Johannes Arz

Syntactic congruences and syntactic algebras

RAIRO. Informatique théorique, tome 17, n° 3 (1983), p. 231-238

<http://www.numdam.org/item?id=ITA_1983__17_3_231_0>
SYNTACTIC CONGRUENCES
AND SYNTACTIC ALGEBRAS (*)

by Johannes Arz (1)

Communicated by J. Berstel


Enfin, en analysant la construction d’une version non-déterministe d’un langage, on trouve une relation supplémentaire entre les algèbres syntactiques et les monoïdes syntactiques.

Abstract. — This paper gives new results about syntactic congruences and algebras. Especially, the notion of separability makes clear the connection between the syntactic algebra, defined by C. Reutenauer, and the algebra of the syntactic monoid, defined by E. Shamir and G. Hotz.

Further, the construction of a nondeterministic version of any language gives another relation between syntactic algebras and syntactic monoids.

INTRODUCTION

It has been introduced by E. Shamir [7] a monoid algebra \( R[M] \) over a ring \( R \), where \( M \) is the set of all so far as possible cancelled expressions of nested parentheses. He proved, that all context free languages are of the form \( L = v^{-1}(K_v) \) for a homomorphism \( v \) and a coset \( K_v \) in \( R[M] \).

G. Hotz has heighted this results by dealing with \( \mathbb{B}[M_D] \), where \( \mathbb{B} \) is the Boolean semiring and \( M_D \) the syntactic monoid of the Dyck language. His main idea was the use of zero divisors computing in his algebra. He called \( \mathbb{B}[M_D] \) the syntactic algebra [3, 4].

Later, C. Reutenauer has defined an algebra over a field \( K \) by introducing a congruence, induced by a given power series, on \( K[X] \), the set of polynomials over \( K \), which he called syntactic algebra, too. In this way, he advanced the theory of formal power series using high sophisticated algebraic technics [5].

In this paper, we compare both concepts by regarding them as algebras over a semiring \( R \) and dealing with the characteristic power series of a language.

(*) Received June 1982, revised December 1982.
(1) Universität des Saarlandes, Fachbereich 10, Angewandte Mathematik und Informatik, 6600 Saarbrücken, R.F.A.
We call $A_L = R[M_L]$ the syntactic algebra and $\mathcal{A}_S = R[X]$ modulo the congruence induced by char $(L)$ the characteristic algebra. Our results are:

- $\mathcal{A}_S$ is a quotient of $A_L$, for any semiring $R$ and language $L$;
- we define the property of a language to be (syntactically) separable and prove: $L$ separable $\Rightarrow \mathcal{A}_S \cong A_L$ for any $R$;
- the converse is true for $R = \mathbb{B}$;
- $\mathcal{A}_S$ is the syntactic monoid of a language $L$ related to some submonoid, where $\bar{L}$ is the nondeterministic version of $L$.
- $\mathcal{A}_S$ is a submonoid of the syntactic monoid of $\bar{L}$.

The autor is very grateful to G. Hotz and Ch. Reutenauer for suggestions and discussions.

1. NOTATIONS AND DEFINITIONS

Let $X$ be a finite alphabet and $X^*$ the free monoid generated by $X$ with neutral element $1$. For each language $L \subseteq X^*$, the syntactic monoid $M_L$ is the quotient of $X^*$ modulo the syntactic congruence $\sim_L$ which is defined (see [2]): for all $w, w' \in X^*$:

$$w \sim w' \iff \text{ for all contexts } u, v \in X^* \text{ holds } (uwv \in L \Leftrightarrow uw'v \in L).$$

The monoid homomorphism $\sigma : X^* \to M_L : = X^*/\sim_L, w \to \bar{w}$, is called the syntactic morphism.

We define $0 : = \{ w \in X^* \mid \text{for all } u, v \in X^* : uwv \notin L \}$. A generalisation of this concept is the restriction of $X^*$ to any submonoid $M$ of $X^*$. The syntactic monoid of a language $L$ related to some submonoid $M \subset X^*$ is (see [2]):

$$\text{Syn}(M, L) : = \{ [x] \mid x \in M \}$$

and:

$$y \in [x] \iff \text{ for all contexts } u, v \in M : (uxv \in L \Leftrightarrow uyv \in L).$$

By definition: $\text{Syn}(X^*, L) = M_L$.

It is easy to see that $\text{Syn}(M, L)$ divides $\text{Syn}(X^*, L)$, that is: There exists a submonoid $\bar{S} \subset \text{Syn}(X^*, L)$ and a monoid epimorphism $\bar{S} \to \text{Syn}(M, L)$.

Let $R$ denotes any semiring. For a monoid $M$, $R[M]$ is the set of all polynomials in $M$ with coefficients in $R$:

$$R[M] : = \{ p = \sum a_w w \mid w \in M, a_w \in R, a_w = 0 \text{ for almost all } w \}$$

R.A.I.R.O. Informatique théorique/Theoretical Informatics
With addition:
\[ \sum a_w w + \sum b_w w = \sum (a_w + b_w) w \]
and multiplication:
\[ \sum a_w w \cdot \sum b_w w = \sum \left( \sum_{w, w = uv} a_u b_v \right) w, \]

\( R[M] \) is a monoid algebra.

In this monoid algebra we identify the zero in \( R \) and the zero in the monoid (if it exists): \( 0_R = 0_M \), that is, we always deal with the contracted algebra of a monoid (see [1]).

We shall often use the following theorem:

**Lemma 0:** Let \( A, B, C \) be monoids (monoid algebras). Let \( f : A \to B, \) \( g : A \to C \) monoid epimorphism (monoid algebra epimorphism):

1) There exists a canonical monoid (algebra) epimorphism \( h : B \to C \), such that \( h \circ f = g \), iff \( f(a) = f(a') \) implies \( g(a) = g(a') \) for all \( a, a' \in A \);
2) \( h \) is an isomorphism iff \( f(a) = f(a') \Leftrightarrow g(a) = g(a') \).

Proof: Define \( h : B \to C \) with \( h(b) = g(a) \) if \( f(a) = b \).

2. **Syntactic Congruences**

Let \( X \) be a finite alphabet and \( L \subseteq X^* \).

**Definition 1:** We write \( R \langle X \rangle \) instead of \( R[X^*] \) and call \( R \langle X \rangle \) the algebra of polynomials over \( R \).

A series \( S = \sum s_w w \), where \( s_w \in R \) and \( w \in X^* \), is called a formal power series in \( X \) with coefficients in \( R \).

The characteristic power series of a formal language \( L \subseteq X^* \) is \( S = \sum s_w w \), where \( s_w = 1 \) iff \( w \in L \) and \( s_w = 0 \) elsewhere. It is sometimes denoted by \( L \).

We define \( (S, r) : R \langle X \rangle \to R \):
\[ p \mapsto (S, p) = (\sum s_w w, \sum a_w w) = \sum s_w a_w \in R. \]

This sum is finite, because \( p \) is a polynomial.

**Definition 2:** Let \( S \) be a formal power series on \( X \). We define a congruence on \( R \langle X \rangle \) by:
\[ p \equiv q \iff (S, rpr') = (S, rqr') \text{ for all } r, r' \in R \langle X \rangle. \]
We give some properties and show, that $\equiv$ is really a congruence satisfying the two operations in $R \langle X \rangle$:

**Lemma 1:** For all $p, q \in R \langle X \rangle$, we have:
1) $p \equiv q, p' \equiv q' \Rightarrow p + p' \equiv q + q', p \cdot p' \equiv q \cdot q'$;
2) $p \equiv q \iff$ for all $u, v \in X^*$ : $(S, upv) = (S, uqv)$;
3) for all $w, w' \in X^*$ : $w \equiv w' \iff w \sim w'$.

*Proof:* 1) is easy to see;
2) follows because of 1) and the linearity of $(S, .)$;
3) $w \sim w' \iff$ for all $u, v \in X^*$ :
   - $uwv \in L \iff u w' v \in L$;
   - $w \equiv w' \iff (S, uwv) = 1$;
   - $w \equiv w' \iff (S, uwv) = (S, u w' v)$;
   - $w \equiv w'$.

**Definition 3:** We define $\mathscr{A}_S := R \langle X \rangle / \equiv$ and call $\mathscr{A}_S$, remembering that it depends on the characteristic power series of a language, the characteristic algebra of $L$ and the algebra homomorphism $\chi : R \langle X \rangle \rightarrow \mathscr{A}_S$ the characteristic homomorphism of $L$.

The syntactic morphism $\sigma : X^* \rightarrow M_L$ extends to an algebra homomorphism (also called) $\sigma : R \langle X \rangle \rightarrow R[M_L]$, defined by linear extension:

$$\sigma : p = \sum a_w w \mapsto \bar{p} = \sum a_w \sigma(w) = \sum a_w \bar{w}.$$

**Definition 4:** We write $A_L$ instead of $R[M_L]$ and call:

$$A_L := \{ \sum a_w \bar{w} \mid \bar{w} \in M_L, a_w \in R, a_w = 0 \text{ for almost all } w \},$$

the syntactic algebra of $L$.

An obvious question is: Is there a relation between the characteristic algebra and the syntactic algebra of a language? A first answer is:

**Lemma 2:** $\mathscr{A}_S$ is a quotient of $A_L$. There exists a monoid algebra epimorphism $\mathcal{C} : A_L \rightarrow \mathscr{A}_S$.

*Proof:* Because $A_L$ is a free $R$ monoid, we may define $\mathcal{C}$ on the basis $\bar{w} \in M_L$:

$$\bar{w} \mapsto \chi(w) \quad \text{if} \quad w \in \sigma^{-1}(\bar{w}).$$

Using Lemma 1,3 we see that $\mathcal{C}$ is a well defined algebra epimorphism.

3. SYNTACTICAL SEPARABILITY

Let $L \subseteq X^*$ and $S$ the characteristic power series of $L$. 

R.A.I.R.O. Informatique théorique/Theoretical Informatics
Let $R = \mathbb{B}$, the Boolean semiring with elements 0 and 1.

**Definition 5:** A language $L$ is (syntactically) separable iff for all $w \in X^*$, $w \neq \emptyset$, exists a context $u, v \in X^*$ such that: (a) $uwv \in L$; (b) for all $w' \in X^*$: $uw'v \in L \Rightarrow w \sim w'$.

In other words, $L$ is separable, if for each class $w \neq \emptyset$ a special context exists that only takes the class $w$ into the language and no other class.

Some examples may help for a better understanding.

**Examples:**
1) the Goldstine language is separable:

$$X = \{ a, b \}; \quad L = \{ abaaba^3 b \ldots a^n b/n \geq 1 \};$$
2) $X = \{ a \};\quad L_1 = a^+ = \{ a^i/i \geq 1 \}$ is not separable; $L_2 = \{ a^{2i}/i \geq 0 \}$ is separable;
3) $X = \{ a, b, c \}; \quad L = \{ aa, ba, ac, bc \}$ is not separable;
4) $L = D$, the Dyck language is not separable;
5) let $\varphi$ a morphism on a group, then $\varphi^{-1}(1)$ is separable.

**Proof:** Hints: In 1) observe that, for instance, choosing $w = aba$ and $w' = abaaba$, $w \sim w'$ and $aba \in L$ and $w' ab \in L$, but $aba b \notin L$. In 4) choose $w = )(( and $w' = )$, for example.

**Lemma 3:** For all $p = w_1 + \ldots + w_m, q = w'_1 + \ldots + w'_m \in \mathbb{B}\langle X \rangle$:

$p \equiv q$ if and only if for all $u, v \in X^*$ the following hold: $uw_i v \in L$ for some $1 \leq i \leq n \Leftrightarrow uw'_j v \in L$ for some $1 \leq j \leq m$.

**Proof:** Follows from the fact that:

$$(S, upv) = 1 \iff uw_i v \in L$$ for some $1 \leq i \leq n$.

Now we give the main result in this section.

**Theorem 1:** $L$ separable $\iff A_L \cong \mathcal{A}_S$; that is, the algebra homomorphism $\varphi$ defined in Lemma 2 is bijective.

**Proof:** " $\Rightarrow "$ Using Lemma 2 it remains to show:

$$\chi(p) = \chi(q) \Rightarrow \sigma(p) = \sigma(q) \text{ for all } p, q \in \mathbb{B}\langle X \rangle.$$
Proof of the claim: From the definition of separability exists a special
context $u, v \in X^*$ such that $uwv \in L$ and from $\chi(p) = \chi(q)$ and Lemma 3 exists
$j \in \{1, \ldots, m\} : uwv \in L$. It follows $w_i \sim w'_j$.

We have shown: To each summand $w_i$ in $p$ exists a summand $w'_j$ in $q$ such
that $\bar{w}_i = \bar{w}'_j$ and vice versa.

Thus $\sigma(p) = \sigma(q)$.

$\Leftarrow$ $L$ being not separable means: there exists $w \in X^*$ such that for all
$u, v \in X^* : uwv \notin L$ or there exists $w' \in X^*$ with the property $uw'v \in L$ and $w' \sim w$.

It follows $\bar{w} + \bar{w}' \neq \bar{w}'$, but for all $u, v \in X^*$:

$$(S, u (w + w') v) = (S, uwv) + (S, uw'v) = (S, uw'v).$$

Remark: Let us assume for a moment, that $B$ is any semiring $R$. It can be
proved by similar arguments that $L$ separable is a sufficient condition for
$A_L \cong \mathcal{A}_S$, but in contrary not a necessary condition: choose $R = \mathbb{Q}$ and $L = a^*$
in example 3) and compute kernel of $\sigma = \text{kernel of } \chi = $

$$\{ \sum x_w w = \sum x_i a' \in \mathbb{Q} \langle a \rangle / x_0 = 0 \text{ and } \sum_{i \geq 1} x_i = 0 \}$$

and therefore $A_L \cong \mathcal{A}_S$.

4. CHARACTERISTIC ALGEBRAS AND SYNTACTIC MONOIDS

Again, let $X$ be a finite alphabet, $L \subseteq X^*$ and $S = \sum_{w \in L} s_w w$. All
homomorphisms considered are monoid homomorphisms.

Let $B$ be the basisring of $\mathcal{A}_S$.

Definition 6: Let $\bar{X} = X \cup \{[, ], +, 0\}; X_0 = X \cup \{0\}$:

$L = \{ [x_{11} + \ldots + x_{1m_1}] [x_{21} + \ldots + x_{2m_2}] \ldots [x_{n1} + \ldots + x_{nm_n}] / x_{ij} \in X^*_0,$

it exists $i_1, i_2, \ldots, i_n : x_{i_1} x_{i_2} \ldots x_{i_n} \in L \}$.

We call $\bar{L}$ the nondeterministic version of $L$ (see [6]).

Example: $L = D$, the Dyck language; $\bar{L}$ the Greibach language.

We want to show, that $\mathcal{A}_S$ the characteristic algebra of $L$ is isomorphic to
the syntactic monoid of $L$ related to some submonoid of $\bar{X}^*$. But there is a
small problem. There is always a zero in $\mathcal{A}_S$, but is there one in the syntactic
monoid? The answer is yes, because we had forced it by adjoining the special
symbol "0" to the alphabet. It is easy to see, that a word "... [0]..." represents
the zero in the syntactic monoid.
Let $M$ be the free submonoid of $X^*$ with basis:
\[ \{ [x] | x \in X^* \setminus \{ [ \] \}, x = x_1 + \ldots + x_n, x_i \in X_0^* \}, \]
where some $x_i$ may be 1.

Let $\tilde{S} = \text{Syn}(M, L)$ the syntactic of $L$ related to $M$: $\tilde{S} = M/\equiv$, where for $w, w' \in M$:
\[ w \equiv w' \iff \text{for all } u, v \in M : uvw \in L \iff uwv \in \tilde{L}. \]

We need the following fact:

**Lemma 4:** For all $w, w' \in \{ [x_1 + \ldots + x_n] / x_i \in X_0^* \}$:
\[ w \equiv w' \iff \text{for all } u, v \in [X^*] \cup \{ 1 \} : uvw \in L \iff uwv \in \tilde{L}. \]

**Proof:** Is easy to see.

Let $\tilde{\sigma} : M \to \tilde{S}$.

There exists a homomorphism $\rho : M \to B \langle X \rangle$, defined on the basis of $M$ by:
\[ [x] = [x_1 + \ldots + x_n] \mapsto (\hat{x}_1 + \ldots + \hat{x}_n) \in B \langle X \rangle, \]
where $\hat{x}_i = 0_B < x_i >$ if 0 is a factor in $x_i$ and $\hat{x}_i = x_i$ otherwise.

It is easy to see, that $\rho$ is a monoid epimorphism.

Let us recall our situation (all mappings are considered as monoid homomorphisms):

\[ \xymatrix{ & M \ar[r]^-{\rho} & B \langle X \rangle \ar[r]^{-\chi} & \tilde{S} \ar[r]^{-\tilde{\sigma}} & S \ar[r]^-{\psi} & A_S } \]

We want to show that there exists a monoid isomorphism $\psi : A_S \to \tilde{S}$.

**Claim:** for $x, y$ in the basis of $M$:
\[ \chi \rho (x) = \chi \rho (y) \iff \tilde{\sigma} (x) = \tilde{\sigma} (y). \]

**Proof of the claim:** Let:
\[ x = [x_1 + \ldots + x_n]; \quad y = [y_1 + \ldots + y_m]; \quad x_i, y_i \in X_0^*. \]

Now $\chi \rho (x) = \chi \rho (y)$
\[ \iff \text{for all } u, v \in X^* : (ux_i v \in L \text{ for some } i \iff uy_j v \in L \text{ for some } j); \]
\[ \iff \text{for all } u, v \in [X^*] \cup \{ 1 \} : (uxv \in \tilde{L} \iff uyv \in \tilde{L}); \]
\[ \iff \text{for all } u, v \in M : (uxv \in \tilde{L} \iff uyv \in \tilde{L}); \]
\[ \iff \tilde{\sigma} (x) = \tilde{\sigma} (y). \]
We have proved:

**Theorem 2:** $A_s \cong \tilde{S}$; there exists a monoid isomorphism between the characteristic algebra of a language and the syntactic monoid of the nondeterministic version of the language related to some submonoid.

In general, $\tilde{S} = \text{Syn} (M, \bar{L})$ divides $\text{Syn} (\bar{X}^*, \bar{L})$, the syntactic monoid. But we can prove still more.

**Lemma 5:** $\text{Syn} (M, \bar{L})$ is a submonoid of $\text{Syn} (\bar{X}^*, \bar{L})$.

*Proof:* Claim: for all $w, w' \in M : w \equiv w' \iff w \sim w'$.

*Proof of the claim:* Observe that for $u, v \in \bar{X}^* \setminus M : uvv \notin \bar{L}$. We have shown:

$$\text{Syn} (M, \bar{L}) = \{ x \in \text{Syn} (\bar{X}^*, \bar{L}) | \text{it exists } y \in M : x \sim y \}.$$  

**Theorem 2:** $A_s \subseteq \text{Syn} (\bar{X}^*, \bar{L})$; the characteristic algebra of a language is a submonoid of the syntactic algebra of the nondeterministic version of the language.

Let $L$ be a contextfree language. Then $\bar{L}$, the nondeterministic version of $L$, is also contextfree.

**Corollary:** For all contextfree languages $L \subseteq X^*$ exist a contextfree language $\bar{L} = (X \cup \{ [ , ], +, 0 \})^*$ such that:

1) $A_{\bar{L}}$ is the syntactic monoid of $\bar{L}$ related to a submonoid $M : (X \cup \{ [ , ], +, 0 \})^*$.

2) $A_{\bar{L}}$ is a submonoid of the syntactic monoid of $\bar{L}$.

**References**


R.A.I.R.O. Informatique théorique/ theoretical Informatics