Discrete Applied Mathematics 12 (1985) 41-56
North-Holland

CLASSES OF REGULAR AND CONTEXT-FREE LANGUAGES OVER COUNTABLY INFINITE ALPHABETS

Friedrich OTTO
Fachbereich Informatik, Universität Kaiserslautern, Postfach 3049, 6750 Kaiserslautern, West Germany

Received 20 December 1983
Revised 5 September 1984

For a countably infinite alphabet Δ, the classes Reg(Δ) of regular languages and CFL(Δ) of context-free languages over Δ are defined by way of an encoding. All the languages contained in these classes are decidable, and these classes do have many properties in common with the class of regular languages Reg(Σ) and the class of context-free languages CFL(Σ), respectively, where Σ is a finite alphabet. In particular, each of these classes can be characterized in a semantical way by a certain type of automata over Δ. Finally, the classes Reg(Δ) and CFL (Δ) are compared to the classes of languages over Δ that are defined by Autebert, Beauquier, and Boasson.

1. Introduction

For a finite alphabet Σ, the class Reg(Σ) of regular languages over Σ has been characterized in many different ways. So when L is a language over Σ, then L is regular, i.e., $L \in \mathrm{Reg}(\Sigma)$, if and only if L is accepted by a finite automaton, if and only if L is represented by a regular expression, if and only if L is generated by a regular grammar, if and only if the syntactic monoid M_L of L is finite [3,4,5]. Furthermore, every regular language is decidable in real time, and the class Reg(Σ) is closed under a large variety of operations, e.g., it is closed under union, intersection, complementation, concatenation, Kleene closure, reversal, GSM mappings, and inverse GSM mappings [5].

The class CFL(Σ) of context-free languages over Σ has also been characterized in several different ways by means of context-free grammars, pushdown automata, and closure properties (cf., e.g., the Chomsky–Schützenberger Theorem [2]). Every context-free language over Σ is decidable by some algorithm from $E_2(\Sigma)$, where $E_k(\Sigma)$ denotes the k-th class of the Grzegorczyk hierarchy of word functions over Σ [9], and the class CFL(Σ) is also closed under various operations [4,5].

In [1] Autebert, Beauquier, and Boasson use several of the characterizations of Reg(Σ) and CFL(Σ) to define classes of languages over Δ, where Δ is a countably infinite alphabet. However, it turns out that none of their classes meets all the nice properties that the corresponding class over Σ has. In particular, each of their classes contains non-recursive languages.

In the present paper we define classes Reg(Δ) and CFL(Δ) from Reg(Σ_2) and

$\mathrm{CFL}(\Sigma_2)$, respectively, by using a specific encoding γ from Δ^* into Σ_2^*. Here Σ_2 denotes a two-letter alphabet. Although at first sight this definition may seem to be rather at will, it turns out that several other encodings yield the same classes. Further, all the languages from $\mathrm{Reg}(\Delta)$ are decidable by algorithms from $E_1(\Delta)$, while those from $\mathrm{CFL}(\Delta)$ are decidable by algorithms from $E_2(\Delta)$. Here $E_k(\Delta)$ denotes the k-th class of the Grzegorczyk hierarchy of word functions over Δ, which is related to the Grzegorczyk hierarchy over Σ_2 by one of the encodings we consider [6].

Then some closure properties and some non-closure properties of $\mathrm{Reg}(\Delta)$ and $\mathrm{CFL}(\Delta)$ are proved. As it turns out, $\mathrm{Reg}(\Delta)$ can be characterized in a syntactical way by certain expressions called Δ-expressions as well as in a semantical way by certain automata called finite Δ-automata. These finite Δ-automata are a direct generalization of finite automata, and they can be considered as a restriction of the Turing machine model Madlener and Otto developed for Δ [6]. In the same manner $\mathrm{CFL}(\Delta)$ can be characterized in a semantical way by certain automata, but we will not proceed this in this paper.

Finally, the classes $\mathrm{Reg}(\Delta)$ and $\mathrm{CFL}(\Delta)$ are compared to the classes defined in [1] giving a good impression of the relative power of the different characterizations of $\mathrm{Reg}(\Sigma)$ and $\mathrm{CFL}(\Sigma)$ when carried over to countably infinite alphabets. As a by-product we get the result that, for each $L \in \mathrm{Reg}(\Delta)$, the syntactic monoid M_L is finite.

Since so many properties of the classes $\mathrm{Reg}(\Delta)$ and $\mathrm{CFL}(\Delta)$ are so close to corresponding properties of the classes $\mathrm{Reg}(\Sigma)$ and $\mathrm{CFL}(\Sigma)$, respectively, we consider the classes $\mathrm{Reg}(\Delta)$ and $\mathrm{CFL}(\Delta)$ as natural generalizations of the classes of regular and context-free languages to countably infinite alphabets.

This paper is organized as follows. In Section 1 several encodings are presented, and some of their properties, that we will use later on, are derived. In Section 2 the class $\mathrm{Reg}(\Delta)$ is defined and investigated, and the same is done in Section 3 for the class $\mathrm{CFL}(\Delta)$. Finally, Section 4 is devoted to comparing the classes $\mathrm{Reg}(\Delta)$ and $\mathrm{CFL}(\Delta)$ to the classes defined in [1].

1. Some specific encodings and their properties

It is assumed that the reader is familiar with the basic concepts of formal language theory as presented in [4] or in [5]. Here some notations and definitions used throughout this paper are given . Then some specific encodings are defined, and we derive some of their properties that we will need later on.

An *alphabet* Σ is a countable (i.e., finite or countably infinite) set whose elements are called *letters*. The *set of words over* Σ is denoted Σ^*, and e denotes the *empty word*. The identity of words is written as $=$, and the *concatenation* of words u and v is simply written as uv. Numerical superscripts are often used to abbreviate words, e.g., a^3 means *aaa*.

In general, $|x|$ denotes the *length* of a word x: $|e|=0$, $|xa|=|x|+1$ for all $x\in\Sigma^*$, $a\in\Sigma$. For a set S, $|S|$ denotes the *cardinality* of S. Each alphabet Σ can be indexed by an initial section of $\mathbb{N}-\{0\}$, i.e., if Σ is finite, then it can be written as $\Sigma=\{s_1,s_2,\ldots,s_n\}$, where $n=|\Sigma|$, and if Σ is infinite, then it can be written as $\Delta=\{s_1,s_2,\ldots,s_i,s_{i+1},\ldots\}$. Hence, we can define the *sum* $\|x\|$ for a word x: $\|e\|=0$, $\|xs_i\|=\|x\|+i$ for all $x\in\Sigma^*$, $s_i\in\Sigma$. Then $\|s_{i_1}s_{i_2}\cdots s_{i_r}\|=\sum_{j=1}^r i_j$, while $|s_{i_1}s_{i_2}\cdots s_{i_r}|=r$.

For all of this paper we fix two alphabets Δ and Σ_2 as follows: $\Delta=\{a_1,a_2,\ldots,a_i,a_{i+1},\ldots\}$, and $\Sigma_2=\{s_1,s_2\}$. In order to be able to compare classes of languages over Δ with classes of languages over Σ_2, we introduce some encodings.

Define the function $c\colon \Delta^*\to\mathbb{N}$ by $c(e)=0$ and

$$c(wa_i)=2^{\|wa_i\|-1}+c(w)\quad \text{for all } w\in\Delta^*,\ a_i\in\Delta.$$

Lemma 1.1 [8]. *The function c is a bijection from Δ^* onto $\mathbb{N}$.*

Proof. It can be seen easily by induction that

$$c(a_{i_1}a_{i_2}\cdots a_{i_r})=2^{\sum_{j=1}^r i_j-1}+2^{\sum_{j=1}^{r-1} i_j-1}+\cdots+2^{i_1-1}.$$

Hence, $c(a_{i_1}a_{i_2}\cdots a_{i_r})=\sum_{k=0}^{l}\mu_k\cdot 2^k$, where $l=\sum_{j=1}^r i_j-1$ $(=\|a_{i_1}a_{i_2}\cdots a_{i_r}\|-1)$, and

$$\mu_k=\begin{cases}1, & \text{if } k\in\left\{p \,\middle|\, p=\sum_{j=1}^{m} i_j-1 \text{ for some } m \text{ with } 1\le m\le r\right\},\\ 0, & \text{otherwise.}\end{cases}$$

Since $i_j\ge 1$ for all j, this implies that c is 1–1 and onto, i.e., c is a bijection from Δ^* onto $\mathbb{N}$. $\square$

Let $\mathrm{bin}\colon\mathbb{N}\to\Sigma_2^*$ be the mapping that, for each integer $n\in\mathbb{N}$, gives the binary representation $\mathrm{bin}(n)$, where s_1 and s_2 are interpreted as 0 and 1, respectively. Define $\tilde{c}\colon\Delta^*\to\Sigma_2^*$ by $\tilde{c}=\mathrm{bin}\circ c$, i.e., $\tilde{c}(w)$ is the binary representation of the integer $c(w)$.

Lemma 1.2. *For all non-empty words $u,v\in\Delta^*$, $\tilde{c}(uv)=\tilde{c}(v)\,\tilde{c}(u)$.*

Proof. From the proof of Lemma 1.1 we immediately derive that

$$\tilde{c}(a_{i_1}a_{i_2}\cdots a_{i_r})=s_2s_1^{i_r-1}s_2s_1^{i_{r-1}-1}\cdots s_2s_1^{i_1-1}.$$

Further, $c(a_k)=2^{k-1}$ implying $\tilde{c}(a_k)=\mathrm{bin}(2^{k-1})=s_2s_1^{k-1}$. Thus,

$$\tilde{c}(a_{i_1}a_{i_2}\cdots a_{i_r})=s_2s_1^{i_r-1}s_2s_1^{i_{r-1}-1}\cdots s_2s_1^{i_1-1}=\tilde{c}(a_{i_r})\,\tilde{c}(a_{i_{r-1}})\cdots\tilde{c}(a_{i_1}),$$

i.e., $\tilde{c}(uv)=\tilde{c}(v)\,\tilde{c}(u)$ for all non-empty words $u,v\in\Delta^*$. $\square$

The function $\alpha_2\colon\Sigma_2^*\to\mathbb{N}$ is defined by $\alpha_2(e)=0$ and $\alpha_2(ws_i)=2\alpha_2(w)+i$ for all

$w \in \Sigma_2^*$, $i = 1, 2$, i.e., a word $w \in \Sigma_2^*$ is simply interpreted as the 2-adic representation of the integer $\alpha_2(w)$. Hence, α_2 is a bijection from Σ_2^* onto $\mathbb{N}$. Let ϱ denote the *reversal function* defined by $\varrho(e) = e$ and $\varrho(ws) = s\varrho(w)$ for all $w \in \Sigma_2^*$, $s \in \Sigma_2$.

Lemma 1.3 *There exists a* GSM *mapping h satisfying the following two conditions for all integers* $n \geq 1$:

(i) $h(\varrho \circ \mathrm{bin}(n)) = \varrho \circ \alpha_2^{-1}(n)$,

(ii) $h^{-1}(\varrho \circ \alpha_2^{-1}(n)) \cap \Sigma_2^* \cdot s_2 = \{\varrho \circ \mathrm{bin}(n)\}$.

Proof. Consider the following algorithm:

Algorithm A

input: a non-empty word $s_{i_r} s_{i_{r-1}} \cdots s_{i_1} s_{i_0} \in \{0,1\}^* \; (= \Sigma_2^*)$;
begin $j := 0$;
    while $j \leq r$ **and** $s_{i_j} = 1$ **do**
    begin $s'_{i_j} := s_1$;
        $j := j + 1$
    end;
    if $j = r + 1$ **then** $m := r$ **else** $m := r - 1$;
    $s'_{i_j} := s_2$;
    $j := j + 1$;
    while $j < r$ **do**
    begin if $s_{i_j} = 1$ **then** $s'_{i_j} := s_2$ **else** $s'_{i_j} := s_1$;
        $j := j + 1$
    end;
    output: $f(s_{i_r} \cdots s_{i_1} s_{i_0}) = s'_{i_m} \cdots s'_{i_1} s'_{i_0}$
end.

Claim. *For all integers* $n \geq 1$, $\mathrm{f}(\mathrm{bin}(n)) = \alpha_2^{-1}(n)$.

Proof. Let n be an integer with $n \geq 1$. If $\mathrm{bin}(n) = 1^r$ for some $r \geq 1$, then $f(\mathrm{bin}(n)) = f(1^r) = s_1^r = \alpha_2^{-1}(n)$, otherwise, $\mathrm{bin}(n) = 1 s_{i_r} \cdots s_{i_1} 0 1^k$ for some $r, k \geq 0$. In the latter case,

$$f(\mathrm{bin}(n)) = s'_{i_r} \cdots s'_{i_1} s_2 s_1^k, \quad \text{where} \quad s'_{i_j} = \begin{cases} s_1, & \text{if } s_{i_j} = 0, \\ s_2, & \text{if } s_{i_j} = 1. \end{cases}$$

Hence, $f(\mathrm{bin}(n)) = \alpha_2^{-1}(n)$ also holds in this case. □

Now it is straightforward to develop a generalized sequential machine B from A such that on input $\varrho \circ \mathrm{bin}(n)$ $(n \geq 1)$ B outputs $\varrho \circ \alpha_2^{-1}(n)$. Let h denote the mapping computed by B. Then for all integers $n \geq 1$, $h(\varrho \circ \mathrm{bin}(n)) = \varrho \circ \alpha_2^{-1}(n)$. Further we have $h^{-1}(\varrho \circ \alpha_2^{-1}(n)) \cap \Sigma_2^* \cdot s_2 = \{(\varrho \circ \mathrm{bin}(n)\}$ for all $n \geq 1$, since $\varrho \circ \mathrm{bin}$ is a bijection from $\mathbb{N} - \{0\}$ onto $\Sigma_2^* \cdot s_2$, and $\varrho \circ \alpha_2^{-1}$ is a bijection from $\mathbb{N}$ onto Σ_2^*. □

With β we denote the function $\alpha_2^{-1} \circ c$. By Lemma 1.1, β is a bijection from Δ^* onto Σ_2^*.

Theorem 1.4 [6]. *The conjugation by β induces a bijection between the linear classes of word functions over Δ and the linear classes of word functions over Σ_2. In particular, $E_n(\Sigma_2) = \beta \circ E_n(\Delta) \circ \beta^{-1}$ for all $n \geq 1$.*

Here $E_n(\Sigma_2)$ ($E_n(\Delta)$) denotes the n-th class of the *Grzegorczyk hierarchy* over Σ_2 (Δ) [6,9]. Further, a class of word functions over Σ_2 (Δ) is called *linear*, if it contains the class $E_1(\Sigma_2)$ ($E_1(\Delta)$), and if it is closed under composition of functions and limited recursion.

Finally we introduce an encoding γ from Δ^* into Σ_2^* by defining

$$\gamma(a_i) = s_1^{i-1} s_2 \quad \text{for all } i \geq 1.$$

2. The class Reg(Δ) of regular languages over Δ

The class Reg(Δ) of regular languages over Δ is defined through the encoding γ. After showing that all the languages in Reg(Δ) are decidable, some closure properties of Reg(Δ) are derived. Finally, a syntactical characterization by means of Δ-expressions and a semantical characterization by means of finite Δ-automata are given for this class.

For a finite alphabet Σ, let Reg(Σ) denote the *class of regular languages over Σ*. Now the class Reg(Δ) of regular languages over Δ is defined as follows. Let L be a subset of Δ^*. Then $L \in \mathrm{Reg}(\Delta)$ if and only if $\gamma(L) \in \mathrm{Reg}(\Sigma_2)$.

Lemma 2.1. *Let L be a subset of Δ^*. Then the following four statements are equivalent:*

(i) $L \in \mathrm{Reg}(\Delta)$,
(ii) $\gamma(L) \in \mathrm{Reg}(\Sigma_2)$,
(iii) $\tilde{c}(L) \in \mathrm{Reg}(\Sigma_2)$,
(iv) $\beta(L) \in \mathrm{Reg}(\Sigma_2)$.

Proof. Statements (i) and (ii) are equivalent by definition of the class Reg(Δ). Let $w = a_{i_r} \cdots a_{i_1} a_{i_0} \in \Delta^+$ $(= \Delta^* - \{e\})$. Then

$$\tilde{c}(w) = \tilde{c}(a_{i_0}) \tilde{c}(a_{i_1}) \cdots \tilde{c}(a_{i_r}) = s_2 s_1^{i_0 - 1} s_2 s_1^{i_1 - 1} \cdots s_2 s_1^{i_r - 1}$$

by Lemma 1.2,

$$\gamma(w) = s_1^{i_r - 1} s_2 \cdots s_1^{i_1 - 1} s_2 s_1^{i_0 - 1} s_2 = \varrho \circ \tilde{c}(w),$$

and

$$\beta(w) = \alpha_2^{-1} \circ c(w) = \varrho \circ h \circ \varrho \circ \mathrm{bin} \circ c(w) = \varrho \circ h \circ \varrho \circ \tilde{c}(w)$$

by Lemma 1.3. Thus, if $e \notin L$, then $\gamma(L) = \varrho \circ \tilde{c}(L)$, and $\beta(L) = \varrho \circ h \circ \varrho \circ \tilde{c}(L)$, and if $e \in L$, then $\gamma(L) = \varrho(\tilde{c}(L) - \{s_1\}) \cup \{e\}$, and $\beta(L) = \varrho \circ h \circ \varrho(\tilde{c}(L) - \{s_1\}) \cup \{e\}$, since $\tilde{c}(e) = s_1$, $\gamma(e) = e$, and $\beta(e) = e$. Now the equivalence of statements (ii) to (iv) follows from well known closure properties of the class $\mathrm{Reg}(\Sigma_2)$. $\square$

From Lemma 2.1 we can immediately derive

Theorem 2.2. *Each language L in Reg(Δ)* is $\mathrm{E}_1(\Delta)$*-decidable.*

Proof. Let L be a language from Reg(Δ). We have to show that the characteristic function χ_L of L is in $E_1(\Delta)$. Since $L \in \mathrm{Reg}(\Delta)$, $\beta(L) \in \mathrm{Reg}(\Sigma_2)$ by Lemma 2.1. This implies in particular that the characteristic function χ_1 of $\beta(L)$ is in $E_1(\Sigma_2)$, and hence, the function $\chi = \beta^{-1} \circ \chi_1 \circ \beta$ is in $E_1(\Delta)$ by Theorem 1.4. But for each $w \in \Delta^*$,

$$\chi(w) = \beta^{-1} \circ \chi_1(\beta(w)) = \begin{cases} \beta^{-1}(s_1) = a_1, & \text{if } \beta(w) \in \beta(L), \\ \beta^{-1}(e) = e, & \text{if } \beta(w) \notin \beta(L), \end{cases}$$

i.e.,

$$\chi(w) = \left.\begin{cases} a_1, & \text{if } w \in L \\ e, & \text{if } w \notin L \end{cases}\right\} = \chi_L(w). \quad \square$$

From Ogden's lemma for regular sets over Σ_2, and from the definition of the class Reg(Δ) we get the following version of Ogden's lemma for languages in Reg(Δ).

Lemma 2.3. *Let L be a language from Reg(Δ).*

(i) *There is an integer n such that, for each word $w \in L$ of length $|w| \geq n$, there exists a partition $w = xyz$ satisfying $1 \leq |y| \leq n$ and $\{xy^m z \mid m \geq 0\} \subseteq L$.*

(ii) *There is an integer $n' \geq 3$ such that, for each word $w \in L$ and each partition $w = w_1 a_i w_2$, where $a_i \in \Delta$ with $i \geq n'$, there exists an integer j satisfying $1 \leq j \leq n' - 2$ and $\{w_1 a_{i+mj} w_2 \mid m \geq -1\} \subseteq L$.*

From part (ii) of Lemma 2.3 we deduce that the language $M = \{a_i a_1 a_i \mid i \geq 1\} \subseteq \Delta^*$ is not contained in the class Reg(Δ). In the following some closure properties and some non-closure properties of the class Reg(Δ) are derived.

Corollary 2.4. (i) *The class* Reg(Δ) *is closed under union, intersection, complementation, concatenation, Kleene closure (= star-operation), and reversal.*

(ii) Reg(Δ) *is not closed under ε-free homomorphisms, projections, or inverse projections.*

(iii) Reg(Δ) *is closed under projections onto Δ-regular subalphabets and inverse projections onto Δ-regular subalphabets.*

Proof. Part (i) follows immediately from corresponding closure properties of

$\mathrm{Reg}(\Sigma_2)$ and from Lemma 2.1. A subalphabet Ω of Δ is Δ-regular, if $\Omega \in \mathrm{Reg}(\Delta)$ holds. By again using well known closure properties of $\mathrm{Reg}(\Sigma_2)$, we can derive part (iii) from Lemma 2.1.

For proving part (ii) let $L_1 = \{a_i \mid i \geq 1\}$ and $L_2 = \{e\}$. Then L_1 and L_2 are contained in $\mathrm{Reg}(\Delta)$. Define a mapping $\varphi : \Delta^* \to \Delta^*$ by $\varphi(a_i) = a_i a_1 a_i$ for all $i \geq 1$. Then $\varphi(L_1) = \{a_i a_1 a_i \mid i \geq 1\} = M \notin \mathrm{Reg}(\Delta)$ implying that the class $\mathrm{Reg}(\Delta)$ is not closed under ε-free homomorphisms. Finally, let Ω be a subalphabet of Δ such that $\Omega \notin \mathrm{Reg}(\Delta)$. Since not all the subsets of $\{s_1^i s_2 \mid i \geq 0\}$ are regular over Σ_2, we see from the definition of $\mathrm{Reg}(\Delta)$ that a subalphabet Ω of this form exists. Now let Π_Ω denote the projection from Δ^* onto Ω^*, i.e.,

$$\Pi_\Omega(a_i) = \begin{cases} a_i, & \text{if } a_i \in \Omega, \\ e, & \text{otherwise.} \end{cases}$$

Then $\Pi_\Omega(L_1) \cap \Delta = \Omega$, and $\Pi_\Omega^{-1}(L_2) = \{w \in \Delta^* \mid \Pi_\Omega(w) = e\} = (\Delta - \Omega)^*$. Thus, $\Omega = \Pi_\Omega(L_1) \cap \Delta$, and $\Omega = (\Pi_\Omega^{-1}(L_2))^c \cap \Delta$, where L^c stands for the complement $\Delta^* - L$ of L. Hence, the class $\mathrm{Reg}(\Delta)$ is neither closed under Π_Ω nor under Π_Ω^{-1}. $\square$

The regular languages over Σ_2 can be described by regular expressions. By restricting our attention to regular languages over Σ_2 that are contained in $(\{s_1\}^* \cdot s_2)^*$ we get an according characterization for the languages in $\mathrm{Reg}(\Delta)$.

Definition 2.5. The *Δ-expressions* and the sets that they denote are defined recursively as follows:

(i) $\emptyset$ is a Δ-expression denoting the empty set.

(ii) ε is a Δ-expression denoting the set $\{e\}$.

(iii) For each $i \geq 1$, a_i is a Δ-expression denoting the set $\{a_i\}$.

(iv) For each $i, j \geq 1$, $B_{i,j}$ is a Δ-expression denoting the set $\{a_{i+kj} \mid k \geq 0\}$.

(v) If r and s are Δ-expressions denoting the sets R and S, respectively, then $(r+s)$, (rs), and (r^*) are Δ-expressions that denote the sets $R \cup S$, RS, and R^*, respectively.

If r is a Δ-expression, then $L(r)$ denotes the set described by r.

From the characterization of regular subsets of $(\{s_1\}^* \cdot s_2)^*$ by regular expressions the following theorem can be derived easily.

Theorem 2.6. *Let L be a subset of Δ^*. The following two statements are equivalent:*

(i) $L \in \mathrm{Reg}(\Delta)$;

(ii) *there exists a Δ-expression r such that $L = L(r)$.*

From this characterization of $\mathrm{Reg}(\Delta)$ we see that $\mathrm{Reg}(\Sigma) = \mathrm{Reg}(\Delta) \cap \mathscr{P}(\Sigma^*)$ for each finite subalphabet Σ of Δ. Here $\mathscr{P}(\Sigma^*)$ denotes the set of all subsets of Σ^*. Further we can derive

Theorem 2.7. *The class* Reg(Δ) *is closed under finite homomorphisms.*

Proof. A homomorphism $\varphi: \Delta^* \to \Delta^*$ is called *finite*, if $\varphi(\Delta)$ is a finite subset of Δ^*. Let φ be a finite homomorphism with $\varphi(\Delta) = \{w_1, w_2, \ldots, w_m\} \subseteq \Delta^*$, and let $L \in \mathrm{Reg}(\Delta)$. We must show that $\varphi(L) \in \mathrm{Reg}(\Delta)$. This is done by induction on the Δ-expression r denoting L.

If $r = \emptyset$, or if $r = \varepsilon$, then $\varphi(L) = L$ implying $\varphi(L) \in \mathrm{Reg}(\Delta)$. If $r = a_i$ for some $i \geq 1$, then $L = \{a_i\}$, and so $\varphi(L) = \{\varphi(a_i)\}$, which clearly is in Reg(Δ). Further, if $r = B_{i,j}$ for some $i, j \geq 1$, then $L = \{a_{i+kj} \mid k \geq 0\} \subseteq \Delta$. Hence, there exists a subset $\{w_1, \ldots, w_p\} \subseteq \varphi(\Delta)$ such that $\varphi(L) = \{w_1, \ldots, w_p\}$. Thus, $\varphi(L) \in \mathrm{Reg}(\Delta)$. Finally, if $r = (r_1 + r_2)$, $r = (r_1 r_2)$, or $r = (r_1^*)$, then $\varphi(L) \in \mathrm{Reg}(\Delta)$ follows from the induction hypothesis applied to r_1 and r_2, respectively, and from the fact that φ is a homomorphism, i.e., $\varphi(uv) = \varphi(u)\varphi(v)$ for all $u, v \in \Delta^*$. □

Finally, we want to carry over the semantical characterization of the regular subsets of $(\{s_1\}^* \cdot s_2)^*$ by means of finite automata to Reg(Δ). For doing so, we define the *class* $A(\Delta)$ *of finite* Δ*-automata* as follows.

Definition 2.8. A *finite* Δ*-automaton* $\mathfrak{A}$ is denoted by a 4-tuple (Q, q_0, F, δ), where Q is a finite set of *states*, $q_0 \in Q$ is the *initial state*, $F \subseteq Q$ is the set of *accepting states*, and δ is the *transition function* mapping $Q \times \Delta$ into Q and satisfying the condition $(*)$:

$$(*) \qquad \forall q \in Q,\ \forall i, j \geq 2: \quad \delta(q, a_i) = \delta(q, a_j).$$

We figure a finite Δ-automaton $\mathfrak{A} = (Q, q_0, F, \delta)$ as a *finite control*, which is in some state from Q, reading a sequence of symbols from Δ written on a tape. In one move the finite Δ-automaton $\mathfrak{A}$ in state q and scanning symbol a_i enters state $\delta(q, a_i)$ and either moves its head one symbol to the right, if $i = 1$, or substitutes a_i by a_{i-1}, *if* $i > 1$. If the cell scanned is empty, then $\mathfrak{A}$ stops.

If the tape inscription of $\mathfrak{A}$ is ua_iv with $u, v \in \Delta^*$ and $a_i \in \Delta$, and if $\mathfrak{A}$ is in state $q \in Q$ scanning the tape cell containing the letter a_i, then this *configuration* of $\mathfrak{A}$ can be described by uqa_iv. Now the behavior of $\mathfrak{A}$ can be defined formally by a function $\mathrm{NEXT}_{\mathfrak{A}}$ that, for each configuration of $\mathfrak{A}$, gives the corresponding successor configuration.

Definition 2.9. Let $\mathfrak{A} = (Q, q_0, F, \delta)$ be a finite Δ-automaton.

(a) For all $u, v \in \Delta^*$, $a_i \in \Delta$, $q \in Q$,

$$\mathrm{NEXT}_{\mathfrak{A}}(uqa_iv) = \begin{cases} ua_1 q'v, & \text{if } i = 1, \\ uq'a_{i-1}v, & \text{if } i > 1, \end{cases}$$

where $q' = \delta(q, a_i)$. Let $\xrightarrow[\mathfrak{A}]{}$ denote the *transition* between configurations induced by $\mathrm{NEXT}_{\mathfrak{A}}$, and let $\xrightarrow[\mathfrak{A}]{*}$ denote the reflexive and transitive closure of $\xrightarrow[\mathfrak{A}]{}$.

(b) Let $L(\mathfrak{A}) = \{w \in \Delta^* \mid \exists q \in F\colon q_0 w \xrightarrow[\mathfrak{A}]{*} a_1^{|w|} q\}$. Then L is the set of words that cause $\mathfrak{A}$ to halt in an accepting state after starting from the configuration $q_0 w$. $L(\mathfrak{A})$ is called the *language accepted by* $\mathfrak{A}$.

The finite Δ-automaton is a restriction of the Turing machine model over Δ as defined in [6]. With $\mathrm{FA}(\Sigma_2)$ we denote the class of deterministic finite automata over Σ_2 as defined in [5]. Then $\mathrm{Reg}(\Sigma_2) = \{L \subseteq \Sigma_2^* \mid \exists \mathfrak{A} \in \mathrm{FA}(\Sigma_2)\colon L = L(\mathfrak{A})\}$. Now we can state the last result of this section.

Theorem 2.10. *Let L be a subset of Δ^*. Then the following two statements are equivalent:*

(i) $L \in \mathrm{Reg}(\Delta)$;

(ii) *there exists a finite Δ-automaton $\mathfrak{A}$ that accepts L.*

Proof. Assume that $\mathfrak{A} = (Q, q_0, F, \delta)$ is a finite Δ-automaton such that $L = L(\mathfrak{A})$. We define a finite automaton $\mathfrak{A}' \in \mathrm{FA}(\Sigma_2)$ as follows: $\mathfrak{A}' = (Q, q_0, F, \delta')$ with $\delta'(q, s_1) = q_1$, if $\delta(q, a_2) = q_1$, and $\delta'(q, s_2) = q_1$, *if* $\delta(q, a_1) = q_1$, for all $q, q_1 \in Q$. It can be seen easily that $L(\mathfrak{A}') \cap (\Sigma_2^* \cdot s_2 \cup \{e\}) = \gamma(L)$. Since $\mathfrak{A}' \in \mathrm{FA}(\Sigma_2)$, $L(\mathfrak{A}') \in \mathrm{Reg}(\Sigma_2)$, and so $\gamma(L) \in \mathrm{Reg}(\Sigma_2)$ implying $L \in \mathrm{Reg}(\Delta)$.

On the other hand, if $L \in \mathrm{Reg}(\Delta)$, then $\gamma(L) \in \mathrm{Reg}(\Sigma_2)$ implying that $\gamma(L) = L(\mathfrak{A}')$ for some finite automaton $\mathfrak{A}' = (Q, q_0, F, \delta') \in \mathrm{FA}(\Sigma_2)$. Define a finite Δ-automaton $\mathfrak{A}$ as follows: $\mathfrak{A} = (Q, q_0, F, \delta)$ with $\delta(q, a_1) = q_1$, if $\delta'(q, s_2) = q_1$, and $\delta(q, a_i) = q_1$, if $\delta'(q, s_1) = q_1$, for all $i \geq 2$ and all $q, q_1 \in Q$. Then $\mathfrak{A} \in A(\Delta)$, and it is straightforward to check that $L = L(\mathfrak{A})$. $\square$

3. The class CFL(Δ) of context-free languages over Δ

Here the encoding γ is used to define the class $\mathrm{CFL}(\Delta)$ of context-free languages over Δ, all of which are decidable. Then some closure properties and some non-closure properties of the class $\mathrm{CFL}(\Delta)$ are proved.

For a finite alphabet Σ, $\mathrm{CFL}(\Sigma)$ denotes the *class of context-free languages over* Σ. Now the class $\mathrm{CFL}(\Delta)$ of context-free languages over Δ is defined as follows. Let L be a subset of Δ^*. Then $L \in \mathrm{CFL}(\Delta)$ if and only if $\gamma(L) \in \mathrm{CFL}(\Sigma_2)$. Obviously, we have $\mathrm{Reg}(\Delta) \subsetneq \mathrm{CFL}(\Delta)$.

Since $\mathrm{CFL}(\Sigma_2)$ is closed under reversal, union, intersection with regular sets, GSM mappings, and inverse GSM mappings, the proof of Lemma 2.1 also applies to context-free languages, thus giving

Lemma 3.1. *Let L be a subset of Δ^*. Then the following four statements are equivalent:*

(i) $L \in \mathrm{CFL}(\Delta)$,

(ii) $\gamma(L) \in \mathrm{CFL}(\Sigma_2)$,
(iii) $\tilde{c}(L) \in \mathrm{CFL}(\Sigma_2)$,
(iv) $\beta(L) \in \mathrm{CFL}(\Sigma_2)$.

Now we can easily show that all the languages in $\mathrm{CFL}(\Delta)$ are decidable. In fact, we prove

Theorem 3.2. *Each language L in* $\mathrm{CFL}(\Delta)$ *is $E_2(\Delta)$-decidable.*

Proof. Let $L \in \mathrm{CFL}(\Delta)$. By Lemma 3.1 this means that $\beta(L) \in \mathrm{CFL}(\Sigma_2)$. The Cocke–Kasami–Younger algorithm (see, e.g. [4]) decides membership for $\beta(L)$ in time $\mathrm{O}(n^3)$ with space $\mathrm{O}(n^2)$. Thus, the characteristic function χ_1 of $\beta(L)$ is in $E_2(\Sigma_2)$, and so the function $\chi = \beta^{-1} \circ \chi_1 \circ \beta$ is in $E_2(\Delta)$ by Theorem 1.4. As in the proof of Theorem 2.2, χ actually is the characteristic function of L implying that L is $E_2(\Delta)$-decidable. □

Before we come to state some of the closure properties and non-closure properties of the class $\mathrm{CFL}(\Delta)$, we want to characterize the context-free subalphabets of Δ.

Theorem 3.3. *Let Ω be a subalphabet of Δ, and let $\Omega_1 = \{a_1^i \mid a_i \in \Omega\}$ be the unary encoding of Ω. Then the following four statements are equivalent:*
(i) *Ω is a Δ-regular subalphabet of Δ, i.e., $\Omega \in \mathrm{Reg}(\Delta)$;*
(ii) *Ω is a Δ-context-free subalphabet of Δ, i.e., $\Omega \in \mathrm{CFL}(\Delta)$;*
(iii) $\Omega_1 \in \mathrm{Reg}(\{a_1\})$,
(iv) $\Omega_1 \in \mathrm{CFL}(\{a_1\})$.

Proof. It is well known that parts (iii) and (iv) are equivalent [4]. Therefore, it suffices to prove the equivalence of (i) and (iii) and of (ii) and (iv), respectively. For that define a homomorphism $\varphi\colon \Sigma_2^* \to \{a_1\}^*$ by $\varphi(s_1) = \varphi(s_2) = a_1$. Now for $\Omega \in \mathrm{CFL}(\Delta)$ $(\mathrm{Reg}(\Delta))$, $\gamma(\Omega) \in \mathrm{CFL}(\Sigma_2)$ $(\mathrm{Reg}(\Sigma_2))$ implying that $\Omega_1 = \varphi(\gamma(\Omega)) \in \mathrm{CFL}(\{a_1\})$ $(\mathrm{Reg}(\{a_1\}))$. On the other hand, if $\Omega_1 \in \mathrm{CFL}(\{a_1\})$ $(\mathrm{Reg}(\{a_1\}))$, then $\varphi^{-1}(\Omega_1) = \{w \in \Sigma_2^* \mid a_{|w|} \in \Omega\} \in \mathrm{CFL}(\Sigma_2)$ $(\mathrm{Reg}(\Sigma_2))$. Now $\gamma(\Omega) = \varphi^{-1}(\Omega_1) \cap (\{s_1\}^* \cdot s_2)$, and so $\gamma(\Omega) \in \mathrm{CFL}(\Sigma_2)$ $(\mathrm{Reg}(\Sigma_2))$ implying $\Omega \in \mathrm{CFL}(\Delta)$ $(\mathrm{Reg}(\Delta))$. □

Hence, the context-free subalphabets of Δ coincide with the regular subalphabets, and they are in 1–1 correspondence with the regular languages over a single-letter alphabet.

Corollary 3.4. (i) *The class* $\mathrm{CFL}(\Delta)$ *is closed under union, concatenation, Kleene closure, intersection with languages from* $\mathrm{Reg}(\Delta)$, *and reversal.*
(ii) $\mathrm{CFL}(\Delta)$ *is not closed under intersection or complementation.*
(iii) $\mathrm{CFL}(\Delta)$ *is not closed under ε-free homomorphisms, finite homomorphisms, projections, or inverse projections.*

(iv) $\mathrm{CFL}(\Delta)$ *is closed under projections onto* Δ*-context-free subalphabets and inverse projections onto* Δ*-context-free subalphabets.*

Proof. Parts (i) and (ii) follow immediately from corresponding properties of $\mathrm{CFL}(\Sigma_2)$ and from Lemmas 2.1 and 3.1, while part (iv) follows from well known closure properties of $\mathrm{CFL}(\Sigma_2)$ together with Theorem 3.3. It remains to prove part (iii).

Let $L_1=\{a_i \mid i\geq 1\}$, $L_2=\{e\}$, and $L_3=\{a_ia_1^i \mid i\geq 2\}$. Then L_1, L_2, and L_3 are in $\mathrm{CFL}(\Delta)$. Further, let A be a non-recursive subset of $\mathbb{N}$, and let $\varphi: \Delta^*\to\Delta^*$ be defined by $\varphi(a_1)=a_1$ and, for all $i\geq 2$,

$$\varphi(a_i)=\begin{cases} a_2, & \text{if } i\in A,\\ a_3, & \text{if } i\notin A.\end{cases}$$

Then $\varphi(\Delta)=\{a_1,a_2,a_3\}$, i.e., φ is an ε-free finite homomorphism. Now $\varphi(L_3)=\{a_2a_1^i \mid i\geq 2 \text{ and } i\in A\}\cup\{a_3a_1^i \mid i\geq 2 \text{ and } i\notin A\}$, and so, for all $i\geq 2$, $i\in A$ if and only if $a_2a_1^i\in\varphi(L_3)$. Thus, the language $\varphi(L_3)$ is non-recursive implying that the class $\mathrm{CFL}(\Delta)$ is neither closed under ε-free homomorphisms nor under finite homomorphisms. Finally, let Ω be a subalphabet of Δ with $\Omega\notin\mathrm{CFL}(\Delta)$, and let Π_Ω denote the projection from Δ^* onto Ω^*. Then $\Omega=\Pi_\Omega(L_1)\cap\Delta$ implying that $\mathrm{CFL}(\Delta)$ is not closed under projections. Further, $\Pi_\Omega^{-1}(L_2)=\{w\in\Delta^* \mid \Pi_\Omega(w)=e\}=(\Delta-\Omega)^*$. Assume that $\Pi_\Omega^{-1}(L_2)\in\mathrm{CFL}(\Delta)$. Then $\Delta-\Omega=\Pi_\Omega^{-1}(L_2)\cap\Delta\in\mathrm{CFL}(\Delta)$ by (i), and hence $\Delta-\Omega\in\mathrm{Reg}(\Delta)$ by Theorem 3.3. Since $\mathrm{Reg}(\Delta)$ is closed under complementation and intersection, this implies that $\Omega\in\mathrm{Reg}(\Delta)$, and hence, $\Omega\in\mathrm{CFL}(\Delta)$, a contradiction. Thus, $\Pi_\Omega^{-1}(L_2)\notin\mathrm{CFL}(\Delta)$ proving that $\mathrm{CFL}(\Delta)$ is not closed under inverse projections. □

Let Σ be a finite subalphabet of Δ. Then for each subset $L\subseteq\Sigma^*$, we have $L\in\mathrm{CFL}(\Sigma)$ if and only if $\gamma(L)\in\mathrm{CFL}(\Sigma_2)$, since the restriction of γ to Σ^* is a homomorphism from Σ^* into Σ_2^*. Thus, $\mathrm{CFL}(\Sigma)=\mathrm{CFL}(\Delta)\cap\mathscr{P}(\Sigma^*)$.

Finally, we want to mention the fact that the class $\mathrm{CFL}(\Delta)$ can also be characterized in a semantical way by automata. Just as finite automata were generalized to finite Δ-automata, one can generalize pushdown automata to Δ-pushdown automata. Then the class $\mathrm{CFL}(\Delta)$ is exactly the class of languages over Δ that are accepted by Δ-pushdown automata. For details see [7].

4. Comparing Reg(Δ) and CFL(Δ) to other classes of languages over Δ

For a finite alphabet Σ, the classes $\mathrm{Reg}(\Sigma)$ of regular languages over Σ and $\mathrm{CFL}(\Sigma)$ of context-free languages over Σ have several nice characteristic properties. For example, a language $L\subseteq\Sigma^*$ is regular, if and only if there exist a finite monoid M, a subset R of M, and a homomorphism $\varphi: \Sigma^*\to M$ such that $L=\varphi^{-1}(R)$ (cf., e.g., [2]), and a language $L'\subseteq\Sigma^*$ is context-free, if and only if for each (finite) subalphabet Σ' of Σ, the set $\Pi_{\Sigma'}(L')$ is in CFL(Σ'). Here $\Pi_{\Sigma'}$ denotes the projection from Σ^* onto Σ'^*.

In [1] Autebert, Beauquier, and Boasson use several of these characteristic properties of Reg(Σ) and CFL(Σ), respectively, to define classes of languages over Δ. After restating their definitions, we compare the classes Reg(Δ) and CFL(Δ) to the classes defined in this way.

First we consider those classes that are derived from properties of Reg(Σ).

Definition 4.1. (i) Rat(Δ) is the family of rational subsets of Δ^*, i.e., Rat(Δ) is the least family of subsets of Δ^*, that contains the set $\emptyset$ and $\{a_i\}$ for all $i \geq 1$, and that is closed under union, concatenation, and Kleene closure [2].

(ii) R-Rat(Δ) is the least family of subsets of Δ^*, that contains all the Δ-regular subalphabets of Δ, and that is closed under union, concatenation, and Kleene closure.

(iii) N-Rat(Δ) is the least family of subsets of Δ^*, that contains all subalphabets of Δ, and that is closed under union, concatenation, and Kleene closure [1].

(iv) A language $L \subseteq \Delta^*$ is in H-Rat, if and only if, for each finite alphabet Σ and each finite homomorphism $\varphi: \Delta^* \to \Sigma^*$, $\varphi(L) \in \mathrm{Reg}(\Sigma)$ [1].

(v) A language $L \subseteq \Delta^*$ is in Π-Rat, if and only if, for each finite subalphabet Σ of Δ, $\Pi_\Sigma(L) \in \mathrm{Reg}(\Sigma)$ [1].

(vi) Rec(Δ) is the family of recognizable subsets of Δ^*, i.e., a language $L \subseteq \Delta^*$ is in Rec(Δ), if and only if there are a finite monoid M, a subset R of M, and a homomorphism $\varphi: \Delta^* \to M$ satisfying $L = \varphi^{-1}(R)$ [1,2].

We have the following chain of inclusions.

Theorem 4.2.

$$\begin{aligned}\bigcup_{\substack{\Sigma \subseteq \Delta \\ \Sigma \text{ finite}}} \mathrm{Reg}(\Sigma) = \mathrm{Rat}(\Delta) &\subsetneq \text{R-Rat}(\Delta) = \mathrm{Reg}(\Delta) \subsetneq \mathrm{Rec}(\Delta) \\ &= \mathrm{N}-\mathrm{Rat}(\Delta) \subsetneq \text{H-Rat} \subsetneq \Pi\text{-Rat}.\end{aligned}$$

Proof. Obviously,

$$\mathrm{Rat}(\Delta) = \bigcup_{\substack{\Sigma \subseteq \Delta \\ \Sigma \text{ finite}}} \mathrm{Reg}(\Sigma).$$

Since $\Delta \in$ R-Rat(Δ), we also have Rat(Δ) $\subsetneq$ R-Rat(Δ) immediately. The class Reg(Δ) is closed under union, concatenation, and Kleene closure implying that R-Rat(Δ) $\subseteq$ Reg(Δ). Now the characterization of Reg(Δ) by Δ-expressions (Theorem 2.6) implies R-Rat(Δ) = Reg(Δ). By Theorem 3.3 there exist subalphabets of Δ that are not Δ-regular, which implies that R-Rat(Δ) $\subsetneq$ N-Rat(Δ). The remaining inclusions are from [1]. □

In particular, we conclude from Theorem 4.2

Corollary 4.3. *The class* Rec(Δ) *contains non-recursive languages.*

On the other hand we have a nice characterization of the languages in Rec(Δ) by means of their syntactic monoids. Before giving this characterization let us first recall the definition of the syntactic monoid for a language $L \subseteq \Delta^*$.

Definition 4.4. For a language $L \subseteq \Delta^*$, the *syntactic congruence* $\approx$ is defined as follows. Let $u, v \in \Delta^*$. Then $u \approx v$ if and only if, for all $x, y \in \Delta^*$, $xuy \in L$ is equivalent to $xvy \in L$. The monoid $M_L = \Delta^*/\approx\, = \{[w]_\approx \mid w \in \Delta^*\}$ is the *syntactic monoid* of L. Here $[w]_\approx$ denotes the congruence class of w.

Theorem 4.5. *Let L be a subset of* Δ^*. *Then the following two statements are equivalent:*

(i) $L \in \mathrm{Rec}(\Delta)$;

(ii) *the syntactic monoid* M_L *of L is finite.*

Proof. Let L be a subset of Δ^*. If $L \in \mathrm{Rec}(\Delta)$, then there are a finite monoid M, a subset R of M, and a homomorphism $\varphi: \Delta^* \to M$ such that $L = \varphi^{-1}(R)$. Define a congruence $\sim$ on Δ^* by $u \sim v$ if and only if $\varphi(u) = \varphi(v)$. Now let $u, v \in \Delta^*$ with $u \sim v$. Then for all $x, y \in \Delta^*$, $xuy \sim xvy$, implying $xuy \in L$ if and only if $xvy \in L$, i.e., $u \approx v$. Hence, the congruence $\sim$ is a refinement of the syntactic congruence $\approx$. Since there are only finitely many congruence classes with respect to $\sim$, there are only finitely many congruence classes with respect to $\approx$. Thus, the syntactic monoid M_L is finite.

Conversely, assume that the syntactic monoid M_L is finite. Define a homomorphism $\varphi: \Delta^* \to M_L$ through $\varphi(a_i) = [a_i]_\approx$, and a subset R of M through $R = \{[w]_\approx \mid w \in L\}$. Obviously, we have $L \subseteq \varphi^{-1}(R)$. On the other hand, if $x \in \varphi^{-1}(R)$, then $\varphi(x) = [x]_\approx = [w]_\approx$ for some $w \in L$, i.e., $x \approx w$. Since $w \in L$, this implies $x \in L$. Thus, $L = \varphi^{-1}(R)$, and so $L \in \mathrm{Rec}(\Delta)$. □

Since Reg(Δ) is a proper subset of Rec(Δ), this gives

Corollary 4.6. *For each language* $L \in \mathrm{Reg}(\Delta)$, *the syntactic monoid* M_L *is finite. However, there are languages* $L \subseteq \Delta^*$ *that have finite syntactic monoids, but that are not in* Reg(Δ).

Now we introduce those classes that are derived from properties of CFL(Σ) [1].

Definition 4.7. (i) A language $L \subseteq \Delta^*$ is in N-Alg(Δ), if and only if there exist a finite alphabet Σ, a language $M \in \mathrm{CFL}(\Sigma)$, and an alphabetic homomorphism $\alpha: \Delta^* \to \Sigma^*$ such that $L = \alpha^{-1}(M)$. Here a homomorphism is called *alphabetic*, if $\alpha(\Delta) \subseteq \Sigma \cup \{e\}$ holds.

(ii) A language $L \subseteq \Delta^*$ is in H-Alg, if and only if, for each finite alphabet Σ and each finite homomorphism $\varphi: \Delta^* \to \Sigma^*$, $\varphi(L) \in \text{CFL}(\Sigma)$.

(iii) A language $L \subseteq \Delta^*$ is in Π-Alg, if and only if, for each finite subalphabet Σ of Δ, $\Pi_\Sigma(L) \in \text{CFL}(\Sigma)$.

Here, we have the following chain of inclusions.

Lemma 4.8. $\text{N-Alg}(\Delta) \subsetneq \text{H-Alg} \subsetneq \Pi\text{-Alg}$.

Proof. Let $L \in \text{N-Alg}(\Delta)$. Then there exist a finite alphabet Σ, a language $M \in \text{CFL}(\Sigma)$, and an alphabetic homomorphism $\alpha: \Delta^* \to \Sigma^*$ such that $L = \alpha^{-1}(M)$. Without loss of generality we may assume that Σ is a subalphabet of Δ implying that $\text{CFL}(\Sigma) \subseteq \text{H-Alg}$. Hence, $M \in \text{H-Alg}$. Since the class H-Alg is closed under inverse alphabetic homomorphisms [1], we have $L = \alpha^{-1}(M) \in \text{H-Alg}$. Thus, $\text{N-Alg}(\Delta) \subseteq \text{H-Alg}$.

Consider the language $L = \{a_i^2 \mid i \geq 1\}$. Then for each finite homomorphism φ, the image $\varphi(L)$ is finite implying that $L \in \text{H-Rat} \subseteq \text{H-Alg}$. On the other hand, $L \notin \text{N-Alg}(\Delta)$ as can be seen easily. Hence, $\text{N-Alg}(\Delta) \subsetneq \text{H-Alg}$. The inclusion $\text{H-Alg} \subsetneq \Pi\text{-Alg}$ is proved in [1]. □

From the proof of Lemma 4.8 we see that H-Rat is not a subclass of N-Alg(Δ). Further non-inclusions are the following.

Lemma 4.9. (i) $\text{N-Alg}(\Delta) \not\subseteq \Pi\text{-Rat}$.

(ii) $\Pi\text{-Rat} \not\subseteq \text{H-Alg}$.

Proof. Take $L_1 = \{a_1^i a_2 a_1^i \mid i \geq 1\}$. Then $L_1 \in \text{N-Alg}(\Delta)$, but $L_1 \notin \Pi\text{-Rat}$ implying that $\text{N-Alg}(\Delta) \not\subseteq \Pi\text{-Rat}$. Let $L_2 = \{a_i^i a_1 a_i^i a_1 a_i^i \mid i \geq 2\}$, and let Σ be a finite subalphabet of Δ. Then

$$\Pi_\Sigma(L_2) = \begin{cases} \{a_i^i a_1 a_i^i a_1 a_i^i \mid i \geq 2 \text{ with } a_i \in \Sigma\} \cup \{a_1^2\}, & \text{if } a_1 \in \Sigma, \\ \{a_i^{3i} \mid i \geq 2 \text{ with } a_i \in \Sigma\} \cup \{e\}, & \text{if } a_1 \notin \Sigma. \end{cases}$$

Since Σ is finite, also $\Pi_\Sigma(L_2)$ is finite, and hence, $L_2 \in \Pi\text{-Rat}$. Define $\varphi: \Delta^* \to \Sigma_2^*$ by $\varphi(a_1) = s_1$, and $\varphi(a_i) = s_2$ for all $i \geq 2$. Then φ is a finite homomorphism with $\varphi(L_2) = \{s_2^i s_1 s_2^i s_1 s_2^i \mid i \geq 2\}$, and so $\varphi(L_2)$ is not in $\text{CFL}(\Sigma_2)$. Thus, $L_2 \notin \text{H-Alg}$ implying that $\Pi\text{-Rat} \not\subseteq \text{H-Alg}$. □

Now we can deduce the following proper inclusions.

Lemma 4.10. (i) $\text{N-Rat}(\Delta) \subsetneq \text{N-Alg}(\Delta)$.

(ii) $\text{H-Rat} \subsetneq \text{H-Alg}$.

(iii) $\Pi\text{-Rat} \subsetneq \Pi\text{-Alg}$.

Proof. A language $L \subseteq \Delta^*$ is in N-Rat(Δ), if and only if there are a finite alphabet

Σ, a language $M \in \mathrm{Reg}(\Sigma)$, and an alphabetic homomorphism $\alpha\colon \Delta^* \to \Sigma^*$ such that $L = \alpha^{-1}(M)$ [1]. Hence, $\text{N-Rat}(\Delta) \subseteq \text{N-Alg}(\Delta)$. Since $\text{N-Alg}(\Delta) \not\subseteq \Pi\text{-Rat}$, this inclusion is proper. Parts (ii) and (iii) are easy consequences of the definitions. $\square$

From Theorem 4.2, from Lemmas 4.8–4.10, and from the fact that $\text{H-Rat} \not\subseteq \text{N-Alg}(\Delta)$ (cf. proof of Lemma 4.8) we get

Lemma 4.11. *The following pairs of classes* (C_1, C_2) *are incomparable:*
(i) $C_1 = \text{N-Alg}(\Delta)$, *and* $C_2 = \text{H-Rat}$;
(ii) $C_1 = \text{N-Alg}(\Delta)$, *and* $C_2 = \Pi\text{-Rat}$;
(iii) $C_1 = \text{H-Alg}$, *and* $C_2 = \Pi\text{-Rat}$.

It remains to determine the relation between the class $\mathrm{CFL}(\Delta)$ and the classes defined in 4.1 and 4.7. So far we only know that $\mathrm{Reg}(\Delta) \subsetneq \mathrm{CFL}(\Delta)$ implying that $\text{R-Rat}(\Delta) \subsetneq \mathrm{CFL}(\Delta)$. On the other hand, since $\mathrm{Rec}(\Delta)$ is a subclass of $\text{N-Rat}(\Delta)$, H-Rat, Π-Rat, $\text{N-Alg}(\Delta)$, H-Alg, and Π-Alg, and since $\mathrm{Rec}(\Delta)$ contains non-recursive languages by Corollary 4.3, $\mathrm{CFL}(\Delta)$ does not include any of these classes. Further, we have

Lemma 4.12. (i) $\mathrm{CFL}(\Delta) \not\subseteq \Pi\text{-Rat}$.
(ii) $\mathrm{CFL}(\Delta) \not\subseteq \text{H-Alg}$.
(iii) $\mathrm{CFL}(\Delta) \subsetneq \Pi\text{-Alg}$.

Proof. Consider $L_1 = \{a_1^i a_2 a_1^i \mid i \geq 1\}$. Then $L_1 \in \mathrm{CFL}(\Delta)$, but $L_1 \notin \Pi\text{-Rat}$, thus showing that $\mathrm{CFL}(\Delta) \not\subseteq \Pi\text{-Rat}$. From the proof of Corollary 3.4 we see that there exists a language $L_3 \in \mathrm{CFL}(\Delta)$ and a finite homomorphism $\varphi\colon \Delta^* \to \{a_1, a_2, a_3\}^*$ such that $\varphi(L_3) \notin \mathrm{CFL}(\{a_1, a_2, a_3\})$. Hence, $L_3 \notin \text{H-Alg}$, and so $\mathrm{CFL}(\Delta) \not\subseteq \text{H-Alg}$. Finally, $\mathrm{CFL}(\Delta)$ is closed under projections onto finite subalphabets. Thus, by the remark following the proof of Corollary 3.4, $\mathrm{CFL}(\Delta) \subseteq \Pi\text{-Alg}$. Since $\mathrm{CFL}(\Delta)$ does not contain Π-Alg, this inclusion is proper. $\square$

Putting all these results together we get the following diagrams:

$$\begin{tikzcd}[column sep=small]
\bigcup\limits_{\substack{\Sigma \subseteq \Delta \\ \Sigma \text{ finite}}} \mathrm{Reg}(\Sigma) = \mathrm{Rat}(\Delta) \subsetneq \text{R-Rat}(\Delta) = \mathrm{Reg}(\Delta) \arrow[dd, dash, "\subsetneq"] & \subsetneq & \mathrm{Rec}(\Delta) = \text{N-Rat}(\Delta) \arrow[d, dash, "\subsetneq"] & \subsetneq & \text{H-Rat} \arrow[d, dash, "\subsetneq"] & \subsetneq & \Pi\text{-Rat} \arrow[d, dash, "\subsetneq"] \\
 & & \text{N-Alg}(\Delta) & \subsetneq & \text{H-Alg} & \subsetneq & \Pi\text{-Alg} \\
\mathrm{CFL}(\Delta) \arrow[urrrrrr, dash, bend right, "\subsetneq"'] & & & & & &
\end{tikzcd}$$

$$\bigcup_{\substack{\Sigma \subseteq \Delta \\ \Sigma \text{ finite}}} \mathrm{Reg}(\Sigma) \subsetneq \bigcup_{\substack{\Sigma \subseteq \Delta \\ \Sigma \text{ finite}}} \mathrm{CFL}(\Sigma) \subsetneq \text{N-Alg}(\Delta) \cap \mathrm{CFL}(\Delta)$$

For all pairs of classes C_1 and C_2, if neither $C_1 \subseteq C_2$ nor $C_2 \subseteq C_1$ can be derived from these diagrams, then C_1 and C_2 are incomparable under set inclusion.

Acknowledgement

The author wants to thank Prof. K.E. Madlener for many helpful discussions regarding the results presented here. He also wishes to thank an anonymous referee for pointing out how to improve several of the very technical proofs contained in a previous version of this paper.

References

[1] J.M. Autebert, J. Beauquier, and L. Boasson, Langages sur des alphabets infinis; Discrete Appl. Math. 2 (1980) 1-20.

[2] J. Berstel, Transductions and context-free languages (Teubner, Stuttgart, 1979).

[3] S. Eilenberg, Automata, Languages, and Machines, Vol. A (Academic Press, New York 1974).

[4] M.A. Harrison, Introduction to Formal Language Theory (Addison-Wesley, Reading, MA, 1978).

[5] J.E. Hopcroft, and J.D. Ullman, Introduction to Automata Theory, Languages, and Computation (Addison-Wesley, Reading, MA, 1979).

[6] K.E. Madlener, and F. Otto, Charakterisierung subrekursiver Teilklassen von Wortfunktionen über abzählbar-unendlichen Alphabeten, Technical Report 3/78, Fachbereich Informatik, Universität Kaiserslautern, 1978.

[7] F. Otto, Kontext-freie Sprachen und Transduktionen über abzählbar-unendlichen Alphabeten, Technical Report 58/82, Fachbereich Informatik, Universität Kaiserslautern, 1982.

[8] F. Schwenkel, Rekursive Wortfunktionen über unendlichen Alphabeten; Z. Math. Logik Grundl. Math. 11 (1965) 133-147.

[9] K. Weihrauch, Teilklassen primitiv-rekursiver Wortfunktionen, Report No. 91, GMD Bonn, 1974.