Combinatorics on Infinite Words*

Juhani Karhumäki and Arto Lepistö

Department of Mathematics &
Turku Centre for Computer Science
University of Turku
20014 University of Turku, Finland
E-mail: karhumak@cs.utu.fi, alepisto@utu.fi

Summary. We consider several problems of infinite words over a finite alphabet. In particular, we describe a few automata-theoretic methods to define infinite words. Properties of infinite words studied in more details are repetition-freeness, periodicity and different kinds of complexity issues. Examples are used to illustrate the power of infinite words in many applications, as well as illustrations of problems from different areas of mathematics.

20.1 Introduction

Combinatorics on words is a relatively new area of discrete mathematics. Although it has connections to numerous branches of mathematics, and indeed much of the early theory is developed implicitly as tools to attack some quite different problems, its basic motivation and drive comes from computer science. This is seen, for instance, in the recent classification of the field in Mathematical Reviews. As illustrated below combinatorics on words is under the main section of Computer Science, but with the very strong mathematical emphasis.

The notion of a word, e.g., a sequence of symbols, is extremely natural in mathematics. Indeed, the representation of a natural number at any base is a word over a finite alphabet. And for a computer any algorithm on numbers is an algorithm on words!

Consequently, it is no surprise that the words has occurred - often implicitly – in mathematical considerations during several centuries. As an illustration we mention that already Gauss, see [16], came to a problem on words, and that Prouhet, see [34], discovered the infinite Thue-Morse word.

A systematic research on words, in fact on infinite words, was initiated by a Norwegian A. Thue almost hundred years ago. In 1906, see [38], he published his first results on repetition-free words, including a construction of

* Supported by the Academy of Finland under the grant 44087
an infinite binary word not containing any cubes, i.e., factors of the form $u^3$. Thue's results were for many decades unnoticed, and many of his results were subsequently rediscovered. Interestingly, Thue seemed to have no motivation – beyond scientific curiosity – for his research. Later very often results on words were obtained as byproducts when looking for tools to attack some other, often very unrelated-looking, problems. Papers to be mentioned after Thue are, for example, [30] and [3].

As a theory, Combinatorics on Words started in 1950's in two places simultaneously and independently. In Russia P.S. Novikov and S. Adian gathered a lot of knowledge when searching for a solution of Burnside Problem for groups, see [1]. In their considerations much of the theory of words was implicit – as it has been throughout the history of the well developed combinatorial group theory, see [26] and [25]. In France, research on words was initiated by M.P. Schützenberger in connection with the theory of codes, see [36].

In coming decades research on words extended rapidly, also geographically. Many remarkable results we revealed, such as the decidability of the satisfiability problem for word equations, see [27] and [33], and the compactness property of word equations, see [2], [17] or [8], just to mention only two jewels.

In 1983 the research, especially the part connected to France, culminated into the first monograph of the field, Combinatorics on Words, see [23]. The book had a very inspiring effect: since that the area has been growing steadily and very fast. A new monograph, Algebraic Combinatorics on Words, see [24], appeared recently, and a biannual conference WORDS has been established. Papers [8] and [6] are two other recent survey articles.

In this presentation we consider a particular part of the theory of Combinatorics on Words, namely that of infinite words. Of course, the theory of infinite words is essentially more complicated than that of finite words. For example, the cardinality of the set of all infinite words (over a nonunary alphabet) is nondenumerable, while that of all finite words is only denumerable. However, many considerations are much neater for infinite words than for fi-