Faculty of Science, Department of Mathematical Sciences, University of Oulu, P.O.Box 3000, FI-90014 University of Oulu, Finland; Infotech Oulu, University of Oulu, P.O.Box 4500, FI-90014 University of Oulu, Finland
Acta Univ. Oul. A 536, 2009
Oulu, Finland

Abstract

In this thesis, we will study very small full trios and full AFLs inside the family of context-free languages. Especially, we are interested in the existence of the smallest nontrivial full trios and full AFLs. This is an old research subject, and it has not been studied much since the 1970s. A conjecture by Autebert *et al.* states that there does not exist a nontrivial minimal full trio inside the family of context-free languages (2) (see also (1)). First, we will show that there does not exist a nontrivial minimal full trio or a nontrivial minimal full AFL with respect to the bounded context-free languages. This result solves another old conjecture stated by Autebert *et al.* (1). Then we will try to generalize our result to also concern unbounded context-free languages. We will make some progress, but the problem still remains open.

Keywords: bounded languages, context-free languages, full AFL, full trio, minimality

Acknowledgements

I wish to express my deepest gratitude to my teacher, Docent Juha Kortelainen, for encouragement and advice. Looking back at my first writings, I really appreciate his patience and effort with my work. With his help, this work has also become financially possible.

I am also grateful to Professors Michel Latteux and Jean Michel Autebert for reading the manuscript and giving excellent comments.

I also wish to express my appreciation to my employer EB and all the relevant people there for their flexibility toward my studies. I also want to thank Infotech Oulu Graduate School for financial support.

Finally, my warmest thanks to my family, especially to my wife Ulla for her support and, most of all, patience. In addition, I sincerely thank her for putting some limits to my working.

Contents

1 Introduction

Noam Chomsky developed the family of context-free languages with an aim of creating a mathematical model for natural languages (Finnish, English, etc.) (8). It has been proved that there are expressions in natural languages that do not belong to any context-free language (19). Context-free languages have anyway risen to a significant role in theoretical computer science.

In the 1960s, Ginsburg and Greibach characterized language families according to their closure properties over some operations on words and languages. After this work, the authors defined the concept of an Abstract Family of Languages (AFL) (11). A full AFL is a family of languages that is closed under six operations: morphism, inverse morphism, intersection with regular languages, union, catenation and catenation closure. Since then, other variations of AFL have also been widely used. A full trio is a family of languages that is closed under the first three operations: morphism, inverse morphism and intersection with regular languages.

In this thesis, we study very small families of context-free languages. The smallness of language families means in our context that a family $\mathcal{L}_1$ is smaller than a family $\mathcal{L}_2$ if and only if $\mathcal{L}_1 \subsetneq \mathcal{L}_2$. The main question is whether a smallest nontrivial full trio or a full AFL exists in the family of context-free languages. Since the family of regular languages, $\mathcal{L}_{REG}$, is a full AFL and all the full trios and full AFLs contain $\mathcal{L}_{REG}$, we exclude $\mathcal{L}_{REG}$ from our study as a trivial case. Thus, we define a minimal full trio (resp. minimal full AFL) to be a smallest full trio (resp. smallest full AFL) that contains nonregular languages. It should be remarked that this definition makes possible that there could exist a single, multiple or no minimal full trios or full AFLs. However, Autebert et al. have stated the following conjecture on the issue.

Conjecture 1.1. *(2) (see also (1)) There does not exist a minimal full trio in the family of context-free languages.*

Moreover, Autebert and Boasson have conjectured that for each nonregular context-free language L_1, L_2 such that $\mathcal{T}(L_1) \subsetneq \mathcal{T}(L_2)$,[1] there exists a nonregular context-free language L_3 such that $\mathcal{T}(L_1) \subsetneq \mathcal{T}(L_3) \subsetneq \mathcal{T}(L_2)$ (1, Conjecture 7)

[1]Here $\mathcal{T}(L)$ denotes the full trio generated by a language L.

(see also (3)). It is easy to see that, if a minimal full trio or a minimal full AFL existed, it had to be principal, i.e. it had to be generated by a single language. Thus, the latter conjecture is stronger than the former one. Autebert et al. have also another analogue conjecture on the issue.

Conjecture 1.2. *(1, Conjecture 10) There does not exist a minimal full trio with respect to the family of bounded context-free languages.*

So far there exist only a few actual results concerning minimal full trios and minimal full AFLs. Berstel and Boasson have proved that the full trios and the full AFLs generated by the languages $S_<$, $S_>$ and $S_{\neq}$ are minimal with respect to 2-bounded context-free languages. The authors have also proved that none of these families are, however, minimal with respect to k-bounded context-free languages when $k \geq 3$ (6). Latteux has shown that the full trio generated by the commutative closure of $S_{\neq}$ is minimal with respect to all commutative languages (17).

While studying small families of context-free languages, it is natural to also study the converse problem; do greatest nontrivial full trios or greatest nontrivial full AFLs exist inside the family of context-free languages? We know that the whole family of context-free language, $\mathcal{L}_{CF}$, is a full AFL. Thus, we define a maximal full trio (resp. maximal full AFL) as a greatest full trio (resp. greatest full AFL) that is a proper subset of $\mathcal{L}_{CF}$. Greibach has shown that a maximal full AFL exists and it is unique (13). In fact, we know that this maximal full AFL is the same as the maximal full trio (5, Proposition 3.4, Theorem 3.5). This maximal full trio is naturally the family containing those context-free languages that do not generate $\mathcal{L}_{CF}$ with the full trio operations.

It has been conjectured that this maximal full AFL is not principal (13).

In Chapter 2, we will give some preliminary definitions and results used in this thesis.

In the third chapter, we will study Conjecture 1.2. For this purpose, we will define a new chain of decreasing language families, $\mathcal{C}_k$, $k \in \mathbb{N}_+$. First, we will see that any bounded context-free language belongs to some family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$. Then we will show that we are able to transform any language from the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ into the family $\mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$ with the full trio operations. Finally, we will show that each family $\mathcal{C}_k$ is a (substitution closed) full AFL. As a result of this deduction, we get a proof to theorem stating that Conjecture 1.2 holds.

In Chapter 4, we will focus on Conjecture 1.1. We will try to solve this problem with a similar chain of decreasing language families. However, this problem turns out to be very difficult, and we will make only some progress.

Chapter 5 summarizes open problems and other interesting issues for future work.

2 Preliminaries

In this chapter, we introduce definitions, notations and some auxiliary results needed in the thesis. After each introduced concept, we will write the symbol usually denoting the concept. Symbols may contain sub- and superscripts.

2.1 Set Theory

The set of all natural numbers is denoted by $\mathbb{N}$. In addition, we set $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$. Let A and B be arbitrary sets. Denote $A \subset B$ if A is a subset of B, and $A \subsetneq B$ if A is a proper subset of B. The complement of the set $A \subset B$ with respect to the set B is the set $\overline{A} = \{a \in B | a \notin A\}$. Let $n \in \mathbb{N}_+$. A set $A \subset \mathbb{N}^n$ is *linear* if there exist $r \in \mathbb{N}$ and $c, p_1, p_2, \ldots, p_r \in \mathbb{N}^n$ such that

$$A = \{c + k_1 p_1 + k_2 p_2 + \cdots + k_r p_r \,|k_1, k_2, \ldots, k_r \in \mathbb{N}\} .$$

The vector c is called *constant*, and the vectors $p_1, p_2, \ldots, p_r$ are called *periods* of the linear set A. A set S is *semilinear* if it is a finite union of linear sets. We call these linear sets as the *linear components* of S. A linear set is *proper* if its periods are linearly independent over $\mathbb{Q}$, the field of rationals.

Theorem 2.1. *(15, Theorem 1) Every semilinear set is a finite union of disjoint proper linear sets.*

A linear set $A \subset \mathbb{N}^n$ is *stratified* if it has a representation using a period set P such that

- each element of P contains at most two nonzero coordinates;
- there do not exist positive integers i, j, k, l with $1 \le i < j < k < l \le n$ and periods $p = (p_1, p_2, \ldots, p_n)$, $q = (q_1, q_2, \ldots, q_n) \in P$ such that $p_i q_j p_k q_l \neq 0$.

Let A be a linear set and $c, p_1, p_2, \ldots, p_n$ its constant and periods, respectively. The *convex closure* of the set A is the set

$$Conv(A) = \{c + l_1 p_1 + l_2 p_2 + \cdots + l_n p_n | l_i \ge 0, l_i \in \mathbb{Q}\} \cap \mathbb{N}^n.$$

Let $A \subset \mathbb{N}^n$ be a linear set and $T \subset \mathbb{N}^n$ an arbitrary set. We say that A is *convex in T* if $A = Conv(A) \cap T$. Call A *convex* if A is convex in $\mathbb{N}^n$. We say that a semilinear set S is *semiconvex in T* if there exists $m \in \mathbb{N}_+$ and linear sets $S_1, S_2, \ldots, S_m$ such that

$$S = \bigcup_{i=1}^{m} S_i = \bigcup_{i=1}^{m} (Conv(S_i) \cap T).$$

2.2 Basic Concepts of Formal Languages

An *alphabet* is a nonempty finite set of abstract symbols. Denote alphabets by symbols Σ, Δ and Γ. Let $\Sigma_\infty = \{a_1, a_2, \ldots\}$ be an infinite set of abstract symbols, where $a_i \neq a_j$ for $i \neq j$. Every alphabet used in this thesis is assumed to be a subset of Σ_∞. We will also denote $\Sigma_n = \{a_1, a_2, \ldots, a_n\}$ for all $n \in \mathbb{N}_+$. The elements of an alphabet are called *letters*. Letters are denoted by a, b, c. A *word* over an alphabet Σ is a finite sequence of symbols of Σ. The set containing all the words of an alphabet Σ is denoted by Σ^*. Letters are written consecutively without any punctuations in words. So $w = b_1 b_2 \cdots b_n$, where $b_i \in \Sigma$ for all $i \in \{1, 2, \ldots, n\}$, is a word over the alphabet Σ. The *length* of a word $w \in \Sigma^*$, denoted by $|w|$, is the number of letters in w. So in the previous example we have $|w| = n$. The number of occurrences of a letter $a \in \Sigma$ in a word w is denoted by $|w|_a$. Thus, we have $|w| = \sum_{a \in \Sigma} |w|_a$. We denote the word that does not contain any letters by ϵ and call it the *empty word*. Latin letters u, v, w, x, y, z are used to denote words.

A *language* is any subset of Σ^*. Languages are denoted by L and K. Single regular languages are denoted by R. For each language L, we define Σ_L as the most compact alphabet such that $L \subset \Sigma_L^*$.

Let n, m be positive integers and $u = b_1 b_2 \cdots b_n$, $v = c_1 c_2 \cdots c_m$ words over Σ. We define a binary operation $(\cdot)$ on Σ^* by $u \cdot v = b_1 b_2 \cdots b_n c_1 c_2 \cdots c_m$. The operation $\cdot$ is called *catenation*. We use a shortened form $u \cdot v = uv$ of the catenation of words. We note that this shortened form is consistent with the definition of a word. The catenation of languages L_1 and L_2 is defined by $L_1 L_2 = \cup_{w_1 \in L_1} \cup_{w_2 \in L_2} \{w_1 w_2\}$. The *power* of a word w is defined by $w^0 = \epsilon$ and $w^k = w w^{k-1}$ for all $k \in \mathbb{N}_+$. Let $k \in \mathbb{N}_+$, $w_1 = v u^k$ and $w_2 = u^k v$. In this case, we may denote $w_1 u^{-k} = v$ and $u^{-k} w_2 = v$. Power can be extended to languages by $L^0 = \{\epsilon\}$ and $L^k = L^{k-1} L$ for all $k \in \mathbb{N}_+$. Define the set $L^* = \cup_{i=0}^{\infty} L^i$. Observe that this definition is consistent with the definition of the set containing all the words of the alphabet, Σ^*. The operation (*) is called *Kleene star* or *catenation closure*. We also define *Kleene plus* by $L^+ = LL^*$. A word $w \in \Sigma^*$ may be equated with the singleton $\{w\}$. Hence, we may express $w^* = \{w\}^* = \{\epsilon, w, w^2, \ldots\}$ and $w^+ = \{w\}^+ = \{w, w^2, \ldots\}$. A *left quotient* and a *right quotient* of a language L with respect to a word w is defined by $w/L = \{w_1 \in \Sigma^* | ww_1 \in L\}$ and $L/w = \{w_1 \in \Sigma^* | w_1 w \in L\}$, respectively.

The catenation is an associative operation. In addition, ϵ is the identity element with respect to the catenation on words. Therefore, $(\Sigma^*, \cdot)$ is a *monoid*. Since every word $w \in \Sigma^*$ may be represented uniquely as a catenation of letters, the monoid $(\Sigma^*, \cdot)$ is *free*, and Σ is the *base* of the monoid. It should be remarked that the set 2^{Σ^*} with respect to the catenation of languages is also a monoid.

Let M and M' be monoids. A function $h : M \to M'$ is a *morphism* if

$$\begin{aligned} h(m_1 m_2) &= h(m_1)h(m_2) \quad \forall m_1, m_2 \in M \qquad \text{and} \\ h(1_M) &= 1'_M, \end{aligned}$$

where 1_M and $1'_M$ are the identity elements of the monoids M and M', respectively. A morphism $h : \Sigma^* \to \Delta^*$ is *alphabetical* (resp. *ϵ-free*) if $h(a) \in \Delta \cup \{\epsilon\}$ (resp. $h(a) \neq \epsilon$) for every $a \in \Sigma$. A *substitution* is a morphism $h : \Sigma^* \to 2^{\Delta^*}$. Morphisms are denoted by letters h and g.

2.3 Bounded Languages

Let $k \in \mathbb{N}_+$. A language L is *k-bounded* if there exist words $w_1, w_2, \ldots, w_k \in \Sigma_L^*$ such that $L \subset w_1^* w_2^* \cdots w_k^*$. A language L is *bounded* if it is k-bounded for some $k \in \mathbb{N}_+$. A language L is *k- strictly bounded* if there exist pairwise different letters $a_1, a_2, \ldots, a_k \in \Sigma_L$ such that $L \subset a_1^* a_2^* \cdots a_k^*$. A language L is *strictly bounded* if it is k- strictly bounded for some $k \in \mathbb{N}_+$. We will denote the set of k- strictly bounded languages by $\mathcal{B}_k$.

Let $\Sigma = \{a_1, a_2, \ldots, a_k\}$. The *Parikh-mapping* is the function $\Phi : \Sigma^* \to \mathbb{N}^k$ for which $\Phi(w) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_k})$ for all $w \in \Sigma^*$. This mapping may also be defined for languages by

$$\Phi(L) = \bigcup_{w \in L} \{\Phi(w)\}.$$

The set $\Phi(L)$ is the *Parikh-image* of a language L. Let $w_1, w_2 \in \Sigma^*$. Then $\Phi(w_1 w_2) = \Phi(w_1) + \Phi(w_2)$ and $\Phi(\epsilon) = \overline{0}$. Thus, the Parikh-mapping is a monoid homomorphism from the monoid $(\Sigma^*, \cdot)$ into the monoid $(\mathbb{N}^k, +)$.

A language L is a *SLIP-language* if $\Phi(L)$ is a semilinear set. Let $L \subset a_1^* a_2^* \cdots a_k^*$ be a SLIP-language and $S_1, S_2, \ldots, S_n$ linear components of the set $\Phi(L)$. Then the languages $L_i = \Phi^{-1}(S_i) \cap a_1^* a_2^* \cdots a_k^*$ are called *linear components of the language* L. Moreover, let us denote $Conv(L_i) = \Phi^{-1}(Conv(S_i)) \cap a_1^* a_2^* \cdots a_k^*$.

Now we state the famous Parikh's theorem.

Theorem 2.2. *(9, Theorem 5.2.1) Every context-free language is a SLIP-language.*

We have the following result as a corollary of Parikh's theorem.

Corollary 2.3. *Every context-free language over one letter alphabet is regular.*

Of course, the reverse of Parikh's theorem is not true in general. However, a well-known result from Ginsburg gives us a necessary and sufficient condition for strictly bounded languages for being context-free.

Theorem 2.4. *(9, Lemma 5.4.2) A strictly bounded language is context-free if and only if its Parikh-image is a finite union of stratified linear sets.*

A SLIP-language $L \subset a_1^* a_2^* \cdots a_k^*$ is *(semi)convex in a language* $K \subset a_1^* a_2^* \cdots a_k^*$ if the set $\Phi(L)$ is (semi)convex in $\Phi(K)$.

2.4 Families of Languages

A *family of languages* is any set of languages that contains at least one nonempty language.[2] Families of languages are denoted by capital calligraphic letters. We denote the family of regular and context-free languages by $\mathcal{L}_{REG}$ and $\mathcal{L}_{CF}$, respectively.[3] Let $\mathcal{L}$ be a family of languages. A substitution $h : \Sigma^* \to 2^{\Delta^*}$ is a *$\mathcal{L}$-substitution* if $h(a) \in \mathcal{L}$ for every $a \in \Sigma$. We call a $\mathcal{L}_{REG}$-substitution also a *regular substitution.* A family of languages $\mathcal{L}$ is said to be *closed under substitution* if the family of languages is closed under $\mathcal{L}$-substitution.

Families of languages are often categorized by their closure properties. Let $\mathcal{L}$ be a family of languages. Then $\mathcal{L}$ is a

[2] Sometimes a family of languages is also required to be closed under copy.

[3] Regular languages are sometimes denoted by *Rat* and $\mathcal{L}_3$ and context-free languages by *Alg* and $\mathcal{L}_2$ in the literature.

– *duo* if it is closed under inverse morphism and intersection with regular languages[4];
– *full trio* if it is a duo and closed under morphism[5];
– *full AFL* if it is a full trio and closed under union, catenation and catenation closure.

In the literature, the terms trio and AFL are also used instead of full trio and full AFL, respectively. A trio (resp. an AFL) is required to be closed under ϵ-free morphism instead of completely arbitrary morphism. This work concentrates completely on full trios and full AFLs. It should be noted that all full AFL operations are not completely independent. If a family of languages is closed under intersection with regular languages, morphism, inverse morphism, union and catenation closure, the family of languages is also closed under catenation (5, Proposition V.4.2). Dependencies of the full AFL operations are studied more deeply in (14) and (10).

Let $\mathcal{T}(\mathcal{L})$ (resp. $\mathcal{F}(\mathcal{L})$, resp. $\mathcal{D}(\mathcal{L})$) denote the smallest full trio (resp. full AFL, resp. duo) containing a language family $\mathcal{L}$. If $\mathcal{L}$ is a singleton, i.e. $\mathcal{L} = \{L\}$, we may write $\mathcal{T}(L)$ (resp. $\mathcal{F}(L)$, resp. $\mathcal{D}(L)$) instead of $\mathcal{T}(\{L\})$ (resp. $\mathcal{F}(\{L\})$, resp. $\mathcal{D}(\{L\})$). In this case, we say that the full trio $\mathcal{T}(L)$ (resp. full AFL $\mathcal{F}(L)$, resp. duo $\mathcal{D}(L)$) is *principal*. A language L is called a *generator* of the full trio $\mathcal{T}(L)$ (resp. full AFL $\mathcal{F}(L)$, resp. duo $\mathcal{D}(L)$).

We say that a language L_1 *dominates rationally* a language L_2 if $\mathcal{T}(L_2) \subset \mathcal{T}(L_1)$. Languages L_1 and L_2 are *rationally equivalent* if $\mathcal{T}(L_1) = \mathcal{T}(L_2)$. Languages L_1 and L_2 are *rationally incomparable* if $\mathcal{T}(L_2) \not\subset \mathcal{T}(L_1)$ and $\mathcal{T}(L_1) \not\subset \mathcal{T}(L_2)$.

Theorem 2.5. *(5, Theorem III.4.1) Let $L \subset \Sigma^*$ and $L' \subset \Delta^*$. Then $L' \in \mathcal{T}(L)$ if and only if there exist an alphabet Γ, a regular language $R \subset \Gamma^*$ and alphabetical morphisms $h : \Gamma^* \to \Delta^*$ and $g : \Gamma^* \to \Sigma^*$ such that $L' = h(g^{-1}(L) \cap R)$.*

Next, we define a sequence of language families that up to a certain extent measure the size of a bounded language. This definition addresses the number of components in a bounded language that can increase without limit simulta-

[4]In the literature, a duo is called sometimes a *cylinder*.
[5]In the literature, a full trio is called sometimes a *rational cone*, or shorter a *cone*.

neously. So we set

$$\mathcal{L}_k = \Big\{ L \subset \Sigma^* \;\Big|\; \exists n \in \mathbb{N}, x_{i,j}, w_{i,j} \in \Sigma^* : L \subset \bigcup_{i=0}^{n} x_{i,0} w_{i,1}^* x_{i,1} w_{i,2}^* \cdots x_{i,k-1} w_{i,k}^* x_{i,k} \Big\}$$

for all $k \in \mathbb{N}$. In addition, we set

$$\mathcal{R}_k = \mathcal{L}_k \cap \mathcal{L}_{REG}$$

for all $k \in \mathbb{N}$. By the properties of bounded regular languages, the language R is in $\mathcal{R}_k$ if and only if R is a finite union of languages of the form $x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k$, where $x_0, w_1, x_1, w_2, \ldots, x_{k-1}, w_k, x_k$ are words.

Naturally, we have relations $\mathcal{L}_k \subset \mathcal{L}_{k+1}$, $\mathcal{R}_k \subset \mathcal{R}_{k+1}$ and $\mathcal{R}_k \subset \mathcal{L}_k$ for all $k \in \mathbb{N}$. Note also that each $\mathcal{L}_k$ contains only bounded languages.

Example 2.6. $a_1 a_2^* a_3^* \cup a_1^* a_2 a_3^* \in \mathcal{R}_2 \setminus \mathcal{R}_1$.

2.4.1 Minimality of Families of Languages

Since all the full trios and full AFLs contain the family of regular languages, it is natural to define minimal families of languages in the following way.

Definition 2.7. A full trio (resp. full AFL, resp. duo) $\mathcal{L}$ is *minimal* if

- $\mathcal{L}$ contains nonregular languages;
- there does not exist a full trio (resp. full AFL, resp. duo) $\mathcal{L}'$ such that $\mathcal{L}'$ contains nonregular languages and $\mathcal{L}' \subsetneq \mathcal{L}$.

If a full trio (resp. full AFL, resp. duo) $\mathcal{L}$ is minimal, we say that $\mathcal{L}$ is a *minimal full trio* (resp. *minimal full AFL*, resp. *minimal duo*).

The next theorem states that each nonregular language in a minimal language family is a generator of the language family.

Theorem 2.8. *A full trio (resp. full AFL, resp. duo) $\mathcal{L}$ is minimal if and only if $\mathcal{L}$ contains nonregular languages and for each $L \in \mathcal{L} \setminus \mathcal{L}_{REG}$, we have $\mathcal{T}(L) = \mathcal{L}$ (resp. $\mathcal{F}(L) = \mathcal{L}$, resp. $\mathcal{D}(L) = \mathcal{L}$).*

Proof. Let us prove the claim concerning full trios. The proofs concerning full AFLs and duos are fully analogue.

Let $\mathcal{L}$ be a minimal full trio. Then there exists a nonregular $L \in \mathcal{L}$. Thus, $\mathcal{T}(L) \subseteq \mathcal{L}$ and $L \in \mathcal{T}(L)$. Hence, $\mathcal{T}(L) = \mathcal{L}$.

Let $\mathcal{L}$ be a full trio that is not minimal. We may assume that $\mathcal{L}$ contains nonregular languages. Then there exists a full trio $\mathcal{L}' \subsetneq \mathcal{L}$ and a nonregular language $L' \in \mathcal{L}'$. Then $\mathcal{T}(L') \subset \mathcal{L}' \subsetneq \mathcal{L}$. □

Corollary 2.9. *Minimal full trios (resp. minimal full AFLs, resp. minimal duos) are principal.*

Let $\mathcal{L}$ be a full AFL and a minimal duo. Then there does not exist a duo $\mathcal{L}'$ such that $\mathcal{L}_{REG} \subsetneq \mathcal{L}' \subsetneq \mathcal{L}$. Thus, there does not exist a full trio $\mathcal{L}''$ or a full AFL $\mathcal{L}'''$ such that $\mathcal{L}_{REG} \subsetneq \mathcal{L}'' \subsetneq \mathcal{L}$ or $\mathcal{L}_{REG} \subsetneq \mathcal{L}''' \subsetneq \mathcal{L}$. Hence, $\mathcal{L}$ is also a minimal full trio and a minimal full AFL.

If $\mathcal{L}_1$ is a minimal duo, $\mathcal{L}_2$ a minimal full trio, $\mathcal{L}_3$ a minimal full AFL and $\mathcal{L}_1 \cap \mathcal{L}_2 \cap \mathcal{L}_3 \not\subset \mathcal{L}_{REG}$, then $\mathcal{L}_1 \subset \mathcal{L}_2 \subset \mathcal{L}_3$.

It is very hard to give a concrete example of the minimal language families. In fact, Autebert et al. have stated the following conjecture on the issue.

Conjecture 2.10. *(2) (see also (1)) There does not exist a minimal full trio in the family of context-free languages.*

Let us state another minimality definition.

Definition 2.11. A full trio (resp. full AFL, resp. duo) $\mathcal{L}$ is *minimal with respect to a language family* $\mathcal{K}$ if

- the family $\mathcal{L} \cap \mathcal{K}$ contains nonregular languages;
- there does not exist a full trio (resp. full AFL, resp. duo) $\mathcal{L}'$ such that $\mathcal{L}' \cap \mathcal{K}$ contains nonregular languages and $\mathcal{L}' \subsetneq \mathcal{L}$.

Theorem 2.8 and Corollary 2.9 may be easily extended to concern minimality with respect to a language family $\mathcal{K}$.

Theorem 2.12. *A full trio (resp. full AFL, resp. duo) $\mathcal{L}$ is minimal with respect to a language family $\mathcal{K}$ if and only if $\mathcal{L} \cap \mathcal{K}$ contains nonregular languages and for each $L \in (\mathcal{L} \cap \mathcal{K}) \setminus \mathcal{L}_{REG}$, we have $\mathcal{T}(L) = \mathcal{L}$ (resp. $\mathcal{F}(L) = \mathcal{L}$, resp. $\mathcal{D}(L) = \mathcal{L}$).*

Proof. Let us prove the claim concerning full trios. The proofs concerning full AFLs and duos are fully analogue.

Let $\mathcal{L}$ be a minimal full trio with respect to a language family $\mathcal{K}$. Then there exists a nonregular $L \in \mathcal{L} \cap \mathcal{K}$. Thus, $\mathcal{T}(L) \subseteq \mathcal{L}$ and $L \in \mathcal{T}(L)$. Hence, $\mathcal{T}(L) = \mathcal{L}$.

Let $\mathcal{L}$ be a full trio that is not minimal with respect to a language family $\mathcal{K}$. We may assume that $\mathcal{L} \cap \mathcal{K}$ contains nonregular languages. Then there exists a full trio $\mathcal{L}' \subsetneq \mathcal{L}$ and a nonregular language $L' \in \mathcal{L}' \cap \mathcal{K}$. Then $\mathcal{T}(L') \subset \mathcal{L}' \subsetneq \mathcal{L}$. □

Corollary 2.13. *Minimal full trios (resp. minimal full AFLs, resp. minimal duos) with respect to a language family $\mathcal{K}$ are principal.*

It is easy to see that, if a full trio (resp. full AFL, resp. duo) is minimal with respect to language families $\mathcal{K}_1$ and $\mathcal{K}_2$, it is also minimal with respect to the language family $\mathcal{K}_1 \cup \mathcal{K}_2$.

Substituting the language family $\mathcal{K}$ by $\mathcal{L}_{CF}$ in Definition 2.11, we see that a full trio (resp. full AFL, resp. duo) $\mathcal{L} \subset \mathcal{L}_{CF}$ is minimal if and only if it is minimal with respect to $\mathcal{L}_{CF}$.

Let us define $S_< = \{a^m b^n | m < n\}$, $S_> = \{a^m b^n | m > n\}$ and $S_{\neq} = \{a^m b^n | m \neq n\}$. Berstel and Boasson have proved the following theorem regarding the minimality of full trios and full AFLs.

Theorem 2.14. *(6, Proposition B) Let L be a nonregular 2-bounded context-free language. Then $\mathcal{T}(S_\theta) \subset \mathcal{T}(L)$ and $\mathcal{F}(S_\theta) \subset \mathcal{F}(L)$ for some $\theta \in \{<,>,\neq\}$.*

Lemma 2.15. *(4, Lemme) Let $L, L' \subset \{a,b\}^*$ be context-free languages. Then $\mathcal{T}(L') \subset \mathcal{T}(L)$ if and only if $\mathcal{F}(L') \subset \mathcal{F}(L)$.*

Since we know that the languages $S_<$, $S_>$ and $S_{\neq}$ are pairwise incomparable (5, Theorem V.7.3), we get the next corollary from Theorem 2.14. The claim concerning full AFLs follows by Lemma 2.15.

Corollary 2.16. *Let $\theta \in \{<,>,\neq\}$. Then the full trio $\mathcal{T}(S_\theta)$ and the full AFL $\mathcal{F}(S_\theta)$ are minimal with respect to the 2-bounded context-free languages.*

2.5 Transducers

A *transducer* is a 6-tuple $M = (Q, \Sigma, \Delta, E, q_-, Q_+)$, where

- Q is a finite set of *states*;
- Σ is an *input alphabet*;

– Δ is an *output alphabet*;
– $E \subset Q \times \Sigma^* \times \Delta^* \times Q$ is a finite set of *transitions*;
– $q_- \in Q$ is an *initialization state*;
– $Q_+ \subset Q$ is a set of *final states.*

Let $p, q \in Q$, $w_1, w_2 \in \Sigma^*$ and $v_1, v_2 \in \Delta^*$. We denote $(p, w_1, v_1) \vdash (q, w_2, v_2)$ if there exist $x \in \Sigma^*$ and $y \in \Delta^*$ such that $w_1 = xw_2$, $v_2 = v_1 y$ and $(p, x, y, q) \in E$. Let $\overset{*}{\vdash}$ be the transitive reflexive closure of the relation $\vdash$. Therefore, $(p_0, w, \epsilon) \overset{*}{\vdash} (p_n, \epsilon, v)$ if and only if there exist $n \in \mathbb{N}$, states $p_1, p_2, \ldots, p_{n-1} \in Q$ and words $w_1, w_2, \ldots, w_n \in \Sigma^*$, $v_1, v_2, \ldots, v_n \in \Delta^*$ such that $w = w_1 w_2 \cdots w_n$, $v = v_1 v_2 \cdots v_n$ and $(p_{i-1}, w_i, v_i, p_i) \in E$ for all $i \in \{1, 2, \ldots, n\}$. If moreover $p_0 = q_-$ and $p_n \in Q_+$, then the word $(p_0, w_1, v_1, p_1)(p_1, w_2, v_2, p_2) \cdots (p_{n-1}, w_n, v_n, p_n) \in E^*$ is a *computation* of the transducer M. Let us denote the set of all the computations of a transducer M by Π_M. Define morphisms $h_M : E^* \to \Sigma^*$ and $g_M : E^* \to \Delta^*$ by $h_M(p, w, v, q) = w$ and $g_M(p, w, v, q) = v$ for all $(p, w, v, q) \in E$. Finally, let us define the function $\tau_M : \Sigma^* \to 2^{\Delta^*}$ for each transducers M by

$$\tau_M(w) = \{g_M(e) \mid h_M(e) = w, e \in \Pi_M\}.$$

The function τ_M is called a *rational transduction.* Rational transductions are denoted by Greek letters τ and ω.

Theorem 2.17. *(5, Theorem III.6.1) Let L and L' be languages. Then $L' \in \mathcal{T}(L)$ if and only if there exists a transducer M such that $\tau_M(L) = L'$.*

3 Very Small Families Generated by Bounded Context-Free Languages

Let us first define new families of context-free languages. These families turn out to be a very interesting and useful tool for examining rational dominating relations of weak bounded context-free languages. Let $k \in \mathbb{N}_+$. Define

$$\mathcal{C}_k = \{L \in \mathcal{L}_{CF} | L \cap x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k \in \mathcal{L}_{REG} \quad \forall x_i, w_i \in \Sigma_L^*\}$$

and $\mathcal{C}_\infty = \cap_{i=1}^{\infty} \mathcal{C}_i$. Clearly $\mathcal{C}_{k+1} \subset \mathcal{C}_k$ for all $k \in \mathbb{N}_+$. It follows directly from the definition that $\mathcal{L}_{REG} \subset \mathcal{C}_\infty$ and $\mathcal{C}_k \subset \mathcal{L}_{CF}$ for all $k \in \mathbb{N}_+$.

We shall make the following observations in this chapter:

Observation 1. each nonregular bounded context-free language belongs to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ for some $k \in \mathbb{N}_+$;

Observation 2. for each $k \in \mathbb{N}_+$, any language in $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ can be transformed into a language in $\mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$ using only full trio operations; and

Observation 3. for each $k \in \mathbb{N}_+$, the family $\mathcal{C}_k$ is a full AFL.

The first and second observations imply that we can proceed any nonregular bounded context-free language downwards in the chain $\mathcal{C}_1 \setminus \mathcal{C}_2, \mathcal{C}_2 \setminus \mathcal{C}_3, \ldots$ as far as we wish using full trio operations. The third observation states that we cannot proceed any language upwards in the same chain even with full AFL operations. Combining these results, we get a proof to the next conjecture.

Conjecture 3.1. *(1) There does not exist a minimal full trio with respect to the bounded context-free languages.*

This chapter is organized as follows. Section 3.1 deals with some basic properties of these families with examples. Two main points of the section are to note that Observation 1 holds and the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ is nonempty for each $k \in \mathbb{N}_+$. In Section 3.2, we shall prove that each full trio generated by a language from the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ contains nonregular $(k+1)$- strictly bounded languages. In Section 3.3, we will study the structure of these $(k+1)$- strictly bounded languages of the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ more deeply. After this work, we are ready to prove in

Section 3.4 that Observation 2 holds. In Section, 3.5 we shall show Observation 3, i.e. that the family $\mathcal{C}_k$ is a full AFL for all $k \in \mathbb{N}_+$. Finally, in Section 3.5 we will give a formal proof for Conjecture 3.1 using the same reasoning pointed out above.

3.1 Examples of Languages in $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$

Let $k \in \mathbb{N}_+$ and $\theta_1, \theta_2, \ldots, \theta_k \in \{<, >, \neq\}$. For each $i \in \{1, 2, \ldots, k\}$, define the language

$$S_i(\theta_1, \theta_2, \ldots, \theta_k) = \left\{ a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k} \,\middle|\, n_1, n_2, \ldots, n_k \in \mathbb{N} \text{ and } \bigvee_{j=1}^{k} n_i \theta_j n_j \right\}.$$

We remark that the condition $n_i \theta_i n_i$ is always false. Thus, this term is included in the definition only for the sake of simplifying the definition.

Example 3.2. Let us consider the language

$$S_1(\neq, >) = \{a_1^{n_1} a_2^{n_2} | n_1 > n_2\} = S_2(<, \neq).$$

For example, $S_1(\neq, >) \cap a_1^* a_2 = a_1 a_1^+ a_2 \in \mathcal{L}_{REG}$ and $S_1(\neq, >) \cap a_1 a_2^* = a_1 \in \mathcal{L}_{REG}$. It is clear that $S_1(\neq, >) \in \mathcal{C}_1 \setminus \mathcal{C}_2$.

Example 3.3. Let us consider the language

$$\begin{aligned} S_1(\neq, \neq, \neq) &= \{a_1^{n_1} a_2^{n_2} a_3^{n_3} | n_1 \neq n_2 \text{ or } n_1 \neq n_3\} \\ &= S_2(\neq, \neq, \neq) = S_3(\neq, \neq, \neq) \\ &= a_1^* a_2^* a_3^* \setminus \{a_1^n a_2^n a_3^n | n \in \mathbb{N}\}. \end{aligned}$$

Now, for instance, $S_1(\neq, \neq, \neq) \cap a_1^* a_2^* a_3 = a_1^* a_2^* a_3 \setminus \{a_1 a_2 a_3\} \in \mathcal{L}_{REG}$. It is easily seen that $S_1(\neq, \neq, \neq) \in \mathcal{C}_2 \setminus \mathcal{C}_3$. Thus, it is quite obvious that $S_1(\neq, \neq, \neq, \neq) \in \mathcal{C}_3 \setminus \mathcal{C}_4$ and furthermore $S_1(\underbrace{\neq, \neq, \ldots, \neq}_{k+1 \text{ times}}) \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ for all $k \in \mathbb{N}_+$ (see Theorem 3.7 for formal proof).

Example 3.4. Set

$$\begin{aligned} L_G &= \left\{ a_2^{i_0} a_1 a_2^{i_1} a_1 \cdots a_2^{i_p} a_1 \,\middle|\, p \in \mathbb{N}, \quad \exists j \in \{0, 1, \ldots, p\} : i_j \neq j \right\} \\ &= (a_2^* a_1)^* \setminus \{a_1 a_2 a_1 a_2^2 a_1 a_2^3 \cdots a_1 a_2^p | p \in \mathbb{N}\}. \end{aligned}$$

Now, for example, $L_G \cap a_2^* a_1 a_2^* a_1 a_2^* a_1 = a_2^* a_1 a_2^* a_1 a_2^* a_1 \setminus \{a_1 a_2 a_1 a_2^2 a_1\} \in \mathcal{L}_{REG}$ and $L_G \cap (a_2^3 a_1)^* = (a_2^3 a_1)^*$. The language L_G is constructed from the language $S_1(\neq,\neq)$ in page 55. At last, from this construction it should be clear that L_G is context-free. Goldstine has proved that every bounded language in $\mathcal{T}(L_G)$ is regular (12).[6] Thus, we can conclude that $L_G \in \mathcal{C}_\infty \setminus \mathcal{L}_{REG}$. We will prove this in Corollary 4.3.

Theorem 3.5. $\mathcal{C}_1 = \mathcal{L}_{CF}$

Proof. Obviously, $\mathcal{C}_1 \subset \mathcal{L}_{CF}$. Let $L \in \mathcal{L}_{CF}$ and $x_0, w_1, x_1 \in \Sigma^*$. Then $L \cap x_0 w_1^* x_1$ is rationally equivalent to the one letter language $\{a^n | x_0 w_1^n x_1 \in L\}$. Thus, by Corollary 2.3, we have $L \cap x_0 w_1^* x_1 \in \mathcal{L}_{REG}$ and $L \in \mathcal{C}_1$. □

Theorem 3.5 will imply that each context-free language belongs either to $\mathcal{C}_\infty$ or exactly one family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$, $k \in \mathbb{N}_+$. In addition, due to Examples 3.2, 3.3 and 3.4, each of the families $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ and $\mathcal{C}_\infty \setminus \mathcal{L}_{REG}$ is nonempty. Thus, we have the chain

$$\mathcal{L}_{CF} = \mathcal{C}_1 \supsetneq \mathcal{C}_2 \supsetneq \cdots \supsetneq \mathcal{C}_\infty \supsetneq \mathcal{L}_{REG}.$$

We will expand a similar chain also inside the family $\mathcal{C}_\infty$ in Chapter 4.

The next intuitively obvious theorem helps us to classify strictly bounded languages into the right family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$.

Theorem 3.6. *Let $n \in \mathbb{N}_+$ and $L \subset a_1^* a_2^* \cdots a_n^*$. Given $k \in \mathbb{N}_+$, the language L is in $\mathcal{C}_k$ if and only if $L \cap a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_n^{\alpha_n} \in \mathcal{L}_{REG}$ for all $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{N} \cup \{*\}$, where $|\{i \in \{1,2,\ldots,n\} | \alpha_i = *\}| = k$.*

Proof. "⇒". This follows directly from the definition of $\mathcal{C}_k$.

"⇐". Assume that $L \notin \mathcal{C}_k$. Then there exist $x_0, w_1, x_1, w_2, \ldots, x_{k-1}, w_k, x_k \in \Sigma_L^*$ such that

$$L' = L \cap x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k \notin \mathcal{L}_{REG}.$$

Since $L \subset a_1^* a_2^* \cdots a_n^*$, we may assume that

$$x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k \subset a_1^* a_2^* \cdots a_n^*.$$

[6] Goldstine actually defined a slightly different form of the language L_G. It is proved on page 56 that these languages generate anyway exactly the same full trio.

We may also assume that $w_i \neq \epsilon$ for all $i \in \{1,2,\ldots,k\}$. Thus, there exist $p_1, r_1, p_2, r_2, \ldots, p_n, r_n \in \mathbb{N}$ such that

$$x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k = a_1^{p_1}(a_1^{r_1})^* a_2^{p_2}(a_2^{r_2})^* \cdots a_n^{p_n}(a_n^{r_n})^*$$

and $|\{i \in \{1,2,\ldots,n\} | r_i > 0\}| = k$. Set

$$\alpha_i = \begin{cases} p_i, & \text{if } r_i = 0, \\ *, & \text{if } r_i > 0 \end{cases}$$

for all $i \in \{1,2,\ldots,n\}$. Then

$$L' = (L \cap a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_n^{\alpha_n}) \cap a_1^{p_1}(a_1^{r_1})^* a_2^{p_2}(a_2^{r_2})^* \cdots a_n^{p_n}(a_n^{r_n})^*.$$

Thus, $L \cap a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_n^{\alpha_n} \notin \mathcal{L}_{REG}$ for some $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{N} \cup \{*\}$ such that $|\{i \in \{1,2,\ldots,n\} | \alpha_i = *\}| = k$. □

Finally, we give a necessary and sufficient condition for our example languages $S_i(\theta_1, \theta_2, \ldots, \theta_{k+1})$ to belong to the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$.

Theorem 3.7. *Let $k \geq 2$ and $i \in \{1,2,\ldots,k+1\}$. Then*

$$S_i(\theta_1, \theta_2, \ldots, \theta_{k+1}) \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$$

if and only if $\theta_j \in \{>, \neq\}$ for all $j \in \{1,2,\ldots,k+1\} \setminus \{i\}$.

Proof. Assume without loss of generality that $i = 1$. The proof of the case $i \neq 1$ is fully analogous.

"$\Leftarrow$". Let $L = S_1(\theta_1, \theta_2, \ldots, \theta_{k+1})$, where $\theta_j \in \{>, \neq\}$ for all $j \in \{2,3,\ldots,k+1\}$. Since $L \notin \mathcal{L}_{REG}$, we have $L \notin \mathcal{C}_{k+1}$. Therefore, it suffices to show that $L \in \mathcal{C}_k$.

Let $m \in \mathbb{N}$. By Theorem 3.6, it is enough to show that

$$L \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{L}_{REG}$$

for all $j \in \{1,2,\ldots,k+1\}$. Assume first that $j = 1$. Then

$$\begin{aligned} & L \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^* \\ = \; & \{a_1^m a_2^{n_2} a_3^{n_3} \cdots a_{k+1}^{n_{k+1}} \mid \bigvee_{l=2}^{k+1} m\theta_l n_l \text{ and } n_2, n_3, \ldots, n_{k+1} \in \mathbb{N}\} \\ = \; & \bigcup_{l=2}^{k+1} \{a_1^m a_2^{n_2} a_3^{n_3} \cdots a_{k+1}^{n_{k+1}} \mid m\theta_l n_l \text{ and } n_2, n_3, \ldots, n_{k+1} \in \mathbb{N}\} \in \mathcal{L}_{REG}. \end{aligned}$$

Assume next that $j \in \{2,3,\dots,k+1\}$. Assume without loss of generality that $j = 2$. We noted earlier that $L \cap a_1^l a_2^* a_3^* \cdots a_{k+1}^*$ is regular for fixed $l \in \mathbb{N}$. Thus, we have

$$\begin{aligned} L \cap a_1^* a_2^m a_3^* a_4^* \cdots a_{k+1}^* &= a_1^{m+1} a_1^* a_2^m a_3^* a_4^* \cdots a_{k+1}^* \\ &\cup \bigcup_{l=0}^{m} \left(L \cap a_1^l a_2^m a_3^* a_4^* \cdots a_{k+1}^* \right) \in \mathcal{L}_{REG}. \end{aligned}$$

Hence, $L \in \mathcal{C}_k$.

"⇒". Suppose that θ_j is the relation $<$ for some $j \in \{2,3,\dots,k+1\}$. Assume again without loss of generality that $j = 2$. Let us consider the language

$$\begin{aligned} L' &= L \cap a_1^* a_3^* a_4^* \cdots a_{k+1}^* \\ &= \left\{ a_1^{n_1} a_3^{n_3} a_4^{n_4} \cdots a_{k+1}^{n_{k+1}} \,\middle|\, \bigvee_{\substack{l=1 \\ l \neq 2}}^{k+1} n_1 \theta_l n_l \vee n_1 < 0 \right\} \\ &= \left\{ a_1^{n_1} a_3^{n_3} a_4^{n_4} \cdots a_{k+1}^{n_{k+1}} \,\middle|\, \bigvee_{l=3}^{k+1} n_1 \theta_l n_l \right\}. \end{aligned}$$

Since $k \geq 2$, at least one condition $n_1 \theta_l n_l$ remains valid, and the language L' is nonregular. Thus, $L \notin \mathcal{C}_k$. The proof is now complete. □

3.2 A Normal Form Representative for Languages in $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$

In this section, we will prove that, for each $k \in \mathbb{N}_+$ and $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$, there exists a $(k+1)$- strictly bounded language $L' \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ such that L' belongs to the full trio generated by L. This result implies that, if we are looking for a language generating a minimal full trio (or a minimal full AFL) inside the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$, we may always restrict to look it inside the family of $(k+1)$- strictly bounded languages.

Let $n \in \mathbb{N}_+$ and $x_0, w_1, x_1, w_2, x_3, \dots, w_n, x_n \in \Sigma^*$. Let us define a transducer $M(x_0, w_1, x_1, w_1, x_2, \dots, x_k, w_k) = (\{p_0, p_1, \dots, p_{n+1}\}, \Sigma_n, \Sigma, E, p_0, \{p_{n+1}\})$, where

$$E = \{(p_i, \epsilon, x_i, p_{i+1}) | i \in \{0,1,\dots,n\}\} \cup \{(p_i, a_i, w_i, p_i) | i \in \{1,2,\dots,n\}\}.$$

If the words $x_0, w_1, x_1, w_2, x_3 \ldots, w_n, x_n$ are clear from the context, we can simplify the notation by setting

$$M = M(x_0, w_1, x_1, w_2, x_3, \ldots, w_n, x_n).$$

Let $L \subset a_1^* a_2^* \cdots a_n^*$. Then

$$\tau_M(L) = \{x_0 w_1^{i_1} x_1 w_2^{i_2} x_3 \cdots w_n^{i_n} x_n | a_1^{i_1} a_2^{i_2} \cdots a_n^{i_n} \in L\}.$$

Therefore, it should be clear that the rational transduction $\tau_M : \Sigma^* \to 2^{\Sigma^*}$ may be regarded as a function $\tau_M : a_1^* a_2^* \cdots a_n^* \to \Sigma^*$. We should also note that, as an inverse of a rational transduction, $\tau_M^{-1} : \Sigma^* \to 2^{\Sigma^*}$ is also a rational transduction.

Lemma 3.8. *Let $n \in \mathbb{N}_+$ and $L \subset x_0 w_1^* x_2 w_2^* x_3 \cdots w_n^* x_n$. Given $k \in \mathbb{N}_+$, the language L is in $\mathcal{C}_k$ if and only if $\tau_M^{-1}(L)$ is in $\mathcal{C}_k$.*

Proof. Assume first that $L \in \mathcal{C}_k$. Let

$$R = z_0 v_1^* z_1 v_2^* \cdots z_{k-1} v_k^* z_k,$$

where $z_0, v_1, z_1, v_2, \ldots, z_{k-1}, v_k, z_k \in \Sigma^*_{\tau_M^{-1}(L)}$. It suffices to show that $\tau_M^{-1}(L) \cap R \in \mathcal{L}_{REG}$. Since $\tau_M^{-1}(L) \subset a_1^* a_2^* \cdots a_n^*$, we may assume that $R \subset a_1^* a_2^* \cdots a_n^*$. Since $L \in \mathcal{C}_k$, we have $L \cap \tau_M(R) \in \mathcal{L}_{REG}$ and

$$\tau_M^{-1}(L) \cap R = \tau_M^{-1}(L) \cap \tau_M^{-1}(\tau_M(R)) \cap R = \tau_M^{-1}(L \cap \tau_M(R)) \cap R \in \mathcal{L}_{REG}.$$

Hence, $\tau_M^{-1}(L) \in \mathcal{C}_k$.

Assume now that $\tau_M^{-1}(L) \in \mathcal{C}_k$. Let

$$R = z_0 v_1^* z_1 v_2^* \cdots z_{k-1} v_k^* z_k,$$

where $z_0, v_1, z_1, v_2, \ldots, z_{k-1}, v_k, z_k \in \Sigma_L^*$. It suffices to show that $L \cap R \in \mathcal{L}_{REG}$. Therefore, we may assume that $R \subset x_0 w_1^* x_2 w_2^* x_3 \cdots w_n^* x_n$. Clearly, there exists a regular language $R' \in \mathcal{R}_k$ such that $\tau_M(R') = R$. Since $\tau_M^{-1}(L) \in \mathcal{C}_k$, we have

$$L \cap R = \tau_M(\tau_M^{-1}(L)) \cap \tau_M(R') = \tau_M(\tau_M^{-1}(L) \cap R') \in \mathcal{L}_{REG}.$$

Therefore, $L \in \mathcal{C}_k$. □

Finally, we are ready to prove the main result of this section. The next theorem will imply that, if we are looking for a language generating a minimal full trio (or a minimal full AFL) inside the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$, we may always restrict to look it inside the family of $(k+1)$- strictly bounded languages.

Theorem 3.9. *Let $k \in \mathbb{N}_+$ and $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$. Then there exists a language $L_1 \in \mathcal{T}(L)$ such that $L_1 \in \mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$.*

Proof. Since $L \notin \mathcal{C}_{k+1}$, we have $x_0, x_1, \ldots, x_{k+1} \in \Sigma^*$ and $w_1, w_2, \ldots, w_{k+1} \in \Sigma^*$ such that

$$L_1 = L \cap x_0 w_1^* x_1 w_2^* \cdots x_k w_{k+1}^* x_{k+1} \notin \mathcal{L}_{REG}.$$

Hence, $L_1 \notin \mathcal{C}_{k+1}$. Since $L \in \mathcal{C}_k$, we have $L_1 \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$. According to Lemma 3.8, we have $\tau_M^{-1}(L_1) \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$, where $\tau_M : a_1^* a_2^* \cdots a_k^* \to \Sigma^*$ is a function defined in the beginning of this section. Since τ_M^{-1} is a rational transduction, we have by Theorem 2.17, $\tau_M^{-1}(L_1) \in \mathcal{T}(L)$. □

Now we are able to reformulate Theorem 2.14 and Corollary 2.16.

Theorem 3.10. *Let $L \in \mathcal{C}_1 \setminus \mathcal{C}_2$. Then $\mathcal{T}(S_\theta) \subset \mathcal{T}(L)$ and $\mathcal{F}(S_\theta) \subset \mathcal{F}(L)$ for some $\theta \in \{<, >, \neq\}$.*

Proof. By Theorem 3.9, there exists a 2-bounded nonregular language $L' \in \mathcal{T}(L)$. By Theorem 2.14, either $\mathcal{T}(S_<) \subset \mathcal{T}(L')$, $\mathcal{T}(S_>) \subset \mathcal{T}(L')$ or $\mathcal{T}(S_{\neq}) \subset \mathcal{T}(L')$. This proves the claim concerning full trios. Since $\mathcal{T}(L) \subset \mathcal{F}(L)$, the claim concerning full AFLs follows by the same reasoning. □

Corollary 3.11. *Let $\theta \in \{<, >, \neq\}$. Then the full trio $\mathcal{T}(S_\theta)$ and the full AFL $\mathcal{F}(S_\theta)$ are minimal with respect to the family $\mathcal{C}_1 \setminus \mathcal{C}_2$.*

3.3 On the Structure of Languages in $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$

In the previous section, we noted that, if we are looking for a language generating a minimal full trio (or a minimal full AFL) inside the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$, we may always restrict to look it inside the family of $(k+1)$- strictly bounded languages. This gives us an excellent starting point for studying the structure of weak languages of the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ in this section. However, general $(k+1)$-strictly bounded languages still provide too complex structure for deeper analysis of the languages. We shall note that every nonregular strictly bounded context-free language can be intersected with a regular language R so that the resulting language is nonregular and semiconvex in R (see Lemma 3.14). Languages that are semiconvex in some regular language turn out to be simple enough so that we will get some results out of the structure of these languages in the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$.

Note that the Parikh image of a strictly bounded regular language may always be represented as a semilinear set whose periods are parallel with a coordinate axis. Thus, the next lemma states that a proper linear set is always convex in a semilinear set whose periods are parallel with a coordinate axis. In other words, a proper linear set is fully characterized by its boundaries and a grid. This is illustrated in Figure 1.

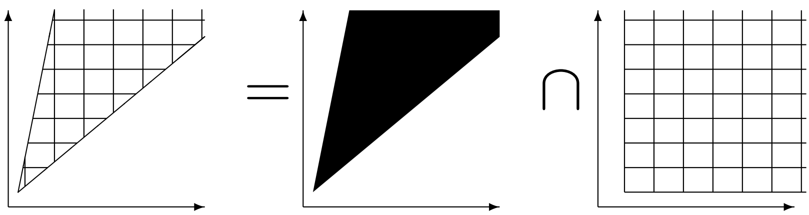

Fig 1. Illustration of Lemma 3.12 in case $k = 2$.

Lemma 3.12. *(16, Lemma 2) Let $k \in \mathbb{N}_+$ and $S \subset \mathbb{N}_+^k$ be a proper linear set. Then there exists a regular language $R \subset a_1^* a_2^* \cdots a_k^*$ such that S is convex in $\Phi^{-1}(R)$.*

Lemma 3.13. *Let $k \in \mathbb{N}_+$ and $L \subset a_1^* a_2^* \cdots a_k^*$ be a SLIP-language such that L is semiconvex in a regular language $R \subset a_1^* a_2^* \cdots a_k^*$. Let $R' \subset a_1^* a_2^* \cdots a_k^*$ be another regular language. Then $L \cap R'$ is semiconvex in $R \cap R'$.*

Proof. Let $n \in \mathbb{N}_+$ and $L_1, L_2, \ldots, L_n$ be linear components of L such that $L = \bigcup_{i=1}^n L_i = \bigcup_{i=1}^n (Conv(L_i) \cap R)$. For each $i \in \{1, 2, \ldots, n\}$, let $n(i) \in \mathbb{N}_+$ and $L_{i,1}, L_{i,2}, \ldots, L_{i,n(i)}$ be linear components of $L_i \cap R'$ such that $L_i \cap R' = \bigcup_{j=1}^{n(i)} L_{i,j}$. Then for each $i \in \{1, 2, \ldots, n\}$, we have

$$\bigcup_{j=1}^{n(i)} Conv(L_{i,j}) \subset Conv(L_i).$$

By the above relation and semiconvexity of L in R, we have

$$\bigcup_{i=1}^{n} \bigcup_{j=1}^{n(i)} Conv(L_{i,j}) \cap R \cap R' \subset \bigcup_{i=1}^{n} Conv(L_i) \cap R \cap R' = L \cap R'.$$

On the other hand, since $L \subset R$ and $L \cap R' = \bigcup_{i,j} L_{i,j} \subset \bigcup_{i,j} Conv(L_{i,j})$, we have

$$L \cap R' = \bigcup_{i,j} Conv(L_{i,j}) \cap R \cap R'.$$

□

The next lemma is a slightly modified version of Lemma 3 from (16). The lemma states that we can factorize every strictly bounded SLIP-language into parts that are semiconvex in some regular language. Moreover, the lemma states that we may do this by intersecting the original SLIP-language with a regular language. This result and Theorem 3.9 will imply that, if we are looking for a language generating a minimal full trio (or a minimal full AFL) inside the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$, we may always restrict to look it inside the family of $(k+1)$-strictly bounded languages that are semiconvex in some regular language. By the semiconvexity, the inner structure of the studied language gets much simpler. Figure 2 illustrates the lemma.

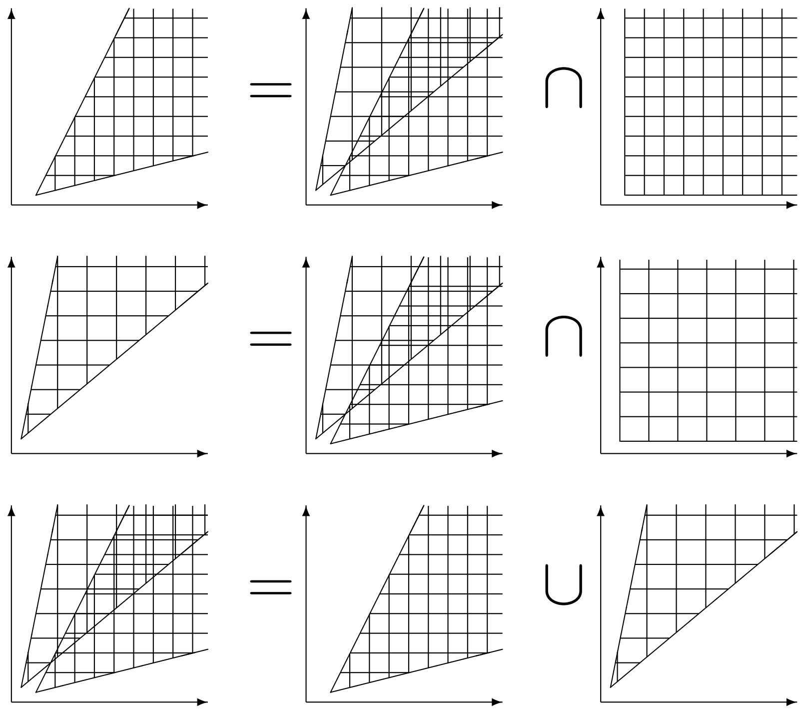

Fig 2. Illustration of Lemma 3.14 in case $k = 2$.

Lemma 3.14. *Let $k \in \mathbb{N}_+$ and $L \subset a_1^* a_2^* \cdots a_k^*$ be a SLIP-language. Then there exist $n \in \mathbb{N}_+$ and regular languages $R_1, R_2, \ldots, R_n \subset a_1^* a_2^* \cdots a_k^*$ such that*

1. *$\Phi(R_i)$ is a linear set for all $i \in \{1,2,\ldots,n\}$;*
2. *$L \cap R_i$ is semiconvex in R_i for all $i \in \{1,2,\ldots,n\}$;*
3. *$L = \bigcup_{i=1}^n L \cap R_i$.*

Proof. Since $\Phi(L)$ is a semilinear set, there exist, by Theorem 2.1, $n \in \mathbb{N}_+$ and proper linear sets $S_1, S_2, \ldots, S_n \in \mathbb{N}^k$ such that $\Phi(L) = \bigcup_{i=1}^n S_i$. For each $i \in \{1,2,\ldots,n\}$, let $L_i \subset a_1^* a_2^* \cdots a_k^*$ be such that $\Phi(L_i) = S_i$. By Lemma 3.12, there exist regular languages $R_1, R_2, \ldots, R_n \subset a_1^* a_2^* \cdots a_k^*$ such that L_i is convex in R_i for each $i \in \{1,2,\ldots,n\}$. For each $I \subset \{1,2,\ldots,n\}$, let $\overline{I} = \{1,2,\ldots,n\} \setminus I$ and

$$R(I) = \left(\bigcap_{i \in I} R_i \right) \cap \left(\bigcap_{i \in \overline{I}} \overline{R_i} \right) \cap a_1^* a_2^* \cdots a_k^*.$$

Since $\bigcup_{I \subset \{1,2,\ldots,n\}} R(I) = a_1^* a_2^* \cdots a_k^*$, we have $L = \bigcup_{I \subset \{1,2,\ldots,n\}} (L \cap R(I))$.

Let $I \subset \{1,2,\ldots,n\}$. If $i \in I$, then by Lemma 3.13, $L_i \cap R(I)$ is semiconvex in $R_i \cap R(I) = R(I)$. If $i \notin I$, then

$$L_i \cap R(I) = Conv(L_i) \cap R_i \cap R(I) = \emptyset.$$

Therefore, for each $I \subset \{1,2,\ldots,n\}$, $L \cap R(I)$ is either empty or semiconvex in $R(I)$.

Finally, it should be obvious that to all $I \subset \{1,2,\ldots,n\}$, we may split the regular language $R(I)$ into regular languages, each of which has a linear Parikh image. □

Let us see an example of the construction of Lemma 3.14.

Example 3.15. Let $L = \{a_1^{2n_1} a_2^{n_2} | n_1 < n_2\} \cup \{a_1^{4n_1} a_2^{n_2} | n_1 > n_2\}$. Then the languages

$$L_1 = L \cap (a_1^2)^* a_2^* \cap (a_1^4)^* a_2^* = \{a_1^{4n_1} a_2^{n_2} | 2n_1 < n_2 \vee n_1 > n_2\}$$

and

$$L_2 = L \cap (a_1^2)^* a_2^* \cap \overline{(a_1^4)^* a_2^*} = \{a_1^{4n_1+2} a_2^{n_2} | 2n_1 + 1 < n_2\}$$

are semiconvex in $(a_1^2)^* a_2^* \cap (a_1^4)^* a_2^* = (a_1^4)^* a_2^*$ and $(a_1^2)^* a_2^* \cap \overline{(a_1^4)^* a_2^*} = a_1^2 (a_1^4)^* a_2^*$, respectively. Moreover, $L = L_1 \cup L_2$.

Lemma 3.16. *Let $k \in \mathbb{N}_+$ and $L \subset a_1^* a_2^* \dots a_k^*$ be a semiconvex language in a language $K \subset a_1^* a_2^* \dots a_k^*$. Then either L or $K \setminus L$ does not contain as a subset any language from the family $\mathcal{R}_k \setminus \mathcal{R}_{k-1}$.*

Proof. This lemma is geometrically obvious since if both L and $K \setminus L$ contained a language from the language family $\mathcal{R}_k \setminus \mathcal{R}_{k-1}$, then these regular languages would overlap infinitely many times. This would break the semiconvexity of L. This is also the idea of the formal proof.

Assume contrariwise that $R_1, R_2 \in \mathcal{R}_k \setminus \mathcal{R}_{k-1}$ are languages for which $R_1 \subset L$ and $R_2 \subset K \setminus L$. Then there exist $n_i, m_i \in \mathbb{N}_+$ and $n_i', m_i' \in \mathbb{N}$ such that

$$a_1^{n_1'} (a_1^{n_1})^* a_2^{n_2'} (a_2^{n_2})^* \cdots a_k^{n_k'} \left(a_k^{n_k}\right)^* \subset R_1 \subset L$$

and

$$a_1^{m_1'} (a_1^{m_1})^* a_2^{m_2'} (a_2^{m_2})^* \cdots a_k^{m_k'} \left(a_k^{m_k}\right)^* \subset R_2 \subset K \setminus L.$$

Without loss of generality, we may assume that $m_i' > n_i'$ for all $i \in \{1, 2, \dots, k\}$. Let $l = \prod_{i=1}^k n_i m_i$. Then

$$a_1^{n_1'} \left(a_1^l\right)^* a_2^{n_2'} \left(a_2^l\right)^* \cdots a_k^{n_k'} \left(a_k^l\right)^* \subset L$$

and

$$a_1^{m_1'} \left(a_1^l\right)^* a_2^{m_2'} \left(a_2^l\right)^* \cdots a_k^{m_k'} \left(a_k^l\right)^* \subset K \setminus L.$$

Define the half-line

$$s(t) = (n_1', n_2', \dots, n_k') + t(m_1' - n_1', m_2' - n_2', \dots, m_k' - n_k'),$$

where $t \in \mathbb{N}_+$. Then $s(jl) \in \Phi(L)$ and $s(jl+1) \in \Phi(K \setminus L) \notin \Phi(L)$ for all $j \in \mathbb{N}$. This contradicts the semiconvexity of $\Phi(L)$ in $\Phi(K)$. $\square$

Lemma 3.17. *Let $k \in \mathbb{N}_+$, $j \in \{1, 2, \dots, k+1\}$ and $L \subset a_1^* a_2^* \cdots a_{k+1}^*$ be a semiconvex language in a regular language $R \subset a_1^* a_2^* \cdots a_{k+1}^*$ such that $\Phi(R)$ is linear and*

$$L \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{L}_{REG}$$

for all $m \in \mathbb{N}$. If for each $m_0 \in \mathbb{N}$, there exists $m \geq m_0$ such that

$$L \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_k \setminus \mathcal{R}_{k-1}, \tag{1}$$

then there exists $m_0' \in \mathbb{N}$ *such that*

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \geq m_0'$.

Proof. Let $p \in \mathbb{N}$ and $L_1, L_2, \ldots, L_p$ be linear components of L such that $L = \bigcup_{i=1}^{p} L_i = \bigcup_{i=1}^{p} (Conv(L_i) \cap R)$. Assume that there exist $m_1, m_2, \ldots, m_{2p+1} \in \mathbb{N}$ such that $m_1 < m_2 < \cdots < m_{2p+1}$,

$$L \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^{m_i} a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_k \setminus \mathcal{R}_{k-1} \tag{2}$$

for all $i \in \{1, 3, 5, \ldots, 2p+1\}$ and

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^{m_i} a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_k \setminus \mathcal{R}_{k-1} \tag{3}$$

for all $i \in \{2, 4, 6, \ldots, 2p\}$. We shall show that this assumption leads to a contradiction. Since by Presumption (1), there exists integers m_i for satisfying Relation (2), Relation (3) must cause the contradiction. This will indicate that there exists $m_0' < m_{2p}$ such that

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \geq m_0'$. Let us now obtain the contradiction.

By Lemma 3.16,

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^{m_i} a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $i \in \{1, 3, 5, \ldots, 2p+1\}$ and

$$L \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^{m_i} a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $i \in \{2, 4, 6, \ldots, 2p\}$. Thus, there exists $n_0 \in \mathbb{N}$ such that

$$\{a_1^{n_1} a_2^{n_2} \cdots a_{j-1}^{n_{j-1}} a_j^{m_i} a_{j+1}^{n_{j+1}} a_{j+2}^{n_{j+2}} \cdots a_{k+1}^{n_{k+1}} \in R | n_1, n_2, \ldots, n_{k+1} \geq n_0\} \subset L$$

for all $i \in \{1, 3, 5, \ldots, 2p+1\}$ and

$$\{a_1^{n_1} a_2^{n_2} \cdots a_{j-1}^{n_{j-1}} a_j^{m_i} a_{j+1}^{n_{j+1}} a_{j+2}^{n_{j+2}} \cdots a_{k+1}^{n_{k+1}} \in R | n_1, n_2, \ldots, n_{k+1} \geq n_0\} \subset R \setminus L$$

for all $i \in \{2, 4, 6, \ldots, 2p\}$. By Relations (2) and (3) and by the linearity of $\Phi(R)$, there exist $n_1, n_2, \ldots, n_{k+1} \geq n_0$ such that

$$a_1^{n_1} a_2^{n_2} \cdots a_{j-1}^{n_{j-1}} a_j^{m_i} a_{j+1}^{n_{j+1}} a_{j+2}^{n_{j+2}} \cdots a_{k+1}^{n_{k+1}} \in R$$

for all $i \in \{1,2,\dots,p\}$. Let us now define the line segment

$$x(t) = (n_1, n_2, \dots, n_{j-1}, t, n_{j+1}, n_{j+2}, \dots, n_{k+1}),$$

where $m_1 \le t \le m_{2p+1}$. The line segment $x(t)$ contains two distinct points from the same convex set $\Phi(Conv(L_i))$ such that their intermediate point belongs to $\Phi(R)$ but does not belong to the set $\Phi(Conv(L_i))$. Therefore, we have a contradiction. Figure 3 illustrates this situation. Thus, the Relation (3) does not hold. Therefore, there exists $m_0' < m_{2p}$ such that

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \ge m_0'$. □

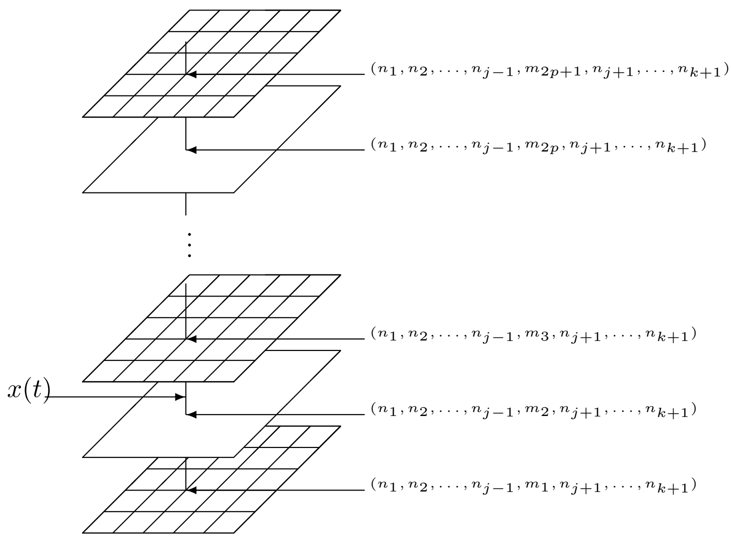

Fig 3. Illustration of the contradiction in the proof of Lemma 3.17. The grid shows points of the set $\Phi(L)$.

Let us generalize the definition of the family $\mathcal{C}_k$ by setting

$$\hat{\mathcal{C}}_k = \{L \in \Sigma^* | L \cap x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k \in \mathcal{L}_{REG} \quad \forall x_i, w_i \in \Sigma_L^*\}$$

for all $k \in \mathbb{N}_+$.

Lemma 3.18. *Let $k \in \mathbb{N}_+$ and $L \subset a_1^* a_2^* \cdots a_{k+1}^*$ be a SLIP-language such that $L \in \hat{\mathcal{C}}_k \setminus \hat{\mathcal{C}}_{k+1}$. Let J be a set of all $j \in \{1,2,\ldots,k+1\}$ for which there exist $m_0 \in \mathbb{N}$ such that*

$$L \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1} \tag{4}$$

for all $m \geq m_0$. Then $\Phi(L)$ contains a period with at least $\min\{k+1, |J|+1\}$ *nonzero coordinates in every representation of $\Phi(L)$ as a finite union of linear sets.*

Proof. Let $L_1, L_2, \ldots, L_n$ be strictly bounded languages such that $\Phi(L_i)$ is a linear set for all $i \in \{1,2,\ldots,n\}$ and $L = \bigcup_{i=1}^n L_i$.

Let $l = |J|$. Without loss of generality, we may assume that $J = \{1,2,\ldots,l\}$. We have two cases: $l < k+1$ and $l = k+1$.

Case 1. We have $l < k+1$. The claim is trivial if $l = 0$. Thus, we may assume that $l \geq 1$. Since $l < k+1$, we have

$$L \cap a_1^* a_2^* \cdots a_l^* a_{l+1}^m a_{l+2}^* a_{l+3}^* \cdots a_{k+1}^* \in \mathcal{R}_k \setminus \mathcal{R}_{k-1} \tag{5}$$

for an arbitrary large number $m \in \mathbb{N}$. Let us denote $J_2 = \{l+2, l+3, \ldots, k+1\}$. By Relation (5), $\Phi(L)$ contains an element e such that

- all the coordinates of J_2 are arbitrary large compared to each coordinate of J in e;
- a coordinate $j \in J$ is arbitrary large compared to a coordinate $j' \in J$ in e, whenever $j > j'$;
- the coordinate $l+1$ is arbitrary large in e (since m is arbitrary large in (5)).

Thus, there exists $r \in \{1,2,\ldots,n\}$ such that

(i) for each $j \in J_2$, there exists a period p_j of $\Phi(L_r)$ such that the coordinate j is nonzero in p_j and all the coordinates of J are zero in p_j;

(ii) for each $j \in J$, there exists a period p_j of $\Phi(L_r)$ such that the coordinate j is nonzero in p_j and all the coordinates of $\{1,2,\ldots,j-1\}$ are zero in p_j;

(iii) there exists a period p_{l+1} of $\Phi(L_r)$ such that the coordinate $l+1$ is nonzero in p_{l+1}.

Let us define the set $J_3 = \{j \in J | \text{the coordinate } j \text{ of } p_{l+1} \text{ is zero}\}$. Assume that the set J_3 is nonempty. Let j' be the smallest element of J_3. By Item (i) of the above list, by pumping periods p_j for all $j \in J_2$, the coordinate j' remains as a constant while all the coordinates of J_2 get arbitrary large values. By Items (ii) and (iii), by pumping the periods p_j for $J_3 \setminus \{j'\}$ and the period p_{l+1}, the coordinate j' remains as a constant while all the coordinates of $J \setminus \{j'\} \cup \{l+1\}$ get arbitrary large values. On the other hand, by Presumption (4), only $k-1$ coordinates of $\Phi(L)$ can have simultaneously arbitrary large values while some coordinate of the set J is fixed. Thus, we have a contradiction. This situation is illustrated in Figure 4. Therefore, the set J_3 is empty. This will indicate that all the coordinates of J are nonzero in p_{l+1}. Thus, p_{l+1} is a period with $l+1$ nonzero coordinates.

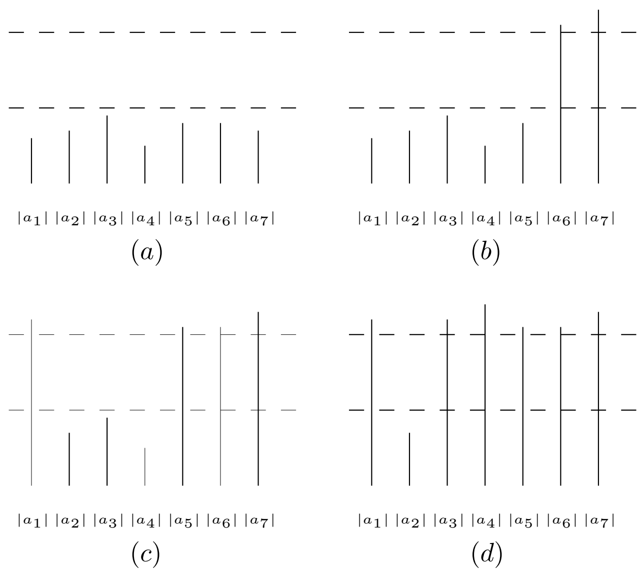

Fig 4. Illustration of the contradiction in the proof of Case 1 of Lemma 3.18. In this example, we have $l=4$, $J=\{1,2,3,4\}$, $J_2=\{6,7\}$ and $J_3=\{2,3,4\}$. Figure (b) shows $\Phi(L_r)$ after pumping the periods p_6, p_7 (the periods p_j for $j \in J_2$). Figure (c) shows $\Phi(L_r)$ after pumping the period p_5 (the period p_{l+1}). Figure (d) shows $\Phi(L_r)$ after pumping the periods p_3, p_4 (the periods p_j for $j \in J_3 \setminus \{j'\}$).

Case 2. We have $l = k+1$. Let $m_0 \in \mathbb{N}$ be a number according to the presumption of the theorem. Therefore

$$L \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1} \tag{6}$$

for all $m \geq m_0$ and $j \in \{1,2,\ldots,k+1\}$.

Let us denote $\mathcal{R}_{-1} = \emptyset$. At this point, it should be noted that since $L \in \hat{\mathcal{C}}_k \setminus \hat{\mathcal{C}}_{k+1}$, we have $L \in \mathcal{L}_{k+1} \setminus \mathcal{L}_k$. Let $J_1 \subset \{1,2,\ldots,k+1\}$ be a set of maximal cardinality such that there exist arbitrary large integers α_j ($j \in \{1,2,\ldots,k+1\} \setminus J_1$) for which

$$L \cap a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_{k+1}^{\alpha_{k+1}} \in \mathcal{R}_{|J_1|} \setminus \mathcal{R}_{|J_1|-1}, \tag{7}$$

where $\alpha_j = *$ for $j \in J_1$. Let $s = |J_1|$. Without loss of generality, we may assume that $J_1 = \{1,2,\ldots,s\}$.

Let us denote $J_2 = \{s+1, s+2, \ldots, k+1\} \setminus J_1$. By Relation (7) (using a similar reasoning that we used in Case 1 for Relation (5)), there exists $r \in \{1,2,\ldots,n\}$ such that

(i) for each $j \in J_1$, there exists a period p_j of $\Phi(L_r)$ such that the coordinate j is nonzero in p_j and all the coordinates of J_2 and $\{1,2,\ldots,j-1\}$ are zero in p_j;

(ii) for each $j \in J_2$, there exists a period p_j of $\Phi(L_r)$ such that the coordinate j is nonzero in p_j.

Assume that there exist $j', j'' \in J_2$ and a period p of $\Phi(L_r)$ such that the coordinate j' of p is zero and the coordinate j'' of p is nonzero. Then keeping the coordinate j' as a constant, the coordinate j'' and all the coordinates of J_1 can increase without limit in $\Phi(L_r)$. This causes a contradiction with the maximality of $|J_1|$. Hence, whenever a coordinate $j \in J_2$ of an arbitrary period p of $\Phi(L_r)$ is nonzero, all the other coordinates of J_2 of the period p are also nonzero.

We show by induction on t that there exists a period p of $\Phi(L_r)$ such that all the coordinates of J_2 and all the coordinates $1,2,\ldots,t$ are nonzero in p for all $t \in \{0,1,\ldots,s\}$. If $t = 0$, then the claim holds by the above conclusion.

Assume now that the induction claim holds for $t = q-1$. Thus, there exists a period p of $\Phi(L_r)$ such that the coordinates $1,2,\ldots,q-1$ and all the coordinates of J_2 are nonzero in p. Let us assume that the coordinate q of p is zero. Then all the coordinates of J_2 and all the coordinates $1,2,\ldots,q-1$ can increase without limit in $\Phi(L_r)$ while the coordinate q stays as a constant. By Item (i) above, the coordinates $q+1, q+2, \ldots, s$ can also increase without limit in $\Phi(L_r)$ while the coordinate q stays as a constant. Thus, keeping the coordinate q as fixed in $\Phi(L_r)$, all the k other coordinates can have arbitrary large values. On the other hand, by (6), only $k-1$ coordinates of $\Phi(L)$ can have simultaneously arbitrary large values while some coordinate is fixed. Thus, the coordinate q of p must also be nonzero. Hence, the induction is extended, and the induction claim holds for $t=s$. Therefore, there exists a period p of $\phi(L_r)$ such that all the $k+1$ coordinates of p are nonzero. $\square$

Let us now give three simple examples of Lemma 3.18. The first one illustrates Case 2 ($l=k+1$) and the second and third one Case 1 ($l<k+1$) of the proof.

Example 3.19. Let $L=\{a_1^n a_2^n a_3^n | n\in\mathbb{N}\}$. We have $L\cap a_1^m a_2^* a_3^* = L\cap a_1^* a_2^m a_3^* = L\cap a_1^* a_2^* a_3^m = \{a_1^m a_2^m a_3^m\}\in\mathcal{R}_0\subset\mathcal{R}_1$ for all $m\in\mathbb{N}$. Therefore, we have $L\in\hat{\mathcal{C}}_2\setminus\hat{\mathcal{C}}_3$ and $l=3$. Lemma 3.18 states that $\Phi(L)$ contains a period with $\min\{3,4\}=3$ nonzero coordinates in every representation of $\Phi(L)$ as a finite union of linear sets.

Example 3.20. Let $L=\{a_1^{n_1}a_2^{n_2}a_3^{n_3}|n_1<n_2\wedge n_1<n_3\in\mathbb{N}\}$. We have $L\cap a_1^m a_2^* a_3^*\in\mathcal{R}_2\setminus\mathcal{R}_1$ and $L\cap a_1^* a_2^m a_3^*, L\cap a_1^* a_2^* a_3^m\in\mathcal{R}_1$ for all $m\in\mathbb{N}$. Therefore, we have $L\in\hat{\mathcal{C}}_2\setminus\hat{\mathcal{C}}_3$ and $l=2$. Lemma 3.18 states that $\Phi(L)$ contains a period with $\min\{3,3\}=3$ nonzero coordinates in every representation of $\Phi(L)$ as a finite union of linear sets.

Example 3.21. Let

$$L=\{a_1^{n_1}a_2^{n_2}a_3^{n_3}a_4^{n_4} \quad | \quad (n_1<n_2\wedge n_1<n_3\wedge n_1<n_4) \\ \vee(n_4<n_1\wedge n_4<n_2\wedge n_4<n_3)\}.$$

It can be easily seen that $L\cap a_1^m a_2^* a_3^* a_4^*, L\cap a_1^* a_2^* a_3^* a_4^m\in\mathcal{R}_3\setminus\mathcal{R}_2$ and $L\cap a_1^* a_2^m a_3^* a_4^*, L\cap a_1^* a_2^* a_3^m a_4^*\in\mathcal{R}_2\setminus\mathcal{R}_1$ for all $m\in\mathbb{N}$. Therefore, $L\in\hat{\mathcal{C}}_3\setminus\hat{\mathcal{C}}_4$ and $l=2$. Lemma 3.18 states that $\Phi(L)$ contains a period with at least $\min\{3,5\}=3$ nonzero coordinates in every representation of $\Phi(L)$ as a finite union of linear sets. It is quite obvious that $\Phi(L)$ contains a period with four nonzero coordinates in every representation of $\Phi(L)$ as a finite union of linear sets. Hence, the statement of the lemma is not optimal in that sense. It is also obvious that $\Phi(L)$ has representations such that $\Phi(L)$ does not contain a period with exactly three nonzero coordinates. Thus, the expression "at least" is necessary in the lemma.

In all three examples above, in every representation of a semilinear set $\Phi(L)$, the set $\Phi(L)$ has contained a period with $k+1$ nonzero coordinates. On the other hand, since our earlier examples have concerned context-free languages, semilinear set $\Phi(L)$ has had a representation such that each period of $\Phi(L)$ does not contain more than two nonzero coordinates. One interesting issue is whether there are any strictly bounded SLIP-languages in the family $\hat{\mathcal{C}}_k \setminus \hat{\mathcal{C}}_{k+1}$ between these two cases.

Open Problem 3.22. *Let $k \geq 3$. Does there exist a $(k+1)$- strictly bounded SLIP-language $L \in \hat{\mathcal{C}}_k \setminus \hat{\mathcal{C}}_{k+1}$ such that*

$$2 < \min\max\{n|\Phi(L) \textit{has a period with } n \textit{ nonzero coordinates}\} < k+1,$$

where min *has been taken over all the representations of $\Phi(L)$?*

The next theorem states that a strictly bounded language $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ that is semiconvex in a regular language $R \subset a_1^* a_2^* \cdots a_{k+1}^*$ is always big, i.e. it contains almost the whole language R. Since by Theorem 3.9 and Lemma 3.14, every language $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ can be transformed to a language of this form with the full trio operations, this theorem gives us important knowledge of the structure of weak languages in the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$. The theorem and its variation (Theorem 3.27) play a crucial role in this thesis; they are used in Section 3.5 for proving that $\mathcal{C}_k$ is closed under morphism and in Section 3.4 for proving that we can transform any language from the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$ into the family $\mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$ with the full trio operations.

Theorem 3.23. *Let $k \geq 2$ be an integer and $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ a semiconvex language in a regular language $R \subset a_1^* a_2^* \cdots a_{k+1}^*$ such that $\Phi(R)$ is linear. Then there exist $m_0 \in \mathbb{N}$ and $j_0 \in \{1, 2, \ldots, k+1\}$ such that*

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \geq m_0$ and $j \in \{1, 2, \ldots, k+1\} \setminus \{j_0\}$.

Proof. Note that since $R \subset a_1^* a_2^* \cdots a_{k+1}^*$ and L is semiconvex in R, we have also $L \subset a_1^* a_2^* \cdots a_{k+1}^*$.

Since $L \in \mathcal{C}_k$, we have $R \setminus L \in \widehat{\mathcal{C}}_k$. Thus

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{L}_{REG}$$

for all $m \in \mathbb{N}$ and $j \in \{1, 2, \ldots, k+1\}$. Let us assume that there exist $j_1, j_2 \in \{1, 2, \ldots, k+1\}$ and arbitrary large integers m_1, m_2 such that

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j_1-1}^* a_{j_1}^{m_1} a_{j_1+1}^* a_{j_1+2}^* \cdots a_{k+1}^* \in \mathcal{R}_k \setminus \mathcal{R}_{k-1}$$

and

$$(R \setminus L) \cap a_1^* a_2^* \cdots a_{j_2-1}^* a_{j_2}^{m_2} a_{j_2+1}^* a_{j_2+2}^* \cdots a_{k+1}^* \in \mathcal{R}_k \setminus \mathcal{R}_{k-1}.$$

Since L is semiconvex in R, the language $R \setminus L$ is also semiconvex in R. Thus, by Lemma 3.17, there exists $m_0 \in \mathbb{N}$ such that

$$L \cap a_1^* a_2^* \cdots a_{j_i-1}^* a_{j_i}^{m} a_{j_i+1}^* a_{j_i+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \geq m_0$ and $i \in \{1,2\}$. If $j_1 \neq j_2$, then by Lemma 3.18, in all representations of $\Phi(L)$ there exists a period p of $\Phi(L)$ such that p contains at least three nonzero coordinates. This causes a contradiction with the context-freeness of L. Thus, $j_1 = j_2$, which concludes the proof. □

Corollary 3.24. *Let $k \geq 2$ be an integer and $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ a semiconvex language in a regular language $R \subset a_1^* a_2^* \cdots a_{k+1}^*$ such that $\Phi(R)$ is linear. Then $R \setminus L$ does not contain any language as a subset from the family $\mathcal{R}_{k+1} \setminus \mathcal{R}_k$.*

Proof. The corollary follows directly from Theorem 3.23. □

Let us now consider two examples of Theorem 3.23.

Example 3.25. Let us consider the language

$$S_1(\neq,\neq,\neq) = \{a_1^{n_1} a_2^{n_2} a_3^{n_3} | n_1 \neq n_2 \vee n_1 \neq n_3\}.$$

We have $a_1^* a_2^* a_3^* \setminus S_1(\neq,\neq,\neq) = \{a_1^{n_1} a_2^{n_2} a_3^{n_3} | n_1 = n_2 = n_3\}$. Thus, all the languages $(a_1^* a_2^* a_3^* \setminus S_1(\neq,\neq,\neq)) \cap a_1^m a_2^* a_3^*$, $(a_1^* a_2^* a_3^* \setminus S_1(\neq,\neq,\neq)) \cap a_1^* a_2^m a_3^*$ and $(a_1^* a_2^* a_3^* \setminus S_1(\neq,\neq,\neq)) \cap a_1^* a_2^* a_3^m$ are singletons, and the index j_0 of Theorem 3.23 may be chosen freely.

Example 3.26. Let us consider the language

$$S_1(\neq,>,>) = \{a_1^{n_1} a_2^{n_2} a_3^{n_3} | n_1 > n_2 \vee n_1 > n_3\}.$$

We have $a_1^* a_2^* a_3^* \setminus S_1(\neq,>,>) = \{a_1^{n_1} a_2^{n_2} a_3^{n_3} | n_1 \leq n_2 \wedge n_1 \leq n_3\}$. Thus, in this case $(a_1^* a_2^* a_3^* \setminus S_1(\neq,>,>)) \cap a_1^m a_2^* a_3^* = a_1^m a_2^m a_2^* a_3^m a_3^* \in \mathcal{R}_2 \setminus \mathcal{R}_1$ and $(a_1^* a_2^* a_3^* \setminus S_1(\neq,>,>)) \cap a_1^* a_2^m a_3^*, (a_1^* a_2^* a_3^* \setminus S_1(\neq,>,>)) \cap a_1^* a_2^* a_3^m \in \mathcal{R}_1$. Hence, the index j_0 of Theorem 3.23 is one.

Next, we state another variation of Theorem 3.23.

Theorem 3.27. *Let $k \geq 2$ be an integer and $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ such that $L \subset a_1^* a_2^* \cdots a_{k+1}^*$. Then there exist $j_0 \in \{1, 2, \ldots, k+1\}$ and $R \in \mathcal{L}_{REG}$ such that $L \cup R \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ and*

$$\overline{L \cup R} \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \in \mathbb{N}$ and $j \in \{1, 2, \ldots, k+1\} \setminus \{j_0\}$.

Proof. By Lemma 3.14, there exists a regular language $R \subset a_1^* a_2^* \cdots a_{k+1}^*$ that has a linear Parikh image such that $L' = L \cap R \notin \mathcal{L}_{REG}$ and L' is semiconvex in R. Thus, $L' \notin \mathcal{C}_{k+1}$. Since $L \in \mathcal{C}_k$, we have also $L' \in \mathcal{C}_k$.

By Theorem 3.23, there exist $m_0 \in \mathbb{N}$ and $j_0 \in \{1, 2, \ldots, k\}$ such that

$$\overline{L'} \cap R \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \geq m_0$ and $j \in \{1, 2, \ldots, k+1\} \setminus \{j_0\}$. Since $\overline{L'} \cap R = \overline{L' \cup \overline{R}} = \overline{L \cup \overline{R}}$, we have

$$\overline{L \cup \overline{R}} \cap a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^* \in \mathcal{R}_{k-1}$$

for all $m \geq m_0$ and $j \in \{1, 2, \ldots, k+1\} \setminus \{j_0\}$.

Finally, let

$$R' = \bigcup_{j=1}^{k+1} \bigcup_{m=0}^{m_0 - 1} a_1^* a_2^* \cdots a_{j-1}^* a_j^m a_{j+1}^* a_{j+2}^* \cdots a_{k+1}^*.$$

Since $L' \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$, $L' \cap \overline{R} = \emptyset$ and $R' \in \mathcal{R}_k$, we have $L \cup \overline{R} \cup R' = L' \cup \overline{R} \cup R' \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$. Hence, the regular language $\overline{R} \cup R'$ satisfies the required conditions. □

Theorem 3.28. *Let $k \in \mathbb{N}_+$ and $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ be such that $L \subset a_1^* a_2^* \cdots a_{k+1}^*$. Then $\Phi(a_1^* a_2^* \cdots a_{k+1}^* \setminus L)$ contains a period with $k+1$ nonzero coordinates in every representation of $\Phi(a_1^* a_2^* \cdots a_{k+1}^* \setminus L)$ as a finite union of linear sets.*

Proof. If $k = 1$, then $a_1^* a_2^* \setminus L$ is not regular, and the claim is trivial. So we may assume that $k \geq 2$.

By Lemma 3.14, there exists a regular language R that has a linear Parikh image such that $L' = L \cap R \notin \mathcal{L}_{REG}$ and L' is semiconvex in R. By Theorem 3.23 and Lemma 3.18, there exists a period of $\Phi(R \setminus L')$ with $k+1$ nonzero coordinates in all representations of $\Phi(R \setminus L')$. Since $R \setminus L' = \overline{L'} \cap R = \overline{L} \cap R = (a_1^* a_2^* \cdots a_{k+1}^* \setminus L) \cap R$, also $\Phi(a_1^* a_2^* \cdots a_{k+1}^* \setminus L)$ contains a period with $k+1$ nonzero coordinates in all of its representations. □

Corollary 3.29. *Let $k \geq 2$ be an integer and $L \in \mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$. Then $\overline{L}$ is not context-free.*

Let $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ be a $(k+1)$- strictly bounded semiconvex language in a regular language R. In Corollary 3.24, we saw that $R \setminus L$ does not contain any language as a subset from the family $\mathcal{R}_{k+1} \setminus \mathcal{R}_k$. However, the next theorem states that $R \setminus L$ always contains a language from the family $\mathcal{L}_{k+1} \setminus \mathcal{L}_k$.

Theorem 3.30. *Let $k \in \mathbb{N}_+$. If $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ is a semiconvex language in a regular language $R \subset a_1^* a_2^* \cdots a_{k+1}^*$, then $R \setminus L \in \mathcal{L}_{k+1} \setminus \mathcal{L}_k$.*

Proof. If $R \setminus L \in \mathcal{L}_k$, then there exists $m_0 \in \mathbb{N}$ such that

$$R \setminus L = (R \setminus L) \cap \left(\bigcup_{i=1}^{k+1} \bigcup_{m=0}^{m_0} a_1^* a_2^* \cdots a_{i-1}^* a_i^m a_{i+1}^* a_{i+2}^* \cdots a_{k+1}^* \right). \tag{8}$$

Since $L \in \mathcal{C}_k$, we have $R \setminus L \in \widehat{\mathcal{C}}_k$. Thus, by Equation (8), we have $R \setminus L \in \mathcal{L}_{REG}$. Then, also $L \in \mathcal{L}_{REG}$, which is a contradiction. □

Thus far, all our example languages $S_i(\theta_1, \theta_2, \ldots, \theta_{k+1})$ have had a property such that $\Phi(S_i(\theta_1, \theta_2, \ldots, \theta_{k+1}))$ has a representation such that each linear component of $\Phi(S_i(\theta_1, \theta_2, \ldots, \theta_{k+1}))$ contains only one period with two nonzero coordinates. One could ask whether this is a global property for all the $(k+1)$-strictly bounded languages in the family $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$. However, this is not the case, as seen in the next example.

Example 3.31. Consider the language

$$L = \{a_1^{n_1} a_2^{n_2} a_3^{n_3} | n_1 > 2n_2 \vee n_1 > 2n_3 \vee n_1 = n_2 + n_3\}.$$

Then $L \in \mathcal{C}_2 \setminus \mathcal{C}_3$. Moreover, L contains a linear component with periods $(1,1,0)$ and $(1,0,1)$ in the most natural representation of $\Phi(L)$. It seems also quite obvious that there exists a linear component with periods $(n_1, n_2, 0)$ and $(n_3, 0, n_4)$, where $n_i > 0$, in every representation of $\Phi(L)$.

Let us define the *commutative closure* of a language L by

$$c(L) = \Phi^{-1}(\Phi(L)).$$

Until now, all our examples of languages in the family $\mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$ have been symmetric in the sense that, if $L \in \mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$, then

$$c(L) \cap a_{\pi(1)} a_{\pi(2)} \cdots a_{\pi(k+1)} \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$$

for all the permutations $\pi : \{1,2,\ldots,k+1\} \to \{1,2,\ldots,k+1\}$. Obviously, the only possibility that

$$c(L) \cap a_{\pi(1)}a_{\pi(2)}\cdots a_{\pi(k+1)} \notin \mathcal{C}_k \setminus \mathcal{C}_{k+1}$$

would be that

$$c(L) \cap a_{\pi(1)}a_{\pi(2)}\cdots a_{\pi(k+1)} \notin \mathcal{L}_{CF}.$$

Let us ask whether this is a general property of the family $\mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$.

Open Problem 3.32. *Let $k \in \mathbb{N}_+$ and $L \in \mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$. Do we have*

$$c(L) \cap a_{\pi(1)}a_{\pi(2)}\cdots a_{\pi(k+1)} \in \mathcal{L}_{CF}$$

for every permutation $\pi : \{1,2,\ldots,k+1\} \to \{1,2,\ldots,k+1\}$?

Let us assume that there exist a language $L \in \mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$ and a permutation $\pi : \{1,2,\ldots,k+1\} \to \{1,2,\ldots,k+1\}$ such that

$$c(L) \cap a_{\pi(1)}a_{\pi(2)}\cdots a_{\pi(k+1)} \notin \mathcal{L}_{CF}.$$

Recalling the definition of a stratified linear set, this implies that in some linear component of L there are two periods that have two different nonzero coordinates. It seems quite natural that then we can intersect the language L by a regular language R so that exponents of two letters become constants eliminating one of these periods and leaving nonregularities of the other period in the resulting language. Then we have a contradiction with the assumption $L \in \mathcal{C}_k$. The next example demonstrates this issue.

Example 3.33. Consider the language

$$L = \{a_1^{n_1}a_2^{n_2}a_3^{n_3}a_4^{n_4} | n_1 > n_2 \vee n_1 > n_3 \vee n_1 > n_4 \vee (2n_1 = n_4 \wedge n_2 = n_3)\}.$$

Then

$$L \cap a_1a_2^*a_3^*a_4^2 = a_1a_3^*a_4^2 \cup a_1a_2^*a_4^2 \cup \{a_1a_2^na_3^na_4^2 | n \in \mathbb{N}\}.$$

Thus, $L \in \mathcal{C}_1 \setminus \mathcal{C}_2$.

We call a language L *commutative* if $L = c(L)$. Latteux has conjectured that a commutative language L is context-free if and only if for all $n \in \mathbb{N}$ and $y_1, y_2, \ldots, y_n \in \Sigma^*$, the language $L \cap y_1^*y_2^*\cdots y_n^*$ is context-free (18). If this conjecture holds, then to solve Open Problem 3.32, it would be enough to prove that $c(L)$ is context-free for every $L \in \mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$.

3.4 Propagation in the Chain

In this section, we will prove in Theorem 3.36 that for each $k \in \mathbb{N}_+$ and $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$, there exists a language $L' \in \mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$ such that L' belongs to the full trio generated by the language L. At the end of this section, we will prove that, for full trios and full AFLs, the minimality with respect to the $(k+1)$-bounded context-free languages is equivalent to the minimality with respect to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$.

Let us sketch the idea of the proof of Theorem 3.36 with a few examples.

Example 3.34. Since

$$\begin{aligned} S_1(\neq,\neq,\neq) &= \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_1 \neq n_2\} \\ &\quad \cup \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_1 \neq n_3\}, \end{aligned}$$

we have $S_1(\neq,\neq,\neq) \in \mathcal{T}(S_1(\neq,\neq))$. Correspondingly,

$$\begin{aligned} S_1(\neq,>,>) &= \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_1 > n_2\} \\ &\quad \cup \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_1 > n_3\} \end{aligned}$$

and

$$\begin{aligned} S_3(>,>,\neq) &= \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_3 > n_1\} \\ &\quad \cup \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_3 > n_2\}. \end{aligned}$$

Thus, we have $S_1(\neq,>,>) \in \mathcal{T}(S_1(\neq,>))$ and $S_3(>,>,\neq) \in \mathcal{T}(S_1(\neq,<))$. By Theorem 3.7, $S_1(\neq,\neq,\neq), S_1(\neq,>,>), S_3(>,>,\neq) \in \mathcal{C}_2 \setminus \mathcal{C}_3$.

Example 3.34 showed us a hint how a language $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ could be transformed into the family $\mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$. We introduce a new letter whose exponent can have arbitrary values. Then we take the union of that language and a language, where the roles of the new letter and some original letter have been changed. We call this trick *copying.* Considering transducers, it is obvious that this kind of copying may be realized with the full trio operations. Unfortunately, we cannot select the copied letter completely freely, as seen in the next example.

Example 3.35. We have

$$\begin{aligned} S_1(\neq,<,<) &= \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_1 < n_2\} \\ &\quad \cup \{a_1^{n_1}a_2^{n_2}a_3^{n_3} | n_1,n_2,n_3 \in \mathbb{N}, n_1 < n_3\}. \end{aligned}$$

But $S_1(\neq,<,<)\cap a_1^*a_2^* = S_1(\neq,<) \in \mathcal{C}_1 \setminus \mathcal{C}_2$.

Example 3.35 showed that we cannot copy whatever letter. In the next theorem, it is shown that Theorem 3.27 (or equivalently Theorem 3.23) defines those letters that can be copied. Thus, there exists at most one letter that cannot be copied.

Theorem 3.36. *Let $k \in \mathbb{N}_+$. For each language $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$, there exists a $(k+2)$- strictly bounded language $L' \in \mathcal{T}(L)$ such that $L' \in \mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$.*

Proof. Let us assume first that $k=1$. By Theorem 3.10, $S_1(\neq,\neq) \in \mathcal{T}(L)$, $S_1(\neq,>) \in \mathcal{T}(L)$ or $S_1(\neq,<) \in \mathcal{T}(L)$. Thus, by Example 3.34, $S_1(\neq,\neq,\neq) \in \mathcal{T}(L)$, $S_1(\neq,>,>) \in \mathcal{T}(L)$ or $S_3(>,>,\neq) \in \mathcal{T}(L)$. In each case, we have a 3- strictly bounded language $L' \in \mathcal{T}(L) \cap \mathcal{C}_2 \setminus \mathcal{C}_3$. Therefore, we may assume that $k \geq 2$.

By Theorem 3.9, we may assume that $L \subset a_1^*a_2^*\cdots a_{k+1}^*$. Let $j_0 \in \{1,2,\ldots,k+1\}$ and R be a number and a regular language according to Theorem 3.27. In addition, let $j \in \{1,2,\ldots,k+1\}\setminus\{j_0\}$. Without loss of generality, we may assume that $j=1$. Let us denote $L' = L \cup R$,

$$\begin{aligned} L_1' &= \{a_0^n a_1^{n_1} a_2^{n_2}\cdots a_{k+1}^{n_{k+1}} \mid n \in \mathbb{N}, a_1^{n_1}a_2^{n_2}\cdots a_{k+1}^{n_{k+1}} \in L'\} \qquad \text{and} \\ L_1'' &= \{a_0^{n_1} a_1^{n} a_2^{n_2} a_3^{n_3}\cdots a_{k+1}^{n_{k+1}} \mid n \in \mathbb{N}, a_1^{n_1}a_2^{n_2}\cdots a_{k+1}^{n_{k+1}} \in L'\}, \end{aligned}$$

where $a_0 \notin \Sigma_{L'}$ is a new letter. Denote also $L_1 = L_1' \cup L_1''$. Since L_1 is obtained from the language L' by copying the letter a_1, it should be clear that $L_1 \in \mathcal{T}(L')$. Moreover, since $L' \notin \mathcal{L}_{REG}$, we have $L_1 \notin \mathcal{L}_{REG}$, and thus, $L_1 \notin \mathcal{C}_{k+2}$. Thus, it suffices to show that $L_1 \in \mathcal{C}_{k+1}$. By Theorem 3.6, we have $L_1 \in \mathcal{C}_{k+1}$ if and only if

$$L_1 \cap a_0^*a_1^*\cdots a_{i-1}^* a_i^m a_{i+1}^* a_{i+2}^* \cdots a_{k+1}^* \in \mathcal{L}_{REG}$$

for all $m \in \mathbb{N}$ and $i \in \{0,1,\ldots,k+1\}$. Since $L' \in \mathcal{C}_k$, we have

$$L_1 \cap a_0^*a_1^*\cdots a_{i-1}^* a_i^m a_{i+1}^* a_{i+2}^* \cdots a_{k+1}^* \in \mathcal{L}_{REG}$$

for all $m \in \mathbb{N}$ and $i \in \{2,3,\ldots,k+1\}$. The cases $i=0$ and $i=1$ are symmetrical. Hence, we need only to prove case $i=1$. Let $m \in \mathbb{N}$ be arbitrary. Let us denote

$$\begin{aligned} L_m' &= L_1' \cap a_0^* a_1^m a_2^* a_3^* \cdots a_{k+1}^* \\ &= \{a_0^n a_1^m a_2^{n_2}\cdots a_{k+1}^{n_{k+1}} \mid n \in \mathbb{N}, a_1^m a_2^{n_2}\cdots a_{k+1}^{n_{k+1}} \in L'\} \end{aligned}$$

and

$$\begin{aligned} L''_m &= L''_1 \cap a_0^* a_1^m a_2^* a_3^* \cdots a_{k+1}^* \\ &= \{a_0^{n_1} a_1^m a_2^{n_2} a_3^{n_3} \cdots a_{k+1}^{n_{k+1}} \mid a_1^{n_1} a_2^{n_2} a_3^{n_3} \cdots a_{k+1}^{n_{k+1}} \in L'\}. \end{aligned}$$

By Theorem 3.27, $a_0^*(\overline{L'} \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^*) \in \mathcal{R}_k$. Thus, there exists a regular language $R_1 \in \mathcal{R}_k$ such that

$$L'_m = a_0^*(L' \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^*) = a_0^* a_1^m a_2^* a_3^* \cdots a_{k+1}^* \setminus R_1.$$

Hence

$$\begin{aligned} & L_1 \cap a_0^* a_1^m a_2^* a_3^* \cdots a_{k+1}^* \\ = \ & L'_m \cup L''_m \\ = \ & ((L'_m \cup L''_m) \cap R_1) \cup ((L'_m \cup L''_m) \cap \overline{R_1}) \\ = \ & (L'_m \cap R_1) \cup (L''_m \cap R_1) \cup (a_0^* a_1^m a_2^* a_3^* \cdots a_{k+1}^* \setminus R_1) \end{aligned}$$

Since $L' \in \mathcal{C}_k$, we have $L'_m, L''_m \in \mathcal{C}_k$. Thus, $L'_m \cap R_1, L''_m \cap R_1 \in \mathcal{L}_{REG}$ and further $L_1 \cap a_0^* a_1^m a_2^* a_3^* \cdots a_{k+1}^* \in \mathcal{L}_{REG}$. Therefore, $L_1 \in \mathcal{C}_{k+1}$. □

We used Theorem 3.27 in the proof of Theorem 3.36. However, it should be remarked that we could have used as well Theorem 3.23 with Lemma 3.14.

Theorem 3.37. *A full trio (resp. full AFL) $\mathcal{L}$ is minimal with respect to the $(k+1)$-bounded context-free languages if and only if it is minimal with respect to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$.*

Proof. Let us prove the claim concerning full trios. The claim concerning full AFLs follows by the same reasoning, observing that, for each language L, we have $\mathcal{T}(L) \subset \mathcal{F}(L)$.

Let the full trio $\mathcal{L}$ be minimal with respect to the $(k+1)$-bounded context-free languages. By Theorem 2.12, it suffices to show that $\mathcal{L} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$ is nonempty and for each $L \in \mathcal{L} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$, we have $\mathcal{T}(L) = \mathcal{L}$.

Let $L \in \mathcal{L}$ be a nonregular $(k+1)$-bounded context-free language. Then $L \notin \mathcal{C}_{k+1}$. Therefore, there exists $l \leq k$ such that $L \in \mathcal{C}_l \setminus \mathcal{C}_{l+1}$. By Theorem 3.36, there exists $L' \in \mathcal{T}(L)$ such that $L' \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$. Thus, $\mathcal{L} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$ is nonempty.

Let $L \in \mathcal{L} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$. Then by Theorem 3.9, there exists a $(k+1)$-bounded nonregular context-free language $L' \in \mathcal{T}(L)$. Since $\mathcal{L}$ is minimal with respect to the $(k+1)$-bounded context-free languages, we have $\mathcal{L} = \mathcal{T}(L') \subset \mathcal{T}(L)$. Thus, $\mathcal{T}(L) = \mathcal{L}$, and $\mathcal{L}$ is minimal with respect to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$.

Let the full trio $\mathcal{L}$ be minimal with respect to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$. By Theorem 2.12, it suffices to show that $\mathcal{L}$ contains nonregular $(k+1)$-bounded context-free languages and for each $(k+1)$-bounded nonregular context-free language $L \in \mathcal{L}$, we have $\mathcal{T}(L) = \mathcal{L}$.

Let $L \in \mathcal{L} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$. Then by Theorem 3.9, there exists a $(k+1)$-bounded nonregular context-free language $L' \in \mathcal{T}(L) \subset \mathcal{L}$.

Let $L \in \mathcal{L}$ be a nonregular $(k+1)$-bounded context-free language. Then $L \notin \mathcal{C}_{k+1}$. Therefore, there exists $l \leq k$ such that $L \in \mathcal{C}_l \setminus \mathcal{C}_{l+1}$. By Theorem 3.36, there exists $L' \in \mathcal{T}(L)$ such that $L' \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$. Since $\mathcal{L}$ is minimal with respect to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$, we have $\mathcal{L} = \mathcal{T}(L') \subset \mathcal{T}(L)$. Thus, $\mathcal{T}(L) = \mathcal{L}$, and $\mathcal{L}$ is minimal with respect to the $(k+1)$-bounded context-free languages. □

3.5 Closure Properties of the Family $\mathcal{C}_k$

In this section, we will prove that the family $\mathcal{C}_k$ is a substitution closed full AFL for each $k \in \mathbb{N}_+$. This will be done in three phases. First, we will prove that $\mathcal{C}_k$ is closed under intersection with regular languages, union and inverse morphism (Theorem 3.38). Closure under morphism is the hardest part of the proof, and it is proved in a separate theorem (Theorem 3.43). Then we shall prove closure under substitution (Theorem 3.44). Closure under catenation and catenation closure follow from the fact that each substitution closed full trio is a full AFL. Finally, as a corollary of these closure properties and Theorem 3.36, we will prove that Conjecture 3.1 holds (Theorem 3.45).

Theorem 3.38. *Let $k \in \mathbb{N}_+$. The families $\mathcal{C}_k$ and $\mathcal{C}_\infty$ are closed under intersection with regular languages, union and inverse morphism.*

Proof. Let us prove the claim concerning the families $\mathcal{C}_k$, where $k \in \mathbb{N}_+$. Therefore, the claim concerning the family $\mathcal{C}_\infty$ follows from the fact that $\mathcal{C}_\infty$ is the intersection of the families $\mathcal{C}_k$.

Let $L_1, L_2 \in \mathcal{C}_k$, $x_0, x_1, \ldots, x_k \in \Sigma^*$ and $w_1, w_2, \ldots, w_k \in \Sigma^*$ be arbitrary. Denote $R = x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k$. Hence, $L_1 \cap R \in \mathcal{L}_{REG}$ and $L_2 \cap R \in \mathcal{L}_{REG}$.

Let us first prove closure under intersection with regular languages. Let R_1 be a regular language. Since $(L \cap R_1) \cap R = (L \cap R) \cap R_1 \in \mathcal{L}_{REG}$, we have

$L \cap R_1 \in \mathcal{C}_k$.

Next, we will prove closure under union. We have

$$(L_1 \cup L_2) \cap R = (L_1 \cap R) \cup (L_2 \cap R).$$

Since $L_1 \cap R$ and $L_2 \cap R$ are regular, the language $(L_1 \cup L_2) \cap R$ is also regular. Therefore, $L_1 \cup L_2 \in \mathcal{C}_k$.

Finally, let us consider closure under inverse morphism. Let $L \in \mathcal{C}_k$ and $h : \Sigma^* \to \Sigma_L^*$ be a morphism. Since $R \subset h^{-1}(h(R))$, we have

$$\begin{aligned} & h^{-1}(L) \cap R \\ = \; & h^{-1}(L) \cap h^{-1}(h(R)) \cap R \\ = \; & h^{-1}(L \cap h(R)) \cap R. \end{aligned}$$

Furthermore, since $L \cap h(R) \in \mathcal{L}_{REG}$, we have $h^{-1}(L) \cap R \in \mathcal{L}_{REG}$. □

Note that we used the assumption $L \in \mathcal{L}_{CF}$ only for ensuring that $\mathcal{L}_{CF}$ is closed under intersection with regular languages, union and inverse morphism. Thus, Theorem 3.38 holds also for a greater family

$$\hat{\mathcal{C}}_k = \{L \in \Sigma^* | L \cap x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k \in \mathcal{L}_{REG} \quad \forall x_i, w_i \in \Sigma_L^*\}.$$

It is easily seen that the family $\hat{\mathcal{C}}_k$ is even closed under intersection. However, a context-free assumption is necessary for closure under morphism, as seen in the example below.

Example 3.39. Let $L = \{a_1^n a_2^n a_3^n | n \in \mathbb{N}\}$ and $h : \{a_1, a_2, a_3\}^* \to \{a_1, a_2\}^*$ be the projection. Obviously, $L \in \hat{\mathcal{C}}_2$ but $h(L) = \{a_1^n a_2^n | n \in \mathbb{N}\} \notin \hat{\mathcal{C}}_2$.

Before proving closure under morphism, we need two auxiliary results for noting that it suffices to prove closure under morphism for strictly bounded languages. After these results, we still have to prove one technical auxiliary lemma.

Theorem 3.40. *(7, Theorem 1, Theorem 2) For each context-free language $L \subset \Sigma^*$, there exists a bounded context-free language $L' \subset \Sigma^*$ such that $\Phi(L') = \Phi(L)$ and $L' \subset L$.*

Lemma 3.41. *Let $k \in \mathbb{N}_+$ and $L \subset \Sigma^*$ be a language such that $L \in \mathcal{C}_k$. Let $h_1 : \Sigma^* \to \Delta^*$ be a morphism such that $h_1(L)$ is strictly bounded. Then there* *exist a strictly bounded language $L' \in \mathcal{C}_k$ and a morphism $h_2 : \Sigma^* \to \Delta^*$ such that $h_2(L') = h_1(L)$.*

Proof. By Theorem 3.40, there exists a bounded context-free language $L_1 \subset \Sigma^*$ such that $\Phi(L_1) = \Phi(L)$ and $L_1 \subset L$. Let $n \in \mathbb{N}$ and $v_1, v_2, \ldots, v_n \in \Sigma^*$ be words such that $L_1 \subset v_1^* v_2^* \cdots v_n^*$. Set $L_2 = L \cap v_1^* v_2^* \cdots v_n^*$. Then $\Phi(L_1) \subset \Phi(L_2) \subset \Phi(L)$. Thus, $\Phi(L_2) = \Phi(L)$. Further $\Phi(h_1(L_2)) = \Phi(h_1(L))$ and since $h_1(L)$ is strictly bounded, we have $h_1(L_2) = h_1(L)$. Define the language $L_2' \subset \Sigma^*$ by

$$L_2' = \{a_1^{m_1} a_2^{m_2} \cdots a_n^{m_n} | v_1^{m_1} v_2^{m_2} \cdots v_n^{m_n} \in L_2\}.$$

Since $L \in \mathcal{C}_k$, we have $L_2 \in \mathcal{C}_k$, and moreover by Lemma 3.8, $L_2' \in \mathcal{C}_k$. Finally, define the morphism $h_2 : \Sigma^* \to \Delta^*$ by $h_2(a_i) = h_1(v_i)$ for all $i \in \{1, 2, \ldots, n\}$. Then $h_2(L_2') = h_1(L_2) = h_1(L)$. □

Lemma 3.42. *Let $L \subset \Delta^*$, $L' \subset \Sigma^*$ be languages and $h : \Sigma^* \to \Delta^*$ a morphism. Then $h(h^{-1}(\overline{L}) \cap L') = h(\overline{h^{-1}(L)} \cap L')$.*

Proof. We have

$$\begin{aligned} h(h^{-1}(\overline{L}) \cap L') &= \{h(w) | w \in h^{-1}(\overline{L}) \cap L'\} \\ &= \{h(w) | h(w) \in \overline{L}, w \in L'\} \\ &= \{h(w) | w \in \overline{h^{-1}(L)} \cap L'\} \\ &= h(\overline{h^{-1}(L)} \cap L'). \end{aligned}$$

□

Theorem 3.43. *Let $k \in \mathbb{N}_+$. The families $\mathcal{C}_k$ and $\mathcal{C}_\infty$ are closed under morphism.*

Proof. Let us prove the claim concerning the families $\mathcal{C}_k$, where $k \in \mathbb{N}_+$. Therefore, the claim concerning the family $\mathcal{C}_\infty$ follows from the fact that $\mathcal{C}_\infty$ is the intersection of the families $\mathcal{C}_k$.

If $k = 1$, then $\mathcal{C}_k = \mathcal{L}_{CF}$, and the claim is trivial. So in the rest of the proof we may assume that $k \geq 2$.

Let $h : \Sigma^* \to \Delta^*$ be a morphism and $L \in \mathcal{C}_k$. Let us assume that $h(L) \notin \mathcal{C}_k$. We shall first make some simplifications.

By Theorem 3.36, there exists a language $L' \in \mathcal{T}(h(L))$ such that $L' \in \mathcal{B}_k \cap (\mathcal{C}_{k-1} \setminus \mathcal{C}_k)$. By Theorem 2.5, there exist an alphabet Γ, morphisms $h_1 : \Gamma^* \to \Delta^*$,

$g_1 : \Gamma^* \to \Sigma^*$ and a regular language $R \subset \Gamma^*$ such that $L' = h_1(g_1^{-1}(L) \cap R)$. Since $\mathcal{C}_k$ is closed under inverse morphism and intersection with regular languages, the language $g_1^{-1}(L) \cap R$ is in $\mathcal{C}_k$. Thus, we may assume that in our original problem $L \in \mathcal{C}_k$ and $h(L) \in \mathcal{B}_k \cap (\mathcal{C}_{k-1} \setminus \mathcal{C}_k)$.

By Lemma 3.41, we may assume that L is strictly bounded. Let n be the smallest integer such that $L \subset a_1^* a_2^* \cdots a_n^*$. Let us decompose the morphism h as a composition of multiple morphisms $h = h_{k+1} \circ h_{k+2} \circ \cdots \circ h_n$, where each morphism h_i reduces the size of the alphabet by one. Let us denote $L_n = L$ and $L_i = h_{i+1}(L_{i+1})$ for all $i \in \{k, k+1, \ldots, n-1\}$. Then there exists $p \in \{k+1, k+2, \ldots, n\}$ such that $L_p \in \mathcal{C}_k$ and $L_{p-1} \notin \mathcal{C}_k$. Thus, by Theorem 3.6, there exists $\alpha_1, \alpha_2, \ldots, \alpha_{p-1} \in \mathbb{N} \cup \{*\}$ such that

$$L'_{p-1} = L_{p-1} \cap a_1^{\alpha_1} a_2^{\alpha_2} \cdots a_{p-1}^{\alpha_{p-1}} \notin \mathcal{L}_{REG},$$

where $|\{i | \alpha_i = *\}| = k$ and $|\{i | \alpha_i \in \mathbb{N}\}| = p-1-k$. Note that since only constants may be mapped to constants, the language L'_{p-1} may be expressed as a finite union of languages

$$h_p(L_p \cap a_1^{\beta_1} a_2^{\beta_2} \cdots a_p^{\beta_p}),$$

where $\beta_1, \beta_2, \ldots, \beta_p \in \mathbb{N} \cup \{*\}$ and $|\{i | \beta_i \in \mathbb{N}\}| \geq p-1-k$. Thus $|\{i | \beta_i = *\}| \leq k+1$. Since $L_p \in \mathcal{C}_k$ and L'_{p-1} is nonregular, we have $|\{i | \beta_i = *\}| = k+1$. Ignoring letters whose powers are constants, we get a morphism that maps a language in the family $\mathcal{B}_{k+1} \cap (\mathcal{C}_k \setminus \mathcal{C}_{k+1})$ to a language in the family $\mathcal{B}_k \cap (\mathcal{C}_{k-1} \setminus \mathcal{C}_k)$. Thus, we may assume that in our original problem $L \subset a_1^* a_2^* \cdots a_{k+1}^*$ and $h(L) \subset a_1^* a_2^* \cdots a_k^*$.

Without loss of generality, we may assume that $h(a_1), h(a_2) \in a_1^*$ and $h(a_{i+1}) \in a_i^*$ for all $i \in \{2, 3, \ldots, k\}$. Let l_1, l_2 be two nonnegative integers such that $h(a_1) = a_1^{l_1}$ and $h(a_2) = a_1^{l_2}$. Now we have made needed simplifications, and we are ready to begin the actual proof.

By Lemma 3.14, there exists a regular language $R \subset a_1^* a_2^* \cdots a_{k+1}^*$ that has a linear Parikh image such that $L \cap R$ is semiconvex in R and $h(L \cap R) \in \mathcal{C}_{k-1} \setminus \mathcal{C}_k$. Let us denote $L_R = L \cap R$.

By Theorem 3.23, there exists $m_0 \in \mathbb{N}$ such that $\overline{L_R} \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^* \cap R \in \mathcal{R}_{k-1}$ or $\overline{L_R} \cap a_1^* a_2^m a_3^* a_4^* \cdots a_{k+1}^* \cap R \in \mathcal{R}_{k-1}$ for all $m \geq m_0$. Without loss of generality, we may assume that $\overline{L_R} \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^* \cap R \in \mathcal{R}_{k-1}$ for all $m \geq m_0$. Therefore, $h(\overline{L_R} \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^* \cap R) \in \mathcal{R}_{k-1}$ for all $m \geq m_0$. Thus, we have for each $m \geq m_0$,

$$\begin{aligned}
& \overline{h(L_R)} \cap h(a_1^m a_2^* a_3^* \cdots a_{k+1}^* \cap R) \\
= \quad & h(h^{-1}(\overline{h(L_R)}) \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^* \cap R) \\
\overset{\text{Lemma 3.42}}{=} \quad & h(\overline{h^{-1}(h(L_R))} \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^* \cap R) \\
\subset \quad & h(\overline{L_R} \cap a_1^m a_2^* a_3^* \cdots a_{k+1}^* \cap R) \in \mathcal{R}_{k-1}. \qquad (9)
\end{aligned}$$

However, since

$$h(a_1^{n_1} a_2^{n_2}) = a_1^{l_1 n_1 + l_2 n_2},$$

we have

$$\bigcup_{j=0}^{l_2-1} h(a_1^j a_2^* a_3^* \cdots a_{k+1}^*) = h(a_1^* a_2^* \cdots a_{k+1}^*). \qquad (10)$$

By Lemma 3.14, there exists a regular language $R' \subset a_1^* a_2^* \cdots a_k^*$ such that $h(L_R) \cap R'$ is semiconvex in R' and $h(L_R) \cap R' \in \mathcal{C}_{k-1} \setminus \mathcal{C}_k$. Let $m \geq m_0$. By Equation (10), there exists $j_0 \in \{0, 1, \ldots, l_2 - 1\}$ such that

$$h(L_R) \cap R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^*) \in \mathcal{C}_{k-1} \setminus \mathcal{C}_k.$$

Since $h(L_R) \cap R'$ is semiconvex in R', by Lemma 3.13, the language

$$h(L_R) \cap R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^* \cap R)$$

is semiconvex in $R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^* \cap R)$. Thus, by Theorem 3.30,

$$\begin{aligned}
& \overline{h(L_R)} \cap R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^* \cap R) \\
= \quad & \left(\overline{h(L_R)} \cup \overline{(R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^* \cap R)}\right) \\
& \cap (R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^* \cap R)) \\
= \quad & \overline{h(L_R) \cap R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^* \cap R)} \\
& \cap (R' \cap h(a_1^{m+j_0} a_2^* a_3^* \cdots a_{k+1}^* \cap R)) \in \mathcal{L}_k \setminus \mathcal{L}_{k-1}.
\end{aligned}$$

This contradicts (9). Thus, $h(L) \in \mathcal{C}_k$, and the proof is complete. □

Finally, we complete the closure results with the following theorem.

Theorem 3.44. *Let $k \in \mathbb{N}_+$. The language families $\mathcal{C}_k$ and $\mathcal{C}_\infty$ are substitution closed full AFLs.*

Proof. Let us prove the claim concerning the families $\mathcal{C}_k$, where $k \in \mathbb{N}_+$. Therefore, the claim concerning the family $\mathcal{C}_\infty$ follows from the fact that $\mathcal{C}_\infty$ is the intersection of the families $\mathcal{C}_k$.

It is well known that each substitution closed full trio is a full AFL (see, e.g., (5, Proposition V.5.5)). Thus, by Theorems 3.38 and 3.43, it suffices to show that $\mathcal{C}_k$ is closed under substitution.

Let $L \subset \Sigma^*$ be a language in the family $\mathcal{C}_k$ and $\sigma : \Sigma^* \to 2^{\Delta^*}$ a $\mathcal{C}_k$-substitution. Since each full trio is closed under regular substitution (5, Example V.5.5), it suffices to show that there exists a regular subsitution $\sigma' : \Sigma^* \to 2^{\Delta^*}$ such that $\sigma'(L) = \sigma(L)$.

Let $R \in \mathcal{R}_k$. Let us define a substitution $\sigma' : \Sigma^* \to 2^{\Delta^*}$ by

$$\sigma'(a) = \sigma(a) \cap F(R) \qquad \forall a \in \Sigma,$$

where

$$F(R) = \{w \in \Sigma^* | \exists x, y \in \Sigma^* : xwy \in R\}.$$

Then we have $\sigma(L) \cap R \subset \sigma'(L)$. Since $\sigma'(L) \subset \sigma(L)$, we have $\sigma(L) \cap R = \sigma'(L) \cap R$. It should be clear that $F(R) \in \mathcal{R}_k$. Thus, σ' is a $\mathcal{R}_k$-substitution. This proves the claim. □

As a direct consequence of Theorems 3.44 and 3.36, we get the following theorem, which proves Conjecture 3.1.

Theorem 3.45. *There does not exist a minimal full trio or a minimal full AFL with respect to the family of bounded context-free languages.*

Proof. Let L be an arbitrary nonregular bounded context-free language. Then $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ for some $k \in \mathbb{N}_+$. By Theorem 3.36, there exists a nonregular bounded language $L' \in \mathcal{T}(L)$ such that $L' \in \mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$. Hence, $\mathcal{T}(L') \subset \mathcal{T}(L)$ and $\mathcal{F}(L') \subset \mathcal{F}(L)$. On the other hand, by Theorem 3.44, $L \notin \mathcal{F}(L')$ and $L \notin \mathcal{T}(L')$. Thus, $\mathcal{T}(L') \subsetneq \mathcal{T}(L)$ and $\mathcal{F}(L') \subsetneq \mathcal{F}(L)$. Hence, $\mathcal{T}(L)$ and $\mathcal{F}(L)$ cannot be minimal with respect to the bounded context-free languages. □

As a corollary of Theorem 3.45, none of the languages $S_i(\theta_1, \theta_2, \ldots, \theta_{k+1})$ can be minimal with respect to the bounded context-free languages. Autebert et al. have stated a conjecture that the full trios generated by the languages

$S_1(\underbrace{\neq,\neq,\ldots,\neq}_{k+1 \text{ times}})$ would be minimal with respect to the $(k+1)$-bounded context-free languages (1). Remembering Theorem 3.7, it is fascinating to expand the question of the conjecture a little bit.

Open Problem 3.46. *Let $k \geq 2$ be an integer, $i \in \{1,2,\ldots,k+1\}$ and $\theta_j \in \{>,\neq\}$ for all $j \in \{1,2,\ldots,k+1\}$. Are the full trios and the full AFLs generated by the languages $S_i(\theta_1,\theta_2,\ldots,\theta_{k+1})$ minimal with respect to the family of $(k+1)$-bounded context-free languages?*

It should be noted that Problem 3.46 is completely solved in case $k=1$. Corollary 2.16 states that the full trios and the full AFLs generated by the languages $S_1(\neq,\theta)$, where $\theta \in \{>,<,\neq\}$, are minimal. Moreover, by Theorem 2.14, these languages generate the only minimal full trios and full AFLs with respect to 2-bounded context-free languages. The situation in higher dimensions is open. So we may extend Problem 3.46 a little bit.

Open Problem 3.47. *Let $k \in \mathbb{N}_+$. Define all the full trios and the full AFLs that are minimal with respect to the $(k+1)$-bounded context-free languages.*

When considering Open Problems 3.46 and 3.47, it should be remembered that, by Theorem 3.37, for full trios and full AFLs, the minimality with respect to the $(k+1)$-bounded context-free languages is equivalent to the minimality with respect to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$.

4 Very Small Families Generated by Unbounded Context-Free Languages

We showed in the previous chapter that there does not exist a minimal full trio or a minimal full AFL with respect to the family of bounded context-free languages. We managed to prove this result since the family $\mathcal{C}_\infty$ does not contain any bounded nonregular language. However, we had a hint in Example 3.4 that the family $\mathcal{C}_\infty$ contains unbounded nonregular languages. That is why we cannot use the chain of the language families $\mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$ ($k \in \mathbb{N}$) to prove that Conjecture 2.10 holds. This motivates us to develop a similar chain of decreasing language families inside $\mathcal{C}_\infty$. Let $k \in \mathbb{N}$.[7] We define

$$\mathcal{C}_\infty^{(k)} = \{L \in \mathcal{C}_\infty \quad | \quad L \cap x' \,(x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k)^* \, x'' \in \mathcal{L}_{REG} \\ \forall x', x'', x_0, w_1, x_1, w_2, \ldots, x_{k-1}, w_k, x_k \in \Sigma_L^*\}$$

and $\mathcal{C}_\infty^{(\infty)} = \cap_{i=0}^{\infty} \mathcal{C}_\infty^{(i)}$. Clearly $\mathcal{C}_\infty^{(k+1)} \subset \mathcal{C}_\infty^{(k)}$ for all $k \in \mathbb{N}$. It follows also directly from the definition that $\mathcal{C}_\infty^{(0)} = \mathcal{C}_\infty$ and $\mathcal{L}_{REG} \subset \mathcal{C}_\infty^{(\infty)}$.

In Section 4.1, we will show that the family $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$ is nonempty for each $k \in \mathbb{N}$. In Section 4.2, we will focus on Conjecture 2.10. We will sketch a proof for the conjecture. However, we will face some problems, and the conjecture will remain as a conjecture.

4.1 The Family $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$ Is Nonempty

In this section, we will show that the full trio $\mathcal{T}(S_1(\neq,\neq))$ contains languages from each family $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$.

Define the regular substitution $\sigma : \Sigma^* \to 2^{\Sigma^*}$ by

$$\sigma(a_i) = \begin{cases} a_2^* a_3^* \cdots a_{k+1}^* a_1, & \text{if } i = 1, \\ a_i, & \text{otherwise.} \end{cases}$$

Define also the rational transduction $\tau : \Sigma^* \to 2^{\Sigma^*}$ by setting

$$\tau(w) = w a_1 (a_2^* a_3^* \cdots a_{k+1}^* a_1)^* \qquad \text{for all } w \in \Sigma^*.$$

[7] It should be remarked that in the previous chapter we used notation $k \in \mathbb{N}_+$, but in this chapter, we assume from now on that $k \in \mathbb{N}$.

Finally, let us denote $\omega = \tau \circ \sigma$. Since a composite function of two rational transductions is a rational transduction, ω is a rational transduction. Let $L \subset a_1^* a_2^* \cdots a_{k+1}^*$. Then

$$\begin{aligned}\omega(L) &= \Big\{ \left(a_2^* a_3^* \cdots a_{k+1}^* a_1\right)^{n_1} a_2^{n_2} a_3^{n_3} \cdots a_{k+1}^{n_{k+1}} a_1 \left(a_2^* a_3^* \cdots a_{k+1}^* a_1\right)^* \\ &\qquad \Big| a_1^{n_1} a_2^{n_2} \cdots a_{k+1}^{n_{k+1}} \in L \Big\} \\ &= \Big\{ (a_2^{i_2^{(0)}} a_3^{i_3^{(1)}} \cdots a_{k+1}^{i_{k+1}^{(0)}} a_1) \cdots (a_2^{i_2^{(p)}} a_3^{i_3^{(p)}} \cdots a_{k+1}^{i_{k+1}^{(p)}} a_1) \\ &\qquad \Big| p \in \mathbb{N}, \exists j \in \{0,1,\ldots,p\} : a_1^j a_2^{i_2^{(j)}} \cdots a_{k+1}^{i_{k+1}^{(j)}} \in L \Big\}.\end{aligned}$$

Define

$$L_G = \omega(S_1(\neq,\neq)) = \Big\{ a_2^{i_0} a_1 a_2^{i_1} a_1 \cdots a_2^{i_p} a_1 \mid p \in \mathbb{N}, \quad \exists j \in \{0,1,\ldots,p\} : i_j \neq j \Big\}.$$

It should be noted that this language is exactly the same as the language in Example 3.4. Correspondingly, it should be noted that, for example,

$$\begin{aligned}\omega(S_1(\neq,\neq,\neq)) &= \Big\{ a_2^{i_0} a_3^{i'_0} a_1 a_2^{i_1} a_3^{i'_1} a_1 \cdots a_2^{i_p} a_3^{i'_p} a_1 \\ &\qquad \Big| p \in \mathbb{N}, \exists j \in \{0,1,\ldots,p\} : i_j \neq j \vee i'_j \neq j \Big\}.\end{aligned}$$

Goldstine has proved that every bounded language in the full trio $\mathcal{T}(L_G)$ is regular (12).[8] We get the same result using Theorem 3.44 and Corollary 4.3, which is proved at the end of this section. Autebert et al. have proved that each nonregular language in the duo $\mathcal{D}(L_G)$ is a generator of $\mathcal{D}(L_G)$, i.e. the duo $\mathcal{D}(L_G)$ is minimal (1). They have also proved that the full trio $\mathcal{T}(L_G)$ is not minimal (1).

Denote $(\neq_1, \neq_2, \ldots, \neq_{k+1}) = (\underbrace{\neq, \neq, \ldots, \neq}_{k+1 \text{ times}})$.

---

[8] Goldstine defined the language L_G by

$$L_{G_1} = \Big\{ a_2^{i_1} a_1 a_2^{i_2} a_1 \cdots a_2^{i_p} a_1 \mid p \in \mathbb{N}_+, \quad \exists j \in \{1,2,\ldots,p\} : i_j \neq j \Big\}.$$

Since

$$L_G = (a_2^* a_1)^+ \setminus \{a_1, a_1 a_2 a_1, a_1 a_2 a_1 a_2^2 a_1, \ldots\}$$

and

$$L_{G_1} = (a_2^* a_1)^+ \setminus \{a_2 a_1, a_2 a_1 a_2^2 a_1, a_2 a_1 a_2^2 a_1 a_2^3 a_1, \ldots\},$$

we have $L_{G_1} = a_1 / L_G$ and $L_G = a_1 L_{G_1} \cup a_2 (a_2^* a_1)^+$. Hence, $\mathcal{T}(L_{G_1}) = \mathcal{T}(L_G)$.

Lemma 4.1. *Let $k \in \mathbb{N}$. Then*

$$\omega(S_1(\neq_1, \neq_2, \ldots, \neq_{k+2})) \in \mathcal{C}_\infty \setminus \mathcal{L}_{REG}.$$

Proof. Let us denote

$$L = \omega(S_1(\neq_1, \neq_2, \ldots, \neq_{k+2}))$$

and

$$\widetilde{L} = (a_2^* a_3^* \cdots a_{k+2}^* a_1)^+ \setminus L.$$

Then

$$\widetilde{L} = \left\{ a_1 a_2 a_3 \cdots a_{k+2} a_1 a_2^2 a_3^2 \cdots a_{k+2}^2 a_1 \cdots a_2^p a_3^p \cdots a_{k+2}^p a_1 \mid p \in \mathbb{N} \right\}. \tag{11}$$

From this expression, it should be clear that $\widetilde{L} \notin \mathcal{L}_{REG}$, and thus, $L \notin \mathcal{L}_{REG}$. Thus, it suffices to show that $L \in \mathcal{L}_\infty$.

Let $n \in \mathbb{N}_+$. We will show that $L \in \mathcal{C}_n$. Let $x_0, w_1, x_1, w_2, \ldots, x_{n-1}, w_n, x_n \in \Sigma^*$ and $R = x_0 w_1^* x_1 w_2^* \cdots x_{n-1} w_n^* x_n$.

Assume that $|\widetilde{L} \cap R| = \infty$. Let $l = |x_0 w_1 x_1 w_2 \cdots x_{n-1} w_n x_n|$. Then there exists an integer $m \geq l$ such that

$$v(m) = a_1 a_2 a_3 \cdots a_{k+2} a_1 a_2^2 a_3^2 \cdots a_{k+2}^2 a_1 \cdots a_2^{n+m} a_3^{n+m} \cdots a_{k+2}^{n+m} a_1 \in \widetilde{L} \cap R. \tag{12}$$

Since $v(m) \in R$, there exist $i_1, i_2, \ldots, i_n \in \mathbb{N}$ such that

$$v(m) = x_0 w_1^{i_1} x_1 w_2^{i_2} \cdots x_{n-1} w_n^{i_n} x_n. \tag{13}$$

By Equation (12), the subword a_2^l occur at least in $(n+1)$ different locations in the word $v(m)$. On the other hand, by Equation (13), this is not possible. Thus, the language $\widetilde{L} \cap R$ is finite. Therefore, we have

$$L \cap R = ((a_2^* a_3^* \cdots a_{k+2}^* a_1)^+ \setminus \widetilde{L}) \cap R = ((a_2^* a_3^* \cdots a_{k+2}^* a_1)^+ \cap R) \setminus (\widetilde{L} \cap R) \in \mathcal{L}_{REG}.$$

Hence, $L \in \mathcal{C}_n$ and $L \in \mathcal{C}_\infty$. □

Theorem 4.2. *Let $k \in \mathbb{N}$. Then*

$$\omega(S_1(\neq_1, \neq_2, \ldots, \neq_{k+2})) \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}.$$

Proof. Let us denote $L = \omega(S_1(\neq_1, \neq_2, \ldots, \neq_{k+2}))$. By Lemma 4.1, we have $L \notin \mathcal{L}_{REG}$, and thus, $L \notin \mathcal{C}_\infty^{(k+1)}$.

Let $x', x'', x_0, w_1, x_2, w_2, \ldots, x_{k-1}, w_k, x_k \in \Sigma^*$ be arbitrary. By Lemma 4.1, it suffices to show that the language

$$\begin{aligned} L' &= L \cap x'(x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k)^* x'' \\ &= L \cap ((a_2^* a_3^* \cdots a_{k+2}^* a_1)^* \cap x'(x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k)^* x'') \end{aligned} \quad (14)$$

is regular. By Equation (14), we may assume that, if two distinct letters occurs in some word w_j, also the letter a_1 occurs in w_j. Thus, there exists $i \in \{2, 3, \ldots, k+2\}$ such that for each $j \in \{1, 2, \ldots, k\}$, we have either $|w_j|_{a_i} = 0$ or $|w_j|_{a_1} > 0$. In either case, between two occurrences of the letter a_1, there are less than

$$n = |x_0 w_1 x_1 w_2 \cdots x_{k-1} w_k x_k|$$

occurrences of the letter a_i in the language $(x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k)^*$. This means that the language

$$\widetilde{L} \cap x'(x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k)^* x''$$

is finite, where $\widetilde{L}$ is the language defined by Equation (11). Since

$$L' = \left((a_2^* a_3^* \cdots a_{k+2}^* a_1)^* \setminus \widetilde{L}\right) \cap x'(x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k)^* x'',$$

we have $L' \in \mathcal{L}_{REG}$. □

Corollary 4.3. $L_G \in \mathcal{C}_\infty^{(0)} \setminus \mathcal{C}_\infty^{(1)}$.

4.2 Outline of the Proof of Conjecture 2.10

In this section, we will sketch a similar proof for Conjecture 2.10 that we had for Conjecture 3.1 in Chapter 3. We will also shortly study why our approach fails in this case.

In order to prove Conjecture 2.10 in the same manner as we proved Conjecture 3.1, we would need to prove following items:

1. Each nonregular language from the family $\mathcal{C}_\infty$ belongs to $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$ for some $k \in \mathbb{N}$;

2. For each $k \in \mathbb{N}$ and $L \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$, there exists a language $L' \in \mathcal{T}(L)$ such that $L' \in \mathcal{C}_\infty^{(k+1)} \setminus \mathcal{C}_\infty^{(k+2)}$;
3. The family $\mathcal{C}_\infty^{(k)}$ is a full trio for all $k \in \mathbb{N}$.

In addition, to prove that there does not exist a minimal full AFL in the family of context-free languages, we would need to prove that $\mathcal{C}_\infty^{(k)}$ is a full AFL for all $k \in \mathbb{N}$.

Let us first consider Item 1. Item 1 is equivalent to the statement $\mathcal{C}_\infty^{(\infty)} = \mathcal{L}_{REG}$. This is proved by very straightforward reasoning.

Theorem 4.4. $\mathcal{C}_\infty^{(\infty)} = \mathcal{L}_{REG}$.

Proof. Let $L \in \mathcal{C}_\infty$ be nonregular and $\Sigma_L = \{a_1, a_2, \ldots, a_n\}$. Then we have $L \subset \Sigma_L^* = \{a_1, a_2, \ldots, a_n\}^* \subset (a_1^* a_2^* \cdots a_n^*)^*$. Thus, $L \notin \mathcal{C}_\infty^{(n)}$ and $\mathcal{C}_\infty^{(\infty)} = \mathcal{L}_{REG}$. □

So we have proved Item 1. Thus, our new chain is applicable for proving Conjecture 2.10. By Theorem 4.4, we now have perfect understanding of the relationships of our family chains. The family chains are illustrated in Figure 5.

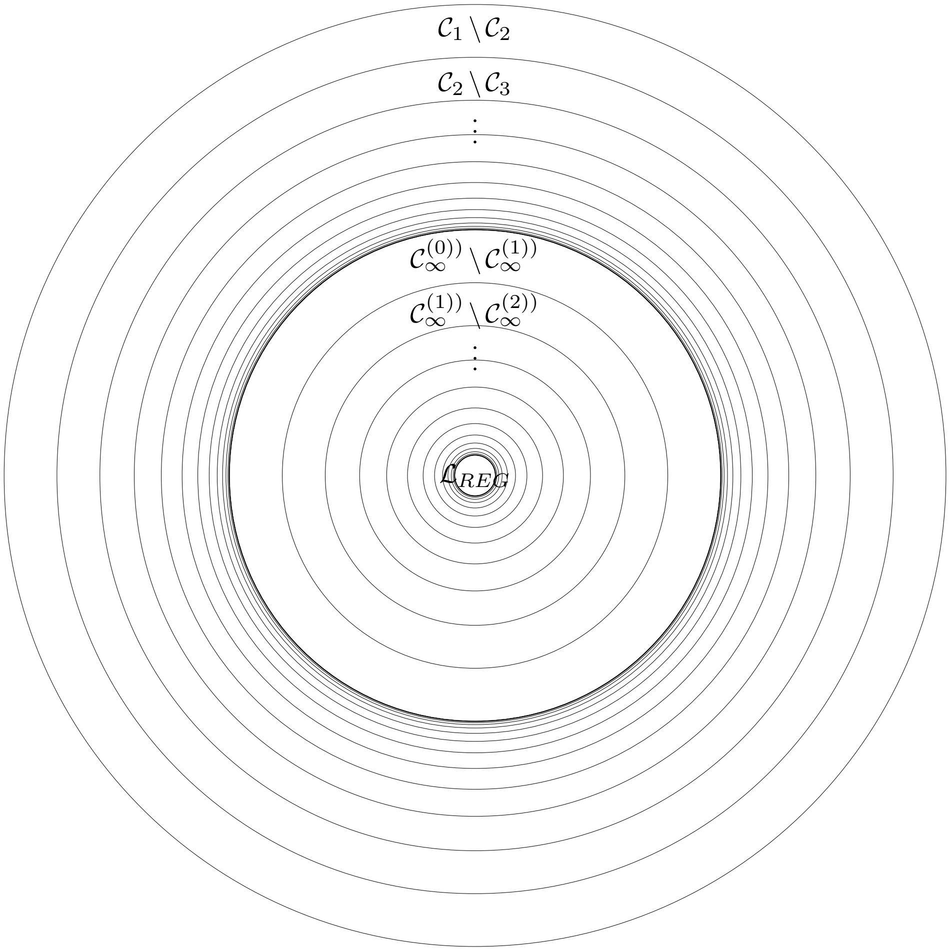

Fig 5. Illustration of the family chains $\mathcal{C}_{k+1}$ and $\mathcal{C}_\infty^{(k)}$. Note that the outermost circle is $\mathcal{L}_{CF}(=\mathcal{C}_1)$.

Let us next consider Item 2. The item is an analogue with Theorem 3.36 that we proved in Section 3.4. The idea of the proof of Theorem 3.36 was to restrict our attention to the family of strictly bounded languages that are semiconvex in a regular language and then to copy one letter. In copying, we introduced a new letter whose exponent can have arbitrary values. Then we took the union of that language and a language, where the roles of the new letter and some original letter had been changed. When dealing with languages of the family $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$, we cannot naturally restrict our attention to the family of strictly bounded languages. Despite that, we will prove in Section 4.2.1 that, for each language $L \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$, there exists a nonregular language L' in the family

$$\widehat{\mathcal{B}}_{k+1} = \{L \subset \Sigma^* | \exists a_1, a_2, \ldots, a_{k+2} \in \Sigma_L : L \subset (a_2^* a_3^* \cdots a_{k+2}^* a_1)^*\}$$

so that $L' \in \mathcal{T}(L)$ and $L' \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$. Thus, we may restrict our attention inside the family $\widehat{\mathcal{B}}_{k+1}$. It feels quite natural that the same copying trick would work also for languages of the family $\widehat{\mathcal{B}}_{k+1} \cap \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$. We actually used this idea already in Theorem 4.2.

Before proving Theorem 3.36, we needed some study of the structure of weak languages of the family $\mathcal{C}_{k+1} \setminus \mathcal{C}_{k+2}$. This was done in Section 3.3. For languages of the family $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$, this study is much more difficult. We would need to redefine our key concepts (e.g., semiconvexity, $\mathcal{L}_k$, $\mathcal{R}_k$), which is not a trivial problem anymore.

Finally, let us consider Item 3. In Section 3.5, we proved that $\mathcal{C}_{k+1}$ is a substitution closed full AFL. Closure under intersection with regular languages, union and inverse morphism was proved by very straightforward reasoning. We will prove these results for the family $\mathcal{C}_\infty^{(k)}$ in Section 4.2.2. Closure under morphism was certainly the hardest part of the proof of Theorem 3.44. Morphism causes even bigger problems for the family $\mathcal{C}_\infty^{(k)}$. In fact, we shall prove in Section 4.2.2 that $\mathcal{C}_\infty^{(k+1)}$ is not closed under morphism. Nevertheless, this result does not necessary break down totally our approach for proving Conjecture 2.10. It is possible that, with slight modifications of the definition of the language family $\mathcal{C}_\infty^{(k)}$, we would achieve closure under morphism without losing the other important properties (Items 1, 2 and 3 above) of the language family.

4.2.1 A Normal Form Representative for Languages in $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$

In this section, we will develop an analogue version of Theorem 3.9 for the family $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$ (Theorem 4.6). The structure of this section is essentially the same as Section 3.2.

At the very beginning, we define a slightly similar family of languages with the family of k- strictly bounded languages. Let us define

$$\widehat{\mathcal{B}}_k = \{L \subset \Sigma^* | \exists a_1, a_2, \ldots, a_{k+1} \in \Sigma_L : L \subset (a_2^* a_3^* \cdots a_{k+1}^* a_1)^*\}.$$

Let $n \in \mathbb{N}$ and $x', x'', x_0, w_1, x_1, w_2, \ldots, x_{n-1}, w_n, x_n \in \Sigma^*$. Let us define a transducer

$$\begin{aligned} & M(x', x_0, w_1, x_1, w_2, \ldots, x_{n-1}, w_n, x_n, x'') \\ = \; & (\{p_0', p_{n+1}', p_0, p_1, \ldots, p_{n+1}\}, \Sigma, \Sigma_{n+1}, E_1 \cup E_2 \cup E_3, p_0', \{p_{n+1}'\}), \end{aligned}$$

where

$$\begin{aligned} E_1 &= \{p_0', \epsilon, x', p_0\} \cup \{(p_0, \epsilon, \epsilon, p_{n+1})\} \cup \{p_{n+1}, \epsilon, x'', p_{n+1}'\} \cup \{(p_{n+1}, \epsilon, \epsilon, p_0)\}, \\ E_2 &= \{(p_i, \epsilon, x_i, p_{i+1}) | i \in \{0, 1, \ldots, n-1\}\} \cup \{(p_n, a_1, x_n, p_{n+1})\} \qquad \text{and} \\ E_3 &= \{(p_i, a_{i+1}, w_i, p_i) | i \in \{1, 2, \ldots, n\}\}. \end{aligned}$$

If the words $x', x'', x_0, w_1, x_1, w_2, \ldots, x_{n-1}, w_n, x_n$ are clear from the context, we may simplify the notation by setting

$$M = M(x', x_0, w_1, x_1, w_2, \ldots, x_{n-1}, w_n, x_n, x'').$$

Let $L \subset (a_2^* a_3^* \cdots a_{n+1}^* a_1)^*$. Then

$$\begin{aligned} \tau_M(L) = \; & \{x' x_0 w_1^{i_{1,1}} x_1 w_2^{i_{1,2}} \cdots x_{n-1} w_n^{i_{1,n}} x_n x_0 w_1^{i_{2,1}} x_1 w_2^{i_{2,2}} \cdots x_{n-1} w_n^{i_{p,n}} x_n x'' \\ & \mid p \in \mathbb{N}_+ \text{ and } a_2^{i_{1,1}} a_3^{i_{1,2}} \cdots a_{n+1}^{i_{1,n}} a_1 a_2^{i_{2,1}} a_3^{i_{2,2}} \cdots a_{n+1}^{i_{p,n}} a_1 \in L\}. \end{aligned}$$

Therefore, it should be clear that the rational transduction $\tau_M(L) : \Sigma^* \to 2^{\Sigma^*}$ may be regarded as a function $\tau_M(L) : (a_2^* a_3^* \cdots a_{n+1}^* a_1)^* \to \Sigma^*$. We should also note that, as an inverse of a rational transduction, τ_M^{-1} is also a rational transduction.

Next, we will prove an analogue version of Lemma 3.8.

Lemma 4.5. *Let $n \in \mathbb{N}$ and $L \subset x'(x_0 w_1^* x_1 w_2^* \cdots x_{n-1} w_n^* x_n)^* x''$. Given $k \in \mathbb{N}$, the language L is in $\mathcal{C}_\infty^{(k)}$ if and only if $\tau_M^{-1}(L)$ is in $\mathcal{C}_\infty^{(k)}$.*

Proof. Assume first that $L \in \mathcal{C}_\infty^{(k)}$. Let

$$R = y'(y_0 v_1^* y_1 v_2^* \cdots y_{k-1} v_k^* y_k)^* y'',$$

where $y', y'', y_0, v_1, y_1, v_2, \ldots, y_{k-1}, v_k, y_k \in \Sigma_{\tau_M^{-1}(L)}$. It suffices to show that $\tau_M^{-1}(L) \cap R \in \mathcal{L}_{REG}$. Since $\tau_M^{-1}(L) \subset (a_2^* a_3^* \cdots a_{n+1}^* a_1)^*$, we may assume that $R \subset (a_2^* a_3^* \cdots a_{n+1}^* a_1)^*$. Since $L \in \mathcal{C}_\infty^{(k)}$, we have $L \cap \tau_M(R) \in \mathcal{L}_{REG}$ and

$$\tau_M^{-1}(L) \cap R = \tau_M^{-1}(L) \cap \tau_M^{-1}(\tau_M(R)) \cap R = \tau_M^{-1}(L \cap \tau_M(R)) \cap R \in \mathcal{L}_{REG}.$$

Hence, $\tau_M^{-1}(L) \in \mathcal{C}_\infty^{(k)}$.

Assume next that $\tau_M^{-1}(L) \in \mathcal{C}_\infty^{(k)}$. Let

$$R = y'(y_0 v_1^* y_1 v_2^* \cdots y_{k-1} v_k^* y_k)^* y'',$$

where $y', y'', y_0, v_1, y_1, v_2, \ldots, y_{k-1}, v_k, y_k \in \Sigma_L$. It suffices to show that $L \cap R \in \mathcal{L}_{REG}$. Therefore, we may assume that $R \subset x'(x_0 w_1^* x_1 w_2^* \cdots x_{n-1} w_n^* x_n)^* x''$. Clearly, there exist $u_1, u_2, \ldots, u_k \in \Sigma^*_{\tau_M^{-1}(L)}$ and $z', z'', z_0, z_1, \ldots, z_k \in \Sigma^*_{\tau_M^{-1}(L)}$ such that

$$\tau_M(z'(z_0 u_1^* z_1 u_2^* \cdots z_{k-1} u_k^* z_k)^* z'') = R.$$

Let us denote $R' = z'(z_0 u_1^* z_1 u_2^* \cdots z_{k-1} u_k^* z_k)^* z''$. Since $\tau_M^{-1}(L) \in \mathcal{C}_\infty^{(k)}$, we have

$$L \cap R = \tau_M(\tau_M^{-1}(L)) \cap \tau_M(R') = \tau_M(\tau_M^{-1}(L) \cap R') \in \mathcal{L}_{REG}.$$

Therefore, $L \in \mathcal{C}_\infty^{(k)}$. □

Finally, we are ready to prove the main result of this section. The next theorem will imply that, if we are looking for a language generating a minimal full trio (or a minimal full AFL) inside the family $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$, we may always restrict to look it inside the family $\widehat{\mathcal{B}}_{k+1}$.

Theorem 4.6. *Let $k \in \mathbb{N}$ and $L \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$. Then there exists a language $L_1 \in \mathcal{T}(L)$ such that $L_1 \in \widehat{\mathcal{B}}_{k+1} \cap (\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)})$.*

Proof. Since $L \notin \mathcal{C}_\infty^{(k+1)}$, there exist $x', x'', x_0, x_1 \ldots, x_{k+1} \in \Sigma^*$ and $w_1, w_2, \ldots, w_{k+1} \in \Sigma^*$ such that

$$L_1 = L \cap x'(x_0 w_1^* x_1 w_2^* \cdots x_k w_{k+1}^* x_{k+1})^* x'' \notin \mathcal{L}_{REG}.$$

Hence, $L_1 \notin \mathcal{C}_\infty^{(k+1)}$. Since $L \in \mathcal{C}_\infty^{(k)}$, we have $L_1 \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$. According to Lemma 4.5, we have $\tau_M^{-1}(L_1) \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$, where $\tau_M : (a_2^* a_3^* \cdots a_{n+1}^* a_1)^* \to \Sigma^*$ is a function defined in the beginning of this section. Since τ_M^{-1} is a rational transduction, we have by Theorem 2.17, $\tau_M^{-1}(L_1) \in \mathcal{T}(L)$. □

4.2.2 Closure Properties of the Family $\mathcal{C}_\infty^{(k)}$

In this section, we will handle closure properties of the family $\mathcal{C}_\infty^{(k)}$.

Theorem 4.7. *Let $k \in \mathbb{N}$. The language family $\mathcal{C}_\infty^{(k)}$ is closed under intersection with regular languages, union and inverse morphism.*

Proof. Let $L_1, L_2 \in \mathcal{C}_\infty^{(k)}$, $x', x'', x_0, x_1, \ldots, x_k \in \Sigma^*$ and $w_1, w_2, \ldots, w_k \in \Sigma^*$ be arbitrary. Let us denote $R = x'(x_0 w_1^* x_1 w_2^* \cdots x_{k-1} w_k^* x_k)^* x''$. Then $L_1 \cap R \in \mathcal{L}_{REG}$ and $L_2 \cap R \in \mathcal{L}_{REG}$.

Let us consider the intersection with regular languages. Let R_1 be a regular language. Since

$$(L_1 \cap R_1) \cap R = (L_1 \cap R) \cap R_1$$

and $L_1 \cap R \in \mathcal{L}_{REG}$, we have $(L_1 \cap R_1) \cap R \in \mathcal{L}_{REG}$ and $L_1 \cap R_1 \in \mathcal{C}_\infty^{(k)}$.

Let us next show closure under union. We have $(L_1 \cup L_2) \cap R = (L_1 \cap R) \cup (L_2 \cap R) \in \mathcal{L}_{REG}$. Thus, $\mathcal{C}_\infty^{(k)}$ is closed under union.

Finally, we will show closure under inverse morphism. Let $L \in \mathcal{C}_\infty^{(k)}$ and $h : \Sigma^* \to \Sigma_L^*$ be a morphism. Since $R \subset h^{-1}(h(R))$, we have

$$\begin{aligned} & h^{-1}(L) \cap R \\ = \; & h^{-1}(L) \cap h^{-1}(h(R)) \cap R \\ = \; & h^{-1}(L \cap h(R)) \cap R. \end{aligned}$$

Furthermore, since $L \cap h(R) \in \mathcal{L}_{REG}$, we have $h^{-1}(L) \cap R \in \mathcal{L}_{REG}$. □

The next theorem shows that, without any modifications, the chain of the language families $\mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$ is not applicable for proving Conjecture 2.10.

Theorem 4.8. *Let $k \in \mathbb{N}$. The family $\mathcal{C}_\infty^{(k+1)}$ is not closed under morphism.*

Proof. Let $k \in \mathbb{N}$. By Theorem 4.2, there exists a language $L \in \mathcal{C}_\infty^{(k+1)} \setminus \mathcal{C}_\infty^{(k+2)}$ such that $L \subset (a_2^* a_3^* \cdots a_{k+2}^* a_1)^*$. Let $\Delta = \{a_1, a_2\}$. Define a morphism $h : \Sigma_L^* \to \Delta^*$ by $h(a_i) = a_2^{i-1} a_1$ for each $i \in \{1, 2, \ldots, k+2\}$. Then h is injective, and we have $L = h^{-1}(h(L))$. Thus, $h(L)$ is nonregular, and moreover, $h(L) \notin \mathcal{C}_\infty^{(1)}$. □

5 Future Work

In this chapter, we will summarize all the open problems we have found so far. In addition, we will pose a few more open issues regarding this thesis.

5.1 Future Work Regarding Chapter 3

As in Chapter 3, we also use the notation $k \in \mathbb{N}_+$ in this section. Our first open problem (Open Problem 3.22) handled the period structure of the Parikh-image of $(k+1)$- strictly bounded SLIP-languages in the family $\hat{\mathcal{C}}_k \setminus \hat{\mathcal{C}}_{k+1}$. The next open problem was the problem whether a $(k+1)$- strictly bounded language $L \in \mathcal{C}_k \setminus \mathcal{C}_{k+1}$ is context-free if we change the order of the letters. This was presented more formally in Open Problem 3.32.

Open Problems 3.46 and 3.47 dealt with the minimality with respect to the $(k+1)$-bounded context-free languages (or equivalently the minimality with respect to $\mathcal{C}_k \setminus \mathcal{C}_{k+1}$). Open Problem 3.46 asked whether the full trios and the full AFLs generated by languages $S_i(\theta_1, \theta_2, \ldots, \theta_k)$ are minimal with respect to the $(k+1)$-bounded context-free languages when $\theta_j \in \{>, \neq\}$ for all $j \in \{1, 2, \ldots, k\} \setminus \{i\}$. Another issue is whether there are any other minimal full trios or minimal full AFLs with respect to the $(k+1)$-bounded context-free languages. Thus, we posed the problem to define all the minimal full trios and full AFLs with respect to the $(k+1)$-bounded context-free languages (Open Problem 3.47).

5.2 Future Work Regarding Chapter 4

As in Chapter 4, we also use the notation $k \in \mathbb{N}$ in this section. To prove Conjecture 2.10 in Section 4.2, we faced two major problems.

Open Problem 5.1. *Let $k \in \mathbb{N}$. Redefine the language family $\mathcal{C}_\infty^{(k)}$ so that $\mathcal{C}_\infty^{(k)}$ would be full trio.*

Open Problem 5.2. *Let $k \in \mathbb{N}$ and $L \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)}$. Does there exist a language $L' \in \mathcal{T}(L)$ such that $L' \in \mathcal{C}_\infty^{(k+1)} \setminus \mathcal{C}_\infty^{(k+2)}$?*

However, to solve Open Problem 5.2, we had a hint that copying a letter of the language $L \in \mathcal{C}_\infty^{(k)} \setminus \mathcal{C}_\infty^{(k+1)} \cap \widehat{B}_{k+1}$ would produce a language in the family $\mathcal{C}_\infty^{(k+1)} \setminus \mathcal{C}_\infty^{(k+2)}$.

Recall the definition of the rational transduction ω in Section 4.1. Let us consider the language

$$S_2(>, \neq, >) = \{a_1^{n_1} a_2^{n_2} a_3^{n_3} \mid n_2 > n_1 \text{ or } n_2 > n_3\}.$$

Then

$$\begin{aligned} & \omega(S_2(>, \neq, >)) \cap a_1^* a_2^* a_3^* a_1 \\ = \; & \{(a_2^* a_3^* a_1)^{n_1} a_2^{n_2} a_3^{n_3} a_1 (a_2^* a_3^* a_1)^* \mid n_2 > n_1 \text{ or } n_2 > n_3\} \cap a_1^* a_2^* a_3^* a_1 \\ = \; & \{a_1^{n_1} a_2^{n_2} a_3^{n_3} a_1 \mid n_2 > n_1 \text{ or } n_2 > n_3\} \notin \mathcal{C}_3. \end{aligned}$$

Thus, the transduction ω does not transfer the language $S_2(>, \neq, >)$ inside the language family $\mathcal{C}_\infty$. Now the problem is whether there exists a language $L \in \mathcal{T}(S_2(>, \neq, >))$ such that $L \in \mathcal{C}_\infty \setminus \mathcal{L}_{REG}$. Let us pose the problem more generally.

Open Problem 5.3. *Does there exist a nonregular context-free language L such that $\mathcal{T}(L)$ does not contain any language from the family $\mathcal{C}_\infty \setminus \mathcal{L}_{REG}$?*

5.3 General Future Work

In this section, we will briefly present a possible nice interface between the approach of this thesis and the theory of iterative pairs. We do not cover the basics of the theory here. A very good description of iterative pairs can be found from the book (5). According to the theory of iterative pairs, for example, the languages

$$\{(a_1^{n_1} a_2^{n_2})^{n_3} a_3^{n_4} | n_1 > n_2 \wedge n_3 > n_4\}$$

and

$$\{(a_1^{n_1} a_2^{n_2})^{n_3} a_3^{n_4} | n_1 < n_2 \wedge n_3 < n_4\}$$

are rationally incomparable since they have different types of iterative pairs. One weakness of the theory of the iterative pairs is that it cannot say anything for so-called *degenerated* iterative pairs. Loosely speaking, degenerated iterative pairs are weak components of context-free languages. However, this thesis covers exactly rational dominating relations of weak context-free languages. Thus, perhaps the theory of iterative pairs may be extended with our approach.

Let us explain this more concretely. For each $i \in \mathbb{N}$, let $h_i : \Sigma_\infty \to \Sigma_\infty$ be a morphism such that $h_i(a_j) = a_{i+j}$ for all $j \in \mathbb{N}_+$. The problem is whether we could develop the theory so that we could say that, for example, the language

$$\{a_1^{n_1} w^{n_2} | n_1 \neq n_2, w \in h_1(L_G)\} h_3(S_1(\neq, >, \neq))$$

does not dominate rationally the language

$$\{a_1^{n_1} w^{n_2} | n_1 \neq n_2, w \in h_1(S_1(\neq, \neq))\} h_3(S_1(\neq, >, \neq))$$

since L_G does not dominate rationally $S_1(\neq, \neq)$.

Another weakness of the theory of iterative pairs is that it can say only that one language does not dominate rationally another language. The theory gives no positive answers. Let us continue our discussion by asking whether we could say that the language

$$\{a_1^{n_1} w^{n_2} | n_1 \neq n_2, w \in h_1(S_1(\neq, \neq))\} h_3(S_1(\neq, >, \neq))$$

dominates rationally the language

$$\{a_1^{n_1} w^{n_2} | n_1 \neq n_2, w \in h_1(L_G)\} h_3(S_1(\neq, >, \neq))$$

since $S_1(\neq, \neq)$ dominates rationally L_G. Further, could we find some "prime context-free languages" so that, by combining these languages, we would obtain a rationally equivalent language for any context-free language.

References

1. Autebert JM, Beauquier J, Boasson L & Latteux M (1980) Very small families of algebraic nonrational languages. Proc. Formal Language Theory, Perspectives and Open Problems, Academic Press, Orlando, 89–107.
2. Autebert JM, Beauquier J, Boasson L & Nivat M (1979) Quelques problèmes ouverts en théorie des langages algébriques. RAIRO Informatique Theorique/Theoretical Informatics 13(4): 363–378.
3. Autebert JM & Boasson L (1980) Generators of cones and cylinders. Proc. Formal Language Theory, Perspectives and Open Problems, Academic Press, Orlando, 49–87.
4. Berstel J (1974) Sur une conjecture de S. Greibach. Acta Informatica 3: 293–295.
5. Berstel J (1979) Transductions and Context-Free Languages. Teubner Studienbücher, Stuttgart.
6. Berstel J & Boasson L (1974) Une suite décroissante de cônes rationnels. Proc. Proceedings of the 2nd Colloquium on Automata, Languages and Programming, Springer-Verlag, London, 383–397.
7. Blattner M & Latteux M (1981) Parikh-bounded languages. Proc. ICALP, 316–323.
8. Chomsky N (1959) On certain formal properties of grammars. Information and Control 2: 137–167.
9. Ginsburg S (1966) The Mathematical Theory of Context-Free Languages. McGraw-Hill, Inc., New York.
10. Ginsburg S (1975) Algebraic and Automata-Theoretic Properties of Formal Languages. Elsevier Science Inc., New York.
11. Ginsburg S & Greibach S (1969) Absract families of languages. Memoirs of the American Mathematical Society 87: 1–32.
12. Goldstine J (1972) Substitution and bounded languages. Journal of Computer and System Sciences 6(1): 9–29.
13. Greibach S (1970) Chains of full AFL's. Mathematical Systems Theory 4(3): 231–242.
14. Greibach S & Hopcroft J (1969) Indepedence of AFL operations. Memoirs of the American Mathematical Society 87: 33–40.
15. Ito R (1969) Every semilinear set is a finite union of disjoint linear sets. Journal of Computer and System Sciences 3(2): 221–231.
16. Kortelainen J (1985) Every commutative quasirational language is regular. Proc. Proceedings of the 12th Colloquium on Automata, Languages and Programming, Springer-Verlag, London, 348–355.
17. Latteux M (1979) Cônes rationnels commutatifs. Journal of Computer and System Sciences 18(3): 307–333.
18. Latteux M (1988) Rational cones and commutations. Proc. IMYCS, 37–54.
19. Shieber S (1985) Evidence against the context-freeness of natural language. Linguistics and Philosophy 8: 333–343.

520. Hilli, Anu (2009) The effect of crop quality and pre-treatment on germination in Scots pine and Norway spruce seeds

521. Kreivi, Marjut (2009) Conservation genetics and phylogeography of endangered boreoarctic seashore plant species

522. Riihijärvi, Jorma (2009) Tietojenkäsittelytieteiden koulutuksen työelämävastaavuus. Esimerkkitapauksena Oulun yliopiston tietojenkäsittelytieteiden laitoksen koulutusohjelma

523. Ilmonen, Jari (2009) Benthic macroinvertebrate and bryophyte assemblages in boreal springs: diversity, spatial patterns and conservation

524. Pujol, François (2009) Experiments on fatty acids chain elongation and glycan flipping in the ER membrane

525. Lampila, Satu (2009) The causes and consequences of population declines of two boreal forest species. The case of the willow tit (*Parus montanus*) and the Siberian flying squirrel (*Pteromys volans*)

526. Anttila, Katja (2009) Swimming muscles of wild, trained and reared fish. Aspects of contraction machinery and energy metabolism

527. Roppola, Katri (2009) Environmental applications of manometric respirometric methods

528. Arhippainen, Leena (2009) Studying user experience: issues and problems of mobile services. – Case ADAMOS: User experience (im)possible to catch?

529. Niemelä, Marika (2009) Biotic interactions and vegetation management on coastal meadows

530. Sørensen, Louise Ilum (2009) Grazing, disturbance and plant soil interactions in northern grasslands

531. Salo, Antti (2009) Expression of lysyl hydroxylases and characterization of a novel disorder caused by mutations in the lysyl hydroxylase 3 gene

532. Vaismaa, Matti (2009) Development of benign synthesis of some terminal ?-hydroxy ketones and aldehydes

533. Meriläinen, Gitte (2009) Structural and enzymological studies of the thiolase enzymes

534. Astorga, Anna (2009) Diversity patterns in marine and freshwater environments. The role of environmental and spatial factors across multiple scales

535. Rivinoja, Antti (2009) Golgi pH and glycosylation

Book orders:
OULU UNIVERSITY PRESS
P.O. Box 8200, FI-90014
University of Oulu, Finland

Distributed by
OULU UNIVERSITY LIBRARY
P.O. Box 7500, FI-90014
University of Oulu, Finland

UNIVERSITY OF OULU P.O.B. 7500 FI-90014 UNIVERSITY OF OULU FINLAND

ACTA UNIVERSITATIS OULUENSIS

ISBN 978-951-42-9273-6 (Paperback)
ISBN 978-951-42-9274-3 (PDF)
ISSN 0355-3191 (Print)
ISSN 1796-220X (Online)

UNIVERSITY of OULU
OULUN YLIOPISTO