Acta Cybernetica **19** (2009) 553–565.

Kleene Revisited by Suschkewitsch

Jean-Marcel Pallo*

Abstract

The aim of this paper is to generalize to nonassociative concatenation the well-known property that the family of left-linear languages is exactly the family of regular languages. For this purpose, we introduce a generalized Kleene star operation.

1 Introduction

It is well-known that the family of left-linear languages is exactly the family of regular languages. This classical result comes from the formal languages theory, which has been developed over free monoids generated by alphabets and equipped with an associative concatenation.

The purpose of this paper is to generalize this statement to the case where the concatenation is no longer associative [13, 14, 15]. The theory of quasigroups [9, 12, 17, 20] is originated from a certain idea of Suschkewitsch [25], which idea we use here for this goal. Given a group $(\mathcal{G}, \star)$, Suschkewitsch observed in 1929 that the proof of Lagrange theorem does not make any use of the associative law: $X \star (Y \star Z) = (X \star Y) \star Z$. This law can be replaced by his more general postulates, $\mathcal{A}$ and $\mathcal{B}$ namely. Postulate $\mathcal{A}$ of [25] can be written as: for all $A, B \in \mathcal{G}$ there is a unique $C \in \mathcal{G}$ such that for all $X \in \mathcal{G}$ we have $(X \star A) \star B = X \star C$. The element C depends upon the elements A and B only and not upon X. If we denote C by $A \circ B$, i.e. $(X \star A) \star B = X \star (A \circ B)$, it is easy to prove that $\circ$ is associative. It has been shown recently in [19] by the author that Postulate $\mathcal{A}$ is a particular case of the concept of relative associativity introduced by Roubaud [21].

Suschkewitsch also considers a special case of Postulate $\mathcal{A}$ which is however more general than the associative law. He states his Postulate $\mathcal{B}$ as: for all $B \in \mathcal{G}$ there is a unique $\widetilde{B} \in \mathcal{G}$ such that for all $X, Y \in \mathcal{G}$ we have $X \star (Y \star B) = (X \star Y) \star \widetilde{B}$. The elements B and $\widetilde{B}$ depend only upon each other. Every B is completely defined by the corresponding $\widetilde{B}$ and conversely. Using Postulate $\mathcal{B}$, it has been shown in [7] that each left-linear language defined with a nonassociative concatenation is a pseudo-regular language, i.e. it can be written by using a generalized Kleene star operation.

*Université de Bourgogne, Département d'Informatique, BP 47870, 21078 DIJON-Cedex, France, E-mail: pallo@u-bourgogne.fr

The converse property (pseudo-regularity implies left-linearity) is established in this paper via a new method. To do this, we are lead to define the inverse $-$ of the operation $\sim$ of Suschkewitsch. Certain coherence problems arise. Then the original groupoid M (i.e. the monoid without associativity) has to be embedded in a larger one. It is absolutely necessary to check that this embedding creates no additional weak associative relation in M, as for example $((xy)(zt))(uv) = (x(yz))((tu)v)$. Afterwards a pseudo-Kleene star operation is defined as a generalization of the standard Kleene star over the monoid. Then we show that the family of left-linear languages defined with a nonassociative concatenation is *exactly* the family of pseudo-regular languages using this new star operation.
Another approach consists of considering trees in place of words, since the notion of a tree strongly reflects nonassociativity. A huge amount of literature relies upon this concept [6].

2 Notations and definitions

Let $V = \{a, b, c, d \dots\}$ be an alphabet, i.e. a finite nonempty set of letters. We denote by $M = M(V)$ the free groupoid over V equipped with a nonassociative concatenation $\bullet$. The symbol $\bullet$ will be omitted as long as no confusion is possible. To write words of M, parentheses cannot be omitted due to nonassociativity: $x(yz) \neq (xy)z$. By λ we denote the empty word of zero letters. The set of all non-empty words over V is denoted M^+.
To the finite alphabet V, we associate the (infinite) set $\mathcal{V}$ which is the set of letters of V with as a superscript an arbitrary number of symbols $-$ and $\sim$.
We denote by $\mathcal{M} = \mathcal{M}(\mathcal{V})$ the free groupoid over $\mathcal{V}$ equipped with the nonassociative concatenation $\bullet$. Let us define $\mathcal{M}^+ = \mathcal{M}(\mathcal{V}) - \{\lambda\}$. Since $M \subset \mathcal{M}$, a word in M is called a real word. We call a word of $\mathcal{M} - M$ a metaword. Given $w \in \mathcal{M}$, we denote by $|w|$ the length of w, i.e. the number of letters of V in w, each letter is counted as many times it occurs. Let us denote by $\mathcal{M}_n$ the set of words such that $|w| = n$. For example, $w^{'} = ((ab)(cd))(e(bd)) \in \mathcal{M}_7$ is a real word and $w^{''} = (e(\bar{\bar{a}}\ \tilde{c}))((a\ \tilde{\tilde{b}}\ )\ \bar{c}) \in \mathcal{M}_6$ is a metaword. The skeleton of a word $w \in \mathcal{M}$ is defined as the real word $sk(w)$ obtained by cancelling all the occurrences of the symbols $-$ and $\sim$. The free word of $w \in \mathcal{M}$ is defined as the word $f(w)$ obtained by cancelling all the occurrences of the symbols $-$, $\sim$, (and). For example, $sk(w^{''}) = (e(ac))((ab)c)$ and $f(w^{''}) = eacabc$.
We call a left (respectively right) word any word where all the open (respectively close) parentheses occur at the beginning (respectively at the end) of the word. For example, $(((((a\ \bar{b})\ \tilde{c})b)a)\ \tilde{c})$ is a left word and $(\tilde{a}\ (b(\bar{a}\ (c(a\ \bar{d}))))$ is a right word.
A weak associative equality of order n is an equation of the form $w^{'} = w^{''}$ where $w^{'}, w^{''} \in \mathcal{M}_n$. For example, $a(b(c(\bar{d}\bar{e}))) = ((ab)c)(\tilde{d}\tilde{e})$ is a weak associative equality of order 5. A weak associative equality is called real if the two words of the

equality are real, i.e. there is no occurrences of $-$ and $\sim$ inside the two words.

Definition 1. *We define the Suschkewitsch algebra $\mathcal{S} = \mathcal{S}(\mathcal{M})$ as the extension of $\mathcal{M}$ with unary operations $-$ and $\sim$ such that the following properties hold for all $P, Q, R \in \mathcal{M}^+$:*

$$\begin{cases} (PQ)R = P(Q\ \bar{R}) \\ P(QR) = (PQ)\ \tilde{R} \\ \ \bar{\tilde{P}}\ =\ \tilde{\bar{P}}\ = P \\ \overline{PQ} = \bar{P}\bar{Q} \\ \widetilde{PQ} = \tilde{P}\tilde{Q} \end{cases}$$

We will establish later that no real words in $\mathcal{S}$ are forced to collapse by the five previous axioms.

3 Preliminary results

Definition 2. *Let us define the relation $\to$ on $\mathcal{S}$ as the smallest preordering, invariant with respect to $-$, $\sim$ and $\bullet$, i.e. if $P \to Q$ then $\bar{P}\to\bar{Q}$, $\tilde{P}\to\tilde{Q}$ and $PR \to QR$, $RP \to RQ$ for all $R \in \mathcal{M}$, and satisfying for all $P, Q, R \in \mathcal{M}^+$:*

$$P(QR) \to (PQ)\ \tilde{R}$$

.

Lemma 1. *We have $AB \to CD$ iff either (1)*

$$\begin{cases} A \to C \\ B \to D \end{cases}$$

or there exists $S \in \mathcal{M}^+$ such that (2)

$$\begin{cases} B \to S\ \bar{D} \\ AS \to C \end{cases}$$

Proof. The conditions are obviously sufficient since $AB \to A(S\ \bar{D}) \to (AS)\ \tilde{\bar{D}}= (AS)D \to CD$.
For proving the necessity, let us consider $A, B, C, D \in \mathcal{M}^+$. We define the relation $\prec$ on $\mathcal{M}$: $AB \prec CD$, iff conditions either (1) or (2) are verified.
$\prec$ is reflexive and invariant with respect to $-$,$\sim$ and $\bullet$. Let us prove the transitivity of $\prec$, i.e. if $AB \prec CD$ and $CD \prec EF$ then we have $AB \prec EF$. Among the four

cases to study, we only detail the following one.
If there exist $S, T \in \mathcal{M}^+$ such that

$$\begin{cases} B \to S\,\bar{D} \\ AS \to C \end{cases}$$

and

$$\begin{cases} D \to T\,\bar{F} \\ CT \to E \end{cases}$$

then there exists $U \in \mathcal{M}^+$ such that

$$\begin{cases} B \to U\,\bar{F} \\ AU \to E \end{cases} .$$

Indeed, $U = S\,\bar{T}$ verifies: $B \to S\,\bar{D} \to S(\bar{T}\,\bar{\bar{F}}) \to (S\,\bar{T})\,\tilde{\bar{\bar{F}}} = U\,\bar{F}$ and $AU = A(S\,\bar{T}) \to (AS)\,\tilde{\bar{T}} = (AS)T \to CT \to E.$ □

Remark 1. The rewrite relation $\to$ is convergent because it is well-known that the rewriting system $x(yz) \leadsto (xy)z$ is convergent. Thus any word w has a unique normal form denoted by $l(w)$ which is a left word.

Lemma 2. *If $A, B, C, D \in \mathcal{M}^+$, assume that:*
(1) $AB \to CD$,
(2) AB is a real word,
(3) CD is a left word,
(4) there exists $S \in \mathcal{M}^+$ such that $B \to S\,\bar{D}$ and $AS \to C$.
Then we can always choose S as a real word.

Proof. $|D| = 1$ i.e. $D \in \mathcal{V}$ since CD is a left word. $|B| \geq 2$ since $|S| \geq 1$. Thus we can write $B = B_1B_2$. If $|B_2| = 1$, then $B_2 = d \in V$ and we can choose $D = \tilde{d}$ and $S = B_1$ which is a real word. If $|B_2| \geq 2$ and if d is the last letter of B_2, we can write $B_2 = (B_{2,k}(\ldots(B_{2,3}(B_{2,2}(B_{2,1}d)^k$. Then
$B_2 = (B_{2,k}(\ldots(B_{2,3}(B_{2,2}B_{2,1})^{k-1}\,\overset{\sim k-1}{d}$ and
$B = B_1B_2 = (B_1(B_{2,k}(\ldots(B_{2,3}(B_{2,2}B_{2,1})^k\,\overset{\sim k}{d}$. We can choose
$S = (B_1(B_{2,k}(\ldots(B_{2,3}(B_{2,2}B_{2,1})^k$ and $D = \overset{\sim k+1}{d}$. Thus S is a real word. □

Lemma 3. *The following rewriting system $\leadsto$:*
(R1) $(xy)z \leadsto x(y\,\bar{z})$
(R2) $\bar{\tilde{x}} \leadsto x$

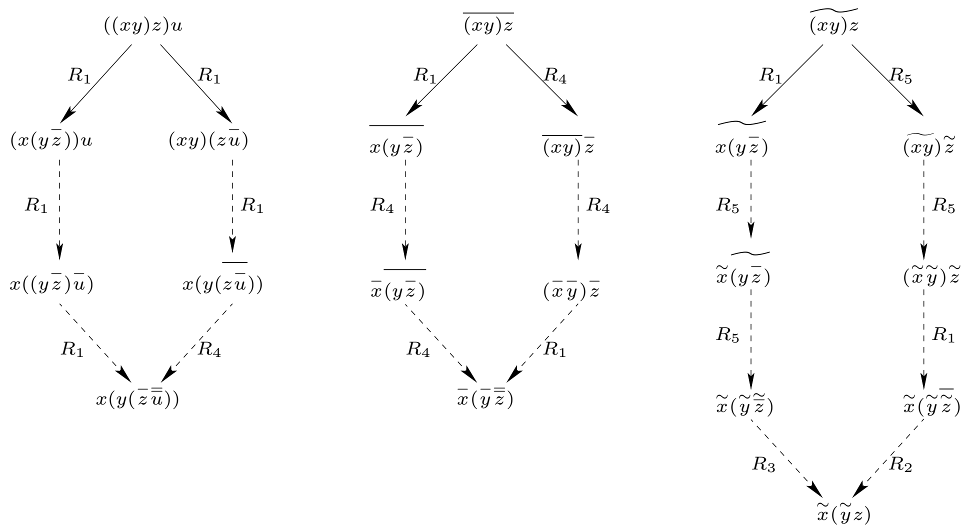

Figure 1: Three convergent critical pairs of $\rightsquigarrow$

(R3) $\widetilde{\overline{x}} \rightsquigarrow x$
(R4) $\overline{xy} \rightsquigarrow \overline{x}\,\overline{y}$
(R5) $\widetilde{xy} \rightsquigarrow \widetilde{x}\widetilde{y}$
is convergent. Thus any term t has a unique normal form which is a right word denoted by $r(t)$.

Proof. Termination is easily proved because $\rightsquigarrow$ is included in the Knuth-Bendix ordering. We choose as precedence relation $- \succ \sim \succ \bullet$ and as weights $wgt(-) = wgt(\sim) = 0$ and $wgt(\bullet) = 1$. The superposition of (R1) on (R1), of (R1) on (R4) and of (R1) on (R5) determines three critical pairs which are convergent: see Figure 1. Thus this rewriting system is convergent [1, 10]. Normal forms are right words. □

4 Nonassociativity

Definition 3. *Given $w \in \mathcal{M}_n$, let us call associahedron $\mathcal{AS}_n(w)$ the diagram which is obtained from w by applying all possible $\rightarrow$ and $\stackrel{-1}{\rightarrow}$ relations.*

See for example the associahedron $\mathcal{AS}_5((x(y(zt)))u)$ in Figure 2.

Theorem 1. *For all n and $w \in \mathcal{M}_n$, the associahedron $\mathcal{AS}_n(w)$ is coherent, i.e. there are no $w', w'' \in \mathcal{AS}_n(w)$ such that $w' \neq w''$ and $sk(w') = sk(w'')$. Therefore*

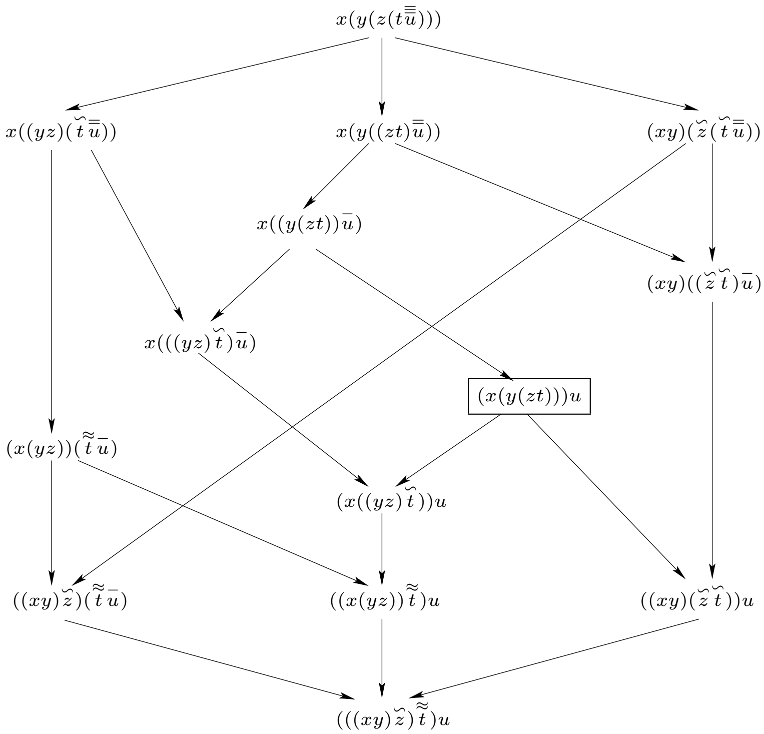

Figure 2: The associahedron $\mathcal{AS}_5((x(y(zt)))u)$

there exist in $\mathcal{AS}_n(w)$ unique words $l(w)$ and $r(w)$ which are respectively left and right words.

Proof. If $n = 4$, the property holds for the five pentagons: see Figure 3. By induction on n, suppose that for $n \geq 5$ and for some $v \in \mathcal{M}_n$ we have in $\mathcal{AS}_n(v)$: $w = AB \to w' = C'D'$ and $w = AB \to w'' = C''D''$ with $|w| = |w'| = |w''| = n$, $w' \neq w''$ and $sk(w') = sk(w'')$. Then we obtain $sk(C') = sk(C'')$ and $sk(D') = sk(D'')$. We apply Lemma 1.
If $A \to C'$, $B \to D'$ and if $A \to C''$, $B \to D''$ then we have $|A| \leq n-1$, $|B| \leq n-1$ and by the induction hypothesis $C' = C''$ and $D' = D''$ since $sk(C') = sk(C'')$ and $sk(D') = sk(D'')$. A contradiction follows since $w' \neq w''$.

If $A \to C'$, $B \to D'$, suppose that there exists $S'' \in \mathcal{S}^+$ such that $B \to S''\,\bar{D}''$ and $AS'' \to C''$. Therefore we have $|B| = |D'| = |S''| + |D''|$ and $|D'| = |D''|$ because $sk(D') = sk(D'')$. A contradiction follows since $w'' \neq \lambda$.

Now let us assume that there exist $S', S'' \in \mathcal{M}^+$ such that $B \to S'\,\bar{D}'$, $AS' \to C'$, $B \to S''\,\bar{D}''$, $AS'' \to C''$. If $D' \neq D''$ then we have $l(D') \neq l(D'')$ by induction because $|D^{'}| \leq n-1$ and $|D^{''}| \leq n-1$. Then $B \rightarrow S^{'}\ \overline{D^{'}} \rightarrow l(S^{'})l(\overline{D^{'}})$ and $B \rightarrow S^{''}\ \overline{D^{''}} \rightarrow l(S^{''})l(\overline{D^{''}})$. We have $sk(l(S^{'})l(\overline{D^{'}})) = sk(l(S^{''})l(\overline{D^{''}}))$ but $l(S^{'})l(\overline{D^{'}}) \neq l(S^{''})l(\overline{D^{''}})$ because $l(D^{'}) \neq l(D^{''})$. Thus, a contradiction follows. Therefore we obtain $D^{'} = D^{''} = D$, $B \rightarrow l(S^{'})D$, $B \rightarrow l(S^{''})D$. Since $sk(l(S^{'})D) = sk(l(S^{''})D)$ and $|B| \leq n-1$, we deduce by induction that $l(S^{'})D = l(S^{''})D$, i.e. $l(S^{'}) = l(S^{''})$. Now let us consider the following diagram:

$$C^{'} \leftarrow AS^{'} \rightarrow Al(S^{'}) = Al(S^{''}) \leftarrow AS^{''} \rightarrow C^{''}.$$

The induction hypothesis can be applied because $|C^{'}| = |C^{''}| \leq n-1$. Then $sk(C^{'}) = sk(C^{''})$ implies $C^{'} = C^{''}$ and a contradiction holds. The existence of $l(w)$ (respectively $r(w)$) is proved using Lemma 1 (respectively Lemma 3). □

Remark 2. The associahedrons $\mathcal{AS}_n(w)$ endowed with the relation $\rightarrow$ are lattices for all n and $w \in \mathcal{M}_n$. This is an immediate consequence of the fact that the skeleton of $\mathcal{AS}_n(w)$ is the well-known *n-th* Tamari lattice. Tamari lattices have been extensively studied for algebraic and combinatorial purposes. A number of references on this subject is available in [18].

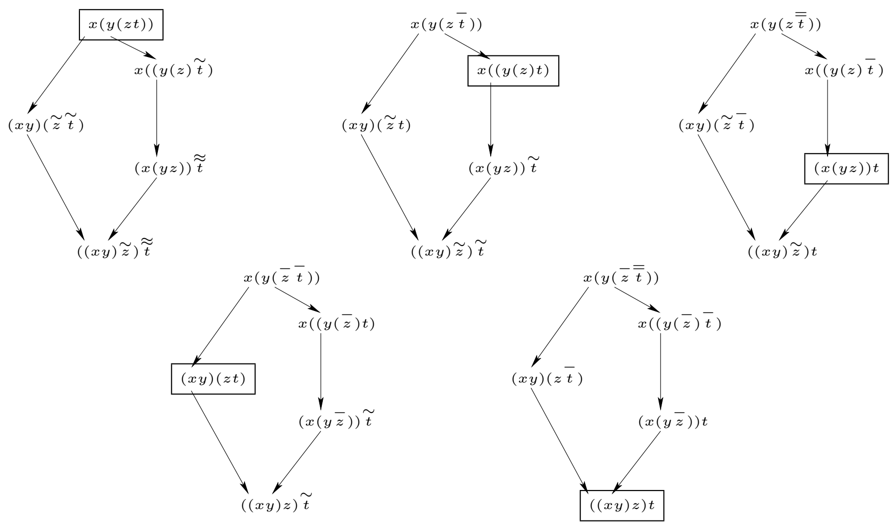

Figure 3: The five associahedrons of size 4

Theorem 2. *The embedding of $\mathcal{M}$ into $\mathcal{S}$ is faithful.*

Proof. Let us prove that the embedding of $\mathcal{M}$ into $\mathcal{S}$ does not create any weak real associativity equality of order n, that is whatever n one cannot find two real words

$w^{'}, w^{''} \in \mathcal{M}_n$ such that $w^{'} = w^{''}$ and $sk(w^{'}) \neq sk(w^{''})$, when applying the five relations of Definition 1.
If $n = 4$, the property holds for the five pentagons: see Figure 3. By induction on n, suppose that for $n \geq 5$ there exist two real words $w^{'} = A^{'}B^{'}, w^{''} = A^{''}B^{''} \in \mathcal{M}_n$ such that $w^{'} = w^{''}$ with $sk(w^{'}) \neq sk(w^{''})$. Following Lemma 3: $r(w^{'}) = r(w^{''}) = v$. In $\mathcal{AS}_n(v)$ there exists a unique left word CD such that $A^{'}B^{'} \to CD$ and $A^{''}B^{''} \to CD$. Since CD is a left word, we have $|D| = 1$. We apply Lemma 1.
If $A^{'} \to C$, $B^{'} \to D$, $A^{''} \to C$, $B^{''} \to D$, then we have $|A^{'}| = |A^{''}| = |C| = n-1$ and the diagram $A^{'} \to C \leftarrow A^{''}$ with $A^{'}$ and $A^{''}$ real, contradicting the induction hypothesis.

If $A^{'} \to C$, $B^{'} \to D$, suppose that there exists $S^{''} \in \mathcal{M}^{+}$ such that $B^{''} \to S^{''} \bar{D}$ and $A^{''}S^{''} \to C$. Since $|D| = 1$ and $B^{'}$ real, then D is a real word, i.e. $D = d \in V$ and thus $B^{''} \to S^{''} \bar{d}$. If the arrow $\to$ is applied to a real word, we can obtain letters with superscript $\sim$ only. Hence, a contradiction follows.

Suppose now that there exist $S^{'}, S^{''} \in \mathcal{M}^{+}$ such that $B^{'} \to S^{'} \bar{D}$, $A^{'}S^{'} \to C$, $B^{''} \to S^{''} \bar{D}$, $A^{''}S^{''} \to C$. Following Lemma 2, since $B^{'}$ and $B^{''}$ are real words, we can choose $S^{'}$ and $S^{''}$ as real words. Then we have $A^{'}S^{'} \to C \leftarrow A^{''}S^{''}$ with $A^{'}S^{'}$ and $A^{''}S^{''}$ which are real words. Since $|A^{'}S^{'}| = |A^{''}S^{''}| = |C| = n-1$, we can apply the inductive hypothesis and we obtain $A^{'}S^{'} = A^{''}S^{''}$. Hence $A^{'} = A^{''}$ and $S^{'} = S^{''} = S$. The diagram $B^{'} \to S\, \bar{D} \leftarrow B^{''}$ follows. Since $|B^{'}| \leq n-1$, the induction hypothesis holds and then $B^{'} = B^{''}$. □

5 Catalan sequences

Definition 4. *If $w \in \mathcal{M}_n$ is a real word, the left word $l(w)$ which is the normal form of w contains only $\sim$ symbols and no $-$ symbols. If $f(w) = x_1x_2 \ldots x_n$, we thus can write $l(w) = ({}^{n-1}x_1x_2)\ \tilde{x}_3^{\,l_1})\ \tilde{x}_4^{\,l_2}) \ldots)\ \tilde{x}_n^{\,l_{n-2}})$. The sequence $L(w) = (l_1, l_2, \ldots, l_{n-2})$ is called the left-sequence of $w \in \mathcal{M}_n$.*

Lemma 4. *If $w^{'} \in \mathcal{M}_n$ and $w^{''} \in \mathcal{M}_m$ with $L(w^{'}) = (l_1^{'}, l_2^{'}, \ldots, l_{n-2}^{'})$ and $L(w^{''}) = (l_1^{''}, l_2^{''}, \ldots, l_{m-2}^{''})$, then the left-sequence of the concatenation of $w^{'}$ and $w^{''}$ is $L(w^{'}w^{''}) = (l_1^{'}, l_2^{'}, \ldots, l_{n-2}^{'}, 0, 1, l_1^{''}+1, l_2^{''}+1, \ldots, l_{m-2}^{''}+1)$.*

Proof. If $f(w^{'}) = x_1x_2 \ldots x_n$ and $f(w^{''}) = y_1y_2 \ldots y_m$, then we have: $l(w^{'}) = ({}^{n-1}x_1x_2)\ \tilde{x}_3^{\,l_1^{'}}) \ldots \tilde{x}_n^{\,l_{n-2}^{'}})$ and $l(w^{''}) = ({}^{m-1}y_1y_2)\ \tilde{x}_3^{\,l_1^{''}}) \ldots \tilde{y}_m^{\,l_{m-2}^{''}})$. Therefore we can write:

$w'w'' = ({}^{n-1}x_1x_2)\ \tilde{x}_3^{\,l_1^{'}}) \ldots \tilde{x}_n^{\,l_{n-2}^{'}})({}^{m-1}y_1y_2)\ \tilde{y}_3^{\,l_1^{''}}) \ldots \tilde{y}_m^{\,l_{m-2}^{''}}) \to$
$({}^{n}x_1x_2)\ \tilde{x}_3^{\,l_1^{'}}) \ldots \tilde{x}_n^{\,l_{n-2}^{'}})({}^{m-2}y_1y_2)\ \tilde{y}_3^{\,l_1^{''}}) \ldots \tilde{y}_m^{\,l_{m-2}^{''}+1}) \to$
$({}^{n+1}x_1x_2)\ \tilde{x}_3^{\,l_1^{'}}) \ldots \tilde{x}_n^{\,l_{n-2}^{'}})({}^{m-3}y_1y_2)\ \tilde{y}_3^{\,l_1^{''}}) \ldots \widetilde{y_{m-1}}^{\,l_{m-3}^{''}+1})\ \tilde{y}_m^{\,l_{m-2}^{''}+1}) \to \ldots \to$
$({}^{n+m-2}x_1x_2)\ \tilde{x}_3^{\,l_1^{'}}) \ldots \tilde{x}_n^{\,l_{n-2}^{'}})y_1)\ \tilde{y}_2)\ \tilde{y}_3^{\,l_1^{''}+1}) \ldots \tilde{y}_m^{\,l_{m-2}^{''}+1}) = l(w^{'}w^{''})$. □

Theorem 3. *An integer sequence $(l_i)_{1 \leq i \leq n-2}$ is the left-sequence of a word of $\mathcal{M}_n$ iff $l_1 \in \{0, 1\}$ and for all $i \in [1, n-3]$: $0 \leq l_{i+1} \leq l_i + 1$.*

Proof. The proof comes by induction on n using Lemma 4. □

The five left-sequences of the words of $\mathcal{M}_4$ are $\{00, 01, 10, 11, 12\}$. The fourteen left-sequences of the words of $\mathcal{M}_5$ are in lexicographic order:

$$\{000, 001, 010, 011, 012, 100, 101, 110, 111, 112, 120, 121, 122, 123\}$$

The left-sequence characterized just above is exhausted among the 66 Catalan sets in [24, p. 222] where it is denoted by (**u**). The right-sequence of w defined from the right word $r(w)$ has been studied in [7] and appears in [24, p. 222] under the notation (**s**).

6 Rational formal power series

We use the classical notations on formal power series described in [2, 11, 22].
Given a semiring $\mathcal{A}$, we denote by $\mathcal{A}[[\mathcal{S}]]$ the set of formal series

$$s = \sum_{\sigma \in \mathcal{S}} < s, \sigma > \sigma$$

where $< s, \sigma > \in \mathcal{A}$.
The sum of two series is defined in the classical manner. The product $s = s^{'} s^{''}$ is defined by $< s, \sigma > = < s^{'}, \sigma^{'} > < s^{''}, \sigma^{''} >$ if $\sigma = \sigma^{'} \sigma^{''}$ and $< s, \sigma > = 0$ otherwise.
$s \in \mathcal{A}[[\mathcal{S}]]$ is proper if the coefficient of the right unit λ (i.e. the constant term of s) vanishes: $< s, \lambda > = 0$.
In this case, the series

$$s^{\star} = \lambda + s + s\ \bar{s} + (s\ \bar{s})\ \bar{s} + ((s\ \bar{s})\ \bar{s})\ \bar{s} + (((s\ \bar{s})\ \bar{s})\ \bar{s})\ \bar{s} + \ldots$$

is defined. We have also

$$s^{\star} = \lambda + s + s\ \bar{s} + s(\bar{s}\ \bar{\bar{s}}) + s(\bar{s}\ (\bar{\bar{s}}\ \bar{\bar{\bar{s}}})) + s(\bar{s}\ (\bar{\bar{s}}\ (\bar{\bar{\bar{s}}}\ \bar{\bar{\bar{\bar{s}}}}))) + \ldots$$

.

Definition 5. *We call $s^{\star}$ the pseudo-Kleene star of the series $s \in \mathcal{A}[[\mathcal{S}]]$.*

Lemma 5. *Let $r, s \in \mathcal{A}[[\mathcal{S}]]$ with s proper. Then the unique solution u of the left-linear equation $u = r + us$ is the series $u = rs^{\star}$.*

Proof. One has $s^\star = \lambda + s^\star\ \bar{s}$ whence $rs^\star = r + r(s^\star\ \bar{s})$ and $rs^\star = r + (rs^\star)\ \overset{\sim}{\bar{s}} = r + (rs^\star)s$. Conversely, from $u = r + us$ it follows that $u = r + (r + us)s = r + rs + (us)s = r + rs + u(s\ \bar{s})$ and inductively $u = r(\lambda + s + s\ \bar{s} + (s\ \bar{s})\ \bar{s} + \ldots + ({}^{n}s\ \bar{s})\ \bar{s})\ldots)\ \bar{s}) + u({}^{n+1}s\ \bar{s})\ \bar{s})\ldots)\ \bar{s})$. Thus going to the limit, one gets $u = rs^\star$ since s is proper. □

7 Kleene theorem

Definition 6. *A formal series is pseudo-rational if it is an element of the smallest subset* $Rat[[\mathcal{S}]]$ *of* $\mathcal{A}[[\mathcal{S}]]$ *containing* $\mathcal{S}$ *and closed for the sum, product and pseudo-Kleene star operation* $\star$.

Definition 7. *A left-linear system of order* N *with pseudo-rational coefficients is a system of the form*

$$u_i = r_i + \sum_{1 \le j \le N} u_j s_{i,j}$$

with $1 \le i \le N$ *where all* $r_i, s_{i,j} \in Rat[[\mathcal{S}]]$.

Theorem 4. *The components of the* N*-tuple solution of a left-linear system with proper pseudo-rational coefficients are pseudo-rational series. Conversely, a pseudo-rational series can be obtained as a component of a* N*-tuple solution of such a system.*

Proof. The proof is done by induction on N. According to Lemma 5, the solution of $u = r + us$ is $u = rs^\star$ which is a pseudo-rational series since $r, s \in Rat[[\mathcal{S}]]$. In a system of order N, u_N is rationally computed from $u_1, u_2, \ldots, u_{N-1}$ and the induction hypothesis is applied.
Conversely, let us prove that the components which are solutions of left-linear systems with pseudo-rational coefficients verify the conditions of Definition 6. Let us denote by u_1 (respectively $u_1^{'}$) the first component of the N-tuple solution (respectively $N^{'}$-tuple solution) of a system $\mathbf{S}$ (respectively $\mathbf{S}^{'}$):

$$\mathbf{S} : u_i = r_i + \sum_{1 \le j \le N} u_j s_{i,j}, 1 \le i \le N$$

and

$$\mathbf{S}^{'} : u_i^{'} = r_i^{'} + \sum_{1 \le j \le N'} u_j^{'} s_{i,j}^{'}, 1 \le i \le N^{'}$$

where all $r_i, r_i^{'}, s_{i,j}, s_{i,j}^{'} \in Rat[[\mathcal{S}]]$.
It is easy to exhibit a system which admits as solution $c_1 u_1 + c_1^{'} u_1^{'}$ with $c_1, c_1^{'} \in \mathcal{A}$.
Now, let $\hat{u}_i = u_1^{'} u_i$. Then

$$u_1^{'} u_i = u_1^{'} r_i + \sum_{1 \le j \le N} u_1^{'} (u_j s_{i,j})$$

and

$$u_1^{'} u_i = u_1^{'} r_i + \sum_{1 \le j \le N} (u_1^{'} u_j) \widetilde{s_{i,j}}$$

Thus $\overset{\wedge}{u_1} = u_1^{'} u_1$ is the first component of the N-tuple solution of the system $\overset{\wedge}{\mathbf{S}}$:

$$\overset{\wedge}{\mathbf{S}} : \overset{\wedge}{u_i} = u_1^{'} r_i + \sum_{1 \le j \le N} \overset{\wedge}{u_j} \widetilde{s_{i,j}}, 1 \le i \le N$$

To conclude, let $\overset{\vee}{u_i} = u_1^{*} u_i$. Then

$$u_1^{\star} u_i = u_1^{\star} r_i + \sum_{1 \le j \le N} u_1^{\star}(u_j s_{i,j})$$

and

$$u_1^{\star} u_i = u_1^{\star} r_i + \sum_{1 \le j \le N} (u_1^{\star} u_j) \widetilde{s_{i,j}}$$

Thus $\overset{\vee}{u_1} = u_1^{\star} u_1 = u_1^{\star} - \lambda$ is the first component of the N-tuple solution of the system $\overset{\vee}{\mathbf{S}}$:

$$\overset{\vee}{\mathbf{S}} : \overset{\vee}{u_i} = u_1^{\star} r_i + \sum_{1 \le j \le N} \overset{\vee}{u_j} \widetilde{s_{i,j}}, 1 \le i \le N$$

□

8 Conclusion

Theorem 4 characterizes à-la-Kleene pseudo-rational series defined with a nonassociative concatenation. The key point in the proof of Theorem 4 is the fact that $-$ and $\sim$ are mutually inverse operations. The axiom $(xy)z = x(y \bar{z})$ allows to factor out r in the solution of the equation $u = r + us$ and therefore to obtain $u = rs^*$ where s^* is defined in terms of $-$. The converse axiom $x(yz) = (xy) \tilde{z}$ allows to show that products and stars of solutions of left-linear systems are yet solutions of certain other left-linear systems that can be computed in terms of $\sim$. The embedding of the original groupoid into the Suschkewitsch algebra creates no weak associative relation. It means that this embedding is faithful.

Theorem 4 is a generalization of the famous Kleene theorem, one of the cornerstones of theoretical computer science. See also [8, 23] for linear languages, [3] for clock languages, [26] for ∞-languages, [4] for a Conway-like approach and [5] for binoid languages.

If $\mathcal{A}$ has the additional structure of a ring (i.e. subtraction is allowed), let us define on $\mathcal{A}[[\mathcal{S}]]$ the bracket: $[s,t] = s \bar{t} - t \bar{s}$. We can easily verify that the Jacobi identity holds: $[s,[t,u]] + [t,[u,s]] + [u,[s,t]] = 0$. This remark could be the starting point of a later study.

Acknowledgments

I am very indebted to the anonymous referees for proofreading my paper.

References

[1] Bachmair, L. *Canonical Equational Proofs.* Birkhäuser, Boston, 1991.

[2] Berstel, J. and Reutenauer, C. *Rational Series and their Languages.* EATCS Monographs on Theoretical Computer Science, vol. 12, Springer, Berlin, 1988.

[3] Bouyer, B. and Petit, A. A Kleene/Büchi-like theorem for clock languages. *J. Autom. Lang. Comb.,* 7:167–186, 2002.

[4] Ésik, Z. and Kuich, W. On iteration semiring-semimodule pairs. *Semigroup Forum*, 75:129 – 159, 2007.

[5] Gazdag, Z. and Németh, Z.L., A Kleene theorem for binoid languages. Submitted to International Journal of Foundations of Computer Science, 2008.

[6] Gécseg, F. and Steinby, M. Tree languages. In Rozenberg, G. and Salomaa, A., editors, *Handbook of Formal Languages*, vol. 3: Beyond Words, pages 1–68. Springer, 1997.

[7] Germain, C. and Pallo, J. Langages rationnels définis avec une concaténation non-associative. *Theor. Comput. Sci.,* 233:217–231, 2000.

[8] Germain, C. and Pallo, J. Linear languages with a nonassociative concatenation. *J. Autom. Lang. Comb.,* 7:311–320, 2002.

[9] Hausmann, B.A. and Ore, O. Theory of quasigroups. *Amer. J. Math.,* 59:983–1004, 1937.

[10] Jantzen, M. Basics of Term Rewriting. In Rozenberg, G. and Salomaa, A., editors, *Handbook of Formal Languages*, vol. 3: Beyond Words, pages 269–313. Springer, 1997.

[11] Kuich, W. Semirings and Formal Power Series: Their Relevance to Formal Languages and Automata. In Rozenberg, G. and Salomaa, A., editors, *Handbook of Formal Languages*, vol. 1: Word, Language, Grammar, pages 609–677. Springer, 1997.

[12] Kunen, K. Quasigroups, loops and associative laws. *J. Algebra,* 185:194–204, 1996.

[13] Kuz'min, E.N. and Shestakov, I.P. Non-Associative Structures. In Kostrikin, A.L. and Shafarevich, I.L., editors, *Chap. II of Algebra VI, Encyclopaedia of Mathematical Sciences,* vol. 57. Springer, 1995.

[14] Lõhmus, J. Preface to the Special Issue on Nonassociative Algebras. Quasigroups and Applications in Physics. *Acta Applic. Math.,* 50:1–2, 1998.

[15] Lõhmus, J., Paal, E. and Sorgsepp L. About nonassociativity in mathematics and physics. *Acta Applic. Math.,* 50:3–31, 1998.

[16] Mac Lane, S. Natural associativity and commutativity. *Rice Univ. Stud.,* 49:28–46, 1963.

[17] Murdoch, D.C. Quasigroups which satisfy certain generalized associative law. *Amer. J. Math.,* 61:509–522, 1939.

[18] Pallo, J.M. Generating binary trees by Glivenko classes on Tamari lattices. *Inform. Process. Lett.,* 85: 235–238, 2003.

[19] Pallo, J.M. Nonassociativity à la Kleene. In Bozapalidis, S. and Rohanis, G., editors, *Proc. Second Conf. Algebraic Informatics,* Lecture Notes in Computer Science, vol. 4728, pages 260–274. Springer, 2007.

[20] Pflugfelder, H.O. Historical notes on loop theory, Comment. *Math. Univ. Carolinae,* 41:359–370, 2000.

[21] Roubaud, J. La notion d'associativité relative. *Math. Sci. Hum.,* 34:43–59, 1970.

[22] Salomaa, A. Formal Languages and Power Series. In J. Van Leeuwen, editor, *In Handbook of Theoretical Computer Science,* vol. B, Chap. 3., pages 103–132. Elsevier, 1990.

[23] Sempere, J.M. On a class of regular-like expressions for linear languages. *J. Autom. Lang. Comb.,* 5 :343–354, 2000.

[24] Stanley, R.P. *Enumerative Combinatorics.* vol. 2, Cambridge University Press, 1999.

[25] Suschkewitsch, A. On a generalization of the associative law. *Trans. Amer. Math. Soc.,* 31:204–214, 1931.

[26] Wilke, T. An algebraic theory for regular languages of finite and infinite words. *Intern. J. Algebra Comput.,* 3: 447–489, 1993.

Received 22nd September 2008