On Block Pumpable Languages*

Christopher Hanrui Chak[1], Rūsiņš Freivalds[2], Frank Stephan[1 3] and Henrietta Tan Wan Yik[1]

[1] Department of Computer Science, National University of Singapore
13 Computing Drive, COM1, Singapore 117417, Republic of Singapore
qornotq@gmail.com, u0909036@u.nus.edu

[2] Institite of Mathematics and Computer Science, University of Latvia
Raiņa bulvāris 29, Riga, LV-1459, Latvia
rusins@cclu.lv

[3] Department of Mathematics, The National University of Singapore
10 Lower Kent Ridge Drive, S17, Singapore 119076, Republic of Singapore
fstephan@comp.nus.edu.sg

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.

Keywords. Block pumping lemma; regular languages; block pumpable languages; automata theory.

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, 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

* 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.

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 $\{P^k(x : i, j) : k \geq 0\}$ 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 univeral 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 (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 (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 chosing 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, one permits to use members of the domain of the structure as constants in the logical language. The present work includes results from the Final Year Projects of Chak [2] and Tan [22].

2 Basic Examples and Helpful Theorems

It is clear, that every regular language is block pumpable. Namely, if p is larger than the number of states of an NFA recognising the language then, when non-deterministically accepting a word x, the run of the automaton has to go through the same state between two different breakpoints and one can therefore pump the part between the two breakpoints.

The following well-known example of $L = 0^*1^*$ illustrates the concept of breakpoints and pumping. An NFA with two states recognises the language as follows: the automaton goes from the first state on 0 to itself and on 1 to the second state; it goes on 1 in the second state to itself; the first state is starting and both states are accepting. In the following it is shown that the block pumping constant is 3: For any word x, say $x = 0000111$, and three breakpoints i, j, k in it, say $(i)00(j)001(k)11$, one sees that only one of the possible segments between two neighbouring breakpoints can contain both symbols, 0 and 1; here these breakpoints are j and k. The segment between the two other neighbouring breakpoints, here i and j, contains only symbols of one type and that can then be pumped. In short, $P^*(0000111 : i, j) \subseteq L$ which means $(00)^*00111 \subseteq L$. In general, if a language is recognised by an NFA with p states then it is block pumpable with constant $p + 1$.

The next example shows that there are block pumpable languages which are not regular. The existence of languages of this type and their properties are the main topic of the present work. In the following, recall that y is a subword of x iff $y = vxw$ for some words v, w and that x is square-free iff x does not have a nonempty subword of the form yy and that x is cube-free iff x does not have a nonempty subword of the form yyy. Similarly, x is square-containing or cube-containing iff x contains a nonempty subword of the form yy or yyy, respectively. For example, 01210212 is square-free, 011010011001 is cube-free and square-containing (it contains the squares 00, 11 and 10011001) and 01001100110011101 is cube-containing (it contains the cube 001100110011).

Example 4. Morse [17], Prouhet [19] and Thue [23] constructed inductively a sequence of words over the alphabet $\{0, 1\}$ by letting

$$\begin{aligned} x_0 &= 0; \\ y_0 &= 1; \\ x_{n+1} &= x_n y_n y_n x_n y_n x_n x_n y_n y_n x_n x_n y_n x_n y_n y_n x_n; \\ y_{n+1} &= y_n x_n x_n y_n x_n y_n y_n x_n x_n y_n y_n x_n y_n x_n x_n y_n. \end{aligned}$$

They showed that every x_n and y_n is cube-free. Furthermore, every x_n and every y_n has length 16^n. The language

$$L = \{z \in \{0, 1\}^* : \forall n\, [z \neq x_n \wedge z \neq y_n]\}$$

is block pumpable and not regular.

Proof. That x_n, y_n are cube-free is known from the literature [17, 19, 23]; indeed, the goal of their original construction was to show that also $x_\infty = \lim_n x_n$ is an infinite cube-free sequence

— though this is not needed for the present proof. The fact that the lengths of the x_n and y_n are both 16^n can be verified by a simple induction; one just needs to count the number of occurrences of words of length 16^n when defining the words x_{n+1}, y_{n+1} of length 16^{n+1}. Therefore, the main task of this proof is to show that L is block pumpable and L is not regular.

First, it will be shown that for any word $x \in L$ and any set I of five breakpoints which split x into $w_0(1)w_1(2)w_2(3)w_3(4)w_4(5)w_5$, either the part $w_0w_1w_2$ before the breakpoint (3) or the part $w_3w_4w_5$ has at most length $|x|/2$. Without loss of generality this is the part before the breakpoint (3). Now the inequalities

$$|x|/2 \leq |w_0w_3w_4w_5| < |w_0w_1w_3w_4w_5| < |w_0w_1w_2w_2w_3w_4w_5| < |w_0w_1w_2w_1w_2w_3w_4w_5| \leq 3|x|/2$$

hold and only one of these words $P^0(x:1,3), P^0(x:2,3), P^2(x:2,3), P^2(x:1,3)$ can have a length which is a power of 16. Thus there is a pair of breakpoints $i,j \in \{1,2,3,4,5\}$ satisfying that $P^0(x:i,j), P^2(x:i,j) \in L$. Furthermore, $x = P^1(x:i,j) \in L$. As all cube-containing words are members of L and every word $P^k(x:i,j)$ with $k \geq 3$ contains a cube, $P^k(x:i,j) \in L$ for all $k \geq 3$. In summary, $P^*(x:i,j) \subseteq L$.

Second, as the elements of the complement of L are all cube-free, the complement of L cannot contain any subset of the form uv^*w although it is infinite. Thus the complement of L is not regular and so L itself is not regular. $\Box$

A slight modification of this example gives a linear block pumpable language which is not regular. Here a linear language is a context-free language where in the grammar every rule is either of the form $A \to v$ or $A \to vBw$ for nonterminals A, B and words v, w over the terminal alphabet. It would have been tempting to directly use the context-free set of non-prefixes of the infinite cube-free sequence provided by Main, Bucher and Haussler [16], however, this set itself is not block pumpable. Thus the example is a slight modification of it.

Example 5. Let $x_0, x_1, \ldots$ as in Example 4 and let

$$L = \{0,1,2\}^* - \{2x_02x_12\ldots 2x_n2 : n \in \mathbb{N}\}.$$

The language L is linear, block pumpable and not regular.

Proof. The word x_{n+1} can be obtained from x_n by replacing each occurrence of 0 by x_0 and each occurrence of 1 by y_1. Furthermore, these words are mirror-images of themselves. Now one considers the language H consisting of all words of the form $a_na_{n-1}\ldots a_2a_12b_1b_2\ldots b_{16n}$ such that all a_m, b_k are in $\{0,1\}$ and that there is an m with either $a_m = 0$ and $b_{16m-15}b_{16m-14}\ldots b_{16m-1}b_{16m} \neq x_1$ or $a_m = 1$ and $b_{16m-15}b_{16m-14}\ldots b_{16m-1}b_{16m} \neq y_1$. This language H does not contain x_n2x_{n+1} for any n. Furthermore, H is a linear language. Now the language L is the union of the following languages:

$$\begin{aligned}
L_1 &= \{u2v2w2z : u,z \in \{0,1,2\}^* \text{ and } v,w \in \{0,1\}^* \text{ and } v2w \in H\};\\
L_2 &= \{u2v2w2z : u,z \in \{0,1,2\}^*, v,w \in \{0,1\}^* \text{ and } |v|\cdot 16 < |w|\};\\
L_3 &= \{u2v2w2z : u,z \in \{0,1,2\}^*, v,w \in \{0,1\}^* \text{ and } |v|\cdot 16 > |w|\};\\
L_4 &= \{z \in \{0,1,2\}^* : z \notin 2x_02\cdot\{0,1,2\}^*\cdot 2 \text{ and } z \neq 2x_02\}.
\end{aligned}$$

It is easy to see that the languages L_1, L_2, L_3 are the concatenation of several regular languages and one linear language and are therefore again linear. Furthermore, L_4 is regular and thus linear. So the language L is the union of linear languages and linear.

The language L is block pumpable with block pumping constant 15: Let $w = w_0 w_1 \ldots w_{15}$ be a word in L with every breakpoint i between w_i and w_{i+1} (for simplicity, the breakpoints are just denoted with numbers). If there is an i with $1 \leq i \leq 12$ and $w_i w_{i+1} \in \{0,1\}^*$ then $P^*(x : i, j) \subseteq L$ for some $j \in \{i+1, i+2\}$ as only for one of the choices for j it can happen that one of $P^0(w : i, j), P^2(w : i, j)$ is of the form $2x_0 2x_1 2 \ldots 2x_n 2$ for some n. In the case that there is no such i, every second block must contain a 2. Hence the block $w_2 \ldots w_5$ contains at least two 2s and $w_2 \ldots w_7$ contains at least one 2 more than $w_2 \ldots w_5$. Assume now that one of $P^0(w : 2, 5)$ and $P^0(w : 2, 7)$ would be $2x_0 2x_1 2 \ldots 2x_n 2$. Then the subword $2x_n 2$ is not modified by pumping, as $w_8 \ldots w_{15}$ contains at least two 2s. Only one of $P^0(w : 2, 5)$ and $P^0(w : 2, 7)$ can be equal to $2x_0 2x_1 2 \ldots 2x_n 2$, the other one cannot, as it does not have n times a 2 before $2x_n 2$. Furthermore, $P^k(w : 2, 5)$ and $P^k(w : 2, 7)$ for $k > 1$ cannot be of equal to $2x_0 2x_1 2 \ldots 2x_n 2$, as they contain some subword $2v2$ twice. So one of the sets $P^*(w : 2, 5)$ and $P^*(w : 2, 7)$ is a subset of L. Hence L is block pumpable. $\Box$

The next examples are used to separate the notions of block cancellable and positively block pumpable from each other. The proof of the positive sides of these examples resembles the one in Example 4 while the negative sides follow well-known techniques about pumping. Note that the alphabet in Example 7 below can be unary, that is, just $\{0\}$. The proof of Example 7 below will use a proposition which is separated out, as it will also be useful in various proofs.

Proposition 6. *Every positively block pumpable language L with block pumping constant p is satisfying the following statements for all words u, v, w:*

1. *If $uv^p w \in L$ then $uv^{p+p!} w \in L$;*
2. *If $uv^k w \in L$ for all $k < p + p!$ then $uv^* w \subseteq L$;*
3. *If $uvw \in L$ and $|v| \geq p$ then there is a word v' with $|v'| = |v| + p!$ and $uv'w \in L$.*

The items 1,2,3 hold also for each positively block pumpable language.

Proof. The first item can be seen by considering the set $I = \{1, 2, \ldots, p\}$ of breakpoints where breakpoint i is placed after uv^i. Then there is a pair i, j of breakpoints such that $uv^i (v^{j-i})^* v^{p-j} w \subseteq L$. The number $r = 1 + p!/(j-i)$ satisfies that the r times pumped word $P^{1+p!/(j-i)}(uv^p w : i, j) = uv^i v^{r(j-i)} v^{p-j} w = uv^{p+p!} w$ is in L.

The second item can be shown by induction. Starting with $h = p + p!$ one shows that the formula $\forall k < h\, [uv^k w \in L]$ implies that $\forall k < h + 1\, [uv^k w \in L]$ by using that the word $uv^{h-p!} w$ is in L and that $h - p! \geq p$. As $\forall k < p + p!\, [uv^k w \in L]$, it then follows that $\forall k < h\, [uv^k w \in L]$ for all $h \geq p + p!$, hence $uv^* w$ is a subset of L.

The third item is shown similarly to the first item by placing the breakpoint i after the i-th symbol of v in the word uvw. Then one does the same as in the first item, only that one does not intend to have any control on which symbols in v are pumped; the only goal is to get $p!$ additional symbols into the word v'. $\Box$

Now Proposition 6 can be used to prove one direction in the next result.

Example 7. The language $L = \{w : |w| \text{ is not a power of } 2\}$ is block cancellable but not positively block pumpable.

Proof. Assume that x is a word and I a set of five breakpoints splitting x into six parts. Now one can find three breakpoints i, j, k such that $i < j < k$ and the part between i and k is at most as long as the half of x. It follows that only one of $P^0(x : i, j)$ and $P^0(x : i, k)$ can have a length which is a power of 2; thus one of these two words is again a member of L and L is block cancellable.

As $p \geq 2$, the number $2^{p!+1} - p!$ is not a power of 2: The inequality $p! < 2^{p!}$ implies the inequality $2^{p!} < 2^{p!+1} - p! < 2^{p!+1}$. Therefore there is a word v of length $2^{p!+1} - p!$ in L and taking u, w to be the empty word, one obtains from Proposition 6 that there is a word $v' \in L$ of length $2^{p!+1}$, in contradiction to the definition of L. Hence the assumption "L is positively block pumpable" is wrong. $\square$

Example 8. The set L consist of all square-containing words over the alphabet $\{0, 1, 2, 3\}$ is positively block pumpable but not block cancellable.

Proof. Kurosaki [14] provided a function φ such that $\varphi(v)$ is of the form uvw where u is obtained from v by exchanging the digits $1, 2$ and w is obtained from v by exchanging the digits $2, 3$. Now starting with the one-digit word $v_0 = 2$ and letting $v_{n+1} = \varphi(\varphi(v_n))$, Kurosaki showed that all these words v_n are square-free. Furthermore, one can easily see that v_{n+1} starts and ends with v_n, as for the first and last block of length $|v_n|$ of v_{n+1}, two digits are exchanged when forming $\varphi(v_n)$ and then exchanged back when forming $\varphi(\varphi(v_n))$ from $\varphi(v_n)$. Therefore every v_{n+1} equals $v_n w_n v_n$ for some word w_n of length $7|v_n|$.

Given a block cancellation constant p, one considers the word $v_p 0 v_p 0$ which is a square and is in L. Furthermore, one puts the k-th breakpoint at the end of the prefix v_k of the word $v_p 0 v_p 0$; so if $p = 3$ then one would split the word into parts v_1, $w_1 v_1$, $w_2 v_1 w_1 v_1$ and $0 v_3 0$. Now, when considering any pair i, j of breakpoints at the ends of the prefixes v_i, v_j of the word, then when omitting the symbols between i, j, the prefix v_j becomes replaced by v_i which is a suffix of v_j. For example, for $p = 3$ the three words obtained by omitting the part between two breakpoints are $v_1 w_2 v_1 w_1 v_1 0 v_3 0$, $v_1 w_1 v_1 0 v_3 0$ and $v_1 0 v_3 0$. As one can also see in the example for $p = 3$, the resulting word is a proper subword $u 0 v_p 0$ of $v_p 0 v_p 0$. As u, v_p are both square-free words on the alphabet $\{1, 2, 3\}$, any square would need to contain a 0; however, as $|v_p| > |u|$ and as a square must contain both zeroes, it must also contain the v_p between them and the part u is shorter than v_p, so a square cannot be formed. It follows that L is not block cancellable.

It is easy to see that L is positively block pumpable: The reason is that whenever i, j is a pair of breakpoints of a word x, then $P^k(x : i, j)$ contains for $k > 1$ at least one square and therefore positive pumping forces the pumped word into L. $\square$

Another related result is that the language of all words with the same number of 0 and 1 is weakly block pumpable but neither block cancellable nor positively block pumpable.

An important tool in all the work on the block pumpable languages is Ramsey Theory. These results are all corollaries of the finite and infinite versions of Ramsey's Theorem for Pairs.

Theorem 9 (Ramsey [20]). *For every $p \in \mathbb{N}$ and every finite set Q of colours, there exists a $p' \in \mathbb{N}$ such that for every ordered set I with p' elements and every function C mapping each pair $i, j \in I$ to a colour $C(i, j) \in Q$ there is a subset $J \subseteq I$ with p elements such that $C(i, j) = C(i', j')$ for all pairs $i, j \in J$ and pairs $i', j' \in J$. Here a pair $i, j \in I$ consists of two elements i and j of I with $i < j$.*

Note that p' depends on p and $|Q|$ and that one can choose $p' \leq (|Q| + 1)^{p \cdot |Q|}$ where for a set Q the notion $|Q|$ refers to the number of elements of Q. The following is an application of Ramsey's Theorem of Pairs to block pumping.

Proposition 10. *Let L be block pumpable with block pumping constant $p \geq 2$.*

1. *There is a constant $p' > p$ such that for every word $x \in L$ and every set I of p' distinct breakpoints in x, there is a subset $J \subseteq I$ consisting of p breakpoints such that $P^*(x : i, j) \subseteq L$ for all $i, j \in J$ and $k \in \mathbb{N}$.*
2. *For every number q there is a constant $p'' > p$ such that for every word x and every finite set Y of q words with $x \cdot Y \subseteq L$ and every set I of p'' distinct breakpoints in x there is a subset $J \subseteq I$ consisting of p breakpoints such that $P^k(x : i, j) \cdot y \in L$ for all breakpoints $i, j \in J$ and all $y \in Y$ and all $k \in \mathbb{N}$.*

Proof. For proving the first part, one chooses $p' = 6^p$. Now one colours every pair i, j of breakpoints in I as follows: If the pair defines a pump, that is, if $P^*(x : i, j) \subseteq L$, then the colour is white else the colour is red. By Ramsey's Theorem there is a monochromatic subset J of I consisting of p breakpoints. If all pairs $i, j \in J$ would be coloured red then the L would not be block pumpable; hence the colour must be white. Thus $P^*(x : i, j) \subseteq L$ for all pairs i, j of breakpoints in J.

For proving the second part, one chooses $p'' = (q + 2)^{p \cdot (q+1)}$. Now one considers any set Y with at most q elements and any x with $x \cdot Y = \{xy : y \in Y\} \subseteq L$ and any set I of p'' breakpoints of x. The colours are now the members of Y plus a special colour #; these colours are linearly ordered with # be the greatest element. For every pair i, j of breakpoints in I one defines that the colour of i, j is the least $z \in Y \cup \{\#\}$ such that $P^*(x : i, j) \cdot z \not\subseteq L$; note that the colour of i, j is always defined as $P^*(x : i, j) \cdot \{\#\} \not\subseteq L$. By Ramsey's Theorem there is now a colour $z \in Y \cup \{\#\}$ and a subset $J \subseteq I$ consisting of p breakpoints such that the colour of all pairs i, j of breakpoints in J is z. One can consider the breakpoints of x also to be breakpoints of xy for all the $y \in Y \cup \{\#\}$. If $z \neq \#$ then $xz \in L$ while $P^*(xz : i, j) = P^*(x : i, j) \cdot z \not\subseteq L$. This contradicts the fact that L is block pumpable. Thus $z = \#$ and $P^*(xy : i, j) = P^*(x : i, j) \cdot y \subseteq L$ for all $y \in Y$, as z is the least word $y \in Y \cup \{\#\}$ with $P^*(xy : i, j) \not\subseteq L$. This completes the proof. □

It is easy to see that this proposition and its proof also hold for the notions of positively block pumpable languages and block cancellable languages. For these, one has in the first part to colour according to $P^+(x : i, j) \subseteq L$ and $P^0(x : i, j) \in L$, respectively. In the second part, one can adjust the proof by choosing the colour to be the least $z \in Y \cup \{\#\}$ with $P^+(x : i, j) \not\subseteq L$

and $P^0(x : i, j) \notin L$, respectively. All other parts of the proof remain the same.

All those proofs which utilise Ramsey's Theorem work in general only for block pumpable languages and its variants but not for weakly block pumpable languages, as Ramsey's Theorem does not state that the minimum and maximum of I are also in J. So Proposition 10 does not hold for the notion of weak block pumpable languages.

The same is true for the below application of Ramsey's Theorem which states that a language is block pumpable iff it is both, block cancellable and positively block pumpable. The corresponding weak version fails: the language $\{(01^n0)^m : n, m \geq 1\} \cup \{0, 1\}^* \cdot \{00, 1\}$ is not weakly block pumpable; however, it satisfies the weak versions of being block cancellable and being positively block pumpable.

Proposition 11. *A language is block pumpable iff it is both, block cancellable and positively block pumpable.*

Proof. By definition, every block pumpable language is block cancellable and positively block pumpable. However, the reverse direction needs a proof which uses Ramsey's Theorem. So for a given language L which is positively block pumpable and block cancellable both with a constant p, let $p' = 9^p$. Now for every word $x \in L$ and every set I of p' breakpoints on x, one assigns to every pair $i, j \in I$, if $P^+(x : i, j) \subseteq L$ then one assigns to i, j the colour white else one assigns to i, j the colour red. Now there is a set $J \subseteq I$ of p breakpoints of x which is monochromatic. At least one pair of breakpoints must have the colour white, as L is positively block pumpable; by being monochromatic, all of these have this property. Furthermore, one of the pairs $i, j \in J$ must satisfy $P^0(x : i, j) \in L$ due to L being block cancellable. It follows that $P^*(x : i, j) \subseteq L$ for this pair and L is block pumpable with the constant p'. $\square$

The following two results dealing with the periodicity of words are useful in various contexts. A word is called *k-periodic* iff for each i, j with $i - j$ being a multiple of k, the symbols at positions i and j in the word are the same.

Theorem 12 (Fine and Wilf [6]). *If a word v is h-periodic and k-periodic and ℓ is the greatest common divisor of h and k and $|v| \geq k + h - \ell$ then v is ℓ-periodic.*

If two words v, w and the greatest common divisor ℓ of $|v|$ and $|w|$ satisfy for some $n, m > 1$ that v^n and w^m have a common prefix of length at least $|v| + |w| - \ell$ then $v, w \in u^*$ for some word u of length ℓ.

Theorem 13 (Lyndon and Schützenberger [15]). *If two words v, w satisfy $vw = wv$ then there is a word u with $v, w \in u^*$.*

A famous language is the language of primitive words, that is, of all words w which are not of the form v^n for any $n > 1$. The open problem whether this language is context-free is subject to intensive investigations; Dömösi and Ito [4] dedicated a whole monograph to the state of the art of this problem. The next result shows that the language of primitive words is also a very natural example of a block cancellable language.

Example 14. The language L of primitive words over an alphabet containing at least 0 and 1 is block cancellable but not positively block pumpable.

Proof. First it will be shown that the language is block cancellable with constant 3^{10}. Given a word $x \in L$ and a set of 3^{10} breakpoints on x, one can find a subset I of 3^8 breakpoints such that $|P^0(x : i, j) \geq 5/6 \cdot |x|$ for all $i, j \in I$; this is due to the fact that there are two breakpoints which have only the distance $|x|/6$ in the word with 3^8 many other breakpoints in between. The set I can be chosen as these 3^8 breakpoints. Now one colours each pair i, j of breakpoints in I as follows:

- the pair i, j is white iff $P^0(x : i, j)$ is primitive;
- the pair i, j is red iff $P^0(x : i, j)$ is of the form zz of some primitive word z;
- the pair i, j is blue iff $P^0(x : i, j)$ is of the form z^n for some word z and some $n \geq 3$.

In the case that there are no white pairs, one can select by Ramsey's Theorem of Pairs a monochromatic subset J of four breakpoints. Let J split x into $x_0(1)x_1(2)x_2(3)x_3(3)x_4$; note that by choice of I it holds that $|x_1x_2x_3| \leq |x_0x_4|/5$. Furthermore, note that for every n and every words v, w it holds that vw is of the form u^n for some u iff the rotation wv of vw is of the form $\tilde{u}^n$ for some word $\tilde{u}$. Therefore one can without loss of generality consider the rotated word $y = x_4x_0(1)x_1(2)x_2(3)x_3(4)$ in place of x; the colouring of the pairs of breakpoints does, as indicated, not change by this; the names $1, 2, 3, 4$ of the breakpoints are kept for notational convenience. Note that $|x_1| \neq |x_2x_3|$ or $|x_1x_2| \neq |x_3|$ due to x_2 containing at least one symbol, without loss of generality $|x_1| \neq |x_2x_3|$.

If the colour of the pairs in J are all *blue* then every of the words $P^0(y : i, j)$ is of the form u^n for some $n \geq 3$. Now one considers the words x_4x_0, $x_4x_0x_1$ and $x_4x_0x_2x_3$. Let u be the shortest word such that $x_4x_0 = u^n$ for some $n \geq 3$. By the Theorem of Fine and Wilf, the words $x_4x_0x_1$ and $x_4x_0x_2x_3$ are also in u^+: their periodicity is at most $|x_0x_4| \cdot 6/5 \cdot 1/3 \leq |x_0x_4| \cdot 2/5$ and the sum of the periodicity $|u|$ of x_4x_0 and the one of the longer word is, in both cases, at most $|x_0x_4| \cdot 11/15$ which is shorter than the common prefix x_0x_4 of all three words. Note that the periodicity is at least $|u|$ by the choice of u and at most the greatest common divisor of $|u|$ and the periodicity of the other words, hence all three words involved are $|u|$-periodic and in u^+. Thus $x_1, x_2x_3 \in u^+$. It follows that $y = x_4x_0x_1x_2x_3 = u^m$ for some $m \geq 5$ and this contradicts the fact that x and y are both primitive words. Thus this case does not occur.

If the colour of the pairs in J are all red then x_4x_0 and $x_4x_0x_1$ and $x_4x_0x_2x_3$ are of the form uu, $uvuv$ and $uwuw$ for some words u, v, w. As u is a prefix of vu and wu and as $|v|, |w| \leq |u|/5$, one has that u is periodic and of the form $z^n\tilde{z}$ for some word z and proper prefix $\tilde{z}$ of z; without loss of generality, z is taken as short as possible. Note that $n \geq 4$. The Theorem of Fine and Wilf gives that there are numbers m_1, m_2 with $v = z^{m_1}$ and $w = z^{m_2}$. Let $\hat{z}$ denote the unique word with $\tilde{z}\hat{z} = z$. One can show that $x_1 = \hat{z}z^{m_1-1}\tilde{z}z^{m_1}$ and $x_2x_3 = \hat{z}z^{m_2-1}\tilde{z}z^{m_2}$; furthermore, by similar arguments as before for the word $x_4x_0x_1x_2$, one can conclude that there is an $m_3 > m_1$ with $x_1x_2 = \hat{z}z^{m_3-1}\tilde{z}z^{m_3}$. As x_1 is a prefix of x_1x_2, it follows that the prefix of length $|z| \cdot (m_1+1)$ of the two words has the forms $\hat{z}z^{m_1-1}\tilde{z}z$ and $\hat{z}z^{m_1}\tilde{z}$ at the same time; in particular $\tilde{z}z = z\tilde{z}$ for the suffix of length $|\tilde{z}z|$ of this word. By the Theorem of Lyndon and Schützenberger [15], either

$\tilde{z}$ is the empty word or z is not primitive, the latter contradicts to the choice of z. So $\tilde{z}$ is the empty word and $u, v, w \in z^+$. It follows that $y = x_4 x_0 x_1 x_2 x_3 = uvw$ is a proper power of z and not primitive; thus also x is not primitive as well. So also this case cannot occur.

Thus there must be a white pair i, j of breakpoints in I and $P^0(x : i, j)$ is a primitive words and in L. It follows that the language of the primitive words is block cancellable with constant 3^{10}.

Second one assumes by way of contradiction that L is positively block pumpable with constant $p \geq 2$. Now the word $0^p 1 0^{p+p!} 1$ is in L, as it is primitive. However, by Proposition 6, also the square $0^{p+p!} 1 0^{p+p!} 1$ would be in L, in contradiction to the choice of L. Hence L cannot be positively block pumpable. $\square$

This proof is not optimised for getting the best constant p. Indeed, one can easily improve it by first using that in the case of a J with all pairs being blue it is enough to have 3 breakpoints in J and secondly use the optimal bounds for Ramsey's Theorem of Pairs. These are that one needs just 9 breakpoints in I to get, in the case that all pairs are coloured red or blue, either a set J of three elements with all pairs coloured blue or a set J of four elements with all pairs coloured in red. Furthermore, given an x, one can cut it into six parts of same length separated by one symbol ($x = v_0 a_1 v_1 a_2 v_2 a_3 v_3 a_4 v_4 a_5 v_5$) such that each breakpoint is inside or neighbouring one v_k and $|v_k| \leq |x|/6$ for all k; having this one can choose I to be the set of breakpoints neighbouring or inside that v_k where there are 9 breakpoints of this type. This can always be achieved if there are at least 49 breakpoints in total. Thus a better upper bound for the constant p is 49; a lower bound is strictly above 3 as witnessed by the three breakpoints in the primitive word 010010(1)1001(2)1001(3): the block cancellations give the words 010 010 and 01001 01001 (twice) which are all not primitive.

3 Closure Properties of Block Pumpable Languages

Regular languages are known for their closure under union, intersection, complementation, concatenation and Kleene star. As there are block pumpable languages which are not regular, it follows from the Theorem of Ehrenfeucht, Parikh and Rozenberg that these are not closed under complement. The next result shows that they nevertheless have various interesting closure properties.

Theorem 15. *Let L, H be block pumpable languages and f be function computed by a transducer. Then the following languages are also block pumpable: 1. $L \cup H$; 2. $L \cap H$; 3. $L \cdot H$; 4. $f(L)$. The same closure properties hold for block cancellable and positively block pumpable languages.*

Proof. Let p be a joint block pumping constant for L and H — one can take the maximum of the two constants if they are different.

1. It is directly clear that $L \cup H$ is also block pumpable with block pumping constant p: If $x \in L$ then every set I of p breakpoints of x contains a pair i, j of breakpoints with $P^*(x : i, j) \subseteq L$; if $x \in H$ then every set I of p breakpoints of x contains a pair i, j of breakpoints with

$P^*(x : i, j) \subseteq H$; in both cases $P^*(x : i, j) \subseteq L \cup H$.

2. By Proposition 10, there is a constant p' such that for every word $x \in L \cap H$ and every set I of p' distinct breakpoints in x there is a subset J of p breakpoints such that $P^*(x : i, j) \subseteq L$ for all $i, j \in J$. Furthermore, as J contains p breakpoints, also one pair of the $i, j \in J$ satisfies that $P^*(x : i, j) \subseteq H$. Thus $P^*(x : i, j) \subseteq L \cap H$ for this pair of $i, j \in J$.

3. Assume that $x \in L$ and $y \in H$. If I is a set of $2p - 1$ breakpoints for xy then one can find a subset J of p breakpoints which lies completely either in the x part or the y part. It follows that there is a pair i, j of breakpoints which satisfies $P^*(xy : i, j) \subseteq L \cdot H$.

4. Let p be the maximum of the block pumping constant of L and the number of states of the non-deterministic transducer computing f. Without loss of generality, the transducer on any transition reads at most one symbol from x and writes at most one symbol from y. Now it is shown that $f(L)$ is block pumpable with block pumping constant p^2; note that this bound requires the transducer to be of the form which processes one symbol per cycle. Given a word $y \in f(L)$ which is the image of some $x \in L$ and given a set I of p^2 breakpoints in y, one can find a corresponding set $\{t_i : i \in I\}$ of breakpoints in x such that there is a non-deterministic run of the transduction of x into y for which each t_i is the position processed by the run in x at the first time point when y is processed up to i; furthermore, let s_i be the state of the automaton when having reached the breakpoints i and t_i in y and x, respectively. Now there is one state s such that $s_i = s$ for at least p breakpoints $i \in I$; so let $J = \{i \in I : s_i = s\}$. Either there are now two $i, j \in J$ with $t_i = t_j$ or, as all t_i, t_j are distinct for pairs of breakpoints in J and L is block pumpable with constant p, there are $i, j \in J$ such that $P^*(x : t_i, t_j) \subseteq L$. In both cases, when the transducer reaches t_i in x and i in y, then it can omit or arbitrarily repeat the part between the breakpoints and afterwards complete the transduction with processing the parts in x after t_j and in y after the breakpoint j. Hence the transducer translates every word $P^k(x : t_i, t_j)$ into the word $P^k(y : i, j)$ and as $P^k(x : t_i, t_j) \in L$, $P^*(y : i, j) \subseteq f(L)$. Thus $f(L)$ is block pumpable. $\square$

The above closure properties also hold for block cancellable and positively block pumpable languages with minor modifications of the proofs — one only has to adjust the types of pumps ($P^*(x : i, j)$ versus $P^0(x : i, j)$ and $P^+(x : i, j)$). However, weakly block pumpable languages are not closed under intersection, for example the set of all words with equally many 0 and 1 as well as 0^*1^* are weakly block pumpable while their intersection $\{0^n1^n : n \in \mathbb{N}\}$ is not. Weakly block pumpable sets satisfy that L^* is weakly block pumpable for every language L; in contrast to this, the other notions are not even closed under Kleene star, as the next result shows.

Theorem 16. *There is a block pumpable language L such that L^* is neither block cancellable nor positively block pumpable.*

Proof. Let $L \subseteq \{0, 1, 2\}^*$ contain all words of the form $2v2w$ such that $v, w \in \{0, 1\}^*$ and at least one of the following conditions hold:

- $v = w$;
- $|v| \neq |w|$;
- v or w contains a cube.

Now it will be shown that L is block pumpable with block pumping constant 6 and that L^* is neither block cancellable nor positively block pumpable.

Let $x \in L$ be given together with six breakpoints. Then at least 3 of the breakpoints are either all inside the v-part (including being between v and a 2) or all inside the w-part (including being between w and the preceding 2). Without loss of generality assume that they are in the v-part, the other case is analogous. Then one has the choice between at least two possible lengths a and b of the pump, therefore one can choose the breakpoints i, j such that in the pumped words $P^0(x : i, j)$ and $P^2(x : i, j)$ the length of the part belonging to v is either $|v| - a$ and $|v| + a$ or $|v| - b$ and $|v| + b$; for one choice, either both $|v| - a, |v| + a$ or both $|v| - b, |v| + b$ are different from $|w|$. Hence one can choose a pair i, j of breakpoints such that $P^0(x : i, j)$ and $P^2(x : i, j)$ are both in L by the second condition. Furthermore, for $k \geq 3$, $P^k(x : i, j)$ contains a cube in the v-part and is in L by the third condition. Thus $P^*(x : i, j) \subseteq L$.

Now L^* contains ε and all words of the form $2v_0 2w_0 2v_1 2w_1 \ldots 2v_n 2w_n$ where for every k either $v_k = w_k$ or $|v_k| \neq |w_k|$ or at least one of v_k, w_k contains a cube. Assume by way of contradiction that L^* would be block pumpable with constant $p \geq 2$. Now one can choose $n = p + 1$ and a length m so large that there are $n + 1$ different cube-free binary words $v_0, v_1, \ldots, v_n$ of length m; let $w_k = v_k$ for all k and $x = 2v_0 2w_0 2v_1 2w_1 \ldots 2v_n 2w_n$ and put the breakpoint number i between $2v_i$ and $2w_i$ — here again the breakpoints are identified with numbers for notational convenience. The word x is in L as $v_k = w_k$ for all k. For any pair of breakpoints i, j, the word $P^0(x : i, j)$ is of the form $2v_0 2w_0 \ldots 2v_{i-1} 2w_{i-1} 2v_i 2w_j 2v_{j+1} 2w_{j+1} \ldots 2v_n 2w_n$ and fails to be in L^*, as $2v_i 2w_j \notin L$; furthermore, the word $P^2(x : i, j)$ $2v_0 2w_0 \ldots 2v_{j-1} 2w_{j-1} 2v_j 2w_i 2v_{i+1} 2w_{i+1} \ldots 2v_n 2w_n$ and fails to be in L^*, as $2v_j 2w_i \notin L$; thus L^* is neither block cancellable nor positively block pumpable. $\square$

4 Multiple Pumping

It is well known that for regular languages L, one can have several independent pumps. For example, if an NFA has p states and a word x is split along $2p + 2$ breakpoints then there is a state visited thrice at certain breakpoints i, j, k; hence the word is twice block pumpable at these breakpoints. The next theorem establishes an analogue of this fact for block pumpable languages.

Theorem 17. *If L is block pumpable there are constants $p, p' \geq 2$ such that for all x, y with $xy \in L$ and all sets I of p breakpoints of x and I' of p' breakpoints of y there are pairs of breakpoints $i, j \in I$ and $i', j' \in I'$ such that $P^*(x : i, j) \cdot P^*(y : i', j') \subseteq L$.*

Proof. Let p' be the pumping constant for L and let $q = p' + p'! + 1$. By Proposition 10, part 2, there is a constant p so large that for every word x and every set Y of q words, if $x \cdot Y \subseteq L$ then for every set I of p breakpoints there is a pair $i, j \in I$ with $P^*(x : i, j) \cdot Y \subseteq L$.

Now, given x, y with $xy \in L$ and a set I of p breakpoints in x and I' of p' breakpoints in y, consider just the I' breakpoints in the part y of xy: As L is block pumpable with constant p', there are two breakpoints $i', j' \in I'$ such that $x \cdot P^*(y : i', j') \subseteq L$. Now let $Y = \{P^k(y : i', j') : k < q\}$. Clearly $x \cdot Y \subseteq L$ and $|Y| \leq q$. By the first paragraph, there is a pair of breakpoints $i, j \in I$

such that $P^*(x : i, j) \cdot Y \subseteq L$.

Consider any fixed k. Now $P^k(x : i, j) \cdot Y \subseteq L$ means that for all $k' < p' + p'!$ it holds that $P^k(x : i, j) \cdot P^{k'}(y : i', j') \in L$; by Proposition 6 this implies that all k' satisfy $P^k(x : i, j) \cdot P^{k'}(y : i', j') \in L$. Hence the inclusion $P^*(x : i, j) \cdot P^*(y : i', j') \subseteq L$ is satisfied and the double pumping property is proven. $\square$

Corollary 18. *If L is block pumpable there are constants p, p', p'' such that for all x, y, z with $xyz \in L$ and all sets I of p breakpoints of x and I' of p' breakpoints of y and I'' of p'' breakpoints of z there are pairs of breakpoints $i, j \in I$ and $i', j' \in I'$ and $i'', j'' \in I''$ such that $P^*(x : i, j) \cdot P^*(y : i', j') \cdot P^*(z : i'', y'') \subseteq L$.*

Similarly one can define the notion of k times block pumpable languages and show that it coincides with the block pumpable languages. All these results transfer to positively block pumpable and block cancellable languages, as the corresponding arguments only need minor adjustments in the choice of the colourings and the handling of the pumping conditions.

However, weakly block pumpable languages behave differently. For example the language of all words with the same number of 0 and 1 is weakly block pumpable but not twice weakly block pumpable. The language

$$\{w \in \{0, 1, 2, 3\}^* : w \in \{0\} \cdot \{1, 2, 3\}^* \cdot \{0\} \Rightarrow w \text{ is square-free}\}$$

is twice weakly block pumpable but not thrice weakly block pumpable. The thrice weakly block pumpable languages coincide with the block pumpable languages. Thus one has here a hierarchy of the once, twice and thrice weak block pumpable languages; for each $k, h \geq 3$, the classes of the k times weakly block pumpable languages and of the h times weakly block pumpable languages coincide.

5 Growth

One of the questions investigated on regular languages is the amount of words of length up to n. So one says that a language L has *polynomial growth* iff there is a polynomial p such that, for all n, L has at most $p(n)$ many words up to length n. A language has *exponential growth* iff there are constants a, b such that, for all n, L has at least 2^n words of length up to $an + b$. Regular languages have either polynomial or exponential growth. Bridson and Gilman [1] as well as Incitti [9] showed this holds even for all context-free languages. This result transfers to block pumpable languages and positively block pumpable languages, the main reason is that all nonregular positively block pumpable languages have exponential growth. This explains why all the examples for nonregular block pumpable languages provided so far have exponential growth.

Theorem 19. *Let L be a nonregular positively block pumpable language. Then L has exponential growth, that is, there are $a, b \in \mathbb{N}$ such that, for every n, there are at least 2^n words of length up to $an + b$ in L.*

Proof. Let $p \geq 2$ be the constant for which L is positively block pumpable. The proof distinguishes two cases such that in Case (a) the language L has exponential growth and in Case (b) the language L is regular.

Case (a). There are words u, v_1, v_2, v_3, v_4, w such that $|v_1| \geq p$, $|v_2| = |v_3| = p!$, $|v_4| \geq p$, $v_2 \neq v_3$ and $u(v_1v_2v_3v_4)^+w \subseteq L$.

Now let $c_0c_1c_2 \ldots c_n$ be any binary word and consider $x_{n+1} = u(v_1v_2v_3v_4)^{n+1}w$. Now one does, inductively, for $m = n, n-1, n-2, \ldots, 2, 1$ the following at the $m+1$-st occurrence of $v_1v_2v_3v_4$ in x_{m+1}: If $c_m = 0$ then one makes by pumping v_4 $p!$ symbols longer else one makes v_1 $p!$ symbols longer, see Proposition 6; the resulting word is called x_m and one can see by induction that $x_m \in L$.

After the words $x_n, x_{n-1}, \ldots, x_1, x_0$ have been made this way, the word x_0 has the length $|uw| + (n+1)(|v_1v_2v_3v_4| + p!)$. Furthermore, the subword of length $p!$ starting from position $|u| + |v_1| + p! + m(|v_1v_2v_3v_4| + p!)$ is either v_3 (in the case that $c_m = 0$) or v_2 (in the case that $c_m = 1$). As $v_2 \neq v_3$, there are 2^{n+1} many words of length $an + b$ where $a = |uv_1v_2v_3v_4w|$ and $b = |v_1v_2v_3v_4| + p!$. So L is a language of exponential growth. Here sample words with $p = 2$, $p! = 2$, $u = 2015$, $v_1 = 22$, $v_2 = 11$, $v_3 = 00$, $v_4 = 33$, $w = 221$ for three possible values of $c_0c_1c_2c_3$:

- Word x_0 for `0011` is `2015 2211003333 2211003333 2222110033 2222110033 221`;
- Word x_0 for `1001` is `2015 2222110033 2211003333 2211003333 2222110033 221`;
- Word x_0 for `0100` is `2015 2211003333 2222110033 2211003333 2211003333 221`.

If $c_m = 0$ then the symbols in x_0 at positions $10m+8, 10m+9$ are 00; if $c_m = 1$ then the symbols in x_0 at positions $10m+8, 10m+9$ are 11.

Case (b). There are no words u, v_1, v_2, v_3, v_4, w such that $|v_1| \geq p$, $|v_2| = |v_3| = p!$, $|v_4| \geq p$, $v_2 \neq v_3$ and $u(v_1v_2v_3v_4)^+w \subseteq L$.

Now assume by way of contradiction that there are $4p$ words $x_1, x_2, \ldots, x_{4p+1}$ of length $p!$, exponents $c_1, c_2, \ldots, c_{4p+1} \geq 1$ and a word x_{4p+2} such that $x_{4m-2} \neq x_{4m-1}$ for all $m \in \{1, 2, \ldots, p\}$ and $x_1^{c_1}x_2^{c_2} \ldots x_{2p+1}^{c_{2p+1}}x_{2p+2} \in L$. Now one can put every breakpoint $i \in \{1, 2, \ldots, p\}$ immediately after $x_1^{c_1}x_2^{c_2} \ldots x_{4i}^{c_{4i}}$ for $i = 1, 2, \ldots, p$ and it follows that there is a pair of breakpoints $i, j \in \{1, \ldots, p\}$ with

$$P^+(x_1^{c_1}x_2^{c_2} \ldots x_{2p}^{c_{2p}}x_{2p+1} : i, j) \subseteq L.$$

Now one lets $u = x_1^{c_1} \ldots x_{4i}^{c_{4i}}$, $v_1 = x_{4i+1}^{c_{4i+1}}x_{4i+2}^{c_{4i+2}-1}$, $v_2 = x_{4i+2}$, $v_3 = x_{4i+3}$, $v_4 = x_{4i+3}^{c_{4i+3}-1}x_{4i+4}^{c_{4i+4}} \ldots x_{4j}^{c_{4j}}$ and $w = x_{4j+1}^{c_{4j+1}} \ldots x_{4p+1}^{c_{4p+1}}x_{4p+2}$. By choice, $u(v_1v_2v_3v_4)^+w \subseteq L$. It is easy to see that the words u, v_1, v_2, v_3, v_4, w satisfy the length constraints from Case (a) and Case (a) would hold in place of Case (b). Thus every word in L is of the form $x_1^{c_1}x_2^{c_2} \ldots x_k^{c_k}x_{k+1}$ for some words $x_1, x_2, \ldots, x_k, x_{k+1}$ and exponents $c_1, c_2, \ldots, c_k \geq 1$ with $k < 4p+2$, $|x_1| = |x_2| = \ldots = |x_k| = p!$, $|x_{k+1}| < p!$ and $x_1 \neq x_2$, $x_2 \neq x_3$, $\ldots$, $x_{k-1} \neq x_k$. There are only finitely many choices for k and $x_1, x_2, \ldots, x_k, x_{k+1}$; it will be shown that, for any fixed choice of $k, x_1, x_2, \ldots, x_k, x_{k+1}$, the set

$$H = \{y \in L : \exists c_1, \ldots, c_k \geq 1\, [y = x_1^{c_1}x_2^{c_2} \ldots x_k^{c_k}x_{k+1}]\}$$

is regular; hence also L as the union of finitely many regular sets will be regular. Thus, from now on, one considers k and $x_1, x_2, \ldots, x_k, x_{k+1}$ as fixed. The goal for the remaining part of the proof is to show that H is the union of finitely many sets of the form

$$H_{d_1,d_2,\ldots,d_k} = \{x_1^{c_1} x_2^{c_2} \ldots x_k^{c_k} x_{k+1} : \forall h \in \{1,2,\ldots,k\} : c_h = d_h \vee d_h \geq p \wedge \exists \ell\, [c_h = d_h + \ell \cdot p!]\}.$$

Note that whenever $x_1^{d_1} x_2^{d_2} \ldots x_k^{d_k} x_{k+1} \in L$ then one can prove by induction similar to the second item of Proposition 6 that $H_{d_1,d_2,\ldots,d_k} \subseteq L$. Thus the inclusion $H_{d_1,d_2,\ldots,d_k} \subseteq H$ is also true.

One defines $H_{d_1,d_2,\ldots,d_k} \equiv H_{e_1,e_2,\ldots,e_k}$ iff for all h, either $d_h = e_h$ or $d_n \geq p$ and $e_h \geq p$ and $p!$ divides $d_h - e_h$. There are $(p + p!)^k$ many equivalence classes of $\equiv$. For every pair of equivalent sets $H_{d_1,d_2,\ldots,d_k}, H_{e_1,e_2,\ldots,e_k}$, one has that $H_{d_1,d_2,\ldots,d_k} \subseteq H_{e_1,e_2,\ldots,e_k}$ iff $d_1 \leq e_1 \wedge d_2 \leq e_2 \wedge \ldots \wedge d_k \leq e_k$.

One can apply Dickson's Lemma [3] to any union of sets $H_{e_1,e_2,\ldots,e_k}$ which are all equivalent to one fixed $H_{d_1,d_2,\ldots,d_k}$: Let

$$I_{d_1,d_2,\ldots,d_k} = \{H_{e_1,e_2,\ldots,e_k} : H_{e_1,e_2,\ldots,e_k} \subseteq H \text{ and } H_{e_1,e_2,\ldots,e_k} \equiv H_{d_1,d_2,\ldots,d_k}\};$$

now Dickson's Lemma [3, 13] implies that there is a finite subset $J_{d_1,d_2,\ldots,d_k} \subseteq I_{d_1,d_2,\ldots,d_k}$ with $\bigcup J_{d_1,d_2,\ldots,d_k} = \bigcup I_{d_1,d_2,\ldots,d_k}$. Since there are only finitely many equivalence classes of $\equiv$, H is the union of finitely many $\bigcup J_{d_1,d_2,\ldots,d_k}$ and thus H is the union of finitely many $H_{e_1,e_2,\ldots,e_k}$.

Each set $H_{e_1,e_2,\ldots,e_k}$ is regular and thus H is regular; as L is the union of finitely many sets of the form H, also L is regular. $\square$

One should also note that in the previous theorem in Case (b), every set $H_{e_1,e_2,\ldots,e_k}$ has polynomial growth and thus, L is the union of finitely many regular sets of polynomial growth and therefore L is also a regular set of polynomial growth. This theorem together with the closure of block pumpable languages under intersection give the following corollary.

Corollary 20. *If L is positively block pumpable and H is a regular language of polynomial growth then $L \cap H$ is regular.*

It is also interesting to study the relation between block pumpable languages and context-free languages in this context. Corollary 20 needs that one replaced the words "regular" by "context-free" in both places, otherwise the intersection of the block pumpable set $\{0,1\}^*$ with the context-free set $\{0^n 1^n : n > 0\}$ of polynomial growth would be a counterexample. So let L be positively block pumpable and H be a context-free set of polynomial growth. Bridson and Gilman [1, Theorem 1.1] showed that H is a subset of a set of the form $w_1^* \cdot w_2^* \cdot \ldots \cdot w_n^*$ for some words $w_1, w_2, \ldots, w_n$. This set has polynomial growth and therefore $L \cap w_1^* \cdot w_2^* \cdot \ldots \cdot w_n^*$ is a regular set. Now $L \cap H = (L \cap w_1^* \cdot w_2^* \cdot \ldots \cdot w_n^*) \cap H$ is the intersection of a regular and context-free set and thus context-free. This establishes the following result.

Corollary 21. *If L is positively block pumpable and H is a context-free language of polynomial growth then $L \cap H$ is context-free.*

Note that this result does not hold for block cancellable languages. Indeed, Example 7 provides a language which for the alphabet $\{0\}$ has polynomial (indeed linear) growth and is block cancellable but is not regular; this language is also weakly block pumpable. Furthermore, the language $\{0^n 1^m : n \neq m\}$ has quadratic growth, is context-free, is not regular and is block cancellable.

6 Block Pumpable Structures and Their Theory

Decidability questions, that is, whether certain questions about languages or structures can be decided by algorithms, lie in the heart of automata theory [10]. A well-known result of automatic structures is that their first-order theory is decidable [7, 8, 12]. Therefore it is natural to ask to which extent one can carry over the results on the decidability of automatic structures to structures which share two weaker aspects of automatic structures, namely, that their domain and relations are firstly block pumpable and secondly recursive. Here note that the domain is always a set of words over a finite alphabet. This approach will now be made more formal.

As usual in the field of automatic structures, one needs the notion of convolution in order to speak about tuples of words. One defines the convolution of finitely many words $u, v, w, \ldots, z$ as a word of all tuples $(u_1, v_1, w_1, \ldots, z_1)\,(u_2, v_2, w_2, \ldots, z_2)\ \ldots\ (u_n, v_n, w_n, \ldots, z_n)$ where n is the maximum length of the words $u, v, w, \ldots, z$ and for each of these words, say for u, the item u_m is the m-th symbol of u in the case that the length of u is at least m and u_m is a special symbol, say $\#$, outside the alphabet of the words in the case that u is shorter than m. So the convolution of 110, 0011221100 and 012101210 is

$$(1,0,0)\,(1,0,1)\,(0,1,2)\,(\#,1,1)\,(\#,2,0)\,(\#,2,1)\,(\#,1,2)\,(\#,1,1)\,(\#,0,0)\,(\#,0,\#).$$

Now one says that a k-ary relation R on a set A of words over a finite alphabet is block pumpable iff

$$\mathit{conv}(R) = \{\mathit{conv}(x_1, x_2, \ldots, x_k) : x_1, x_2, \ldots, x_k \in A \wedge R(x_1, x_2, \ldots, x_k) \text{ holds}\}$$

is block pumpable as a set on the larger alphabet used by the convolution. A function is block pumpable iff its graph — viewed as a relation — is block pumpable. These definitions are analogous to the case of automatic structures where one defines that a relation is automatic iff the set of the convoluted tuples satisfying the relation is regular. Furthermore, one says that a relation R is effective block pumpable iff $\mathit{conv}(R)$ is besides being block pumpable also a recursive set (also known as a decidable set). This permits to give the following definition summarising also the previous ones.

Definition 22. A structure $(A, R_1, R_2, \ldots, R_k)$ is called effectively block pumpable iff the domain A is a recursive and block pumpable set of words over a finite alphabet and for each relation R_h and its arity n, $\mathit{conv}(R_h)$ is a block pumpable and recursive subset of the n-fold convolution $\{\mathit{conv}(w_1, w_2, \ldots, w_n) : w_1, w_2, \ldots, w_n \in A\}$ of A.

First note that every automatic structure is effectively block pumpable: A and for every of its relations R the set $\mathit{conv}(R)$ are block pumpable; furthermore, every automatic relation is also recursive.

Second, here an example how one can use automatic relations restricted to an effectively block pumpable domain in order to make effectively block pumpable structures. One can take $A = \{w \in \{0,1\}^* : w$ is cube-free or w does not have a length which is a power of $16\}$ and $<_{lex}$ as the lexicographical order (which is automatic). Now the corresponding relation on $A \times A$ is given as $\{\mathit{conv}(v, w) : v, w \in A \wedge v <_{lex} w\}$ and is block pumpable as it is the intersection of three block pumpable sets: $\{conv(v,w) : v,w \in \{0,1\}^* \wedge v <_{lex} w\}$, $\{conv(v,w) : v,w \in \{0,1\}^* \wedge v \in A\}$ and $\{conv(v,w) : v,w \in \{0,1\}^* \wedge w \in A\}$. It is also clear that A and the relation R are both recursive, so (A,R) is an effectively block pumpable structure.

Theorem 23 below will show that some fragment of the existential first-order theory of effectively block pumpable structures is decidable. The reader should be aware of the fact that "decidability" only requires that "there is an algorithm" but does not say how to make the algorithm. Therefore, when making the algorithm in the proof, one can use other algorithms like the decision algorithms for the various sets and relations involved and also constants like an upper bound on the block pumping constants of all the basic relations in the effectively block pumpable structure. The next result will stand in contrast to Theorem 24 below which will provide an effectively block pumpable structure with an undecidable first-order theory; such a structure is not isomorphic to any automatic structure, as automatic structures always have a decidable first-order theory.

Theorem 23. *If $(A, R_1, R_2, \ldots, R_k)$ is an effective block pumpable structure then the set of all existentially quantified formulas over disjunctions of conjunctions of the relations is decidable — here the entities in the relations can either be fixed members of the domain A (given as words over the finite alphabet) or the variables quantified over.*

Proof. Let ϕ be an existential formula of the required form, say ϕ is $\exists x_1, x_2, x_3\, [R_1(x_1,x_2) \wedge R_1(x_1,x_3) \wedge R_2(x_2,1100) \vee R_1(x_2,x_3) \wedge R_2(x_1,x_1)]$ where $\wedge$ binds stronger than $\vee$. Now one considers all the sets which are obtained by applying exactly one relation symbol, in the current example these are the following sets:

$$\begin{aligned}
B_{1,1} &= \{conv(x_1,x_2,x_3) : x_1,x_2,x_3 \in A \wedge R_1(x_1,x_2)\};\\
B_{1,2} &= \{conv(x_1,x_2,x_3) : x_1,x_2,x_3 \in A \wedge R_1(x_1,x_3)\};\\
B_{1,3} &= \{conv(x_1,x_2,x_3) : x_1,x_2,x_3 \in A \wedge R_2(x_2,1100)\};\\
B_{2,1} &= \{conv(x_1,x_2,x_3) : x_1,x_2,x_3 \in A \wedge R_1(x_2,x_3)\};\\
B_{2,2} &= \{conv(x_1,x_2,x_3) : x_1,x_2,x_3 \in A \wedge R_2(x_1,x_1)\}.
\end{aligned}$$

So every set of this type contains one relation over members of A with some entries of the relation referring to components of the tuples in the set and others to constants. These sets are effective block pumpable and that one can compute an upper bound on the pumping constant, provided that an upper bound p on all pumping constants for A and $R_1, R_2, \ldots, R_k$ is known. There are three issues which have to be taken care off.

1. Unused components: In $B_{1,1}$ the third component x_3 does not show up in the relation. Hence one has to use that $B_{1,1}$ can be seen as the intersection of the sets

   $$\{conv(x_1,x_2,x_3) : x_1,x_2 \in A \wedge R_1(x_1,x_2) \wedge x_3 \in \Sigma^*\}$$

   and

   $$\{conv(x_1,x_2,x_3) : x_1,x_2 \in \Sigma^* \wedge x_3 \in A\}$$

where Σ is the underlying alphabet and adding components from Σ^* does not increase the pumping constant. So both sets are block pumpable with the constant p (or less) and they are also recursive. The pumping constant of the intersection can be computed by following the proof of Theorem 15 and using known bounds for the finite Ramsey Theorem of pairs. Furthermore, it is obvious that the intersection of recursive sets is recursive.

2. Multiple use of entries: The set $B_{2,2}$ has a relation which uses the entry twice. When pumping any entry of the form $conv(x_1, x_1)$, the resulting pumped word is again of this form. Hence one can conclude that the sets
   - $\{conv(x_1, x_1) : x_1 \in A \wedge R_2(x_1, x_1)\}$,
   - $\{x_1 : x_1 \in A \wedge R_2(x_1, x_1)\}$ and
   - $\{conv(x_1, x_2, x_3) : x_1 \in A \wedge R_2(x_1, x_1) \wedge x_2, x_3 \in \Sigma^*\}$

   are all block pumpable with block pumping constant p. Now, as in the previous case of $B_{1,1}$, one can form the intersection of the last of these three sets with sets just requiring that the components x_2 and x_3, respectively, are in A in order to obtain the set $B_{1,1}$; the set $B_{1,1}$ is also effective block pumpable and one can compute an upper bound on the block pumping constant.
3. Constant parameters in the relation: The set $B_{1,3}$ has a constant parameter in the relation. Given a constant $a_1 a_2 \ldots a_q \in A$ of length q, the set

$$\{conv(x_2, a_1 a_2 \ldots a_q) : x_2 \in A \wedge R(x_2, a_1 a_2 \ldots a_q)\}$$

   has then the block pumping constant $p+q$. The reason is that if one considers $p+q$ breakpoints in a convolution of an $x_2 \in A$ and $a_1 a_2 \ldots a_q$, then p of these breakpoints are at the end of the constant or behind and then one can find a pump using two of the last p breakpoints; this pump does not change the constant in the expression. Once having this constant, one can then conclude that also

$$\{conv(x_1, x_2, x_3) : x_1, x_2, x_3 \in A \wedge R_2(x_2, 1100)\}$$

   is block pumpable and one can compute a block pumping constant for it. Furthermore, also sets of this type are recursive.

The use of the three principles above permits to compute for each primitive set $B_{1,1}, B_{1,2}, B_{1,3}, B_{2,1}, B_{2,2}$ a block pumping constant and it is also clear from the above considerations that these sets are recursive and thus effective block pumpable. Now the intersections

$$B_1 = B_{1,1} \cap B_{1,2} \cap B_{1,3} \text{ and } B_2 = B_{2,1} \cap B_{2,2}$$

are also effective block pumpable and one can compute the corresponding block pumping constants, let r be their maximum. Now ϕ is true iff there is a tuple in $B = B_1 \cup B_2$ which is in one of the sets B_1 and B_2. Using the fact that one can pump down such a tuple while keeping it in the set until it has at most the length r, one can conclude that ϕ holds iff B_1 or B_2 have a member of length up to r; as both sets B_1 and B_2 are recursive and as one can obtain the decision procedure from decision procedures of the domain and the relations, one has an algorithm to decide whether ϕ is true. Hence the positive existential theory of the given structure is decidable. $\square$

While the positive existential theory of effective block pumpable structures is decidable, this is no longer true if one permits to use the positive existential universal theory where one is permitted to use universal quantifiers after the existential ones. Here again one is permitted to use arbitrary elements of the structure as parameters.

Theorem 24. *There is an effective block pumpable structure (A, R_1, R_2) such that its positive existential universal theory is undecidable.*

Proof. The structure (A, R_1, R_2) and the formulas in question are defined as follows (where x_0, x_1, x_2 are variables used in the first-order logic of the structure and a is a parameter from A):

- $\phi(a)$ is the positive $\exists\forall$-formula $\phi(a) = \exists x_1 \in A\, \forall x_2 \in A\; [R_1(a, x_1) \wedge R_2(x_2, x_1)]$;
- $A = \{v2w : v, w \in \{0,1\}^*$ and if $v = y_n$ and $w = y_m$ then either n is not in the halting problem K or the time to enumerate n into K is different from $m\}$ with $y_0, y_1, \ldots$ being the cube-free words from Example 4;
- $R_1(x_0, x_1)$ is true iff x_0, x_1 are in A and x_0 ends with 2 and x_1 properly extends x_0;
- $R_2(x_2, x_1)$ is true iff x_1, x_2 are in A and x_2 is different from the word obtained by removing the last symbol from x_1.

The set A is clearly recursive. The block pumping constant of A is at most 7. Assume that a word $w \in A$ is given and that it is split by the breakpoints into $w_0\, w_1\, w_2\, w_3\, w_4\, w_5\, w_6\, w_7$. Assume first that the position of 2 in w is in w_4 or before. In this case, w_5 or $w_5 w_6$ is a pump: If y_n is the part before the 2 then there is only one word y_m (if any) such that $y_n 2 y_m \notin A$. As $y_n 2 y_m$ is cube-free and as $w_0 \ldots w_4 w_7$, $w_0 \ldots w_4 w_6 w_7$, $w_0 \ldots w_4 w_5 w_5 w_6 w_7$ and $w_0 \ldots w_4 w_5 w_6 w_5 w_6 w_7$ are all distinct, only one of them can be $y_n 2 y_m$ and so at least one of the sets $w_0 w_1 \ldots w_4 (w_5)^* w_6 w_7$ or $w_0 w_1 \ldots w_4 (w_5 w_6)^* w_7$ is a subset of A. Otherwise the 2 occurs in w_5 or w_6 or w_7. One of the subwords $w_1 w_2$ and $w_3 w_4$ is shorter than the half of the length of $w_0 w_1 w_2 w_3 w_4$, say $w_1 w_2$. Now, it follows from the length constraints — as outlined in the proof of Example 4 — that only one of the words $w_0 w_2 \ldots w_7$, $w_0 (w_1)^2 w_2 \ldots w_7$, $w_0 w_3 \ldots w_7$, $w_0 (w_1 w_2)^2 w_3 \ldots w_7$ has before the 2 a word which is of the form y_n for some n. Furthermore, when the pump occurs at least three times then the corresponding front part cannot be of the form y_n for any n, as all y_n are cube free. It follows that one of the languages $w_0 w_1^* w_2 \ldots w_7$ and $w_0 (w_1 w_2)^* w_3 \ldots w_7$ is a subset of A. Thus A is block pumpable.

The relations R_1 and R_2 are restrictions of automatic relations to $A \times A$ and therefore effective block pumpable relations.

If $n \notin K$ then every word properly extending $y_n 2$ in A satisfies that also after omitting the last symbol the corresponding word is still in A. Thus there is for every x_1 with $R_2(a, x_1)$ also an $x_2 \in A$ with $R_2(x_2, x_1)$. So $\phi(y_n 2)$ is not satisfied. If $n \in K$ then let m be the time to enumerate n into K and let $x_1 = y_n 2 y_m 1$. As $y_n 2 y_m \notin A$, every $x_2 \in A$ is not obtained by omitting the last symbol of x_1 and therefore $R_1(y_n2, x_1)$ and $R_2(x_2, x_1)$ are both true for every $x_2 \in A$. So $\phi(y_n2)$ is satisfied. Thus the $\exists\forall$-theory of the structure is undecidable, as it permits to solve the halting problem K. $\Box$

Note that a similar proof can also be made in order to show that the existential theory of block pumpable structures with negation is undecidable. Furthermore, for the positive side, the algorithms in Theorem 23 generalise to block cancellable structures but not to positively block pumpable structures.

7 Conclusion

The overall goal is to show that block pumpable languages are a quite natural class with interesting closure properties: They are closed under union, intersection, concatenation and images of transduced functions. They admit multiple pumping. Furthermore, the block pumpable languages are either regular or have exponential growth, which implies that the intersection of any block pumpable language with regular sets of polynomial growth is regular.

A setback is that many of the proofs to show these closure properties heavily rely on Ramsey's Theorem which causes exponential or worse increase in the pumping constants. In contrast to this, the pumping constants grow for the corresponding operations of regular sets only moderately either in a linear or quadratic fashion. Thus it is indeed interesting to ask the following question.

Open Problem 25. *Let $p \geq 2$ be a given constant, let L, H be two block pumpable languages with pumping constant p and let f be a function computed by a nondeterministic transducer with p states.*

What are the optimal bounds (in dependence of p) for the pumping constants of $L \cap H$ and $f(L)$, respectively? What is the optimal pumping constant (in dependence of p) for having two simultaneous pumps in L?

In the case of regular languages, there is a close connection between the number of states of an NFA recognising the set and the block pumping constant, as the number of states of the NFA is an upper bound for the block pumping constant. However, while the number of states of the NFA gives an exponential upper bound on the number of states of an NFA for the complement, there is no upper bound at all for block pumping constants of regular languages. More precisely, given any $p \geq 4$, the language

$$\{w \in \{0,1\}^* : w \text{ is cube-free or } |w| \neq p-2\}$$

has the block pumping constant 6 while its complement contains only words of length $p-2$ and one needs the block pumping constant to be p as p is the first constant where one cannot place p distinct breakpoints into a word of length $p-2$.

In the context of multiple pumping, one has for regular sets the following variant: If L is recognised by an NFA with p states, there for every word $x \in L$ and every set I of $2p+1$

breakpoints of x, there are three distinct breakpoints $i, j, k \in I$ cutting x into words v, w, y, z such that $v \cdot w^* \cdot y^* \cdot z \subseteq L$. One can ask whether the same holds for block pumpable languages in general.

Open Problem 26. *If L is block pumpable, is there a constant p such that for every word $x \in L$ and every set I of p breakpoints of x, there are three distinct breakpoints $i, j, k \in I$ cutting x into words v, w, y, z such that $v \cdot w^* \cdot y^* \cdot z \subseteq L$?*

This problem is also open for positively block pumpable structures; it can, however, be resolved for block cancellable languages by using Ramsey's Theorem.

As indicated, the results in Section 5 hold for positively block pumpable languages and for block pumpable languages. Most of the results do not hold for block cancellable languages, as indicated by the examples provided. However, the following question is open.

Open Problem 27. *Are there block cancellable languages which neither have polynomial nor exponential growth?*

Automatic structures satisfy that all involved sets and relations (coded as convolutions) are regular; thus all relations and sets involved are block pumpable. It is indeed possible to generalise the concept to block pumpable structures which shares some of the decidability properties with automatic structures, although the corresponding results are much weaker and hold only for the existential theory. Though several of these structures are constructed in the proofs, there is a lack of really interesting structures which are block pumpable but not automatic; as it is not possible to precisely say what "interesting" means, this is more a motivation for further research than a clearly stated open problem. However, the following question might also be interesting and is more concrete.

Open Problem 28. *If $f : A \to B$ is a block pumpable function from a block pumpable set A to a block pumpable set B, is f the restriction of an automatic function (from words to words) to the domain A?*

The main results of this paper are not only obtainable for block pumpable languages but also for the variants of positively block pumpable languages (except the decidability of the existential theory of the effective versions of corresponding structures) and block cancellable languages (except for the growth behaviour). Therefore these two concepts might also be useful in various contexts when block pumpable is difficult to obtain. In particular it might be easier to construct interesting effectively block cancellable structures than interesting effectively block pumpable structures, as there are more natural block cancellable sets.

Acknowledgments. The authors would like to thank Henning Fernau, Rupert Hölzl, Sanjay Jain and the anonymous referees for correspondence, discussions and useful comments.

References

1. Martin Bridson and Robert Gilman. Context-free languages of sub-exponential growth. *Journal of Computer and System Sciences*, 64(2):308-310, 2002.
2. Christopher Hanrui Chak. *Pumping Lemma and Automatic Functions.* BComp Dissertation (Final Year Project), School of Computing, National University of Singapore, 2013.
3. Leonard Eugene Dickson. Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *American Journal of Mathematics*, 35(4):413–422, 1913.
4. Pál Dömösi and Masami Ito. *Context-Free Languages and Primitive Words*. World Scientific, 2015.
5. Andrzej Ehrenfeucht, Rohit Parikh and Grzegorz Rozenberg. Pumping lemmas for regular sets. *SIAM Journal on Computing*, 10:536–541, 1981.
6. Nathan J. Fine and Herbert S. Wilf. Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16:109–114, 1965.
7. Bernard R. Hodgson. *Théories Décidables par Automate Fini.* Ph.D. thesis, Département de mathématiques et de statistique, Université de Montréal, 1976.
8. Bernard R. Hodgson. Décidabilité par automate fini. *Annales des Sciences Mathématiques du Québec*, 7(1):39–57, 1983.
9. Roberto Incitti. The growth function of context-free languages. *Theoretetical Computer Science*, 255:601–605, 2001.
10. John E. Hopcroft, Ravjeev Motwani and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Third edition. Addison Wesley, 2007.
11. Jeffrey Jaffe. A necessary and sufficient pumping lemma for regular languages. *ACM SIGACT News*, 10(2):48–49, January 1978.
12. Bakhadyr Khoussainov and Anil Nerode. Automatic presentations of structures. *Springer LNCS*, 960:367–392, 1995.
13. Joseph Kruskal. The theory of well-quasi-ordering: A frequently discovered concept. *Journal of Combinatorial Theory, Series A*, 13(3):297–305, 1972.
14. Tetsuo Kurosaki. Direct definition of a ternary infinite square-free sequence. *Information Processing Letters*, 106(5):175–179, 2008.
15. Roger Lyndon and Marcel-Paul Schützenberger. The equation $a^M = b^N c^P$ in a free group. *Michigan Mathematical Journal*, 9:289–298, 1962.
16. Micheal G. Main, Walter Bucher and David Haussler. Applications of an infinite square-free co-CFL. *Theoretical Computer Science*, 49:113–119, 1987.
17. Marston Morse. Recurrent geodesics on a surface of negative curvature. *Transactions of the American Mathematical Society*, 22:84–100, 1921.
18. Anton Nijholt. An annotated bibliography of pumping. *Bulletin of the EATCS*, 17:34–53, June 1982.
19. Eugéne Prouhet. Mémoire sur quelques relations entre les puissances de nombres. *Comptes Rendus des Scéances de l'Académie des Sciences Paris Séries I*, 33:225, 1851.
20. Frank P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, 30:264–286, 1930.

21. Rudolph Sommerhalder. Classes of languages proof against regular pumping. *RAIRO Informatique théorique*, 14:169–180, 1980.
22. Henrietta Tan Wan Yik. *Properties of pumpable languages.* BComp Dissertation (Final Year Project), School of Computing, National University of Singapore, 2014.
23. Axel Thue. Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln. *Norske Videnskabers Selskabs Skrifter, I, Mathematisch-Naturwissenschaftliche Klasse (Kristiania)*, 10, 34 pages, 1914.
24. Stefano Varricchio. A pumping condition for regular sets. *SIAM Journal on Computing*, 26(3):764–771, 1997.