An EILENBERG Theorem for ∞-Languages

Thomas Wilke
Institut für Informatik und Praktische Mathematik
Christian-Albrechts-Universität zu Kiel
e-mail: tw@informatik.uni-kiel.dbp.de

Abstract. We use a new algebraic structure to characterize regular sets of finite and infinite words through recognizing morphisms. A one-to-one correspondence between special classes of regular ∞-languages and pseudovarieties of right binoids according to EILENBERG's theorem for regular sets of finite words is established. We give the connections to semigroup theoretical characterizations and classifications of regular ω-languages, and treat concrete classes of ∞-languages in the new framework.

1 Introduction

The motivation of the present work is twofold:

In [3, p. 95] BRUNO COURCELLE is concerned with a general notion of recognizability in arbitrary algebraic structures. He points out that regularity of ω-languages "is not algebraic since the corresponding sets of infinite [...] words are not recognized with respect to any (known) algebraic structure".

In [6] and [5] JEAN-PIERRE PÉCUCHET deals with semigroup-defined classes of regular ω-languages. He asks for a description of these classes in terms of closure properties with respect to language theoretical operations.

In this paper we develop an algebraic theory of regular sets of finite and infinite words (∞-languages), which is based on the new notions 'right binoid', 'RAMSEY condition', and '∞-variety '.

Right binoids are an algebraic structure that is modelled on the set of all finite and infinite words over an alphabet. Together with morphisms that satisfy the so-called RAMSEY condition they seem to repair the lack of a recognizing structure for regular ω- and ∞-languages.

∞-varieties are classes of regular ∞-languages that are closed with respect to inverse images of language morphisms, boolean combinations, and quotients. In connection with an EILENBERG Theorem for ∞-languages which we prove, they give an answer to the question raised by JEAN-PIERRE PÉCUCHET. Similar to EILENBERG's original theorem ([4, p. 195, th. 3.2s]), the one formulated here establishes a one-to-one correspondence between ∞-varieties and pseudovarieties of special ('+-generated') right binoids.

The paper falls into two parts. In the first part we introduce right binoids and the RAMSEY condition in order to define recognizability (both section 2), and prove the EILENBERG Theorem for ∞-languages (section 3).

In the second part we relate right binoids and ∞-varieties to the semigroup approach to the characterization and classification of regular ω-languages (section 4), deal with some concrete classes of regular ∞-languages (section 5), and show how to compute with ∞-languages and right binoids (section 6).

Nearly all proofs are, due to space limitations, omitted here. They are contained in [15]. More detailed sketches of the proofs and examples of proofs can be found in [14].

Notations. Let A be an alphabet throughout this paper. Set $A^\infty := A^+ \cup A^\omega$ and $A^\pi := \{u \cdot v^\omega | u, v \in A^+\}$, i.e., A^π is the set of all ultimately periodic words over A. A set $L \subseteq A^\infty$ is called an ∞-language over A. It is said to be regular if its $+$-part $L_+ := L \cap A^+$ and its ω-part $L_\omega := L \cap A^\omega$ are regular.

Remark. Strictly speaking a language is a pair (A, L) that consists of an alphabet A and a set L of words over the alphabet. If the alphabet is clear from the context, it will not be mentioned explicitly and L stands also for the pair (A, L). In that case A will be referred to as A(L).

2 RAMSEY Condition and Right Binoids

The aim of this section is to give an algebraic characterization of regular ∞-languages. We start with a characterization through congruences over finite and infinite words (∞-congruences). Then we turn to a new algebraic structure (right binoid) that takes over the role of the semigroups they play in the case of sets of finite words. We try to motivate the additional property (RAMSEY condition) which we impose on saturating congruences resp. recognizing morphisms.

Definition 1 (∞-Congruence) *An* ∞-congruence *over A is an equivalence relation $\approx \subseteq (A^+ \times A^+) \cup (A^\omega \times A^\omega)$ such that*

$$x \cdot x' \approx y \cdot y' \qquad\qquad x^\omega \approx y^\omega$$

hold for every choice of words $x, y \in A^+, x', y' \in A^\infty$ with $x \approx y$ and $x' \approx y'$.

Set $\approx_\pi := (A^+ \times A^+) \cup (A^\pi \times A^\pi) \cup (A^\omega \setminus A^\pi \times A^\omega \setminus A^\pi)$. Then the language A^π regarded as a language over A is saturated by $\approx_\pi$ (it is a union of $\approx_\pi$-classes), $\approx_\pi$ is of finite index, but A^π is known to be not regular if A has at least two elements.

Hence we do not get a characterization of regular ∞-languages in terms of finite, saturating congruences as known from $+$-languages. The problem arises from the non-ultimately periodic words that cannot be expressed in terms of concatenation and ω-iteration, and thus are not affected by the definition of ∞-congruences. The following definition remedies the situation.

Definition 2 (RAMSEY Condition for Congruences) *An ∞-congruence over A satisfies the* RAMSEY condition *if*

$$\prod X \approx X_0^\omega$$

holds for every infinite sequence $X: \omega \to A^+$ of $\approx$-equivalent words $X_0, X_1, \ldots$

Then we have

Theorem 1 *An ∞-language L over A is regular if and only if there is a* RAMSEY *congruence over A that is of finite index and saturates L.*

Sketch of proof. For one direction, a saturating RAMSEY congruence of finite index is derived from a recognizing finite automaton similar to [13, proof of Th. 2.1, p. 139]. For the other direction we observe that every ω-language that is saturated by a given RAMSEY congruence of finite index is saturated (in the sense of [1]) by the restriction of the given congruence to the finite words, and use [13, Lemma 2.2, p. 139] ∎

Consider again A^π and the saturating ∞-congruence $\approx_\pi$. We verify that $\approx_\pi$ is not a RAMSEY congruence: let u be a non-ultimately periodic word. Then the decomposition $u_0, u_1, u_2, \ldots$ of u in its letters is a decomposition in a sequence of $\approx_\pi$-equivalent words, but u is not equivalent to $u_0{}^\omega$, which is ultimately periodic.

Closer analysis of the algebraic structure of A^∞ together with concatenation and ω-iteration leads to

Definition 3 (Right Binoid) *A quintuple $B = (C, D, \cdot_+, \cdot_\omega, {}^\omega)$ that is formed from two sets C and D, two binary operations $\cdot_+ : C \times C \to C$ and $\cdot_\omega : C \times D \to D$, and a unary operation ${}^\omega : C \to D$ is a* right binoid *iff*

- *the two binary operations are associative, i.e., they satisfy the equations $u \cdot_+ (v \cdot_+ w) = (u \cdot_+ v) \cdot_+ w$ and $u \cdot_+ (v \cdot_\omega w) = (u \cdot_+ v) \cdot_\omega w$ for $u, v \in C$, $w \in C$ resp. D, and*
- *the unary operation satisfies the equations*

$$(x \cdot_+ y)^\omega = x \cdot_\omega (y \cdot_+ x)^\omega \tag{1}$$
$$(\underbrace{x \cdot_+ \ldots \cdot_+ x}_{n \text{ times}})^\omega = x^\omega \tag{2}$$

for every choice of elements $x, y \in C$ and every $n \in \omega \setminus \{0\}$.

C is called the *$+$-part* of B, denoted by B_+, and D is called the *ω-part*, denoted by B_ω. The binary operations are called *$+$-product* resp. *ω-product*, while the unary operation is called *iteration*. Equation 1 is called *rotation rule*, Equation 2 is called *iteration rule*.

For the sake of simplicity we shall write $\cdot$ or just nothing instead of $\cdot_+$ and $\cdot_\omega$, for there is no danger of misunderstanding. In order to save brackets and operation symbols we introduce a new symbol: by $s \nearrow t$ we mean $s \cdot t^\omega$.

We consider A^∞ as the right binoid $(A^+, A^\omega, \cdot_+, \cdot_\omega, {}^\omega)$ with the ordinary concatenation as products and the ordinary ω-iteration of finite words as iteration, and call it the *alphabetic right binoid* over A. For $u, v \in A^+$ we have $u \nearrow v = uv^\omega$, and for $+$-languages U, V we define $U \nearrow V = UV^\omega$.

With regard to the RAMSEY condition for congruences we define

Definition 4 (RAMSEY Condition for Morphisms) *A right binoid morphism $f : A^\infty \to B$ satisfies the* RAMSEY condition *if the equality*

$$f(\prod X) = f(X_0)^\omega$$

holds for every infinite sequence $X : \omega \to A^+$ of finite words with $f(X_0) = f(X_1) = \ldots$

Notice that the definition requires that the domain of the morphism under consideration is an alphabetic right binoid.

Then we have

Proposition 1 *An ∞-language L over A is regular if and only if it is recognized by a* RAMSEY *morphism $f: A^\infty \to B$ into a finite right binoid B in the sense that $L = f^{-1}(f(L))$. We then also say that L is* recognized *by B* ∎

According to the syntactic congruence of $+$-languages we define an ∞-congruence over finite and infinite words for ∞-languages.

Definition 5 (Syntactic Congruence) *Let L be a language over A. Finite words x, $y \in A^+$ are equivalent with respect to L if the equivalences*

$$uxv \in L \iff uyv \in L \qquad\qquad u(xw)^\omega \in L \iff u(yw)^\omega \in L$$

hold for all words $u, w \in A^*$, $v \in A^* \cup A^\omega$.

Infinite words x, $y \in A^\omega$ are equivalent if the equivalence

$$ux \in L \iff uy \in L$$

holds for all words $u \in A^*$.

$\approx_L$ is called the syntactic congruence *of L.*

We have the following

Theorem 2 *1. Let L be an ∞-language over A.*

(a) The relation $\approx_L$ is an ∞-congruence.

(b) $\approx_L$ is the coarsest ∞-congruence saturating L, and L is regular if and only if $\approx_L$ is a RAMSEY *congruence of finite index.*

2. The syntactic congruence $\approx_L$ of a regular ∞-language can be computed from a recognizing finite automaton (either with MULLER, RABIN *or* BÜCHI *acceptance condition for ω-words).*

Hints to the proof. (1.a) and the first part of (1.b) are straightforward. The second part of (1.b) is based on the first part and Theorem 1. For the proof of (2.) compare with [7, Th. 3.4, p. 83] ∎

The right binoid quotient of A^∞ modulo an ∞-congruence $\approx$ is denoted by $A^\infty/\!\approx$. Then we have

Corollary 1 *Let L be an ∞-language over A.*

1. *L is regular if and only if $A^\infty/\!\approx_L$ is finite and the canonical morphism $f: A^\infty \to A^\infty/\!\approx_L$ with $f(u) = [u]_{\approx_L}$ satisfies the* RAMSEY *condition.*

2. *If L is regular then $A^\infty/\!\approx_L$ is the smallest right binoid (up to isomorphism, the 'divides'-relation as ordering) that recognizes L* ∎

3 ∞-Varieties and EILENBERG Correspondence

In this section we present the main theorem, but before we have to give some definitions that lead to the new notions of '∞-varieties' and 'varieties of right binoids'.

Let A and B be alphabets. A function $f: A^\infty \to B^\infty$ is a *language morphism* if every single-letter word goes onto a finite word, and the image of every arbitrary word is the concatenation of the images of the sequence of its letters, i.e.,

$$f(x) = f(x_0)f(x_1)f(x_2)\cdots$$

for every $x \in A^\infty$.

Let L be a language over A and $x \in A^*$, $y \in A^\omega$. Then we define the *quotients* of L with respect to x resp. y:

$$\begin{aligned} x_+^{-1}L &:= \{y \in A^+ | xy \in L\} \\ x_\omega^{-1}L &:= \{y \in A^\omega | xy \in L\} \\ Ly^{-1} &:= \{x \in A^+ | xy \in L\} \\ x^{-\omega}L &:= \{y \in A^+ | (xy)^\omega \in L\} \end{aligned}$$

Definition 6 (∞-Variety) *A class $\mathcal{L}$ of regular languages is called an* ∞-variety *if it is closed under boolean combinations, inverse images of language morphisms, and quotients. The class of all ∞-varieties is denoted by* **LVar**.

In order to obtain uniqueness in the EILENBERG Theorem, we restrict the considerations to $+$-generated right binoids: a right binoid B is *$+$-generated* if there exist elements $b, c \in B_+$ for every element $a \in B_\omega$ such that $a = b \nearrow c$.

Definition 7 (Variety of Right Binoids) *A class of finite $+$-generated right binoids is a* variety of right binoids *if it is closed with respect to finite direct products, homomorphic images, and $+$-generated subbinoids. The class of all varieties of right binoids is denoted by* **BVar**.

Every intersection of a non-empty system of varieties of right binoids is a variety too. Hence for every class $\mathbf{B}$ of $+$-generated right binoids, the class

$$\langle \mathbf{B} \rangle := \bigcap \{\mathbf{C} | \mathbf{C} \text{ is a variety of right binoids and } \mathbf{B} \subseteq \mathbf{C}\}$$

constitutes the smallest variety including $\mathbf{B}$. We say that $\langle \mathbf{B} \rangle$ is the variety of right binoids generated by $\mathbf{B}$.

Lemma 1 *If* $\mathbf{B}$ *is a variety of right binoids then the class $\{L | \exists\, B \in \mathbf{B}\ B \text{ recognizes } L\}$ is an ∞-variety* ∎

Then we have the main theorem

Theorem 3 (EILENBERG Theorem) *The functions* **bl** *and* **lb** *defined by*

$$\mathbf{bl}: \begin{cases} \mathbf{LVar} \to \mathbf{BVar} \\ \mathcal{L} \mapsto \langle \{A^\infty/\approx_L | L \in \mathcal{L} \wedge \mathrm{A}(L) = A\} \rangle \end{cases} \qquad \mathbf{lb}: \begin{cases} \mathbf{BVar} \to \mathbf{LVar} \\ \mathbf{B} \mapsto \{L | \exists\, B \in \mathbf{B}\ B \text{ recognizes } L\} \end{cases}$$

form a pair of bijective functions, i.e.:

$$\mathbf{bl} \circ \mathbf{lb} = \mathrm{id}_{\mathbf{BVar}} \qquad \mathbf{lb} \circ \mathbf{bl} = \mathrm{id}_{\mathbf{LVar}}$$

Hints to the proof. In general follow the proof given in [10, p. 32 ff.] for the case of monoids and regular $*$-languages. An important difference concerns the fact that A^∞ is not freely generated by A if A has more than one element. The following lemma states the property of A^∞ that is used instead.

Lemma 2 *If B is a right binoid with finite $+$-part and $f_0: A \to B_+$ is a mapping, there exists a unique* RAMSEY *morphism $f: A^\infty \to B$ the restriction of which on A coincides with f_0.*

Hints to the proof. Follow the considerations in the proof of [7, Lemme 2.2, p. 76 ff.] ∎

4 Regular ω-Languages and Semigroups

Every regular ω-language is a regular ∞-language and conversely the ω-part of a regular ∞-language is a regular ω-language by definition. Therefore it is reasonable to compare the known algebraic descriptions of regular ω-languages with the description of regular ∞-languages presented in this paper.

As stated in the introduction, there is no natural algebraic characterization of regular sets of infinite words. Since semigroups are canonical in the case of regular $+$-languages and regular ω-languages can be represented in terms of regular $+$-languages with $\cup$, $\cdot$, and $^\omega$, semigroups have also been used to give algebraic characterization of regular sets of infinite words, and varieties of semigroups have been used to classify them.

In the following we will show the connections between right binoids and ∞-languages on the one hand and semigroups and ω-languages on the other. The first subsection deals with the notions of 'saturation' and 'recognizability', the second subsection treats an approach that arose from MCNAUGHTON's theorem.

4.1 Saturation and Recognizability

Let L be an ω-language over A and $f: A^+ \to S$ a semigroup morphism into a finite semigroup. Then L is said to be *saturated* (see [1] and [9]) by f resp. S if one of the inclusions $U \subseteq L$ and $U \subseteq A^\omega \setminus L$ holds for every language U of the form $f^{-1}(s) \nearrow f^{-1}(t)$ for $s, t \in S$.

From a given finite semigroup S we will construct a right binoid S^∞ that recognizes an ω-language L if and only if S saturates L.

For this purpose we need some definitions. A pair $(s, e) \in S \times S$ is called a *linked pair* if it satisfies $se = s$ and $ee = e$. Two linked pairs (s, e) and (t, f) are *conjugated* if $\exists u, v \in S^1\; e = uv \wedge f = vu \wedge s = tv \wedge t = su$; and we write $(s, e) \equiv (t, f)$. We have that $\equiv$ is an equivalence relation on the set $\mathrm{P}(S)$ of all linked pairs.

Now we can define the quintuple $S^\infty = (S, \mathrm{P}(S)/\equiv, \cdot, \cdot_\omega, {}^\omega)$ with $s \cdot_\omega [(t, e)]_\equiv := [(st, e)]_\equiv$ and $s^\omega := [(s, s^\pi)]_\equiv$ where s^π denotes the unique idempotent in the cyclic semigroup generated by s in S. Then we have the following

Theorem 4 *Let S be a finite semigroup and L an ω-language.*

1. *The operations of S^∞ are well-defined and S^∞ is a finite right binoid.*

2. *Every +-generated right binoid with +-part S is a homomorphic image of S^∞.*

3. *L is saturated by S if and only if L is recognized by S^∞.*

S^∞ *is called the* free right binoid *of S* ∎

Let S be a variety of semigroups. Then we consider the class S^s of regular ω-languages that are saturated by elements of S. We try to describe this class through an ∞-variety.

First we observe that the class of right binoids consisting of all +-generated right binoids with +-part in S is a variety of right binoids. Thus we have

Corollary 2 *For every variety* S *there is an ∞-variety $\mathcal{L}$ such that an ω-language L belongs to* S^s *if and only if L belongs $\mathcal{L}$* ∎

Let us turn to the notion of 'recognizability' which is also known for ω-languages (see [9]): let L, f and S be like above. A language $f^{-1}(s) \nearrow f^{-1}(e)$ with $(s, e) \in \mathrm{P}(S)$ is called *f-simple*. L is said to be *recognized* by f resp. S if L is a union of f-simple languages. For a given variety S, the class S^ω consists of all ω-languages that are recognized by an element of S.

There are strong connections between recognizability in this sense and transition semigroups of nondeterministic BÜCHI automata (see [6]). The problem with this notion in the context of right binoids is illustrated by the fact that a given semigroup may recognize a language but not its complement (like a BÜCHI automaton that accepts a language but not its complement no matter which states are supposed to be final).

There are even examples of varieties S of semigroups for which S^ω is not closed under boolean operations (see [7, th. 3.2, p. 57]). Thus we cannot expect a result like Corollary 2 in general here. However, we can use the corollary to obtain similar results in special situations. For example, if S contains only $\mathcal{D}$-trivial semigroups, we have $\mathrm{S}^s = \mathrm{S}^\omega$ (see [5]). Hence, by Corollary 2, there exists an ∞-variety whose ω-languages coincide with S^ω $(= \mathrm{S}^s)$.

4.2 Limes Languages

The MCNAUGHTON Theorem says that an ω-language is regular if and only if it is a boolean combination of languages of the form $\lim(U)$ with regular +-languages U. Therefore we consider for a given finite semigroup S the languages which can be written as boolean combinations of languages $\lim(U)$ where U is a +-language recognized by S. And with a given variety S of semigroups we associate the class $\overrightarrow{\mathrm{S}}$ of regular ω-languages that are boolean combinations of languages of the form $\lim(U)$ with regular +-languages U that are recognized by elements of S.

Again we start with a construction that converts a semigroup into a right binoid. Let S be a finite semigroup, and let $\mathcal{R}$ denote GREEN's relation ($x\mathcal{R}y \Leftrightarrow \exists a, b \in S^1(xa = y \wedge yb = x)$). Then we can define the quintuple $\overrightarrow{S} = (S, S/\mathcal{R}, \cdot, \cdot_\omega, {}^\omega)$ with $s \cdot_\omega [t]_\mathcal{R} := [st]_\mathcal{R}$ and $s^\omega := [s^\pi]_\mathcal{R}$. We have

Theorem 5 *Let S be a finite semigroup and L an ω-language.*

1. *The operations of $\overrightarrow{S}$ are well-defined and $\overrightarrow{S}$ is a finite right binoid.*

2. *If S is $\mathcal{R}$-trivial then the following equivalence holds: L is a boolean combination of languages* $\lim(U)$ *with* $+$*-languages U recognized by S if and only if L is recognized by* $\overrightarrow{S}$.

$\overrightarrow{S}$ *is called the* $\mathcal{R}$-binoid *of* S ∎

From the preceding theorem we obtain as a consequence

Corollary 3 *Let* S *be a variety of* $\mathcal{R}$*-trivial semigroups. Then there exists an* ∞*-variety* $\mathcal{L}$ *such that L belongs to* $\overrightarrow{\mathbf{S}}$ *if and only if L belongs to* $\mathcal{L}$ ∎

It is still open whether there are classes of type $\overrightarrow{\mathbf{S}}$ that cannot be described by ∞-varieties in such a way. But again, we can apply Corollary 2 in special cases. For example, if S is closed under SCHÜTZENBERGER product we have $\mathbf{S}^{s} = \overrightarrow{\mathbf{S}}$ (see [8], [9, Prop. 1.3, p. 137]). Hence, by Corollary 2, there exists an ∞-variety whose ω-languages form the class $\overrightarrow{\mathbf{S}}$ $(= \mathbf{S}^{s})$.

5 Concrete Classes of ∞-Languages and Defining Equations

The topic of this section is the study of concrete classes of regular ∞-languages in the framework of the new concept of right binoids and ∞-varieties. In order to demonstrate that the concept of ∞-varieties is flexible we examine examples that are differently motivated. The first subsection deals with topologically defined, the second subsection with syntactically defined classes of languages.

5.1 Topological Classes inside the BOREL Hierarchy

Let us consider A^{∞} in this subsection not as an algebraic structure but as a metric space with the distance function

$$\operatorname{dist}(u,v) := \begin{cases} 0 & \text{if } u = v \\ 2^{-\min\{i \mid u_i \neq v_i\}} & \text{if } u \neq v, u, v \in A^{\omega} \\ 1 & \text{otherwise} \end{cases}$$

Following the usual practice we shall denote the classes of the BOREL hierarchy of the induced topology by $\mathcal{G}$, $\mathcal{F}$, $\mathcal{G}_{\delta}$, $\mathcal{F}_{\sigma}$, $\mathcal{G}_{\delta\sigma}, \ldots$ (see [13, p. 152 ff.])

It is known that every regular language is in the boolean closure of $\mathcal{F}_{\sigma}$ (which coincides by symmetry with the boolean closure of $\mathcal{G}_{\delta}$). We are therefore interested in the lower levels of the hierarchy. We focus on the classes $\mathcal{F} \cap \mathcal{G}$ and $\mathcal{F}_{\sigma} \cap \mathcal{G}_{\delta}$, which are closed with respect to boolean combinations, and hence are candidates for ∞-varieties.

The main result is

Theorem 6 *1. The regular $\mathcal{F} \cap \mathcal{G}$-languages form an ∞-variety. A regular ∞-language belongs to $\mathcal{F} \cap \mathcal{G}$ if and only if it is recognized by a finite right binoid that satisfies the equation*

$$v^{\pi} w \nearrow x \;=\; v^{\pi} y \nearrow z \tag{3}$$

2. *The regular $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$-languages form an ∞-variety. A regular ∞-language belongs to $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ if and only if it is recognized by a finite right binoid that satisfies the equation*

$$(x^\pi y^\pi)^\pi \nearrow x \;=\; (x^\pi y^\pi)^\pi \nearrow y \tag{4}$$

Hints to the proof. The proof of (1.) is based on the fact that a regular language L over A belongs to $\mathcal{F} \cap \mathcal{G}$ iff $L_\omega = U \nearrow A$ with a finite language U of finite words. The proof of (2.) rests on a result due to LUDWIG STAIGER and KLAUS WAGNER (see [12, Satz 5, p. 387]): a regular set is in the boolean closure of $\mathcal{F}$ if and only if it is in the intersection of $\mathcal{F}_\sigma$ and $\mathcal{G}_\delta$ ∎

5.2 Piecewise- and Weakly Piecewise-testable Languages

As an example of syntactically defined ∞-languages we consider piecewise-testable and weakly piecewise-testable languages.

We recall some definitions: let k denote a finite, non-zero cardinal. A $+$-language of the form $A^*a_1A^*\dots A^*a_lA^*$ with $l \in \{1,\dots,k\}$ is called piecewise k-simple over A. A $+$-language is piecewise k-testable if it is a boolean combination of piecewise k-simple languages. An ω-language is piecewise k-testable if it is a boolean combination of languages of the form $A^* \nearrow U$ and $U \nearrow A$ where U is piecewise k-simple.

We call an ω-language weakly piecewise k-testable if it is a boolean combination of languages of the form $U \nearrow A$ where U is piecewise k-simple, thus the membership of an ω-word depends no longer on the subwords occurring infinitely often.

An ∞-language L is (weakly) piecewise k-testable if L_+ is piecewise k-testable and L_ω is (weakly) piecewise k-testable. By $w\mathcal{P}t_k$ and $\mathcal{P}t_k$ we denote the appropriate classes of languages. A language is (weakly) piecewise-testable if it belongs to $\mathcal{P}t := \bigcup_{k>0} \mathcal{P}t_k$ resp. $w\mathcal{P}t := \bigcup_{k>0} w\mathcal{P}t_k$.

The main result is

Theorem 7 *1. The classes of the (weakly) piecewise (k-)testable languages form ∞-varieties.*

2. *An ∞-language is weakly piecewise (k-)testable if and only if it is piecewise (k-)testable and in $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$, i.e.,*

$$w\mathcal{P}t = \mathcal{P}t \cap F_\sigma \cap \mathcal{G}_\delta \qquad\qquad w\mathcal{P}t_k = \mathcal{P}t_k \cap F_\sigma \cap \mathcal{G}_\delta$$

Hints to the proof. The starting point of an algebraic proof is the definition of (weakly) piecewise (k-)equivalence of finite and infinite words. Then (1.) is immediate. The inclusions from left to right in (2.) are straightforward, for the reverse inclusions Equation 4 is used ∎

IMRE SIMON showed ([11]) that the piecewise-testable $+$-languages correspond to the variety of $\mathcal{J}$-trivial semigroups. Form this, one obtains a description of these languages by equations. JEAN-PIERRE PÉCUCHET shows that these equations describe piecewise-testable ω-languages too. Thus we obtain immediately defining equations for piecewise-testable ∞-languages and in combination with Equation 4 also equations for weakly piecewise-testability.

The first part of the theorem can easily be transferred to locally testable languages (see [7]) and to locally threshold-testable languages (see [2]), but the question about the characterization of weak testability through $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$-sets is still open in both cases.

6 Practical Computation

In order to show how to calculate with right binoids and ∞-languages, we examine an example. This shows that the weakly piecewise testable ω-languages and the regular $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$-languages do not form a class of type S^s.

The alphabet $\mathrm{A} = \{\mathrm{a}, \mathrm{b}\}$ is fixed. Consider the ω-language L given through $\mathrm{L} = \mathrm{a}^\omega + \mathrm{a}^*\mathrm{ba}^\omega$. L is weakly piecewise 2-testable as seen by the following equivalent description $\mathrm{L} = \mathrm{A}^\omega \setminus (\mathrm{A}^*\mathrm{bA}^*\mathrm{bA}^* \nearrow \mathrm{A})$

$\approx_\mathrm{L}$ consists of the classes a^+, $\mathrm{a}^*\mathrm{ba}^*$, $\mathrm{a}^*\mathrm{ba}^*\mathrm{bA}^*$, a^ω, $\mathrm{a}^*\mathrm{ba}^\omega$ and $\mathrm{a}^*\mathrm{ba}^*\mathrm{bA}^\omega$. For a proof verify the following equivalences for a given +- resp. ω-word x

$$\begin{aligned}
uxv \nearrow z \in L &\Leftrightarrow z \in a^+ \wedge ((x \in \mathrm{a}^+ \wedge uv \in \mathrm{a}^* + \mathrm{a}^*\mathrm{ba}^*) \vee (x \in \mathrm{a}^*\mathrm{ba}^* \wedge uv \in \mathrm{a}^*))\\
u \nearrow xv \in L &\Leftrightarrow xv \in \mathrm{a}^+ \wedge u \in \mathrm{a}^* + \mathrm{a}^*\mathrm{ba}^*\\
ux \in L &\Leftrightarrow (x \in \mathrm{a}^\omega \wedge u \in \mathrm{a}^* + \mathrm{a}^*\mathrm{ba}^*) \vee (x \in \mathrm{a}^*\mathrm{ba}^\omega \wedge u \in \mathrm{a}^*)
\end{aligned}$$

Now we compute the f-simple languages of the canonical morphism $\mathrm{f}\colon \mathrm{A}^+ \to \mathrm{A}^+/\sim_\mathrm{L}$ where $\sim_L$ denotes the restriction of $\approx_L$ to finite words (that always coincides with André Arnold's syntactic congruence for ω-languages):

$$\begin{aligned}
\mathrm{a}^+ \nearrow \mathrm{a}^+ &= \mathrm{a}^\omega\\
\mathrm{a}^*\mathrm{ba}^* \nearrow \mathrm{a}^+ &= \mathrm{a}^*\mathrm{b} \nearrow \mathrm{a}\\
\mathrm{a}^*\mathrm{ba}^*\mathrm{bA}^* \nearrow \mathrm{a}^+ &= \mathrm{a}^*(\mathrm{ba}^*)^+ \nearrow \mathrm{a}\\
\mathrm{a}^*\mathrm{ba}^*\mathrm{bA}^* \nearrow \mathrm{a}^*\mathrm{ba}^*\mathrm{bA}^* &= \mathrm{a}^* \nearrow \mathrm{ba}^*
\end{aligned}$$

From this we can derive that $\sim_\mathrm{L}$ saturates the language $\mathrm{L}' := \mathrm{a}^* \nearrow \mathrm{ba}^*$ which is obviously not weakly piecewise testable. Hence by [7, Th. 3.3, p. 83] we have

Corollary 4 *1. The class of the weakly piecewise-testable languages and the classes of the weakly piecewise k-testable languages are not of type* S^s *for a variety* S *of semigroups and* $k \geq 2$.

2. The regular languages in $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ *do not form a class of languages of type* S^s *for a variety* S *of semigroups* ■

At the end we show that $(\mathrm{A}^+/\sim_\mathrm{L})^\infty$ does not satisfy Equation 4, where $\mathrm{A}^\infty/\!\approx_\mathrm{L}$ does, as we would expect.

Equation 4 is equivalent to $(x^\pi y^\pi)^\pi \nearrow x^\pi = (x^\pi y^\pi)^\pi \nearrow y^\pi$. Therefore it remains to check wether the equation $xy \nearrow x = xy \nearrow y$ holds for all choices of idempotent elements x, y in the +-part of $(\mathrm{A}^+/\sim_\mathrm{L})^\infty$ resp. $\mathrm{A}^\infty/\!\approx_\mathrm{L}$.

The classes [a] and [bb] constitute the only idempotent elements in $\mathrm{A}^+/\sim_\mathrm{L}$ resp. $(\mathrm{A}^\infty/\!\approx_\mathrm{L})_+$.

Thus $(A^+/\sim_L)^\infty$ does not satisfy Equation 4, since we have that the two linked pairs $([a]_{\sim_L}[bb]_{\sim_L}, [a]_{\sim_L})$ and $([a]_{\sim_L}[bb]_{\sim_L}, [bb]_{\sim_L})$ are not conjugated. This follows from the fact that the corresponding f-simple languages are disjoint (see further above).

If $x = y$ the equation holds trivially. Since the +-part of $A^\infty/\approx_L$ is commutative, Equation 4 holds in $A^\infty/\approx_L$, because we have $a \cdot bb \nearrow bb \approx_L a \cdot bb \nearrow a$.

Discussion

We have shown that right binoids together with a strong type of morphism lead to an adequate notion of recognizability of ∞-languages, which even gives the possibility of stating a suitable EILENBERG Theorem. The additional property demanded of recognizing morphisms compensates the lack of an infinite product (which could hardly be handled in a constructive manner anyway).

It turns out that there is a close connection between recognizability (in the sense of saturation) of ω-languages by semigroups and recognizability of ∞-languages in terms of right binoids. But as the example of the weakly piecewise-testable languages shows, right binoids constitute a finer means for classification.

Future research could deal with other important classes of regular languages in the new framework (I consider for instance logically defined subsystems of propositional temporal logic). Furthermore investigations in the theory of the structure of right binoids is necessary (perhaps modelled on the structure theory of finite semigroups), and finally we need an efficient algorithm that computes the syntactic binoid of a given language.

Acknowledgements. I would like to thank in particular WOLFGANG THOMAS for supervising the diploma thesis [15] this paper is based on. Also from his critical reading the paper gained a lot in clarity.

I am grateful to ANDRÉ ARNOLD and JEAN-ERIC PIN for fruitful discussions, from which section 6 arose, and to DAMIAN NIWINSKY for his suggestions that improved the readability of sections 4 and 5. I met all three on the Workshop 'Logics and Recognizable Sets' of the ESPRIT Basic Research Action 3166 'ASMICS' held in Dersau, October 8–10, 1990. JEAN-ERIC PIN also suggested to give a cleaner definition of right binoids from the view of universal algebra, and pointed out the notion of linked pair.

References

[1] A. Arnold. A syntactic congruence for rational ω-languages. *Theoretical Computer Science*, 39:333–335, 1985.

[2] D. Beauquier and J. E. Pin. Factors of words. In *Proceedings on the 16th ICALP*, pages 63–79, Berlin, 1989. Springer. LNCS 372.

[3] B. Courcelle. On recognizable sets and tree automata. In H. Aït Kaci and M. Nivat, editors, *Resolution of Equations in Algebraic Structures*, volume 1, pages 93–126. Academic Press, 1989.

[4] S. Eilenberg. *Automata, Languages and Machines*, volume B. Academic Press, 1976.

[5] J.-P. Pécuchet. Etude syntaxique des parties reconnaissables de mots infinis. In L. Kott, editor, *Proceedings on the 13th ICALP*, pages 294–303. Springer, 1986. LNCS 226.

[6] J.-P. Pécuchet. Variétés de semigroupes et mots infinis. In G. Vidal-Naquet, editor, *Proceedings on the 3rd Symposium on Theoretical Aspects of Computer Science*, pages 180–191, Berlin, Heidelberg, New York, 1986. Springer. LNCS 210.

[7] J.-P. Pécuchet. *Automates boustrophedons, langages reconnaissables de mots infinis et variétés de semigroupes.* PhD thesis, Univ. Rouen, 1987.

[8] D. Perrin. Variétés de semigroupes et mots infinis. *Comptes Rendues Academie de Science Paris*, 295:595–598, 1982.

[9] D. Perrin. Recent results on automata and infinite words. In *Mathematical Foundation of Computer Science*, pages 134–148, Berlin, Heidelberg, New York, 1984. Springer. LNCS 176.

[10] J.-E. Pin. *Varieties of Formal Languages.* North Oxford Academic Press, London, 1986.

[11] I. Simon. Piecewise testable events. In *Proceedings of the 2nd GI Conference*, pages 214–222, Berlin, Heidelberg, New York, 1975. Springer Verlag. LNCS 33.

[12] L. Staiger and K. Wagner. Automatentheoretische Charakterisierungen topologischer Klassen regulärer Folgenmengen. *Elektronische Informationsverarbeitung und Kybernetik*, 10:379–392, 1974.

[13] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, pages 134–191. Elsevier Science Publishers B.V., 1980.

[14] Th. Wilke. An EILENBERG theorem for ∞-languages. In *Proceedings of the Workshop 'Logics and Recognizable Sets'*, Rep. 9004, Inst. f. Informatik u. Prakt. Math., Univ. Kiel, 1990.

[15] Th. Wilke. *Eine* EILENBERG-*Korrespondenz für reguläre ∞-Sprachen.* Diploma thesis, RWTH Aachen, 1990.