INFORMATIQUE THÉORIQUE ET APPLICATIONS

M. BEAUDRY

Finite idempotent groupoids and regular languages

Informatique théorique et applications, tome 32, nº 4-6 (1998), p. 127-140

<http://www.numdam.org/item?id=ITA_1998__32_4-6_127_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 32, n° 4-5-6, 1998, pp. 127 à 140)

FINITE IDEMPOTENT GROUPOIDS AND REGULAR LANGUAGES (*)

by M. BEAUDRY (1)

Communicated by J. E. PIN

Abstract. – *We investigate the language-recognizing power of those finite groupoids whose multiplication monoid belongs to the variety* $\mathbf{A}_1$ *of the idempotent monoids. We find that they recognize a strict subset of the star-free languages. We also consider groupoids which always contain an identity (quasimonoids); within this restriction, we show the existence of a chain of strict inclusions between the languages classes defined by specifying that the multiplication monoid belongs to varieties* $\mathbf{J}_1$, $\mathbf{R}_1$, *and* $\mathbf{A}_1$. © Elsevier Paris

Résumé. – *On s'intéresse ici aux langages que peuvent reconnaître les groupoïdes dont le monoïde multiplicatif appartient à la variété* $\mathbf{A}_1$ *des monoïdes idempotents. On démontre que ceux-ci reconnaissent un sous-ensemble strict de la classe des langages sans étoile. On étudie ensuite les groupoïdes contenant une identité (quasimonoïdes); on montre alors des relations d'inclusion stricte entre les classes de langages définies en spécifiant que le monoïde multiplicatif appartient aux variétés* $\mathbf{J}_1$, $\mathbf{R}_1$ *et* $\mathbf{A}_1$. © Elsevier Paris

1. INTRODUCTION

The deep correspondence between the classifications of finite semigroups and monoids, of regular word languages and of word congruences of finite index has been extensively investigated, with considerable success (see [9, 15]). An analogous correspondence between finite algebras, classes of regular tree languages, and of tree congruences of finite index has also been studied (see [17, 16, 18]). In both cases, classifications in terms of varieties were developed and the varieties lattices of algebras, of languages, and of congruences were proved to be isomorphic.

Binary algebras, also called *groupoids,* also find applications in the study of word languages: indeed the finite groupoids recognize exactly the class of the

(*) Received March 1997; accepted September 1997; final version June 1998.
(1) Département de Mathématiques et d'Informatique, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1 Canada.

0988-5004/98/04-05-06/© Elsevier, Paris

context-free languages (a result found independently in [10] and [3]) under the following convention. A language $L \subseteq A^*$ is recognized by groupoid G iff there exist a monoid homomorphism $\phi : A^* \to G^*$ and a subset $F \subset G$, such that $w \in L$ if, and only if $w\phi$ can evaluate to an element in F. When the operation is nonassociative, the outcome of the evaluation of $w\phi$ varies depending on the way this word of G^* is parenthesized, and the question consists in asking whether there exists a "successful" parenthetization for $w\phi$.

This can be regarded as but another way of saying that the context-free languages are exactly those generated by invertible context-free grammars in Chomsky normal form. Finite groupoids have also been shown, under a different name, to be the syntactic algebras of regular tree languages (a generalization of the parenthesis languages [13]), in the special case where the trees are binary and the interior nodes are labeled with the groupoid's operation; the words in a context-free language are exactly those obtained by taking the frontier (the ordered sequence of leaf labels) of the elements of an appropriately defined regular tree language.

Note that the notion of a syntactic groupoid does not exist for context-free languages. In fact, there exist "universal" groupoids, able to recognize all CFL's (this is a rewording of a result of Greibach [11]). Nevertheless, links can be established between classes of groupoids and classes of languages, and research has been done in this direction. In a first approach, ad hoc examples of groupoids whose recognition power coincides with several significant subclasses of CFL's were presented in [3, 14]. Another approach was initiated and partially explored in [8, 2]; it consists in classifying groupoids in terms of the properties of their multiplication monoid (denoted $\mathcal{M}(G)$; for definition see Subsection 2.3), and to look at the languages recognized by the groupoids in a given class. An obvious case is the *quasigroups,* which are those groupoids such that $\mathcal{M}(G)$ is a group: it was shown recently that they recognize exactly the open regular languages [8, 4]. The study of the aperiodic groupoids, namely those for which $\mathcal{M}(G)$ is a group-free monoid, began when it was observed that there exist aperiodic universal groupoids [6]; it was then shown in [2] that groupoids such that $\mathcal{M}(G)$ is $\mathcal{J}$-trivial with threshold 2 are powerful enough to recognize the regular languages and only them, that this property extends to all groupoids with $\mathcal{M}(G) \in \mathbf{DA}$, and that as soon as one looks beyond $\mathbf{DA}$ (i.e., into varieties containing the Perkins monoid BA_2), there exist groupoids which recognize non-regular languages, and even an "almost hardest" context-free language: "hardest" under nonerasing homomorphism reduction, "almost" in the sense that the language does not contain the empty word. The question whether the monoid variety generated by BA_2 contains the multiplication monoid of a universal groupoid is still open.

In this article, we investigate further the case where $\mathcal{M}(G)$ is aperiodic. More precisely, we extend the study to the *idempotent groupoids*, namely those for which $\mathcal{M}(G)$ is aperiodic of threshold one. (*Note: An alternate definition would consist in deciding that a groupoid is idempotent iff its operation is idempotent. We prefer the definition in terms of the multiplication monoid in order to stay consistent with the choice made in [7, 2] for the more general class of the aperiodic groupoids.*) We show that they recognize a strict subset of the star-free languages. Concentrating next on special cases, we investigate classes of idempotent groupoids defined in terms of subvarieties of $\mathbf{A_1}$ and we look at classes defined in terms of idempotent *quasimonoids,* which are idempotent groupoids containing an identity. Our results are summarized in Figure 2, at the end of Section 4.

2. PRELIMINARIES

A *groupoid* is a binary algebra; we denote both the set and the algebra by G, and the operation either by "·" or by concatenation of the arguments. In this paper, all groupoids are finite. An element 1 is an *identity* in the groupoid if for all x, $1x = x1 = x$. An element $\perp$ is *absorbing* if for all x, $\perp x = x \perp = \perp$. We call *quasimonoid* a groupoid containing an identity; *monoids* are associative quasimonoids. We work within the usual classification of finite algebras into *varieties* [1], classes closed under division and finite direct product.

A groupoid can be used as a language-recognizing device. If G is nonassociative, a word $x \in G^+$ can yield different values depending on the order in which it is evaluated (its parenthetization); given $x, y \in G^*$, we denote by $x \rightsquigarrow y$ the statement that x can evaluate to y (partial evaluation when y has length 2 or more). Further, for $t \in G$, let $L(G; t) = \{w \in G^* | w \rightsquigarrow t\}$. A language $L \subseteq A^*$ is recognized by G if there exist a monoid homomorphism ϕ from A^* to G^* and a subset F of G, such that a word $w \in A^*$ belongs to L if, and only if $w\phi \rightsquigarrow f$ for at least one $f \in F$. In other words, L is recognized by G iff there exist $F \subseteq G$ and $\phi : A^* \to G^*$ such that $L = [\bigcup_{f \in F} L(G; f)]\phi^{-1}$.

(1) Also commonly referred to as pseudovarieties.

Given $w \in G^*$ and $f \in G$ such that $w \rightsquigarrow f$, the corresponding parenthetization of w can be represented as a binary tree (an *evaluation tree*) where each node is labeled with a value from G: the leaves with the letters of w, the root with f, and each interior node with the outcome of the evaluation of the subtree rooted at this node.

The varieties of monoids we work with in this article lie within the variety $\mathbf{A}$ of the aperiodic monoids. We will mention or work with the following:

- the variety $\mathbf{A_1}$ of all idempotent monoids, namely those which satisfy the identity $xx = x$;
- the variety $\mathbf{R_1} \subset \mathbf{A_1}$ of all monoids which satisfy the identity $xyx = xy$;
- the variety $\mathbf{L_1} = \mathbf{R_1^r}$ defined by the identity $xyx = yx$;
- the variety $\mathbf{J_1} = \mathbf{R_1} \wedge \mathbf{L_1}$ of the $\mathcal{J}$-trivial idempotent monoids;
- the variety $\mathbf{ACom}$ of the aperiodic and commutative monoids;
- the variety $\mathbf{J} \wedge \mathbf{A_2}$ of the $\mathcal{J}$-trivial monoids of threshold 2.

Relationships of inclusion between these varieties are depicted in Figure 1.

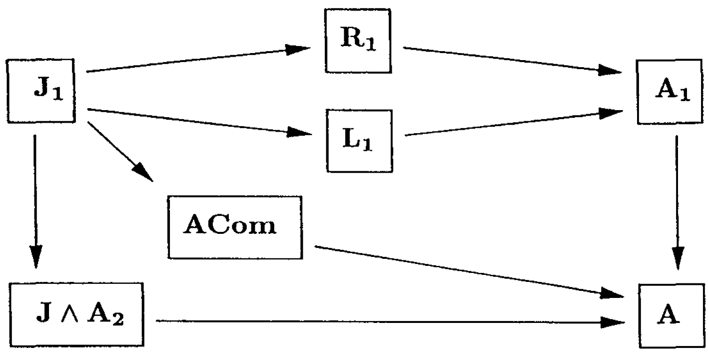

Figure 1. – Varieties of aperiodic monoids.

With a groupoid G we associate the set $\mathcal{A}(G) = \{R(a), L(a) | a \in G\}$ of functions from G to itself, where $bR(a) = ba$ (resp. $bL(a) = ab$) is the *multiplication to the right* by a (resp. to the left) in the groupoid (see [5]). This defines an automaton $\mathcal{Q}(G)$ with G as set of states, $\mathcal{A}(G)$ as alphabet of input symbols, and transitions defined by $bL(a) = aR(b) = c$ if $c = ab$. The *multiplication semigroup* $\mathcal{S}(G)$ is the transformation semigroup of $\mathcal{Q}(G)$; the *multiplication monoid* $\mathcal{M}(G)$ is $\mathcal{S}(G)$ plus the identity function. Define also the submonoids $\mathcal{R}(G)$ generated by $R(G) = \{R(g) | g \in G\}$ and

$\mathcal{L}(G)$ generated by $L(G) = \{L(g) | g \in G\}$. A groupoid is *aperiodic* (resp. *idempotent*) iff its multiplication monoid is group-free (resp. idempotent).

Given a variety $\mathbf{V}$ of monoids, we denote by $\mathcal{L}(\mathbf{V})$ the class of the (regular) languages recognized by monoids belonging to $\mathbf{V}$, by $\mathcal{G}(\mathbf{V})$ the class of all groupoids whose multiplication monoid belongs to $\mathbf{V}$ and by $\mathbb{L}(\mathbf{V})$ (resp. $\mathbb{L}(\mathbf{1},\mathbf{V})$) the class of all languages recognized by groupoids (resp. quasimonoids) of $\mathcal{G}(\mathbf{V})$. For any given variety $\mathbf{V}$ of monoids, it is not hard to see that the corresponding class $\mathcal{G}(\mathbf{V})$ of groupoids is a variety of groupoids; also, for any two groupoids H_1 and H_2, $\mathcal{M}(H_1 \times H_2)$ divides $\mathcal{M}(H_1) \times \mathcal{M}(H_2)$. For a variety $\mathbf{V}$ of monoids, the class of languages $\mathbb{L}(\mathbf{V})$ is readily seen to be closed under finite union and inverse homomorphic image; whether it is also closed under other operations may depend on $\mathbf{V}$. We now quote a basic property of the multiplication monoid of a monoid.

PROPOSITION 2.1: *For any monoid M, we have $\mathcal{R}(M) \cong M$, $\mathcal{L}(M) \cong M^r$, where M^r denotes the reverse of M. Furthermore, $ab = ba$ for any $a \in \mathcal{R}(M)$ and $b \in \mathcal{L}(M)$, hence $\mathcal{M}(M) = \mathcal{R}(M)\mathcal{L}(M)$.*

Proof: Since M is associative, we have $nR(p)L(m) = mnp = nL(m)R(p)$ for any $m, n, p \in M$. Further, associativity implies $mR(n)R(p) = mnp = mR(np)$, so that $\mathcal{R}(M) = R(M)$, and the same holds for $\mathcal{L}(M)$ and $L(M)$. It is not hard to verify that the monoids M and $\mathcal{R}(M)$ (resp. M^r and $\mathcal{L}(M)$) are isomorphic. □

As a consequence, it is not true that for any monoid M and variety $\mathbf{V}$, $M \in \mathbf{V} \Rightarrow \mathcal{M}(M) \in \mathbf{V}$. For example, consider the monoid $R_1 = \{1, a, b\}$ with $aa = ab = a$ and $ba = bb = b$, which belongs to variety $\mathbf{R}_1$. A simple calculation verifies that $\mathcal{M}(R_1)$ is in variety $\mathbf{R}_1 \vee \mathbf{L}_1$, outside of $\mathbf{R}_1$.

We conclude this section with a short digression on the properties of groupoids having a commutative multiplication monoid. All such groupoids are associative: $f(gh) = gR(h)L(f) = gL(f)R(h) = (fg)h$, for any f, g, h. They are not necessarily commutative: consider for example the set $N = \{a, b, c, \perp\}$ with operation defined by $ab = c$ and $xy = \perp$ for all other x, y; this is the syntactic semigroup of the language $\{ab\}$. Note however that in these semigroups, all expressions of length at least 3 are commutative, which follows from $abc = aR(b)R(c) = aR(c)R(b) = acb$ and $abc = cL(b)L(a) = cL(a)L(b) = cab$. Some consequences of this are: that all quasimonoids of $\mathcal{G}(\mathbf{ACom})$ are themselves commutative monoids; that all semigroups of $\mathcal{G}(\mathbf{ACom})$ are idempotent central (that is, such that for all elements e, m, if $e = e^2$, then $em = me$); and finally that any language in $\mathbb{L}(\mathbf{ACom})$ is the union of a language of $\mathcal{L}(\mathbf{ACom})$ consisting of words of length at least 3, with an arbitrary set of words of length ≤ 2.

3. ARBITRARY IDEMPOTENT GROUPOIDS

We are able to prove some basic properties of the languages recognizable by idempotent groupoids; we summarize the main ones as follows.

THEOREM 3.1: $\mathcal{L}(\mathbf{A_1}) \subset \mathbb{L}(\mathbf{1}, \mathbf{A_1}) \subset \mathbb{L}(\mathbf{A_1}) \subset \mathcal{L}(\mathbf{A})$.

Proof: That all regular languages recognizable by an idempotent monoid belong to $\mathbb{L}(1, \mathbf{A_1})$ is a direct consequence of Proposition 2.1: if monoid M is idempotent, then so are $\mathcal{R}(M) \cong M$ and $\mathcal{L}(M) \cong M^r$; hence for any $m, n, p \in M$,

$$m[R(n)L(p)][R(n)L(p)] = mR(n)R(n)L(p)L(p) = mR(n)L(p).$$

The class $\mathbb{L}(\mathbf{1}, \mathbf{A_1})$ also contains languages outside of $\mathcal{L}(\mathbf{A_1})$, which will be proved in the next section (Propositions 4.5 and 4.7).

An element e in a groupoid is *idempotent* iff $e = ee$. Let G be an idempotent quasimonoid; having $1 \in G$ and, $g = 1 \cdot g$ for any $g \in G$, we can write $g = 1R(g) = 1R(g)R(g) = (1 \cdot g) \cdot g = g \cdot g$; therefore all elements of an idempotent quasimonoid are idempotent. If $1 \notin G$, then G may contain nonidempotent elements; however a product hk is always idempotent. To see this, observe that $hk = g$ implies $hR(g) = hg = h(hk) = kL(h)L(h) = kL(h) = hk = g$; then $gg = gR(g) = hR(g)R(g) = hR(g) = g$. Therefore $G = H \cup I$ with $H = \{g \in G | gg \neq g\}$ and $I = \{g \in G | gg = g\}$, and a language L recognized by G is expressible as $L = L_H \cup L_I$, where L_H is a set of one-character words, while by [2] L_I is a regular set recognized by G with an accepting subset consisting exclusively of idempotent elements; L_I is infinite, empty, or equal to $\{\lambda\}$, where λ is the empty word. Hence the singleton $\{a\}$ is an example of a language in $\mathbb{L}(\mathbf{A_1}) - \mathbb{L}(\mathbf{1}, \mathbf{A_1})$.

We prove in Lemma 3.3 the inclusion $\mathbb{L}(\mathbf{A_1}) \subseteq \mathcal{L}(\mathbf{A})$. This inclusion is strict: observe for example that a groupoid of $\mathcal{G}(\mathbf{A_1})$ cannot recognize the language $\{ab\}$ (which belongs to $\mathbb{L}(\mathbf{ACom})$, see Section 2) under any homomorphism ϕ, for having $(ab)\phi \rightsquigarrow g$ for some accepting g would imply that g is idempotent, hence that $(ab)^k\phi \rightsquigarrow g$ for any $k \geq 1$ □

Consider a language $L \in \mathbb{L}(\mathbf{A_1})$, decomposed as above into $L = L_H \cup L_I$. A necessary condition on L_I can be deduced from the idempotency of the accepting elements: in the syntactic monoid of L_I the accepting subset P

must satisfy $m \in P \Rightarrow m^k \in P$, $\forall k \geq 1$. Proposition 3.2 states a more powerful condition.

PROPOSITION 3.2: *Let $K \in \mathbb{L}(\mathbf{A}_1)$; then for any word z of length at least 2, for any decomposition of $z \in K$ into $z = xy$ and for any $j, k \geq 0$, it holds that $x^{j+2} y^{k+2} \in K$.*

Proof: Assume that K is recognized by an idempotent groupoid G through some homomorphism ψ. Let $f \in G$ be an accepting value for K. Consider a word $g_1 \ldots g_r = z\psi$, $r \geq 2$, such that $z\psi \rightsquigarrow f$. In the evaluation tree for an accepting computation of $z\psi$ into f, the computations done along a path starting from the leaf labelled with g_i (for some $1 \leq i \leq r$) and leading to the root (which is labelled with the output value f) can be described with the expression

$$\gamma[a_0 b_1 a_1 \cdots b_s a_s] = f,$$

where $\gamma = g_i$ for some $1 \leq i \leq r$, each $a_k \in \mathcal{R}(G)$ comes from the evaluation of elements to the right of γ (from the values g_t, $t > i$) and each $b_j \in \mathcal{L}(G)$ comes from the evaluation of elements to the left of γ, i.e. from the values g_t, $t < i$.

By idempotency of $\mathcal{M}(G)$, $h = gm \Rightarrow hm = h$ for any $g, h \in G$ and $m \in \mathcal{M}(G)$. Then, $gh = hL(g) = hL(g)L(g) = g(gh)$, and therefore $gh = gR(gh)$. Consider now the case where a_0 is nonempty, i.e. the path begins by taking the output $\alpha \in G$ of a nonempty right subtree; then we have $a_0 = R(\alpha)a_0'$, where $a_0' \in \mathcal{M}(G)$; from the above we obtain

$$\gamma[R(\gamma\alpha)a_0' b_1 a_1 \cdots b_s a_s] = f = f[R(\gamma\alpha)a_0' b_1 a_1 \cdots b_s a_s],$$

and therefore

$$\gamma[a_0 b_1 a_1 \cdots b_s a_s][R(\gamma\alpha)a_0' b_1 a_1 \cdots b_s a_s] = f,$$

which implies $[g_1 \cdots g_{i-1}]^2 [g_i \cdots g_r]^2 \rightsquigarrow f$. Now we use a property derived from the inclusion of $\mathbf{A}_1$ in $\mathbf{DA}$ and [15, chapter 3, exercise 5.9]: for any elements $a_1, \ldots, a_t$ in an idempotent monoid and any $w \in \{a_1, \ldots a_t\}^*$, we have $a_1 \cdots a_t = a_1 \cdots a_t w a_1 \cdots a_t$. With $j, k \geq 0$, this gives

$$\begin{aligned}&[R(\gamma\alpha)a_0' b_1 a_1 \cdots b_s a_s][R(\gamma\alpha)a_0' b_1 a_1 \cdots b_s a_s]\\ &\quad = [R(\gamma\alpha)a_0' b_1 a_1 \cdots b_s a_s]\,[b_1 \cdots b_s]^j\,[R(\gamma\alpha)a_0' a_1 \cdots a_s]^k\\ &\quad \times [R(\gamma\alpha)a_0' b_1 a_1 \cdots b_s a_s],\end{aligned}$$

which leads to

$$\begin{aligned}\gamma[a_0b_1a_1\cdots b_sa_s][R(\gamma\alpha)a_0'b_1a_1\cdots b_sa_s][b_1\cdots b_s]^j[R(\gamma\alpha)a_0'a_1\cdots a_s]^k\\ \times [R(\gamma\alpha)a_0'b_1a_1\cdots b_sa_s]\\ = f[b_1\cdots b_s]^j[R(\gamma\alpha)a_0'a_1\cdots a_s]^k[R(\gamma\alpha)a_0'b_1a_1\cdots b_sa_s]\\ = \gamma[a_0b_1a_1\cdots b_sa_s][b_1\cdots b_s]^j[R(\gamma\alpha)a_0'a_1\cdots a_s]^k\\ \times [R(\gamma\alpha)a_0'b_1a_1\cdots b_sa_s] = f,\end{aligned}$$

which implies $[g_1\cdots g_{i-1}]^{j+2}[g_i\cdots g_r]^{k+2} \rightsquigarrow f$. When $g_i\cdots g_{i-1} = x\psi$ for some prefix x of z, this leads to $x^{j+2}y^{k+2}\psi \rightsquigarrow f$. In the other case where a_0 is empty, i.e. the path can be described as $\gamma[b_1a_1\cdots b_sa_s] = f$, let $b_1 = L(\beta)b_1'$ and observe that

$$\gamma L(\beta) = \beta R(\gamma) = \beta R(\gamma)R(\gamma) = (\beta\gamma)\gamma = \gamma L(\beta)R(\gamma);$$

from this a reasoning analogous to the above can be developed. □

LEMMA 3.3: *The languages of* $\mathbb{L}(\mathbf{A}_1)$ *are star-free*

Proof: Let $L \subset A^*$ belong to $\mathbb{L}(\mathbf{A}_1)$; from [2] we know that it is a regular language, so let $h : A^* \to M$ be its syntactic morphism. Then it suffices to show that M contains no nontrivial group. Let $\{1, s^1, ..., s^{p-1}\}$ be the largest group contained in M and consider a word $w \in A^*$ such that $wh = s$. Then Proposition 3.2 applied to the preimage of s by h and to $w^p = ww^{p-1} \in K$ leads to a contradiction unless $p = 1$ and the group is trivial. □

4. SUBCLASSES OF $\mathbb{L}(\mathbf{A}_1)$

Obvious special cases of idempotent monoids are the three smallest varieties of nontrivial idempotent monoids, namely $\mathbf{J}_1$, $\mathbf{R}_1$ and $\mathbf{L}_1$. We are able to separate the class $\mathbb{L}(\mathbf{J}_1)$ from $\mathbb{L}(\mathbf{R}_1)$ and $\mathbb{L}(\mathbf{L}_1)$; however the relationship of $\mathbb{L}(\mathbf{R}_1)$ and $\mathbb{L}(\mathbf{L}_1)$ between themselves and with $\mathbb{L}(\mathbf{A}_1)$ remains unresolved. Next we concentrate on the special case of the quasimonoids, for which we can obtain more substantial information on the corresponding classes of languages. We state our results in Theorems 4.1 and 4.2. We also give two significant examples of languages recognized by quasimonoids of $\mathcal{G}(\mathbf{R}_1)$, Propositions 4.5 and 4.7.

THEOREM 4.1: $\mathbb{L}(\mathbf{J}_1) \subset \mathbb{L}(\mathbf{R}_1) \cap \mathbb{L}(\mathbf{L}_1)$.

Proof: We know from Section 2 that the groupoids of $\mathcal{G}(\mathrm{J}_1)$ are associative. We now show that they are also commutative. Indeed, for any two elements a, b:

$$ab = aR(b) = aR(b)R(b) = (ab)b = a(bb) = bL(b)L(a) = bL(a)L(b) = bab,$$

and

$$ba = aL(b) = aL(b)L(b) = b(ba) = (bb)a = bR(b)R(a) = bR(a)R(b) = bab.$$

It is easily verified that $\mathcal{G}(\mathbf{J}_1)$ consists of the variety $\mathbf{J}_1$ of the commutative idempotent monoids, plus those threshold-two commutative semigroups not containing an identity (consider for example the syntactic semigroup of the language $\{a\}$).

The inclusion $\mathbb{L}(\mathbf{J}_1) \subset \mathbb{L}(\mathbf{R}_1) \cap \mathbb{L}(\mathbf{L}_1)$ can thus be proved by showing that the language $K = b\{a,b\}^* a\{a,b\}^*$, not recognizable by a commutative semigroup, belongs to $\mathbb{L}(\mathbf{R}_1) \cap \mathbb{L}(\mathbf{L}_1)$. The tables below describe a groupoid $G_R \in \mathcal{G}(\mathbf{R}_1)$ which recognizes K with accepting subset $\{f\}$ and homomorphism defined by $a \mapsto a$, $b \mapsto b$, and another groupoid $G_L \in \mathcal{G}(\mathbf{L}_1)$ which recognizes K with accepting subset $\{d\}$ and homomorphism defined by $a \mapsto a$, $b \mapsto b$. We leave it to the reader to verify both assertions. □

Table 1

	a	b	f	$\bot$
a	$\bot$	$\bot$	$\bot$	$\bot$
b	f	$\bot$	f	$\bot$
f	f	f	f	$\bot$
$\bot$	$\bot$	$\bot$	$\bot$	$\bot$

The table of G_R

	a	b	c	d
a	a	c	c	c
b	d	b	d	d
c	c	c	c	c
d	d	d	d	d

The table of G_L

We move to the special case of quasimonoids, for which we are able to prove the following.

THEOREM 4.2: $\mathcal{L}(\mathbf{J}_1) = \mathbb{L}(\mathbf{1}, \mathbf{L}_1) \subset \mathbb{L}(\mathbf{1}, \mathbf{R}_1) \subset \mathbb{L}(\mathbf{1}, \mathbf{A}_1)$.

We prove $\mathbb{L}(\mathbf{1}, \mathbf{L}_1) \subseteq \mathcal{L}(\mathbf{J}_1)$ (Lemma 4.3) and $\mathbb{L}(\mathbf{1}, \mathbf{R}_1) \neq \mathbb{L}(\mathbf{1}, \mathbf{A}_1)$ (Lemma 4.4); all other inclusions are trivial.

LEMMA 4.3: $\mathbb{L}(\mathbf{1}, \mathbf{L}_1) \subseteq \mathcal{L}(\mathbf{J}_1)$.

Proof: We show that the quasimonoids of $\mathcal{G}(\mathbf{L_1})$ are commutative, idempotent and associative, hence belong to $\mathbf{J_1}$. Consider two elements a, b of a quasimonoid belonging to $\mathcal{G}(\mathbf{L_1})$. Then $a(ab) = bL(a)L(a) = bL(a) = ab$; meanwhile, $(ab)a = 1R(a)R(b)R(a) = 1R(b)R(a) = ba$ by the identity $xyx = yx$ which defines $\mathbf{L_1}$. Analogous reasonings prove $(ba)a = ba$ and $a(ba) = ab$. Now, by $xyx = yx$,

$$ba = (ba)a = (ba)(a(ba)) = (ba)(ab) = ((ab)a)(ab) = a(ab) = ab,$$

therefore the groupoid is commutative. In other words, $R(x) = L(x)$ for all x. Next, let a, b, c be three groupoid elements, and let $g = (ab)c$ and $h = a(bc)$; observe that $aR(g) = ag = a((ab)c) = ((ab)c)a = 1R(a)R(b)R(c)R(a) = 1R(b)R(c)R(a) = (bc)a = h$.
Meanwhile,

$$h = aR(g) = aR(g)R(g) = hR(g) = hg = gh.$$

Now, since $h = (cb)a$ and $g = c(ba)$, a similar reasoning leads to $cR(h) = g$ and $g = cR(h) = gh = hg$, hence $g = h$; therefore the groupoid is associative. □

Given a language $L \subseteq A^*$, define the set $\Sigma(L) = \bigcup_{w \in L}\{a \in A | w \in A^*aA^*\}$. We say that language L is a *zero* iff $L = \Sigma(L)^* L \Sigma(L)^*$. If $L \subseteq A^*$ is a zero, then for any monoid homomorphism $\psi : B^* \to A^*$, the language $L\psi^{-1}$ is also a zero.

LEMMA 4.4: $\mathbb{L}(\mathbf{1}, \mathbf{R_1}) \subset \mathbb{L}(\mathbf{1}, \mathbf{A_1})$.

Proof: Consider a groupoid $G \in \mathcal{G}(\mathbf{R_1})$, hence such that $\mathcal{M}(G)$ satisfies the identity $xyx = xy$, which defines $\mathbf{R_1}$. Let $g_1, ..., g_r$, $f \in G$ be such that $g_1 \cdots g_r \rightsquigarrow f$; for each g_i, $1 \le i \le r$, $g_i = g_iR(g_i) = g_iL(g_i)$ because G is a quasimonoid (see the proof of Theorem 3.1), and there exists $m \in \mathcal{M}(G)$ such that $g_im = f$. Therefore, $f = g_iR(g_i)m = g_iR(g_i)mR(g_i) = g_imR(g_i) = fR(g_i)$, and similarly $f = fL(g_i)$. If G recognizes a language $L \subset A^*$ with accepting value f under homomorphism ϕ, then for any word w such that $w\phi \rightsquigarrow f$ and letter a in w, let $a\phi = h_1 \cdots h_k$: it holds that $f = fR(h_1) = \cdots = fR(h_k)$ and therefore $(wa)\phi \rightsquigarrow f$ and $(aw)\phi \rightsquigarrow f$. Hence the subset $\{w \in A^* | w\phi \rightsquigarrow f\}$ of L is a zero, and all languages of $\mathbb{L}(\mathbf{1}, \mathbf{R_1})$ are finite unions of zeros. Now consider the language $a\{a, b\}^*$: it is not a union of zeros, but it is recognized by the idempotent monoid R_1 (see Section 2), and thus belongs to $\mathbb{L}(\mathbf{1}, \mathbf{A_1})$. □

It is neither true that all star-free zeros belong to $\mathbb{L}(\mathbf{1}, \mathbf{R_1})$, nor that the languages of $\mathbb{L}(\mathbf{1}, \mathbf{R_1})$ are all recognizable by idempotent monoids. We substanciate this statement with two examples, Propositions 4.5 and 4.7.

PROPOSITION 4.5: *The language $L_{aa} = \{a,b\}^*\{aa,bb\}\{a,b\}^*$ belongs to $\mathbb{L}(\mathbf{1}, \mathbf{R_1})$.*

Proof: Define $G = \{1, a_1, a_2, b_1, b_2, f, \perp\}$ with operation given by the "rules"

- $a_2 \cdot a_1 = b_2 \cdot b_1 = f$;
- $x \cdot f = f \cdot x = f$ for all $x \neq \perp$;
- everything not specified above evaluates to $\perp$.

Verify that G recognizes L_{aa} with accepting subset $\{f\}$ under the homomorphism: $a \mapsto a_1a_2, b \mapsto b_1b_2$, and that $\mathcal{M}(G) \in \mathbf{R}_1$. □

Since $L_{aa} \notin \mathcal{L}(\mathbf{J_1})$ (its syntactic monoid is actually the Perkins monoid BA_2), this proposition implies $\mathbb{L}(\mathbf{1}, \mathbf{L_1}) \subset \mathbb{L}(\mathbf{1}, \mathbf{R_1})$. It may also suggest that for every $w \in A^*$ the star-free zero A^*wA^* belongs to $\mathbb{L}(\mathbf{1}, \mathbf{R_1})$, but this is not true: a counter-example is the language $\{a,b\}^*baab\{a,b\}^*$. That it is not recognizable by a quasimonoid of $\mathcal{G}(\mathbf{R}_1)$ is verified by applying on the word $baab$ the following "pumping" property, which we prove for arbitrary groupoids in $\mathcal{G}(\mathbf{R}_1) \cup \mathcal{G}(\mathbf{L_1})$.

PROPOSITION 4.6: *Consider a language $L \subseteq A^*$, recognized by a groupoid of $\mathcal{G}(\mathbf{R}_1) \cup \mathcal{G}(\mathbf{L}_1)$; for any $x, y, z \in A^*$ and $k \geq 2$, if $xy^2z \in L$, then $xy^kz \in L$.*

Proof: Let L be recognized by G under homomorphism ϕ; let f be an accepting element and $xyyz\phi \rightsquigarrow f$. We assume that $|y| \geq 2$, since otherwise the statement is obvious. Observe now that in any evaluation tree of $xyyz\phi$, there is at least one subtree whose frontier has a prefix or suffix of length $|y\phi|$ which is a factor of $yy\phi$. Denote by w the factor found in this way and observe that it is a cyclic permutation of $y\phi$. Assume w.l.o.g. that w is a suffix: let $g_1 \cdots g_r$ be the frontier of the subtree, let $g_i \cdots g_r = w$ for $i = r - |y\phi| + 1$, and let h be the value the subtree's output. If $i = 1$, then $hh = h$ implies that $ww \rightsquigarrow h$; therefore an extra factor w can be inserted at the appropriate place in $xyyz\phi$, so that $xy^3z\phi \rightsquigarrow f$; the case $k = 2$ implies the claim for $k \geq 3$. Else $i > 1$; consider then in the subtree the leaf-root path starting at the leaf carrying value g_{i-1}. Evaluation along this path can be described as

$$g_{i-1}[a_0b_1a_1 \cdots b_sa_s] = h,$$

where $b_1, ..., b_s \in \mathcal{R}(G)$ come from the evaluation of $g_i \cdots g_r$ and $a_0, ..., a_s \in \mathcal{L}(G)$ from the evaluation of $g_1 \cdots g_{i-1}$. If $\mathcal{M}(G) \in \mathbf{R}_1$, then $[a_0 b_1 a_1 \cdots b_s a_s] = [a_0 b_1 a_1 \cdots b_s a_s] b_1 \cdots b_s$ and $g_{i-1}[a_0 b_1 a_1 \cdots b_s a_s] b_1 \cdots b_s = h$, which can be interpreted as the insertion of an extra factor w, correctly "attached" at the root of the subtree. Else if $\mathcal{M}(G) \in \mathbf{L}_1$, then $[a_0 b_1 a_1 \cdots b_s a_s] = b_1 \cdots b_s [a_0 b_1 a_1 \cdots b_s a_s]$ and $g_{i-1} b_1 \cdots b_s [a_0 b_1 a_1 \cdots b_s a_s] = h$, which can be seen as the insertion of an extra factor w, this time immediately at the right of g_{i-1}. In both cases, if $xyyz\phi \rightsquigarrow f$, then $xy^3 z\phi \rightsquigarrow f$. $\square$

PROPOSITION 4.7: *For any alphabet A and characters $b_1, ..., b_p \in A$, the language $L = A^* b_1 A^* \cdots A^* b_p A^*$ belongs to $\mathbb{L}(\mathbf{1}, \mathbf{R}_1)$.*

Proof: We proceed by induction on p. For the basis, observe that the language $A^* b A^*$ belongs to $\mathcal{L}(\mathbf{J}_1)$. Next, assume that two subsets L_1 and L_2 of A^* are recognized respectively by G_1 and G_2 through the homomorphisms ϕ_1 and ϕ_2, with accepting subsets F_1 and F_2. Assume that the G_i's both belong to $\mathcal{G}(\mathbf{R}_1)$. Now define a new groupoid G whose set of values is $G_1 \times G_2$, with elements denoted by $\langle g_1, g_2 \rangle$, plus three new elements, 1, f, and $\perp$. Define the operation on G as follows:

- element 1 is the identity and $\perp$ is absorbing;
- for all $g_1 \in G_1$ and $g_2 \in G_2$ such that at least one of g_1 and g_2 is not accepting, for any h and k, $\langle g_1, h \rangle \langle k, g_2 \rangle = \langle g_1 k, h g_2 \rangle$;
- for all $g_1 \in G_1$ and $g_2 \in G_2$ such that both g_1 and g_2 are accepting, for any h and k, $\langle g_1, h \rangle \langle k, g_2 \rangle = f$;
- for all $u \in G_1 \times G_2$, $uf = fu = f$;
- $ff = f$.

Define $\varphi : A^* \rightarrow G$ as the homomorphism which satisfies $a\varphi = \langle a\phi_1, a\phi_2 \rangle$. Thus, if a word belongs to $A^* L_1 L_2 A^*$, then there exists a way to evaluate the image by φ of its appropriate factor into either f or an element of the form $\langle g_1, h \rangle$; similarly, there is another factor to its right which can evaluate to f or to $\langle k, g_2 \rangle$, where $g_1 \in F_1$ and $g_2 \in F_2$. Since both L_1 and L_2 are zeros by the induction hypothesis, then these two factors of w can be taken to be contiguous, and thus $w\varphi$ evaluates to f. In the other direction, the groupoid is custom-made so that $w\varphi$ cannot evaluate to f unless it has a factor in L_1 followed later by a factor in L_2.

These remains to show that $\mathcal{M}(G) \in \mathbf{R}_1$. A simple calculation [1] shows that it suffices to verify the condition $xyx = xy$ on the monoid's generators. Since $\mathcal{M}(G_1 \times G_2)$ divides $\mathcal{M}(G_1) \times \mathcal{M}(G_2)$, and since 1 and $\perp$ behave

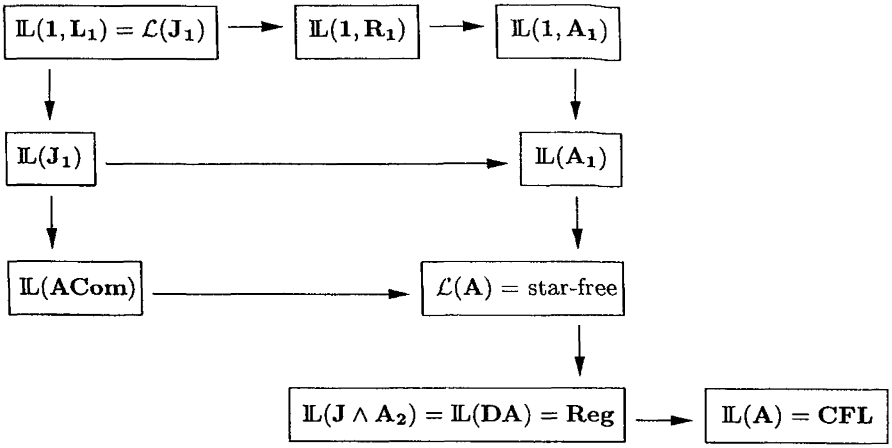

Figure 2. – Relationships between language classes

trivially, we concentrate on $R(f)$ and $L(f)$ and their interaction with the other generators. We verify that for any $g \in G$ other than $\perp$, any x other than $R(\perp)$ and $L(\perp)$, and any y in $\{R(f), L(f)\}$, it holds that $gxyx = gxy = f$ and $gyxy = gyx = f$. □

The star-free zeros discussed in Proposition 4.7 actually are piecewise-testable, but all piecewise-testable zeros do not belong to $\mathbb{L}(\mathbf{1}, \mathbf{R_1})$. Consider for instance

$$B^*b(\lambda \cup a \cup aa \cup aaaa^+)bB^* \cup B^*ba^*baB^+,$$

with $B = \{a, b\}$; verify that since $baab$ is in the language, then so must $baaab$ by Proposition 4.6.

5. COMMENTS

In this article, we have extended the results obtained in [2], which described two steps in a hierarchy of classes of languages recognized by aperiodic groupoids. We have investigated the language-recognizing power of those groupoids whose multiplication monoid is idempotent; the inclusion relationships we have obtained are summarized in Figure 2. Our main result is that the regular languages they recognize are a strict subset of the star-free languages. We also investigated some subcases of idempotent groupoids and thereby identified strict subclasses of $\mathbb{L}(\mathbf{A_1})$.

Our results leave many questions open. For instance, it is still unresolved whether the classes $\mathbb{L}(\mathbf{R_1})$ and $\mathbb{L}(\mathbf{L_1})$ are strict subsets of $\mathbb{L}(\mathbf{A_1})$; in particular, all languages of $\mathbb{L}(\mathbf{A_1})$ known so far to the author satisfy Proposition 4.6. If these classes are different, then there may well exist an infinite chain of classes $\mathbb{L}(\mathrm{V})$, one for each variety $\mathbf{R_1} \subset \mathbf{V} \subset \mathbf{A_1}$ of idempotent monoids. Also, except for the simpler cases of $\mathbb{L}(\mathbf{ACom})$ and $\mathbb{L}(\mathbf{J_1})$, the language classes discussed here still lack a precise characterization; questions pertaining to the closure properties they satisfy, and to possible combinatorial descriptions independent of any reference to groupoids, remain open.

The author thanks an anonymous referee for improvements to the proof of Proposition 4.7. Work reported on in this article was supported by NSERC grant OGP0089786 and FCAR grant 91-ER-0642.

REFERENCES

1. M. Beaudry, Characterization of idempotent monoids, *Information Processing Letters,* 1989, *31*, pp. 163-166.
2. M. Beaudry, Languages recognized by finite aperiodic groupoids, *Proc. 13th STACS, LNCS,* 1996, *1046*, pp. 113-124.
3. F. Bédard, F. Lemieux and P. McKenzie, Extensions to Barrington's *M*-program model, *Theor. Comp. Sc.,* 1993, *107*, pp. 31-61.
4. M. Beaudry, F. Lemieux and D. Thérien, Finite loops recognize exactly the regular open languages, *Proc. 24th ICALP, LNCS* 1256, 1997, pp. 110-120.
5. R. H. Bruck, "A survey of binary systems", Springer-Verlag, 1966.
6. H. Caussinus, Un groupoïde permettant de caractériser SAC^1, Manuscript, 1993.
7. H. Caussinus, Contributions à l'étude du non-déterminisme restreint, thèse de doctorat, Université de Montréal, Montréal, 1996.
8. H. Caussinus and F. Lemieux, The Complexity of Computing over Quasigroups, *Proc. FST & TCS,* 1994, pp. 36-47.
9. S. Eilenberg, "Automata, Languages and Machines, Vol. B", Academic Press, 1976.
10. F. Gécseg and M. Steinby, "Tree Automata", Akadémiai Kiadó, Budapest, 1984.
11. S. Greibach, The Hardest Context-Free Language, *SIAM J. Comp.,* 1973, *2*, pp. 304-310.
12. G. Lallement, "Semigroups and Combinatorial Applications", Addison-Wesley, 1979.
13. R. McNaughton, Parenthesis Grammars, *J. ACM,* 1967, *14*, pp. 490-500.
14. A. Muscholl, Characterizations of LOG, LOGDCFL and NP based on groupoid programs, Manuscript, 1992.
15. J.-É. Pin, "Variétés de langages formels", Masson, 1984.
16. M. Steinby, A theory of tree language varieties, in Tree Automata and Languages, M. Nivat and A. Podelski Eds., North-Holland, 1992, pp. 57-82.
17. W. Thomas, Logical aspects in the study of tree languages, in 9th Coll. on Trees in Algebra and Programming, B. Courcelle Ed., Cambridge University Press, 1984, pp. 31-51.
18. T. Wilke, Algebras for classifying regular tree languages and an application to frontier testability, *Proc. 20th ICALP, LNCS,* 1993, *700*, pp. 347-358.