Theoretical Computer Science 609 (2016) 272–285

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On block pumpable languages ☆

Christopher Hanrui Chak [a], Rūsiņš Freivalds [b], Frank Stephan [a,c,*], Henrietta Tan Wan Yik [a]

[a] Department of Computer Science, National University of Singapore, 13 Computing Drive, COM1, Singapore 117417, Republic of Singapore
[b] Institute of Mathematics and Computer Science, University of Latvia, Raiņa bulvāris 29, Riga, LV-1459, Latvia
[c] Department of Mathematics, The National University of Singapore, 10 Lower Kent Ridge Drive, S17, Singapore 119076, Republic of Singapore

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form 22 June 2015
Accepted 1 October 2015
Available online 9 October 2015
Communicated by Z. Esik

Keywords:
Block pumping lemma
Regular languages
Block pumpable languages
Automata theory

ABSTRACT

Ehrenfeucht, Parikh and Rozenberg gave an interesting characterisation of the regular languages called the block pumping property. When requiring this property only with respect to members of the language but not with respect to nonmembers, one gets the notion of block pumpable languages. It is shown that these block pumpable are a more general concept than regular languages and that they are an interesting notion of their own: they are closed under intersection, union and homomorphism by transducers; they admit multiple pumping; they have either polynomial or exponential growth.

1. Introduction

Many variants of the pumping lemma have been studied extensively for both, the regular and the context-free languages, see Nijholt [18] for an overview and textbooks like the ones of Hopcroft, Motwani and Ullman [10] and of Dömösi and Ito [4] for the general relation between regular languages, context-free languages and the various pumping lemmas as well as for the automata-theoretic background. While the standard pumping lemma does not characterise the regular sets [21], Jaffe [11] was the first to publish a version of the pumping lemma which characterises the regular sets. This characterisation can be seen as a direct coding of the conditions of the Myhill–Nerode Theorem. Ehrenfeucht, Parikh and Rozenberg [5] provided a more natural characterisation of regular languages which was based on the idea of block pumping: A language L is regular iff there is a number p such that for all words of length $p-1$ or more, if one sets p breakpoints which split the words into blocks, then there are two of the breakpoints satisfying that the blocks of the word between the breakpoints can be omitted or arbitrarily often repeated without changing the membership of the word in L. For understanding this better, one should first formalise the notion of breakpoints and blocks.

Given a word x and a pair of breakpoints i, j with $i < j$ (i strictly before j), let u denote the part of x before i, v the part of x between i and j and w the part of x after j. The k times pumped version uv^kw of x where the part between i and j is repeated k times (in place of originally 1 time) is denoted as $P^k(x:i,j)$; furthermore, let $P^*(x:i,j)$ denote

☆ R. Freivalds' research was supported by the project 271/2012 from the Latvian Council of Science and by the project ERAF Nr. 2DP/2.1.1.1/13/APIA/VIAA/027. F. Stephan was supported in part by NUS grants R146-000-181-112 and R146-000-184-112.

* Corresponding author.
E-mail addresses: qornotq@gmail.com (C.H. Chak), rusins@cclu.lv (R. Freivalds), fstephan@comp.nus.edu.sg (F. Stephan), u0909036@u.nus.edu (Tan Wan Yik).

http://dx.doi.org/10.1016/j.tcs.2015.10.008
0304-3975/

$\{P^k(x:i,j):k \geq 0\}$ and $P^+(x:i,j)$ denote $\{P^k(x:i,j):k \geq 1\}$. So $P^0(x:i,j) = uv$, $P^1(x:i,j) = x$ and $P^*(x:i,j) = uv^*w$. Furthermore, note that breakpoints can, but do not need to be, coincide with the beginning and the end of the words. A set I of p breakpoints $k_1, k_2, \ldots, k_p$ splits the word x into blocks $w_0, w_1, \ldots, w_p$ where w_0, w_p can be empty and all other blocks contain at least one symbol, that is, every two different breakpoints sit at different positions. Thus universal quantifications over sets of p breakpoints are vacuously true for words shorter than $p-1$ symbols. Furthermore, it is always assumed that a pair i, j of breakpoints satisfies $i < j$ (in the ordering of appearance of the breakpoints). Furthermore, often breakpoints are identified with numbers in ascending order, so that one could write 00(1)00(2)01(3)00(4) for four breakpoints $1, 2, 3, 4$ in a word 00000100 spaced out by two symbols; $P^*(00000100:1,3)$ is $00(0001)^*00$. Using this formalism, one can state the Theorem of Ehrenfeucht, Parikh and Rozenberg as follows, where $\overline{L}$ denotes the complement of L.

Theorem 1. *(See Ehrenfeucht, Parikh and Rozenberg [5].) A language L is regular iff there is a constant p such that for every word x and every set I of p breakpoints of x there is a pair i, j of breakpoints in I with either $P^*(x:i,j) \subseteq L$ or $P^*(x:i,j) \subseteq \overline{L}$.*

Ehrenfeucht, Parikh and Rozenberg were able to show an even stronger version of their result, namely that L is regular iff there is a constant p such that for every word x and every set I of p breakpoints of x there is a pair i, j of breakpoints in I with $L(P^1(x:i,j)) = L(P^0(x:i,j))$. Here, for a language L, if x is in L then the expression $L(x)$ evaluates to 1 else it evaluates to 0. They called the so defined property the block cancellation property. Related to this is the positive block pumping property and they asked whether the below theorem is true. The affirmative answer was provided by Varricchio [24] sixteen years after Ehrenfeucht, Parikh and Rozenberg raised the question.

Theorem 2. *(See Varricchio [24].) A language L is regular iff there is a constant p such that for every word x and every set I of p breakpoints of x there is a pair i, j of breakpoints in I with either $P^+(x:i,j) \subseteq L$ or $P^+(x:i,j) \subseteq \overline{L}$.*

The goal of the present work is to restrict these three notions, as done in the original pumping lemma, to address only members of the language L but to be silent on members of $\overline{L}$. This gives rise to three natural notions, among these the notion of block pumpable will be the most important; however, the other two notions share various properties of this notion as well.

Definition 3. A language L is called *block pumpable* iff there is a constant $p \geq 2$ such that for every $x \in L$ and every set I of p breakpoints of x there is a pair $i, j \in I$ such that $P^*(x:i,j) \subseteq L$.

A language L is called *positively block pumpable* iff there is a constant $p \geq 2$ such that for every $x \in L$ and every set I of p breakpoints of x there is a pair $i, j \in I$ such that $P^+(x:i,j) \subseteq L$.

A language L is called *block cancellable* iff there is a constant $p \geq 2$ such that for every $x \in L$ and every set I of p breakpoints of x there is a pair $i, j \in I$ such that $P^0(x:i,j) \in L$.

The constant p in the above definition is called the pumping constant of L; often one does not chose the optimal (smallest) pumping constant but that pumping constant for which the corresponding proof (for example, that the given language L is block pumpable with constant p) can be done most easily.

Note that one could weaken the notion of block pumpable to *weakly block pumpable* by only considering sets of breakpoints I which include the left end and the right end of the word. This weakening makes it possible to satisfy the requirement by choosing the ends of the words as breakpoints i, j and then $P^*(x:i,j) = x^*$; thus languages like $\{x \in \Sigma^*$: all symbols in Σ occur in x in the same number$\}$ are weakly block pumpable and many of the properties of the class of block pumpable languages are not shared by the class of weakly block pumpable languages. Hence the notions block pumpable, positively block pumpable and block cancellable are more natural choices.

The next section will provide a comparison of the notion blockpumpable with related notions and it will provide examples of sets which separate out the notions; the section will furthermore provide some background on useful combinatorial theorems like Ramsey's Theorem and the Theorem of Fine and Wilf which are helpful in obtaining the results of this and subsequent sections. In particular Ramsey's Theorem will be applied in various proofs of this paper, including the result that a language is block pumpable iff it is positively block pumpable and also block cancellable.

Section 3 will then show that the block pumpable languages have nice closure properties: their unions, intersections and concatenations are again block pumpable, furthermore their images under transductions are block pumpable as well. Section 4 will show that multiple pumping is possible, that is, one can split words in L into two parts such that each part can be pumped independently of the other part. Section 5 will show that block pumpable languages fall into two groups: those of polynomial growth which are always regular and those of exponential growth which can also be nonregular. While the results of Section 3 and Section 4 also apply to block cancellable and positively block pumpable languages, the results of Section 5 only generalise to positively block pumpable languages and fail for block cancellable languages.

Section 6 will introduce the concept of effective block pumpable structures and will show that the existential positive theory of effective block pumpable structures is decidable, however, larger fragments of the first-order theory fail to be decidable. Here a block pumpable structure consists of only finitely many relations and all relations and sets involved are recursive and that every relation given as a set of convoluted tuples satisfying the relation is block pumpable; furthermore,