Fundamenta Informaticae 25 (1996) 217-239
IOS Press

TOWARDS AN ALGEBRAIC THEORY OF CONTEXT-FREE LANGUAGES

J. BERSTEL*
LITP, Institut Blaise Pascal Université Pierre et Marie Curie 4 place Jussieu
F-75252 Paris Cedex 05 France

L. BOASSON
LITP Univers. é Denis Diderot 2 place Jussieu F-75251 Paris Cedex 05 France

Abstract

The purpose of the paper is to present the implications of a new definition of context-free languages. The main interests of this approach are first to allow full formal proofs and second to enlight the crucial role of rational closures. The proposed formalism is based on a theorem of Wechler. It is very near to the "radical algebras" of Conway.

1 Introduction

In teaching general results on context-free or algebraic languages one frequently meets the difficulty to explain to students some construction which is not complex but somewhat lengthy to write down explicitly. This is the case for instance for the grammar for the intersection of a context-free and a rational language.

This difficulty is well-known, and there have been several attempts to present the theory of formal languages in a different framework : regular tree grammars, Hotz algebras [9], or the description through Kolmogorov complexity [10]. Conway [3] uses systems of equations and inequalities His radical algebras, even if they are introduced in a different way, are closely related to our stable algebras.

The present paper is along the lines of these attempts. Its aim is to show how the theory of algebraic languages is organized when one takes as definition not arbitrary context-free grammars, but grammars in Greibach normal form, and reformulate them according to a weak version of a theorem of Wechler [13] : a language is context-free if and only if it belongs to an algebra of languages that is finitely generated and stable under left quotient.

We intend to prove some of the main results of the theory of context-free languages starting with this characterization, and without using any other result about these languages. On the contrary, we admit the theory of finite automata and rational languages ([4, 12]).

We shall try to show that this approach results in a double benefit. First, we get a great number of standard results about algebraic languages without effort, and with very

*Partially supported by the PRC "Mathématiques et Informatique" and by ESPRIT BRA working group 6317 – ASMICS 2

short full proofs (we get no new result) Next, our approach highlights the ubiquity of rational closure as a basic operation, in various formulations. Up to now, our approach has limitations: we are unable to obtain as precise pumping lemmas as Ogden's one. Also, the theory of deterministic languages seems difficult to express in our formalism: push-down automata are easily introduced, but by nature they are always real-time

The paper is organized as follows: in the first section, we review some properties of quotients. In section 2, we define context-free algebras and give some examples. The next section contains those closure properties which are straightforward Section 4 present an extension of stable algebras which will be used in the next section devoted to various normal forms Closure under morphism and, more generally, under rational transductions, is the object of Section 6 Pumping lemmas are studied in Section 7 Then, we prove a theorem of Chomsky-Schützenberger and, in the last section, a theorem of Shamir

The paper is self-contained Notation is the usual one, as found in Salomaa [12] and Eilenberg [4].

2 Quotients

For easy reading and writing, we denote union by the symbol $+$. Given an alphabet A, one defines for $w \in A^*$ and $X \subset A^*$, the *left quotient* by

$$w^{-1}X = \{v \in A^* \mid wv \in X\}$$

Symmetrically, one defines the right quotient

$$Xw^{-1} = \{v \in A^* \mid vw \in X\}.$$

The following computation rules are clear:

$$\begin{aligned}(uv)^{-1}X &= v^{-1}(u^{-1}X)\\ u^{-1}(X+Y) &= u^{-1}X + u^{-1}Y\\ u^{-1}(X \cap Y) &= u^{-1}X \cap u^{-1}Y\end{aligned}$$

and for a letter $a \in A$

$$\begin{aligned}a^{-1}(XY) &= (a^{-1}X)Y + (X \cap \{\varepsilon\})(a^{-1}Y)\\ a^{-1}(X^*) &= (a^{-1}X)X^*\end{aligned}$$

The first of these formulas can be written as:

$$a^{-1}(XY) = \begin{cases}(a^{-1}X)Y & \text{if } \varepsilon \notin X;\\ (a^{-1}X)Y + a^{-1}Y & \text{otherwise}\end{cases}$$

The notion of quotient is extended to languages by:

$$Y^{-1}X = \bigcup_{y \in Y} y^{-1}X = \{u \in A^* \mid Yu \cap X \neq \emptyset\}$$

One has

$$\begin{aligned}(Y+Z)^{-1}X &= Y^{-1}X + Z^{-1}X\\ (YZ)^{-1}X &= Z^{-1}(Y^{-1}X)\\ (Y^*)^{-1}X &= X + Y^{-1}((Y^*)^{-1}X) = X + (Y^*)^{-1}(Y^{-1}X)\end{aligned}$$

The last formula comes from the fact that $Y^* = \varepsilon + Y^*Y = \varepsilon + YY^*$ Finally,

$$Z^{-1}(XY) = (Z^{-1}X)Y + (X^{-1}Z)^{-1}Y$$

and

$$X = \sum_{a \in A} a(a^{-1}X) + X \cap \{\varepsilon\}$$

All these formulas also hold for right quotients.

3 Context-free Algebras

3.1 Definition

An *algebra* $\mathcal{A}$ over an alphabet A is a set of languages over A closed under union and product, and containing all finite subsets.

Let $\mathcal{T}$ be a set of subsets of A^*. We denote by $\textsc{Pol}(\mathcal{T})$ the set of *polynomials* over $\mathcal{T}$, that is the set of languages that can be expressed as finite unions of products (*monomials*) of the form

$$T_1 T_2 \cdots T_k$$

with $k \geq 0$ and for $0 \leq i \leq k$, either $T_i \in \mathcal{T}$ or $T_i = \{a\}$ for some letter $a \in A$ Thus, $\mathcal{A}$ is an algebra if and only if $\mathcal{A} = \textsc{Pol}(\mathcal{A})$. If $\mathcal{A} = \textsc{Pol}(\mathcal{T})$, one says that $\mathcal{A}$ is *generated* by $\mathcal{T}$ and that $\mathcal{T}$ is a *basis* of $\mathcal{A}$.

An algebra $\mathcal{A}$ is of *finite type* if it has a finite basis. If $\mathcal{T} = \{T_1, \ldots, T_n\}$ is a finite basis of $\mathcal{A}$, then we also write $\mathcal{A} = \textsc{Pol}(T_1, \ldots, T_n)$

An algebra $\mathcal{A}$ is *stable* if it is closed under left quotient, that is if

$$u \in A^*, X \in \mathcal{A} \Rightarrow u^{-1}X \in \mathcal{A}.$$

The verification of stability of an algebra is easy due to the following rule:

Rule. *An algebra $\mathcal{A}$ with basis $\mathcal{T}$ is stable if and only if $a^{-1}T \in \mathcal{A}$ for any letter $a \in A$ and any $T \in \mathcal{T}$.*

Proof. By the first formula of the preceding section, it suffices to consider the quotient by a letter. Next, the formulas $a^{-1}(X + Y) = a^{-1}X + a^{-1}Y$ et $a^{-1}(XY) = (a^{-1}X)Y + (X \cap \{\varepsilon\})(a^{-1}Y)$ show how to reduce the quotient to the quotient of a monomial and eventually of a generator ■

We call *context-free algebra* any algebra that is stable and of finite type. A language L is *algebraic* if and only if it belongs to some context-free algebra $\mathcal{A}$. We say then that $\mathcal{A}$ is a (context-free) algebra *for* L Thus, the set of algebraic languages over the alphabet A is the union of all context-free algebras over A

3.2 Examples

Example 0. Wechler's theorem [13] in its version for languages, is the following:

Theorem 3.1. *A language L over A is context-free iff L belongs to a finitely generated stable algebra over A*

Thus, algebraic languages coincide with context-free languages in the usual sense. For languages, the proof is not difficult when one uses Greibach normal form.

EXAMPLE 1. The *language of Lukasiewicz* over the alphabet $A = \{a, b\}$ satisfies the equation $L = b + aLL$. It follows that $a^{-1}L = L^2$ and $b^{-1}L = \{\varepsilon\}$. Consequently, the algebra $\mathrm{Pol}(L)$ is stable; it is a context-free algebra for L.

EXAMPLE 2 Over the alphabet $A = \{a, b\}$, consider the language

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

Setting $U = b^*$, one gets

$$\begin{aligned} a^{-1}L &= Lb \\ b^{-1}L &= U \\ a^{-1}U &= \emptyset \\ b^{-1}U &= U \end{aligned}$$

thus $\{L, U\}$ generates a stable algebra for L.

EXAMPLE 3 Still over $A = \{a, b\}$, consider the language

$$L = \{a^n b^m \mid n, m > 0,\ n \neq m\}$$

Set $U = b^+$ and $V = a^+$. Then

$$\begin{array}{lll} a^{-1}L = Ub + Vb + Lb & a^{-1}U = \emptyset & a^{-1}V = \varepsilon + V \\ b^{-1}L = \emptyset & b^{-1}U = \varepsilon + U & b^{-1}V = \emptyset \end{array}$$

Thus the algebra $\mathrm{Pol}(L, U, V)$ is stable. We shall see later how to get rid of the intermediate variables U and V by authorizing also the star operation (semi-stable algebras).

EXAMPLE 4 The algebra $\mathrm{Pol}(\emptyset)$ of finite languages is clearly context-free

EXAMPLE 5. Every rational language K over A is algebraic. Indeed, the set $\{u^{-1}K \mid u \in A^*\}$ is finite, and therefore generates an algebra of finite type which of course is stable.

EXAMPLE 6. Assume that a grammar G is given in Greibach normal form (with ε-productions added when needed). Then, the associated system of equation gives effectively the expressions of the left quotients of the components of the solution as polynomials in the components of the solution. They determine uniquely and effectively the components, provided one remembers, for each component, whether it contains the empty word or not (see also Corollary 3 4 below).

EXAMPLE 7 In view of the preceding remark, the membership problem is decidable for algebraic languages. Indeed, in order to check whether $u \in L$, it suffices to check whether the polynomial $u^{-1}L$ has a constant term ε.

3.3 Proper Basis

The special and frequently disturbing role played by the empty word in context-free languages is considerably weakened here by the fact that $a^{-1}(L \cup \{\varepsilon\}) = a^{-1}L$ for any language L and any letter a. The presence or absence of the empty word has almost no influence on the relations that express the quotients of elements of a basis as polynomials in this basis. Consider for example the language L of the example 2 above, without the empty word:

$$L' = \{a^n b^m \mid 0 < n \leq m\} = U' + aLb$$

with $U' = b^+$. Since $U' = Ub$, one has $L', U' \in \mathrm{POL}(L, U)$. Conversely, since $L = L' + \varepsilon$ and $U = U' + \varepsilon$, one has $L, U \in \mathrm{POL}(L', U')$. Thus $\{L, U\}$ and $\{L', U'\}$ generate the same algebra.

A family of languages $\mathcal{T}$ is *proper* if $\varepsilon \notin T$ for all $T \in \mathcal{T}$.

LEMMA 3.2 *Every context-free algebra possesses a finite proper basis*

Proof. Let $\mathcal{A}$ be a context-free algebra and let $\mathcal{T}$ be a finite basis of $\mathcal{A}$. For $T \in \mathcal{T}$, denote by $T' = T - \{\varepsilon\}$ the language without the empty word, and let $\mathcal{T}' = \{T' \mid T \in \mathcal{T}\}$. Since

$$T' = \sum_{a \in A} a(a^{-1}T)$$

one has $T' \in \mathrm{POL}(\mathcal{T})$ for all T'. Conversely, $T = T' + (T \cap \{\varepsilon\})$, whence $T \in \mathrm{POL}(\mathcal{T}')$ for all T'. This shows that $\mathrm{POL}(\mathcal{T}) = \mathrm{POL}(\mathcal{T}')$. ∎

3.4 Homomorphisms of algebras

Let A be an alphabet and let $\mathcal{A}$ and $\mathcal{B}$ be two stable algebras over A. A mapping

$$\alpha : \mathcal{A} \to \mathcal{B}$$

is an *algebra homomorphism* if $\alpha(\{w\}) = \{w\}$ for all $w \in A^*$ and if, for $X, Y \in \mathcal{A}$ and $a \in A$,

$$\alpha(X + Y) = \alpha(X) + \alpha(Y)$$
$$\alpha(XY) = \alpha(X)\alpha(Y)$$
$$\alpha(a^{-1}X) = a^{-1}\alpha(X)$$

Clearly, a homomorphism is determined by its values on a basis $\mathcal{T}$ of $\mathcal{A}$.

EXAMPLE 1. Let $\mathcal{T}$ be a basis of $\mathcal{A}$, and define α on $\mathcal{T}$ by $\alpha(T) = \emptyset$ for all $T \in \mathcal{T}$. The image of $\mathcal{A}$ by α is the algebra of finite languages.

EXAMPLE 2. Let $\mathcal{A}$ be a context-free algebra over A with basis $\mathcal{T}$, and let $\mathcal{T}' = \{T' \mid T \in \mathcal{T}\}$ where $T' = T + \{\varepsilon\}$ for $T \in \mathcal{T}$. The mapping $\alpha : T' \mapsto T$ is a homomorphism of the algebra $\mathrm{POL}(\mathcal{T}')$ over $\mathcal{A}$ because $a^{-1}T' = a^{-1}T$.

An algebra homomorphism $\alpha : \mathcal{A} \to \mathcal{B}$ is a *renaming* (with respect to $\mathcal{T}$) if α is a bijection of some basis $\mathcal{T}$ of $\mathcal{A}$ onto a basis $\mathcal{T}'$ of $\mathcal{B}$.

PROPOSITION 3.3. *Let $\mathcal{A}$ and $\mathcal{A}'$ be two stable algebras, let $\mathcal{T}$ be a basis of $\mathcal{A}$ and let $\alpha : \mathcal{A} \to \mathcal{A}'$ be a renaming with respect to $\mathcal{T}$. If*

$$T \cap \{\varepsilon\} = \alpha(T) \cap \{\varepsilon\}$$

for all $T \in \mathcal{T}$, then $\alpha(X) = X$ for all $X \in \mathcal{A}$, and $\mathcal{A} = \mathcal{A}'$

Another formulation of this proposition is the following:

COROLLARY 3.4. *A stable algebra is completely determined by those elements of a basis that contain the empty word ε and by the quotients of the elements of this basis by letters.*

Proof. We prove the proposition by showing, by induction on the integer $n \geq 0$ that

$$X \cap A^n = \alpha(X) \cap A^n, \qquad (X \in \mathcal{A})$$

For $n = 0$, this results from the hypotheses by considering

$$\begin{aligned}(X+Y) \cap \{\varepsilon\} &= (X \cap \{\varepsilon\}) + (Y \cap \{\varepsilon\})\\ (XY) \cap \{\varepsilon\} &= (X \cap \{\varepsilon\})(Y \cap \{\varepsilon\})\end{aligned}$$

Next, since

$$X = \sum_{a \in A} a(a^{-1}X) + (X \cap \{\varepsilon\})$$

on obtains for $n \geq 0$

$$X \cap A^{n+1} = \sum_{a \in A} a(a^{-1}X \cap A^n)$$

By the induction hypothesis and from $\alpha(a^{-1}X) = a^{-1}\alpha(X)$, one gets

$$a^{-1}X \cap A^n = \alpha(a^{-1}X) \cap A^n = a^{-1}\alpha(X) \cap A^n$$

whence

$$X \cap A^{n+1} = \sum_{a \in A} a(a^{-1}\alpha(X) \cap A^n) = \alpha(X) \cap A^{n+1}.$$ ■

Observe that, in the terminology of grammars and systems of equations, we just proved that the system of equations associated to a grammar in Greibach normal from (even with ε-productions) has a unique solution.

4 Elementary Operations

In this section are gathered together those closure properties of context-free languages which can be proved in an elementary way. For others, an extension of the definition will appear to be useful.

4.1 Rational Closure

Proposition 4.1. *If L and L' are algebraic languages over A, then $L + L'$ and LL' are algebraic over A.*

Proof. If

$$L \in \mathrm{Pol}(T_1, \ldots, T_n) \quad \text{and} \quad L' \in \mathrm{Pol}(S_1, \ldots, S_m)$$

then

$$L + L', LL' \in \mathrm{Pol}(T_1, \ldots, T_n, S_1, \ldots S_m)$$

and it is straightforward, by the Rule, that $\mathrm{Pol}(T_1, \ldots, T_n, S_1, \ldots S_m)$ is stable. ■

Proposition 4.2. *If L is algebraic over A, then L^* is algebraic over A.*

Proof. Let $\mathcal{A} = \mathrm{Pol}(T_1, \ldots, T_n)$ be a stable algebra containing L. Then the algebra

$$\mathcal{B} = \mathrm{Pol}(T_1, \ldots, T_n, L^*)$$

contains of course L^*. To show that $\mathcal{B}$ is stable, observe that

$$a^{-1}L^* = (a^{-1}L)L^*$$

and since $a^{-1}L \in \mathcal{A}$, one has $a^{-1}L^* \in \mathcal{B}$. ■

We denote by $\textsc{Rat}(\mathcal{T})$ the *rational closure* of $\mathcal{T}$, i. e. the smallest family of languages containing $\mathcal{T}$ and the finite languages, and closed under union, product and star. A family of languages $\mathcal{L}$ is *rationally closed* if $\mathcal{L} = \textsc{Rat}(\mathcal{L})$. A consequence of the preceding propositions is :

Corollary 4.3. *The family of algebraic languages is rationally closed.* ■

4.2 Length-preserving Morphisms

Proposition 4.4. *Let $f : A^* \to B^*$ be a length-preserving morphism. If $\mathcal{A}$ is a context-free algebra over A^*, then $f(\mathcal{A}) = \{f(X) \mid X \in \mathcal{A}\}$ is a context-free algebra over B. In particular, the image of an algebraic language under a length-preserving morphism is an algebraic language.*

Proof. Let $\mathcal{B} = f(\mathcal{A}) = \{f(X) \mid X \in \mathcal{A}\}$. It is of course an algebra of finite type. It suffices to show that $\mathcal{B}$ is stable. We prove that for $X \in \mathcal{A}$ and $b \in B$,

$$b^{-1}f(X) = \sum_{f(a)=b} f(a^{-1}X) \tag{4.1}$$

Indeed, if $bw = f(x)$ with $x \in X$, then $x = ay$ with $f(a) = b$ and $f(y) = w$, whence $w \in f(a^{-1}X)$. Conversely, if $w \in f(a^{-1}X)$, then $w = f(y)$, for some y such that $ay \in X$. Thus $f(ay) = bw$ and $w \in b^{-1}f(X)$.

Consequently, each quotient $b^{-1}f(X)$ is a finite sum of polynomials, and thus is itself a polynomial. ■

4.3 Intersection with a Rational Language

Proposition 4.5. *The intersection of an algebraic language and a rational language is an algebraic language.*

Proof. Let L be an algebraic language, and let $\mathcal{A} = \textsc{Pol}(M_1, \ldots, M_n)$ be a context-free algebra for L.

Let K be a rational language, recognized by a finite deterministic complete automaton $\mathbb{A} = (Q, i, T, \cdot)$ with initial state i, final states T and next-state function $\cdot$. Let $K_{p,q}$ denote the set of words recognized by the automaton $\mathbb{A}$ when taking p as the initial and q as the final state. Since

$$L \cap K = \sum_{t \in T} L \cap K_{i,t}$$

it suffices to prove that every $L \cap K_{p,q}$ is algebraic.

Consider for this the algebra $\mathcal{B}$ generated by the languages $M_j \cap K_{p,q}$, for $0 \le j \le n$ and $p, q \in Q$. Observe first that, for all X, Y,

$$\begin{aligned}(X + Y) \cap K_{p,q} &= (X \cap K_{p,q}) + (Y \cap K_{p,q})\\ XY \cap K_{p,q} &= \sum_{r \in Q} (X \cap K_{p,r})(Y \cap K_{r,q})\end{aligned}$$

Consequently, all languages of the form $M \cap K_{p,q}$, for $M \in \mathcal{A}$, belong to $\mathcal{B}$. In particular $L \cap K_{p,q} \in \mathcal{B}$ for all $p, q \in Q$.

It remains to prove that $\mathcal{B}$ is stable. Since

$$a^{-1}(M \cap K_{p,q}) = a^{-1}M \cap a^{-1}K_{p,q} = a^{-1}M \cap K_{p \cdot a,q}$$

and $a^{-1}M \in \mathcal{A}$, one has $a^{-1}M \cap K_{p\,a,q} \in \mathcal{B}$. ■

It is not difficult to check that most of the usual "syntactic" proofs of the intersection property can be rephrased in the framework of context-free algebras. Our claim is that rephrasing simplifies the verification.

4.4 Syntactic Substitution and Inverse Projection

Let A and B be two disjoint alphabets and $X \subset A^*$, $Y \subset B^*$. The *syntactic substitution* of Y in X is the language

$$Z = X \uparrow Y$$

defined by

$$Z = \sum\{a_1 Y a_2 Y \cdots a_n Y \mid n \geq 0,\ a_1, \dots, a_n \in A,\ a_1 \cdots a_n \in X\}$$

In particular, $X \uparrow \{\varepsilon\} = X$ et $\{\varepsilon\} \uparrow Y = \{\varepsilon\}$. The following identities are easily verified. For $a \in A$ and $b \in B$:

$$\begin{aligned} a^{-1}(X \uparrow Y) &= Y((a^{-1}X) \uparrow Y) \\ b^{-1}(X \uparrow Y) &= \emptyset \\ (X + X') \uparrow Y &= X \uparrow Y + X' \uparrow Y \\ (XX') \uparrow Y &= (X \uparrow Y)(X' \uparrow Y) \end{aligned}$$

PROPOSITION 4.6. *If X is algebraic over A and Y is algebraic over B, then $X \uparrow Y$ is algebraic over $A \cup B$.*

Proof. Let $\mathcal{A} = \text{POL}(\mathcal{T})$ and $\mathcal{B} = \text{POL}(\mathcal{S})$ be context-free algebras for X and Y respectively. The algebra $\mathcal{C}$ generated by

$$\mathcal{U} = \mathcal{T} \cup \mathcal{S} \cup \{T \uparrow Y \mid T \in \mathcal{T}\}$$

is of course of finite type and contains $X \uparrow Y$. Now $a^{-1}(T \uparrow Y)$ is a polynomial in the languages of $\mathcal{U}$, and consequently $\mathcal{C}$ is stable. ■

Let A be an alphabet, let B be a subset of A and set $C = A \setminus B$. Let π be the projection of A^* onto A^*, defined by $\pi(b) = b$ for $b \in B$ et $\pi(c) = \varepsilon$ for $c \in C$.

PROPOSITION 4.7. *If L is an algebraic language over B, then $\pi^{-1}(L)$ is algebraic over A.*

Proof. One has $\pi^{-1}(L) = C^*(L \uparrow C^*)$, and the result follows from proposition 4.6. ■

In view of general results on decomposition of morphisms (see e. g. [1]), the closure properties of this section imply closure of algebraic languages under nonerasing morphism, and it suffices to show closure under (elementary) projection to get closure under general morphisms, rational transductions and substitutions.

It appears that closure under projection is much more difficult to prove, and is post-poned to Section 6.

5 Semi-stable Algebras

An algebra $\mathcal{A}$ is *semi-stable* if $\text{RAT}(\mathcal{A})$ is stable.

PROPOSITION 5.1. *Let $\mathcal{A}$ be an algebra, and let $\mathcal{T}$ be a basis of $\mathcal{A}$. The following conditions are equivalent :*

(*i*) *the algebra $\mathcal{A}$ is semi-stable;*
(*ii*) *for all words u and all $X \in \mathcal{A}$, one has $u^{-1}X \in \textsc{Rat}(\mathcal{A})$;*
(*iii*) *for each letter a,and each $T \in \mathcal{T}$, one has $a^{-1}T \in \textsc{Rat}(\mathcal{A})$.*

Proof. The implications *(i)⇒(ii)* and *(ii)⇒(iii)* are clear It suffices to prove the implication *(iii)⇒(i)*. We check first that $a^{-1}X \in \textsc{Rat}(\mathcal{A})$ for all $X \in \textsc{Rat}(\mathcal{A})$ and all letters a. Let $\mathcal{S}$ be the family of languages defined by

$$\mathcal{S} = \{S \in \textsc{Rat}(\mathcal{A}) \mid a^{-1}S \in \textsc{Rat}(\mathcal{A}) \text{ for each letter } a \in A\}$$

By hypothesis, $\mathcal{T} \subset \mathcal{S}$. Let us show that $\mathcal{S} = \textsc{Rat}(\mathcal{A})$ Indeed, if $S_1, S_2 \in \mathcal{S}$, then $S_1 + S_2, S_1S_2 \in \mathcal{S}$. Finally, if $S \in \mathcal{S}$, then $a^{-1}(S^*) = (a^{-1}S)S^*$. Both factors of this product are in $\textsc{Rat}(\mathcal{A})$, whence $S^* \in \mathcal{S}$ Thus, the family $\mathcal{S}$ is rationally closed, and $\textsc{Rat}(\mathcal{A}) = \textsc{Rat}(\mathcal{T}) \subset \mathcal{S} \subset \textsc{Rat}(\mathcal{A})$, whence $\mathcal{S} = \textsc{Rat}(\mathcal{A})$. This shows that $a^{-1}X \in \textsc{Rat}(\mathcal{A})$ pour $X \in \textsc{Rat}(\mathcal{A})$.

To prove the implication, consider now $X \in \textsc{Rat}(\mathcal{A})$ et $u \in A^*$. We show that $u^{-1}X \in \textsc{Rat}(\mathcal{A})$ by induction on $|u|$. Set $u = va$, with a a letter. Then $u^{-1}X = a^{-1}(v^{-1}X)$. By induction hypothesis, $Y = v^{-1}X \in \textsc{Rat}(\mathcal{A})$, and by the preceding argument, $a^{-1}Y \in \textsc{Rat}(\mathcal{A})$ ∎

Note that the algebra $\textsc{Rat}(\mathcal{A})$ is not of finite type in general, even if $\mathcal{A}$ is of finite type. On the contrary, some sub-algebras of finite type may be constructed explicitly This will be done now.

Let e be a rational expression. We associate to e a finite set of languages denoted $\textsc{St}(e)$ (for *Stars* of e) inductively as follows:

- If e is a singleton or $e = \emptyset$, then $\textsc{St}(e) = \emptyset$;
- $\textsc{St}(e_1 \cdot e_2) = \textsc{St}(e_1 + e_2) = \textsc{St}(e_1) \cup \textsc{St}(e_2)$;
- $\textsc{St}(e^*) = \textsc{St}(e) \cup \{L(e^*)\}$, where $L(e)$ is the language denoted by e

For example, for the expression $e = ((a^* + b)^*(c^*d))^*$, one has

$$\textsc{St}(e) = \{((a^* + b)^*(c^*d))^*, (a^* + b)^*, a^*, c^*\}$$

The definition extends to a set E of rational expressions by

$$\textsc{St}(E) = \bigcup_{e \in E} \textsc{St}(e).$$

We denote by $L(E)$ the set of languages $L(e)$, for $e \in E$. By construction, $L(e)$ is contained in $\textsc{Pol}(\textsc{St}(E))$. Recall that by convention $\textsc{Pol}(\mathcal{L})$ contains all finite languages.

Lemma 5 2. *Let $\mathcal{T}$ be a family of languages, and let E be a set of rational expressions over $\mathcal{T}$. If $a^{-1}T \in L(E)$ for each letter a and each $T \in \mathcal{T}$, then*

$$\mathcal{A} = \textsc{Pol}(\mathcal{T} \cup \textsc{St}(E))$$

is a stable algebra

Proof. In view of the Rule, it suffices to prove that $a^{-1}X \in \mathcal{A}$ for all $X \in \mathcal{T} \cup \textsc{St}(E)$. If $X \in \mathcal{T}$, this results from the hypothesis. If $X \in \textsc{St}(E)$, one has $X = R^*$ for some language $R \in \mathcal{A}$, and $a^{-1}X = (a^{-1}R)X$. Since the number of stars in the expression denoting R is smaller than the corresponding number for X, an induction completes the proof. ∎

PROPOSITION 5.3. *Let $\mathcal{A} = \text{POL}(\mathcal{T})$ be a semi-stable algebra with finite basis $\mathcal{T}$, and let*

$$\mathcal{S} = \text{SI}(\{a^{-1}T \mid a \in A, T \in \mathcal{T}\}).$$

Then the algebra $\mathcal{B}$ generated by $\mathcal{T} \cup \mathcal{S}$ is context-free.

Proof. Since $\mathcal{A}$ is semi-stable, the languages $a^{-1}T$, for $a \in A, T \in \mathcal{T}$, are in $\text{RAT}(\mathcal{A})$. Let E be a finite set of rational expressions denoting these languages, and set $\mathcal{S} = \text{SI}(E)$. By the lemma, the algebra $\mathcal{B}$ generated by $\mathcal{T} \cup \mathcal{S}$ is stable, whence context-free ■

COROLLARY 5.4 *A language is algebraic if and only if it belongs to a semi-stable algebra of finite type* ■

6 Normal Forms

There exist many normal forms for context-free grammars Among the most well-known, there are the Greibach quadratic normal form [6] and the two-sided normal form [11, 8, 5] In this Section, we show how to derive, in a simple and constructive manner, these normal forms in the framework of context-free algebras Moreover, they will be useful to prove the closure of algebraic languages under morphism

Let $\mathcal{A}$ be a stable algebra and let $\mathcal{T}$ be a basis of $\mathcal{A}$. For each generator $T \in \mathcal{T}$ and each letter $a \in A$, the language $a^{-1}T$ is a polynomial over $\mathcal{T}$ The *degree* of $\mathcal{T}$ is, by definition, the maximum of the degrees of the polynomials $a^{-1}T$. The degree of a polynomial is the maximum of the degrees of its monomials and the degree of a monomial is the number of letters and elements of the basis that composes it.

6.1 Greibach Quadratic Normal Form

PROPOSITION 6 1. *Every context-free algebra possesses a (proper) basis of degree at most 2.*

Proof. Let $\mathcal{A} = \text{POL}(\mathcal{T})$ be a stable algebra of degree $k > 2$. Without loss of generality, we suppose that $\{a\} \in \mathcal{T}$ for each letter $a \in A$ and that the basis is proper. Consider now the finite set

$$\mathcal{S} = \{T_1 \cdots T_h \mid T_i \in \mathcal{T},\ 1 \leq h \leq k-1\}$$

For $a \in A$ and $M = T_1 \cdots T_h \in \mathcal{S}$, one has $a^{-1}M = (a^{-1}T_1)T_2 \cdots T_h$ because the basis is proper Each $a^{-1}T_i$ is a polynomial in $\mathcal{T}$ of degree at most k. Thus, $a^{-1}M$ is a polynomial of degree at most $k+h-1 \leq 2k-2$ in $\mathcal{T}$ Now, each monomial of degree at most $2k-2$ in $\mathcal{T}$ can be written as a product of two monomials of degree at most $k-1$ in $\mathcal{T}$, and thus also as a monomial of degree at most 2 in $\mathcal{S}$. Hence, $a^{-1}M$ is a polynomial over $\mathcal{S}$ of degree at most 2, showing that $\mathcal{S}$ is a basis of degree at most 2. ■

6.2 Bistable Algebras

In this section, it will be necessary to distinguish left and right quotients, and it is convenient to add the prefix *left* to the previously introduced notations An algebra $\mathcal{A} = \text{POL}(\mathcal{T})$ is called *right stable* if $u \in A^*$, $X \in \mathcal{A}$ imply $Xu^{-1} \in \mathcal{A}$. Similarly, $\mathcal{A}$ is *right semi-stable* if $\text{RAT}(\mathcal{A})$ is right stable An algebra is *bistable* if it is both right stable and left stable.

PROPOSITION 6.2 *Every finitely generated left stable algebra is contained in a finitely generated bistable algebra.*

Proof. Let $\mathcal{A}$ be a left stable algebra of finite type and let $\mathcal{T}$ be a finite basis of $\mathcal{A}$. We can assume that $\mathcal{T}$ is proper and contains the languages reduced to a letter. We show first that $\mathcal{A}$ is right semi-stable. For this, let $T \in \mathcal{T}$ and $a \in A$. One has

$$T = \sum_{b \in A} bP_{b,\mathcal{T}}$$

where $P_{b,\mathcal{T}} = b^{-1}T$ is a polynomial in $\mathcal{T}$. It follows that

$$Ta^{-1} = \sum_{b \in A} b(P_{b,\mathcal{T}}a^{-1})$$

and consequently that Ta^{-1} is the sum of a polynomial in $\mathcal{T}$ and of a sum of productions of the form

$$bT_1 \cdots T_h(T_0a^{-1})$$

with $T_i \in \mathcal{T}$ for $0 \leq i \leq h$ and $b \in A$. This proves that the set of languages Ta^{-1}, for $T \in \mathcal{T}$, satisfies a proper system of right-linear equations with coefficients in $\mathcal{A}$. Thus $Ta^{-1} \in \mathrm{RAT}(\mathcal{A})$.

It follows from (the symmetric statement of) proposition 5.3 that the algebra generated by

$$\mathcal{T} \cup \mathrm{SI}(\{Ta^{-1} \mid a \in A,\ T \in \mathcal{T}\})$$

is right stable. It is elementary to check that this algebra is also left stable ∎

We illustrate this construction and the following by a running example, namely the language of Lukasiewicz.

EXAMPLE. The language of Lukasiewicz, denoted here by S, is a generator of the algebra $\mathrm{POL}(S)$ which is stable because

$$a^{-1}S = SS$$
$$b^{-1}S = \varepsilon$$

To compute the right form, we observe that $S = a(a^{-1}S) + b(b^{-1}S) = aSS + b$, whence

$$Sa^{-1} = aS(Sa^{-1}) + (ba^{-1})$$
$$Sb^{-1} = aS(Sb^{-1}) + (bb^{-1})$$

This implies that $Sa^{-1} = \emptyset$ and $Sb^{-1} = (aS)^*$. Thus we introduce a new element in the set of generators, namely $R = (aS)^*$. One obtains

$$Ra^{-1} = (aS)^*a(Sa^{-1}) = \emptyset$$
$$Rb^{-1} = (aS)^*a(Sb^{-1}) = RaR$$

whence the relations (we report only those with a nonempty result):

$$Sb^{-1} = R$$
$$Rb^{-1} = RaR$$

Observe that this corresponds to the grammar

$$S \to Rb$$
$$R \to RaRb + \varepsilon$$

According to the proof of the proposition, this algebra is also (left) stable, and indeed one gets:

$$\begin{aligned} a^{-1}R &= SR \\ b^{-1}R &= \emptyset \end{aligned}$$

Corollary 6.3. *Left stable algebras of finite type and right stable algebras of finite type define the same families of languages.* ∎

Corollary 6.4. *If L is an algebraic language, then its reversal is also algebraic.* ∎

6.3 Hotz Quadratic Normal Form

We call "Hotz Quadratic Normal Form" the two-sided quadratic normal form of a grammar. Let now $\mathcal{A}$ be a bistable algebra of finite type with basis $\mathcal{T}$. For each $T \in \mathcal{T}$, and all $a, b \in A$, the language $a^{-1}Tb^{-1}$ is a polynomial in $\mathcal{T}$. The maximum of the degrees of these polynomials is the *bidegree* of $\mathcal{T}$.

Example (cont'd). The polynomials $a^{-1}Tb^{-1}$ are computed as follows:

$$\begin{array}{ll} a^{-1}Sa^{-1} = SSa^{-1} = \emptyset & a^{-1}Ra^{-1} = SRa^{-1} = \emptyset \\ a^{-1}Sb^{-1} = SSb^{-1} = SR & a^{-1}Rb^{-1} = (SR)b^{-1} = S(Rb^{-1}) + Sb^{-1} = SRaR + R \\ b^{-1}Sa^{-1} = \emptyset & b^{-1}Ra^{-1} = \emptyset \\ b^{-1}Sb^{-1} = \emptyset & b^{-1}Rb^{-1} = \emptyset \end{array}$$

whence

$$\begin{aligned} a^{-1}Sb^{-1} &= SR \\ a^{-1}Rb^{-1} &= SRaR + R \end{aligned}$$

this leads to the two-sided Hotz grammar :

$$\begin{aligned} S &\to aSRb + b \\ R &\to aSRaRb + aRb + \varepsilon \end{aligned}$$

Proposition 6.5. *Every bistable algebra of finite type possesses a basis of bidegree at most 2.*

Proof. The proof is in two steps. Let $\mathcal{A}$ be a bistable algebra of finite type, and let $\mathcal{T}$ be a proper basis of $\mathcal{A}$. By repeating the proof of prop 6.1 simultaneously for the sets $a^{-1}T$ and the sets Ta^{-1}, one may suppose that the basis $\mathcal{T}$ has degree at most 2 both for the left quotients and for the right quotients. This means that the languages $a^{-1}T$ and Ta^{-1} can be expressed as polynomials of degree 2 in $\mathcal{T}$. It results that for $T \in \mathcal{T}$ and $a, b \in A$, $a^{-1}Tb^{-1} = Pb^{-1}$ for some polynomial P of degree at most 2, and that Pb^{-1} is a polynomial of degree at most 3 in $\mathcal{T}$.

Now let $\mathcal{S}$ be the set of languages composed of $\mathcal{T}$ and of all monomials of degree 2 over $\mathcal{T}$. Clearly, $a^{-1}Tb^{-1}$, for $T \in \mathcal{T}$, is a polynomial of degree 2 at most 2 in $\mathcal{S}$. Let now $S = T_1T_2$, with $T_1, T_2 \in \mathcal{T}$. Since $\mathcal{T}$ is proper, one has

$$a^{-1}Sb^{-1} = (a^{-1}T_1)(T_2b^{-1})$$

The right term of the equation is a polynomial of degree at most 4 in $\mathcal{T}$, hence a polynomial of degree at most 2 in $\mathcal{S}$. ∎

EXAMPLE (cont'd) In order to obtain a basis of bidegree at most 2 from

$$\begin{aligned} a^{-1}Sb^{-1} &= SR \\ a^{-1}Rb^{-1} &= SRaR + R \end{aligned}$$

we turn first the following algebra into a left quadratic and right quadratic one:

$$\begin{aligned} a^{-1}S &= SS \\ b^{-1}S &= \varepsilon \\ Sb^{-1} &= R \\ Rb^{-1} &= RaR \\ a^{-1}R &= SR \end{aligned}$$

For this, set (for instance) $T = Ra$. One gets $b^{-1}T = Tb^{-1} = \emptyset$ and $a^{-1}T = SRa + \varepsilon$, $Ta^{-1} = R$. The quadratic system has the form:

$$\begin{aligned} a^{-1}S &= SS \\ b^{-1}S &= \varepsilon \\ Sb^{-1} &= R \\ Rb^{-1} &= TR \\ a^{-1}R &= SR \\ a^{-1}T &= ST + \varepsilon \\ Ta^{-1} &= R \end{aligned}$$

For the two-sided form, one gets

$$\begin{aligned} a^{-1}Sb^{-1} &= SR \\ a^{-1}Rb^{-1} &= STR + R \\ a^{-1}Ta^{-1} &= SR \end{aligned}$$

To eliminate the monomial which is not quadratic, set (for instance) $U = ST$. Then one obtains $b^{-1}Ua^{-1} = R$ and $a^{-1}Ua^{-1} = (a^{-1}ST)a^{-1} = SSR$, which leads to consider the language $V = SS$, for which $a^{-1}Vb^{-1} = VR$ et $b^{-1}Vb^{-1} = R$ The adjunction of these two languages to the set of generators yields the desired system, namely:

$$\begin{aligned} a^{-1}Sb^{-1} &= SR & a^{-1}Ua^{-1} &= VR \\ a^{-1}Rb^{-1} &= UR + R & b^{-1}Ua^{-1} &= R \\ a^{-1}Ta^{-1} &= SR & a^{-1}Vb^{-1} &= VR \\ & & b^{-1}Vb^{-1} &= R \end{aligned}$$

(and this of course gives in turn a two-sided quadratic grammar).

6.4 Monomial Algebra

A basis $\mathcal{T}$ of a context-free algebra $\mathcal{A}$ over A is *monomial* if $a^{-1}T$ is a monomial for any $a \in A$ and any $T \in \mathcal{T}$. An algebra is monomial (of degree k) if it admits a monomial basis (of degree k).

PROPOSITION 6.6 *Any context-free algebra is the image through a length-preserving homomorphism of a context-free monomial algebra with same degree.*

Proof Let $\mathcal{A}$ be a context-free algebra over A, and let $\mathcal{T}$ be a proper basis of $\mathcal{A}$ of degree k. We assume that the languages formed by a single letter belong to $\mathcal{T}$. We denote by $\mathcal{M}$ the set of monomials of degree at most k.

Let B be a new alphabet defined by

$$B = \{(a, T, M) \mid a \in A,\ T \in \mathcal{T},\ M \in \mathcal{M}\}$$

and let $\pi : B^* \to A^*$ be the length-preserving morphism defined by $\pi((a, T, M)) = a$. Consider now the algebra $\mathcal{A}'$ generated by the family $\mathcal{T}' = \{T' \mid T \in \mathcal{T}\}$ of those subsets of B^* defined by

$$T' \cap \{\varepsilon\} = T \cap \{\varepsilon\}$$

and by

$$(a, T, M)^{-1} S' = \begin{cases} \emptyset & \text{if } T \neq S \text{ or } M \text{ is not a monomial in } a^{-1}T, \\ M' & \text{otherwise,} \end{cases}$$

where, for each monomial $M = T_1 \cdots T_n$ de $\mathcal{A}$, we denote $M' = T_1' \cdots T_n'$. By construction, the algebra $\mathcal{A}'$ is monomial. Let $T'' = \pi(T')$ for $T' \in \mathcal{T}'$ and $\mathcal{A}'' = \pi(\mathcal{A}')$ By proposition 4.4, the algebra $\mathcal{A}''$ is context-free and, translating formula (4.1) gives, for $a \in A$ and $T' \in \mathcal{T}'$,

$$\begin{aligned} a^{-1}T'' &= \sum_{\pi(b)=a} \pi(b^{-1}T') \\ &= \sum \pi\{(a, S, M)^{-1}T' \mid M \in \mathcal{M},\ S \in \mathcal{T}\} \\ &= \sum \pi\{(a, T, M)^{-1}T' \mid M \text{ monomial of } a^{-1}T\} \\ &= \sum \{M'' \mid M \text{ monomial of } a^{-1}T\} \end{aligned}$$

where, as above, $M'' = \pi(M')$. Furthermore, $T'' \cap \varepsilon = T' \cap \varepsilon = T \cap \varepsilon$. Hence, due to corollary 3.4, we get $T'' = T$ for all T, so that $\mathcal{A}'' = \mathcal{A}$ ■

7 Morphism

In this section, we prove that algebraic languages are closed under morphism, rational transduction and substitution. We begin by the closure under projection. Let A be an alphabet, b a letter not in A, and let $B = A \cup \{b\}$. We denote π the projection of B^* on A^* : it is defined by $\pi(b) = \varepsilon$ and $\pi(a) = a$ for $a \in A$

PROPOSITION 7.1 *If K is an algebraic language over B, then $\pi(K)$ is algebraic over A*

The proof of this proposition needs some preparation. For any language X over B, we denote $\hat{X} = (b^*)^{-1}X$. Let $a \in A$. Then

$$a^{-1}\pi(X) = \pi(a^{-1}\hat{X}) \tag{7.1}$$

Indeed, if $w \in a^{-1}\pi(X)$, then $aw \in \pi(X)$ and there exists an integer $n \geq 0$ and a word v such that $b^n av \in X$ and $\pi(v) = w$. Thus, $v \in (b^*a)^{-1}X = a^{-1}\hat{X}$ The other inclusion is proved in the same way.

LEMMA 7.2. *Let $X_1, \ldots, X_k$ be languages over B. Then*

$$(b^*)^{-1}(X_1 \cdots X_k) = \hat{X}_1 X_2 \cdots X_k + \sum_{i=1}^{k-1} \langle X_1 \cdots X_i, b^* \rangle \, \hat{X}_{i+1} X_{i+2} \cdots X_k \qquad (7.2)$$

where

$$\langle X, R \rangle = \begin{cases} \emptyset & \text{if } X \cap R = \emptyset \\ \{\varepsilon\} & \text{otherwise} \end{cases}$$

Proof. The formula is proved by induction on k, and it is enough to prove it for $k = 2$. It can be then deduced from the general formula :

$$Z^{-1}(XY) = (Z^{-1}X)Y + (X^{-1}Z)^{-1}Y$$

In our case, $Z = b^*$, and we have $X^{-1}Z \neq \emptyset$ if and only if $b^* \cap X \neq \emptyset$, so that $X^{-1}Z = Z$. ∎

LEMMA 7.3. *Let $\mathcal{B}$ be a bistable algebra with basis $\mathcal{T}$. The languages $\hat{T} = (b^*)^{-1}T$, for $T \in \mathcal{T}$, belong to* RAI($\mathcal{B}$).

Proof. The basis $\mathcal{T}$ can be assumed to be proper; it can be assumed too to contain all the languages reduced to a single letter. Let $T \in \mathcal{T}$. As $\mathcal{T}$ is proper,

$$T = \sum_{a \in A} (Ta^{-1})a$$

As $b^* = \varepsilon + bb^*$, we get

$$\hat{T} = T + (bb^*)^{-1}T = T + (b^*)^{-1}(b^{-1}T) = T + \sum_{a \in A} (b^*)^{-1} U_{b,a} a$$

where $U_{b,a} = b^{-1}Ta^{-1}$. $U_{b,a}$ is a polynomial over $\mathcal{T}$. Using formula 7.2, the language $(b^*)^{-1}U_{b,a}$ is the sum of a polynomial over $\mathcal{T}$ and of a sum of products of the form $\hat{T}_1 T_2 \cdots T_h$ where $T_1, \ldots, T_h \in \mathcal{T}$. In other words, the languages $\hat{T}$, for $T \in \mathcal{T}$, satisfy a system of left-linear and proper equations with coefficients in $\mathcal{B}$. Hence, they belong to RAI($\mathcal{B}$) ∎

LEMMA 7.4. *Let $\mathcal{B}$ be a bistable algebra of finite type over B. Then $\pi(\mathcal{B}) = \{\pi(X) \mid X \in B\}$ is a semi-stable algebra over A*

Proof. Clearly, $\pi(\mathcal{B})$ is an algebra. Let $\mathcal{T}$ be a basis of $\mathcal{B}$, let $T \in \mathcal{T}$ and $a \in A$. By (7.1), we get $a^{-1}(\pi(T)) = \pi(a^{-1}\hat{T})$. By lemma 7.3, the language $\hat{T}$, and therefore the language $a^{-1}\hat{T}$ belongs to à RAI($\mathcal{B}$). It follows that $a^{-1}(\pi(T)) \in$ RAI($\pi(\mathcal{B})$) $= \pi$(RAI($\mathcal{B}$)). ∎

Proof of proposition 7.1. The proof is now immediate : if K is in a context-free algebra $\mathcal{B}$, by proposition 6.2, it lies too in a bistable algebra of finite type $\mathcal{B}'$, so that, by the previous lemma, $\pi(K)$ is in the semi-stable algebra $\pi(\mathcal{B}')$, hence $\pi(K)$ is algebraic by corollaire 5.4. ∎

COROLLARY 7.5. *The family of algebraic languages is closed under rational transduction*

Proof. It is well known ([4]) that any rational transduction can be decomposed in a sequence of inverse projections, an intersection with a rational language and a sequence of projections. ∎

Morphisms and inverse morphisms are rational transductions, hence :

Corollary 7 6 *The family of algebraic languages is closed under morphism and under inverse morphism* ∎

Corollary 7 7 *The family of algebraic languages is closed under substitution*

Proof It is well known ([1]) that an algebraic substitution can be realized through the syntactical substitution, finite unions and rational transductions ∎

8 Iteration

It is not surprising that an iteration lemma as general as Ogden's lemma is difficult to prove in this context The following proposition, even if it does not deal with iteration, is very simple to prove and allows to establish that some languages are not algebraic

Proposition 8 1 *For any algebraic language L over A, there exists an integer C such that for any word u in A^*, either $u^{-1}L$ is empty, or $u^{-1}L$ contains a word of length at most $C \cdot |u|$*

Proof Let $\mathcal{A}$ be a context-free algebra for L, and let $\mathcal{T}$ be a proper basis for $\mathcal{A}$ of degree k We may assume that L belongs to $\mathcal{T}$ For any $T \in \mathcal{T}$, let h_T be an integer such that T contains a word of length at most h_T, and let $h = \max_{T \in \mathcal{T}} h_T$ For any word $u \in A^*$, the polynomial $u^{-1}T$ has degree at most $|u| \cdot k$ Hence, if $u^{-1}T$ is not empty, it contains a word of length at most $|u| \cdot C$, with $C = k \cdot h$ ∎

Example The language $\{a^n b^{n^2} \mid n \geq 1\}$ is not algebraic More generally, a language of the form $\{a^n b^{f(n)} \mid n \geq 1\}$, where f grows faster than an affine linear function, can never be algebraic

We now come to iteration properties We prove :

Theorem 8 2 *For any algebraic language L over A, there exists an integer C such that any word u in L of length $|u| \geq C$ can be factorized $u = \alpha x \beta y \gamma$, with $0 < |x|$ and $|\alpha x| \leq C$, and $\alpha x^n \beta y^n \gamma \in L$ for all $n \geq 0$*

Proof Let $\mathcal{A}$ be a context-free algebra for L By proposition 6 6, we may assume that $\mathcal{A}$ has a proper monomial basis $\mathcal{T}$ of degree 2. We may assume too that the letters are in the basis, and that $L \in \mathcal{T}$ We begin by a lemma :

Lemma 8 3 *Let $u \in A^*$ be a word of length n and let $M = TN$ be a monomial, with $T \in \mathcal{T}$ For $0 \leq i \leq n$, let u_i be the prefix of u of length i, and let $M_i = u_i^{-1}M$ If* $\deg(M_i) \geq \deg(M)$ *for all i, then*

$$M_i = (u_i^{-1}T)N \qquad (i = 0, \ldots, n)$$

Proof The lemma holds for $i = 0$ because $u_i = \varepsilon$ Assume $M_i = (u_i^{-1}T)N$ As $\deg(M_i) > \deg(N)$, we have $\deg(u_i^{-1}T) > 0$, and setting $u_{i+1} = u_i a$,

$$M_{i+1} = u_{i+1}^{-1}M = a^{-1}((u_i^{-1}T)N) = (a^{-1}(u_i^{-1}T))N = (u_{i+1}^{-1}T)N$$

This proves the lemma ∎

Let now k be the cardinal of $\mathcal{T}$ The number of monomials of degree at most k is $k \cdot \frac{k^k - 1}{k-1}$ Let $u \in L$ a word of length $n \geq C$, with $C = 1 + k \cdot \frac{k^k-1}{k-1}$ We set $u = a_1 \cdots a_n$, with $a_i \in A$, and

$$u_i = a_1 \cdots a_i, \quad M_i = u_i^{-1}L \qquad (0 \leq i \leq n)$$

Hence, $M_n = \{\varepsilon\}$ and $\deg(M_i) \geq 1$ for $i = 0, \ldots, n-1$.

If the monomials M_i are of degree at most k, then, by the choice of C, there exist two indices $0 \leq r < s \leq n$ such that $M_r = M_s$. Then there exists a factorization $u = \alpha x \gamma$ such that $\alpha^{-1}L = M_r$, $x^{-1}M_r = M_r$, $\gamma^{-1}M_r = \{\varepsilon\}$, and it follows $\alpha x^* \gamma \subset L$.

So, we may assume now that there exists monomials of degree greater than k, and hence, of degree $k+1$. Let λ_{k+1} be the smallest index $j \leq n$ such that $\deg(M_j) = k+1$, and for $r = 1, \ldots, k$, let λ_k be the greatest index $j < \lambda_{k+1}$ such that $\deg(M_j) = r$. We then have

$$\begin{aligned} &\deg(M_{\lambda_1}) = 1 \\ &\deg(M_i) > r \quad \text{for} \quad \lambda_r < i \leq \lambda_{r+1} \end{aligned} \tag{8.1}$$

We set

$$P_r = M_{\lambda_r} = T_r N_r \qquad (r = 1, \ldots, k+1)$$

with $T_r \in \mathcal{T}$ and N_r a monomial (of degree $r-1$). Let $\lambda_0 = 0$, $T_0 = L$, $N_0 = \{\varepsilon\}$ and $P_0 = M_0$. Finally, let

$$v_r = a_{\lambda_r+1} \cdots a_{\lambda_{r+1}} \qquad (0 \leq r \leq k)$$

By construction, we have $P_{r+1} = v_r^{-1}P_r$, and it follows from (8.1) and from the lemma that

$$v_r^{-1}P_r = (v_r^{-1}T_r)N_r \qquad (0 \leq r \leq k)$$

so that $(v_r^{-1}T_r)N_r = T_{r+1}N_{r+1}$. As $\deg(P_{r+1}) = 1 + \deg(P_r)$, there exist elements $Q_{r+1} \in \mathcal{T}$ such that

$$v_r^{-1}T_r = T_{r+1}Q_{r+1}, \quad N_{r+1} = Q_{r+1}N_r \qquad (0 \leq r \leq k)$$

The choice of k implies that there exist two indices $0 \leq r < s \leq k$ such that $T_r = T_s$. So, we get

$$(v_r v_{r+1} \cdots v_{s-1})^{-1} T_r = T_r Q_s \cdots Q_{r+1}$$

Set

$$u = \alpha x w$$

with

$$\alpha = v_0 \cdots v_{r-1}, \; x = v_r v_{r+1} \cdots v_{s-1}, \; w = (\alpha x)^{-1} u$$

By construction, we have $\alpha^{-1}L = T_r N_r$, $x^{-1}T_r = T_r Q$, with $Q = Q_s \cdots Q_{r+1}$, so that $\varepsilon = u^{-1}L = w^{-1}T_r Q N_r$. But, this implies that $w \in T_r Q N_r$, and that there exists a factorization $w = \beta y \gamma$, with $\beta \in T_r$, $y \in Q$, $\gamma \in N_r$. It follows immediately that $\alpha x^n \beta y^n \gamma \in L$ for all n. ■

9 Chomsky-Schützenberger's Theorem

The aim of this section is to prove the existence of a family $(S_n)_{n \geq 0}$ of algebraic languages such that :

Theorem 9.1 (Chomsky-Schützenberger) *For any algebraic language L, there exist an integer n, a rational language K and a length-preserving morphism β such that*

$$L = \beta(S_n \cap K)$$

Proof The language L being algebraic, there exists a bistable algebra $\mathcal{A}$ with a monomial quadratic basis $\mathcal{T}$ and a length-preseving morphism θ such that $L = \theta(T_0)$, for some $T_0 \in \mathcal{T}$ (proposition 6.6). Hence, it suffices to prove the result for a language belonging to a proper bistable monomial quadratic basis

Hence, let L be an algebraic language over a fixed alphabet E, and let $\mathcal{A}$ be a context-free algebra for L Let $\mathcal{T}$ be a finite monomial proper bistable basis of $\mathcal{A}$

We first prove that the language L is the homomorphic image of a language that we qualify as "very simple". In a second step, we prove that, to this "very simple" language, is naturally associated a rational language K. In the final step, we show that there exists a language S_n such that $S_n \cap K$ is exactly the previous "very simple" language. Let us call *residue* any monomial $a^{-1}T$, for $a \in E$ and $T \in \mathcal{T}$. We index the quadratic residues from 1 to n, the linear residues from 1 to m and the ("constant") residues which are formed of the empty word from 1 to p To this indexing, we associate the alphabets

$$\begin{array}{ll} A = \{a_i \mid 1 \le i \le n\} & \bar{A} = \{\bar{a}_i \mid 1 \le i \le n\} \\ B = \{b_i \mid 1 \le i \le m\} & \bar{B} = \{\bar{b}_i \mid 1 \le i \le m\} \\ C = \{c_i \mid 1 \le i \le p\} & \bar{C} = \{\bar{c}_i \mid 1 \le i \le p\} \end{array}$$

Next, we construct as many disjoint copies of the alphabet E as there are elements in the basis $\mathcal{T}$. If E_T is such a copy, we consider the alphabet $D_T = \{a_T \mid a \in T \cap E\}$ The union of the alphabets D_T, for $T \in \mathcal{T}$ is denoted D We then set

$$Z = A \cup B \cup C, \quad \bar{Z} = \bar{A} \cup \bar{B} \cup \bar{C}, \quad \hat{Z} = Z \cup \bar{Z} \cup D$$

We now define a context-free algebra $\mathcal{B}$ over $\hat{Z}$ and a length preserving morphism α from $\hat{Z}^*$ to E^* as follows The algebra $\mathcal{B}$ is generated by the basis $\mathcal{T}' = \{T' \mid T \in \mathcal{T}\}$, with

$$\left.\begin{array}{l} a_i^{-1}T'\bar{a}_i^{-1} = G'D' \\ \alpha(a_i) = x, \ \alpha(\bar{a}_i) = y \end{array}\right\} \iff \text{the } i\text{th quadratic residue is } x^{-1}Ty^{-1} = GD$$

$$\left.\begin{array}{l} b_j^{-1}T'\bar{b}_j^{-1} = U' \\ \alpha(b_j) = x, \ \alpha(\bar{b}_j) = y \end{array}\right\} \iff \text{the } j\text{th linear residue is } x^{-1}Ty^{-1} = U$$

$$\left.\begin{array}{l} c_k^{-1}T'\bar{c}_k^{-1} = \varepsilon \\ \alpha(c_k) = x, \ \alpha(\bar{c}_k) = y \end{array}\right\} \iff \text{the } k\text{th constant residue is } x^{-1}Ty^{-1} = \varepsilon$$

Finally, we set $\alpha(e_T) = e$ for $e \in T \cap E$.

By construction, in the algebra $\mathcal{B}$, each letter of $Z \cup \bar{Z}$ lies in a single residue. A letter from A defines a quadratic residue, a letter from B a linear residue and a letter from C a constant residue. We call "very simple" the elements of the basis $\mathcal{B}$.

Moreover, for any $T' \in \mathcal{T}'$, we have $\alpha(T') = T$ by corollairy 3 4. So,

$$L = \alpha(T_0')$$

where T_0' is a very simple language.

Consider now the algebra $\mathcal{B}$ In the sequel, we write $\mathcal{T}$ instead of $\mathcal{T}'$. With these notations, we have $L = \alpha(T_0)$ To each language $T \in \mathcal{T}$ are naturally associated two alphabets

$$\begin{array}{r} \mathrm{deb}(T) = \{z \in \hat{Z} \mid z^{-1}T \ne \emptyset\} \subset A \cup B \cup C \cup D \\ \mathrm{fin}(T) = \{z \in \hat{Z} \mid Tz^{-1} \ne \emptyset\} \subset \bar{A} \cup \bar{B} \cup \bar{C} \cup D \end{array}$$

Note that

$$\bigcup_{T\in\mathcal{T}} (\mathrm{deb}(T) \cup \mathrm{fin}(T)) = \hat{Z}.$$

Observe also that, in a word $u \in V$, with $V \in \mathcal{T}$,

- a letter $a \in \mathrm{deb}(T) \cap A$ is followed by a letter in $\mathrm{deb}(G)$ if the corresponding quadratic residue is $a^{-1}T\bar{a}^{-1} = GD$;
- a letter $\bar{a} \in \mathrm{fin}(T) \cap \bar{A}$ follows a letter in $\mathrm{fin}(D)$ if the corresponding quadratic residue is $a^{-1}T\bar{a}^{-1} = GD$;
- a letter $b \in \mathrm{deb}(T) \cap B$ is followed by a letter in $\mathrm{deb}(U)$ if the corresponding linear residue is $b^{-1}T\bar{b}^{-1} = U$;
- a letter $\bar{b} \in \mathrm{fin}(T) \cap \bar{B}$ follows a letter in $\mathrm{fin}(U)$ if the corresponding linear residue is $b^{-1}T\bar{b}^{-1} = U$;
- a letter $c \in \mathrm{deb}(T) \cap C$ is followed by a letter $\bar{c}$ if the corresponding constant residue is $c^{-1}T\bar{c}^{-1} = \varepsilon$;
- a letter $\bar{c} \in \mathrm{fin}(T) \cap \bar{C}$ follows a letter c if the corresponding constant residue is $c^{-1}T\bar{c}^{-1} = \varepsilon$.

So, to each letter z in Z we can associate a set of letters, denoted $\mathrm{succ}(z)$, contained in $Z \cup D$ and composed of those letters that can follow z in a word. Similarly, to each letter $\bar{z}$ in $\bar{Z}$ we can associate a set of letters, denoted $\mathrm{pred}(\bar{z})$, contained in $\bar{Z} \cup D$ and composed of those letters that can be followed by $\bar{z}$ in a word. These (local) rules define a rational language R over $\hat{Z}$. We denote R_T the rational language over $\hat{Z}$ formed by those words in R beginning by a letter in $\mathrm{deb}(T)$ and ending by a letter in $\mathrm{fin}(T)$. By construction, $T \subset R_T$. In particular, if $K = R_{T_0}$, we know that $T_0 \subset K$, and we so obtain the announced result of this second step.

Finally, it should be remarked that

$$\begin{aligned} z^{-1}R_T \neq \emptyset &\iff z \in \mathrm{deb}(T) \\ R_T z^{-1} \neq \emptyset &\iff z \in \mathrm{fin}(T) \end{aligned}$$

Over the alphabet $\hat{Z}$, we now define an algebra $\mathcal{C}$ of monomial bistable quadratic proper basis $\{S\}$ given by :

$$\begin{aligned} a_i^{-1}S\bar{a}_i^{-1} &= SS \qquad (1 \le i \le n) \\ b_j^{-1}S\bar{b}_j^{-1} &= S \qquad (1 \le j \le m) \\ c_k^{-1}S\bar{c}_k^{-1} &= \varepsilon \qquad (1 \le k \le p) \end{aligned}$$

with $S \cap \hat{Z} = D$. We will prove that

$$S \cap R_T = T, \qquad \text{for all } \; T \in \mathcal{T}$$

In order to do so, set $S_T = S \cap T$ for $T \in \mathcal{T}$, and let us compute

$$z^{-1}S_T\bar{z}'^{-1} = z^{-1}S\bar{z}'^{-1} \cap z^{-1}R_T\bar{z}'^{-1}$$

First, we observe that to get a nonempty first term in the intersection, we have to choose $z \in A \cup B \cup C$ and $z = z'$. Similarly, to avoid the second term of the intersection being empty, we must chose $z \in \mathrm{deb}(T)$ (so that $\bar{z} \in \mathrm{fin}(T)$).

So, we just have to compute the elements of the form

$$z^{-1}S_T\bar{z}^{-1} = z^{-1}S\bar{z}^{-1} \cap z^{-1}R_T\bar{z}^{-1}$$

with

$$z \in \mathrm{deb}(T) \cap (A \cup B \cup C).$$

Case 1 - $z \in A$, that is $z = a$ Then

$$a^{-1}S_T\bar{a}^{-1} = SS \cap a^{-1}R_T\bar{a}^{-1}$$

and $u \in a^{-1}R_T\bar{a}^{-1}$ iff $u \in R$ and u begins by $\mathrm{deb}(G) = \mathrm{succ}(a)$ and u ends by $\mathrm{fin}(D) = \mathrm{pred}(\bar{a})$ Hence,

$$a^{-1}S_T\bar{a}^{-1} = \{vw \in R \mid v \in S \cap \mathrm{deb}(G)\hat{Z}^*,\ w \in S \cap \hat{Z}^*\mathrm{fin}(D)\}$$

From this, as v and w lie in S and as v ends by $\mathrm{fin}(G)$ and as w begins by $\mathrm{deb}(D)$,we deduce that $v \in S_G$ and $w \in S_D$ This means :

$$a^{-1}S_T\bar{a}^{-1} = S_GS_D.$$

Case 2 - $z \in B$, that is $z = b$. Then

$$b^{-1}S_T\bar{b}^{-1} = S \cap b^{-1}R_T\bar{b}^{-1}$$

and $u \in b^{-1}R_T\bar{b}^{-1}$ iff $u \in R$ and u begins by $\mathrm{deb}(U) = \mathrm{succ}(b)$ and u ends by $\mathrm{fin}(U) = \mathrm{pred}(\bar{b})$. This means $u \in R_U$ and

$$b^{-1}S_T\bar{b}^{-1} = S_U.$$

Case 3 - $z \in C$, that is $z = c$. Then

$$c^{-1}S_T\bar{c}^{-1} = \{\varepsilon\} \cap c^{-1}R_T\bar{c}^{-1}$$

As $c\bar{c} \in R_T$ because $\bar{c} \in \mathrm{succ}(c)$ and $c \in \mathrm{pred}(\bar{c})$), we get

$$c^{-1}S_T\bar{c}^{-1} = \varepsilon.$$

Finally, we remark that the set of letters in D which belong to S_T is the same than the one which belongs to T So, the languages T and S_T are the same because of corollary 3 4. It then follows :

$$S \cap R_T = T, \qquad \text{for all} \quad T \in \mathcal{T}$$

and in particular $S \cap K = T_0$.

In order to complete the proof of the theorem, we just note that, if A, B, C and D are of different cardinalities, we may enlarge each alphabet by new letters. The rational language K will not allow them to appear in any word. So, we may assume that S is defined over an alphabet

$$\hat{Z}_n = Z_n \cup \bar{Z}_n \cup D_n$$

with D_n of cardinality n, and $Z_n = A_n \cup B_n \cup C_n$, and A_n, B_n, C_n of cardinality n. It is this language S over this alphabet that we denote S_n and which is used in the theorem. ∎

Note : It is now possible to prove that there exists a rational transduction τ_n such that $\tau_n(S_n) = S$ where S is the language over $\{a, b\}$ defined by $a^{-1}S\bar{a}^{-1} = SS$ and $b \in S$. This would establish that S is a full generator of the rational cone of context-free languages.

10 Shamir's Theorem

In this section, we prove a theorem originally established by Shamir. This theorem is very similar to the "hardest language" theorem of Greibach (see below). Let $\mathcal{A}$ be a context-free algebra over an alphabet A with a finite proper basis $\mathcal{T}$. We assume that each letter of A is an element of $\mathcal{T}$.

Let Z be an alphabet in bijection with $\mathcal{T}$ through $\beta : Z \to \mathcal{T}$ It is extended to an application β from the set $\mathcal{B}$ of the finite languages over Z in $\mathcal{A}$ as follows. If $t_1, \ldots, t_n \in Z$, then

$$\beta(t_1 \cdots t_n) = \beta(t_n) \cdots \beta(t_1)$$

that is to say

$$\beta(uv) = \beta(v)\beta(u)$$

for words $u, v \in Z^*$. Next, for $p, q \in \mathcal{B}$,

$$\beta(p \cup q) = \beta(p) + \beta(q)$$

As $\mathcal{T}$ is a proper basis of $\mathcal{A}$, the application β is a surjection onto $\mathcal{A}$.

Furthermore, let $\bar{Z} = \{\bar{z} \mid z \in Z\}$ be a disjoint copy of Z, and let $\hat{Z} = Z \cup \bar{Z}$. We consider, over $\hat{Z}^* \cup \{0\}$, where 0 is a new element, the usual one-sided Dyck reduction, defined by

$$\mathrm{red}(\varepsilon) = \varepsilon$$

and, for $m \in \hat{Z}^*$ and $t \in Z$, by

$$\mathrm{red}(mt) = \mathrm{red}(m)t$$

$$\mathrm{red}(m\bar{t}) = \begin{cases} p & \text{if } \mathrm{red}(m) = pt \\ 0 & \text{otherwise.} \end{cases}$$

We denote

$$D^* = \{m \in \hat{Z}^* \mid \mathrm{red}(m) = \varepsilon\}$$

the one-sided Dyck language. For each $a \in A$ et $T \in \mathcal{T}$, we denote $p_{a,t}$ an element of $\mathcal{B}$ (hence a finite subset of Z^*) such that

$$\beta(t) = T \quad \text{and} \quad a^{-1}T = \beta(p_{a,t})$$

in such a way that

$$a^{-1}\beta(t) = \beta(p_{a,t})$$

Finally, to each letter $a \in A$, we associate the finite subset of $\hat{Z}^*$:

$$\sigma(a) = \bigcup_{t \in Z} \bar{t}\, p_{a,t}$$

Observe that, almost by definition, the following holds : $\mathrm{red}(t\sigma(a)) = p_{a,t}$. More generally :

LEMMA 10.1. *Let $a \in A$ et $q \in \mathcal{B}$. Then*

$$\beta(\mathrm{red}(q\sigma(a))) = a^{-1}\beta(q)$$

Proof. By linearity, it is enough to prove the lemma for q composed of only one word w. Let $w = vt$, with $v \in Z^*$ and $t \in Z$. Then

$$w\sigma(a) = vt \bigcup_{s \in Z} \bar{s}\, p_{a,s}$$

so that

$$\operatorname{red}(w\sigma(a)) = vp_{a,t}$$

Consequently,

$$\begin{aligned}\beta(\operatorname{red}(w\sigma(a))) &= \beta(vp_{a,t}) = \beta(p_{a,t})\beta(v) \\ &= (a^{-1}\beta(t))\beta(v) = a^{-1}(\beta(t)\beta(v)) \\ &= a^{-1}\beta(vt) = a^{-1}\beta(w)\end{aligned}$$

The equality $(a^{-1}\beta(t))\beta(v) = a^{-1}(\beta(t)\beta(v))$ holds because the basis $\mathcal{T}$ is proper ∎

The application σ is extended in a substitution from A^* in the finite subsets of $\hat{Z}^*$. Then, the following holds :

Corollary 10.2. *Let $u \in A^*$ and $T \in \mathcal{T}$ Let $t \in Z$ such that $\beta(t) = T$ Then*

$$\beta(\operatorname{red}(t\sigma(u))) = u^{-1}T$$

Proof. By induction on the length of u, the case of a letter being settled by the above lemma. If $u = va$, with $a \in A$, then

$$\operatorname{red}(t\sigma(v)\sigma(a)) = \operatorname{red}(\operatorname{red}(t\sigma(v))\sigma(a))$$

By the previous lemma, denoting $q = \operatorname{red}(t\sigma(v))$, we have

$$\beta(\operatorname{red}(t\sigma(v))) = a^{-1}\beta(q)$$

and, by induction, $\beta(q) = v^{-1}T$. Hence the result. ∎

Proposition 10.3. *Let $u \in A^*$ and $T \in \mathcal{T}$ Then*

$$u \in T \iff t\sigma(u) \cap D^* \neq \emptyset.$$

Proof. $u \in T$ if and only if $\varepsilon \in u^{-1}T$. By the previous corollary, this is equivalent to $\varepsilon \in \beta(\operatorname{red}(t\sigma(u)))$ As the basis is proper, this holds only if $\varepsilon \in \operatorname{red}(t\sigma(u))$ which, in turn, means that $t\sigma(u) \cap D^* \neq \emptyset$ ∎

From this result follows

Theorem 10.4. (Shamir's theorem) *A language L over A is algebraic if and only if there exists an alphabet $\hat{Z}$ and a finite substitution σ from A^* in $\hat{Z}^*$ such that*

$$u \in L \iff t\sigma(u) \cap D^* \neq \emptyset$$

where $t \in \hat{Z}$ and D^* is the one-sided Dyck language over Z

Proof. The proposition is in fact the direct way. The converse is immediate : just note that $L = \tau^{-1}(D^*)$, where τ is the rational transduction which transforms u in $t\sigma(u)$. ∎

Remark 1. This theorem can be restated to obtain the "hardest context-free language" H_n In order to achieve this, just replace the finite subset $\sigma(a) = \{x_1, \ldots, x_n\}$ by the word $\sigma'(a) = \#x_1 + x_2 + \cdots + x_n\#$, where $\#$ and $+$ are new letters. So σ becomes a homomorphism and the condition $t\sigma(u) \cap D^* \neq \emptyset$ reads $\#t\#\sigma'(u) \in H_n$.

REMARK 2. If, in the above remark, the letter $+$ is considered as an addition and the letter $\#$ as parenthesis, we get the formulation given by Hotz [8].

REMARK 3. It is possible to consider the set $\mathcal{B}$ as a context-free algebra by defining, for $a \in A$ and $t \in Z$, the formal quotient by $a^{-1}t = p_{a,t}$. We are then given an algebra homomorphism as defined previously

Acknowledgements

Thanks to J.-M. Autebert, B. Courcelle and M. Latteux for their careful reading of a first draft of this paper.

References

[1] J. BERSTEL, *Transductions and Context-Free Languages*, Teubner Verlag, 1979

[2] N. CHOMSKY ET M. P. SCHÜTZENBERGER, The algebraic theory of context-free languages, in : P. Braffort, S. Hirschberg (eds.), *Computer Programming and Formal Systems*, North-Holland, Amsterdam, 1070, 116–121.

[3] J. H. CONWAY, *Regular Algebra and Finite Machines*, Chapman and Hall, 1971.

[4] S. EILENBERG, *Automata, Languages and Machines*, Vol. A, Academic Press, 1974

[5] J. ENGELFRIET, An elementary proof of double Greibach normal form, *Inform. Proc. Letters* **44** (1992), 291–293

[6] S. A. GREIBACH, Erasable context-free languages, *Inform. Control* 4 (1975), 301–326.

[7] M. HARRISON, *Introduction to Formal Language Theory*, Addison-Wesley, 1978.

[8] G. HOTZ, Normal-form transformations of context-free grammars, *Acta Cybernet.* **4** (1978), 65–84.

[9] G. HOTZ, A representation theorem of infinite dimensional algebras and applications to language theory, *J. Comput. Syst. Sci.* **33** (1986), 423–455.

[10] M. LI, P. M. B. VITÁNYI, A new approach to formal language theory by Kolmogorov complexity, in :*Proc. 16th ICALP*, Lect. Notes Comp. Sci. , Vol 372, 1989, 506–520.

[11] D. J. ROSENKRANTZ, Matrix equations and normal forms for context-free grammars *J. Assoc. Math. Comput.* **14** (1967), 501–507.

[12] A. SALOMAA, *Formal Languages*, Academic Press, 1973.

[13] H. WECHLER, Characterization of rational and algebraic power series, *Rairo Informatique théorique* **17**, (1983), 3–11.