An Eilenberg Theorem for \( \omega \)-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 \( \omega \)-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 \( \omega \)-languages, and treat concrete classes of \( \omega \)-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 \( \omega \)-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 \( \omega \)-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 (\( \omega \)-languages), which is based on the new notions 'right binoid', 'Ramsey condition', and '\( \omega \)-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 \( \omega \)-and \( \omega \)-languages.

\( \omega \)-varieties are classes of regular \( \omega \)-languages that are closed with respect to inverse images of language morphisms, boolean combinations, and quotients. In connection with an Eilenberg Theorem for \( \omega \)-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 \( \omega \)-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 \( \omega \)-languages (section 3).
In the second part we relate right binoids and \( \infty \)-varieties to the semigroup approach to the characterization and classification of regular \( \omega \)-languages (section 4), deal with some concrete classes of regular \( \infty \)-languages (section 5), and show how to compute with \( \infty \)-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 \( A^\star := \{u \cdot v^\omega \mid u,v \in A^+\} \), i.e., \( A^\star \) is the set of all ultimately periodic words over \( A \). A set \( L \subseteq A^{\infty} \) is called an \( \infty \)-language over \( A \). It is said to be regular if its \( + \)-part \( L_+ := L \cap A^+ \) and its \( \omega \)-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 \( \infty \)-languages. We start with a characterization through congruences over finite and infinite words (\( \infty \)-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 (\( \infty \)-Congruence) An \( \infty \)-congruence over \( A \) is an equivalence relation \( \approx \subseteq (A^+ \times A^+) \cup (A^\omega \times A^\omega) \) such that

\[
x \cdot x' \approx y \cdot y' \quad \text{and} \quad x^\omega \approx y^\omega
\]

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

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

Hence we do not get a characterization of regular \( \infty \)-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 \( \omega \)-iteration, and thus are not affected by the definition of \( \infty \)-congruences. The following definition remedies the situation.

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

\[
\prod X \approx X_0^\omega
\]

holds for every infinite sequence \( X : \omega \rightarrow A^+ \) of \( \approx \)-equivalent words \( X_0, X_1, \ldots \).