Theoretical Computer Science 412 (2011) 805–821

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Büchi context-free languages☆

Zoltán Ésik, Szabolcs Iván *

Department of Computer Science, University of Szeged, Hungary

ARTICLE INFO

Article history:
Received 26 October 2009
Received in revised form 2 November 2010
Accepted 9 November 2010
Communicated by M. Ito

Keywords:
Context-free languages
Countable words
Ordinals
Buchi acceptance condition

ABSTRACT

We define context-free grammars with Büchi acceptance condition generating languages of countable words. We establish several closure properties and decidability results for the class of Büchi context-free languages generated by these grammars. We also define context-free grammars with Müller acceptance condition and show that there is a language generated by a grammar with Müller acceptance condition which is not a Büchi context-free language.

1. Introduction

A word over an alphabet Σ is an isomorphism type of a labeled linear order. In this paper, in addition to finite words and ω-words, we also consider words whose underlying linear order is any countable linear order, including scattered and dense linear orders, cf. [26].

Finite automata on ω-words were introduced by Büchi [11,12]. He used automata to prove the decidability of the monadic second-order theory of the ordinal ω. Automata on ω-words have since been extended to automata on ordinal words beyond ω, cf. [13,14,2,33,34], to words whose underlying linear order is not necessarily well-ordered, cf. [4,10], and to automata on finite and infinite trees, cf. [17,29,25]. Many decidability results have been obtained using the automata theoretic approach, both for ordinals and other linear orders, and for first-order and monadic second-order theories in general.

Countable words were first investigated in [16], where they were called "arrangements". It was shown that any arrangement can be represented as the frontier word (i.e., the sequence of leaf labels) of a possibly infinite labeled binary tree. Moreover, it was shown that words definable by finite recursion schemes are exactly those words represented by the frontiers of regular trees. These words were called regular in [8]. Courcelle raised several problems that were later solved in the papers [21,30,7]. In [30], it was shown that it is decidable for two regular trees whether they represent the same regular word. In [21], an infinite collection of regular operations has been introduced and it has been shown that each regular word can be represented by a regular expression. Complete axiomatizations have been obtained in [5] and [7] for the subcollections of the regular operations that allow for the representation of the regular ordinal words and the regular scattered words, respectively. Complete axiomatization of the full collection of the regular operations has been obtained in [8], where it is also proved that there is a polynomial time algorithm to decide whether two regular expressions represent the same regular word. Regular expressions representing languages (i.e., sets) of scattered countable words and languages of possibly dense words with no upper bound on the size of the words have been proposed in [10,4], where Kleene theorems

☆ Research supported by grant no. K 75249 from the National Foundation of Hungary for Scientific Research and by the TÁMOP-4.2.1/B-09/1/KONV-2010-0005 program of the Hungarian National Development Agency.

* Corresponding author. Tel.: +36 62546396; fax: +36 62546397.
E-mail address: szabivan@inf.u-szeged.hu (S. Iván).

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2010.11.026

have been established stating that a language of infinite words is recognizable by a finite automaton iff it can be represented by a regular expression. For a connection to monadic second-order logic see [3].

In addition to automata and expressions (or terms) built by certain operations, a third common way of representing languages of finite words is by generative grammars. Context-free grammars have been used to generate languages of ω-words in [15] and in [9,23]. However, we are not aware of any work on context-free grammars as a device generating languages of countable words possibly longer than ω, except for the recent [18] that deals only with linear grammars.

In this paper, we consider languages of countable words generated by context-free grammars equipped with a Büchi-type acceptance condition, called BCFG's. A BCFG is a system $G = (N, \Sigma, P, S, F)$, where (N, Σ, P, S) is an ordinary context-free grammar and $F \subseteq N$ is the set of designated nonterminals. A derivation tree t of a grammar G is a possibly infinite tree whose vertices are labeled in the set $N \cup \Sigma \cup \{\varepsilon\}$, so that each vertex is labeled by a nonterminal in N, a letter in the terminal alphabet Σ, or by the empty word ε. The labeling is locally consistent with the rules contained in P in the usual way. Moreover, it is required that each derivation tree satisfies the "Büchi condition F", i.e., on each infinite path of t at least one designated nonterminal has to occur infinitely many times. The frontier of a derivation tree t determines a countable word w over the alphabet $N \cup \Sigma$. When w is a word over the terminal alphabet Σ and the root of t is labeled by the start symbol S, we say that w is contained in the Büchi context-free language generated by G. The language class BCFL consists of all such Büchi context-free languages.

It is well-known (see e.g., [20]) that ordinary context-free languages of finite words are precisely the frontier languages of sets of finite trees recognizable by finite tree automata. Tree automata over infinite trees have been introduced in [25]. Just as automata over ω-words, a tree automaton may be equipped with different acceptance conditions such as the Büchi and Müller acceptance conditions, or the Rabin, Streett and parity conditions, cf. [24,31,32]. In the setting of ω-words, these conditions are equally powerful (at least for nondeterministic automata). Nevertheless, some yield more succinct representation than others, or have different algorithmic properties. On the other hand, in the setting of infinite trees, the Büchi acceptance condition is strictly less powerful than the Müller acceptance condition which is equivalent to the Rabin, Streett, and parity conditions, cf. [24,31,32]. While in the present paper we are mainly concerned with the Büchi condition for generating context-free languages of countable words, we still show that the Müller condition is strictly more powerful also in the setting of countable words. While Büchi context-free languages can be characterized as the frontiers of Büchi recognizable tree languages, and similarly for Müller context-free languages, this result is not immediate from the tree case. A more detailed study of Müller context-free languages is left for future research.

The paper is organized as follows. In Section 2, we recall some notions and results for linear orders and words. Then, in Section 3, we define BCFG's and give several examples of Büchi context-free languages generated by BCFG's. Section 4 is devoted to elimination of useless nonterminals, ε-productions, chain productions, etc. Then, in Section 5, we establish the closure of BCFL's with respect to substitution and derive many other closure properties from this result. In Section 6, we establish several decidability results. Among others, we show that it is decidable whether the BCFL generated by a BCFG consists of well-ordered, or scattered, or dense words. In each case, a polynomial time algorithm is found. Moreover, we show that for every BCFL of scattered words there is a finite bound n such that all words in the language are of "rank" at most n. (Our notion of rank is similar to the Hausdorff rank of linear orders, cf. [26].) This implies that the set of all well-ordered countable words is not a BCFL. Then, in Section 7, we compare BCFL's of ω-words with the context-free languages of ω-words of Cohen and Gold [15]. Section 8 is devoted to an undecidability result: It is undecidable for a BCFG with a 1-letter terminal alphabet whether it generates the universal language of all countable words. Then, in Section 9, we introduce MCFG's, i.e., context-free grammars with Müller acceptance condition, and show that there is a language over the 1-letter alphabet that can be generated by an MCFG but which is not a BCFL. Section 10 contains some concluding remarks and some suggestions for further research directions. An extended abstract of this paper appeared in [19].

Notation. The ordered set of nonnegative integers is denoted $\mathbb{N}_0$, and $\mathbb{N}$ stands for the positive integers. The ordered set of rationals is denoted $\mathbb{Q}$.

2. Linear orders and words

In this section we recall some concepts for linear orders and words. A good reference on linear orders is [26].

A partial order, or partial ordering is a set P equipped with a (partial) order relation usually denoted $\leq$. We sometimes write $x < y$ if $x \leq y$ and $x \neq y$. A linear order is a partial order $(P, \leq)$ whose order relation is total, so that $x \leq y$ or $y \leq x$ for all $x, y \in P$. A countable (finite or infinite, respectively) linear order is a linear order which is a countable (finite or infinite, respectively) set. When $(P, \leq)$ and $(Q, \leq)$ are linear orders, an isomorphism (embedding, respectively) $(P, \leq) \to (Q, \leq)$ is a bijection (injection, respectively) $h : P \to Q$ such that $x \leq y$ implies $h(x) \leq h(y)$ for all $x, y \in P$. When two linear orders are isomorphic, we also say that they have the same order type (or isomorphism type).

Below when there is no danger of confusion, we will denote a linear order just by $P, Q, \ldots$. Suppose that P is a linear order. Then any subset X of P determines a sub-order of P whose order relation is the restriction of the order relation of P to X. Note that the inclusion function $X \hookrightarrow P$ is an embedding of X into P. When in addition X is such that for all $x, y \in X$ and $z \in P$, $x < z < y$ implies that $z \in X$, then we call X an interval. In particular, for any $x, y \in P$, the set $[x, y] = \{z : x \leq z \leq y\}$ is an interval.

We recall that a linear order $(P, \leq)$ is a *well-order* if each nonempty subset of P has a least element, and is *dense* if it has at least two elements and for any $x < y$ in P there is some z with $x < z < y$.[1] A *quasi-dense* linear order is a linear order $(P, \leq)$ containing a dense linear sub-order, so that P has a subset P' such that $(P', \leq)$ is a dense order. Finally, a *scattered* linear order is a linear order which is not quasi-dense.

It is clear that every finite linear order is a well-order, every well-order is a scattered order, and every dense order is quasi-dense. It is well-known that up to isomorphism there are four countable dense linear orders, the rationals $\mathbb{Q}$ with the usual order, $\mathbb{Q}$ endowed with a least or a greatest element, and $\mathbb{Q}$ endowed with both a least and a greatest element.

An ordinal is an order type of a well-order. The finite ordinals n are the isomorphism types of the finite linear orders. As usual, we denote by ω the least infinite ordinal, which is the order type of the finite ordinals, and of the natural numbers $\mathbb{N}_0$ equipped with the usual order. The order type of $\mathbb{Q}$ will be denoted η.

When τ and τ' are order types, we say that $\tau \leq \tau'$ if there is an embedding of a linear order of type τ into a linear order of type τ'. The relation $\leq$ defined above is a linear order of the ordinals.

We define several operations on linear orders. First, the reverse $(P, \leq')$ of a linear order $(P, \leq)$ is defined by $x \leq' y$ iff $y \leq x$, for all $x, y \in P$. We will sometimes denote the reverse order $(P, \leq')$ by P^r. It is clear that the reverse of a scattered linear order is scattered, and the reverse of a dense linear order is dense.

Suppose that P and Q are linear orders. Then the sum $P + Q$ is the linear order on the disjoint union of P and Q such that P and Q are intervals of $P + Q$ and $x \leq y$ holds for all $x \in P$ and $y \in Q$. There is a more general notion. Suppose that I is a linear order and for each $i \in I$, P_i is a linear order. Then the generalized sum

$$P = \sum_{i \in I} P_i$$

is obtained by replacing each point i of I with the linear order P_i. Formally, the generalized sum P is the linear order on the disjoint union $\biguplus_{i \in I} P_i$ equipped with the order relation such that each P_i is an interval and for all $i, j \in I$ with $i < j$, if $x \in P_i$ and $y \in P_j$ then $x < y$. The generalized sum gives rise to a product operation. Let P and Q be linear orders, and for each $y \in Q$, let P_y be an isomorphic copy of P. Then $P \times Q$ is defined as the linear order $\sum_{y \in Q} P_y$. Note that this linear order is isomorphic to the linear order on the cartesian product of P and Q equipped with the order relation $(x, y) \leq (x', y')$ iff $y < y'$ or ($y = y'$ and $x \leq x'$).

Lemma 2.1 ([26]). *Any scattered generalized sum of scattered linear orders is scattered. Similarly, any well-ordered generalized sum of well-orders is a well-order. Every quasi-dense linear order is a dense generalized sum of (nonempty) scattered linear orders.*

Thus, when I is a scattered linear order and for each $i \in I$, P_i is a scattered linear order, then so is $\sum_{i \in I} P_i$, and similarly for well-orders. And if P is a quasi-dense linear order, then there is a dense linear order D and (nonempty) scattered linear orders P_x, $x \in D$ such that P is isomorphic to $\sum_{x \in D} P_x$.

The above operations preserve isomorphism, so that they give rise to corresponding operations $\tau + \tau'$ and $\tau \times \tau'$ on order types. In particular, the sum and product of two ordinals is well-defined (and is an ordinal). The reverse of an order type τ will be denoted $-\tau$. The ordinals are also equipped with the exponentiation operation, cf. [26].

An alphabet Σ is a finite nonempty set. A word over an alphabet Σ is a labeled linear order, i.e., a system $u = (P, \leq, \lambda)$, where $(P, \leq)$ is a linear order, sometimes denoted $\mathrm{dom}(u)$, and λ is a labeling function $P \to \Sigma$. The underlying linear order $\mathrm{dom}(\varepsilon)$ of the empty word ε is the empty linear order. We say that a word is finite (infinite or countable, respectively), if its underlying linear order is finite (infinite or countable, respectively). An isomorphism of words is an isomorphism of the underlying linear orders that preserves the labeling. Embeddings of words are defined in the same way. We usually identify isomorphic words. We will say that a word u is a subword of a word v if there is an embedding $u \hookrightarrow v$. When in addition the image of the underlying linear order of u is an interval of the underlying linear order of v we call u a factor of v.

The order type of a word is the order type of its underlying linear order. Thus, the order type of a finite word is a finite ordinal. A word whose order type is ω is called an ω-word.

Let $\Sigma = \{a, b\}$. Some examples of words over Σ are the finite word aab which is the (isomorphism class of the) 3-element labeled linear order $\{0 < 1 < 2\}$ whose points are labeled a, a and b, in this order. Examples of infinite words are a^ω and $a^{-\omega}$, whose order types are ω and $-\omega$, respectively, such that each point is labeled a. For another example, consider the linear order $\mathbb{Q}$ of the rationals and label each point a. The resulting word of order type η is denoted a^η. More generally, let Σ contain the (different) letters $a_1, \ldots, a_n$. Then up to isomorphism there is a unique labeling of the rationals such that between any two points there are n points labeled $a_1, \ldots, a_n$, respectively. The resulting word is denoted $(a_1, \ldots, a_n)^\eta$, cf. [21].

The reverse of a word $u = (P, \leq, \lambda)$ is $u^r = (P, \leq', \lambda)$, where $(P, \leq')$ is the reverse of $(P, \leq)$. Suppose that $u = (P, \leq, \lambda)$ and $v = (Q, \leq, \lambda')$ are words over Σ. Then their concatenation (or product) uv is the word over Σ whose underlying linear order is $P + Q$ and whose labeling function agrees with λ on points in P, and with λ' on points in Q. More generally, when I is a linear order and u_i is a word over Σ with underlying linear order $P_i = \mathrm{dom}(u_i)$, for each $i \in I$, then the generalized concatenation $\prod_{i \in I} u_i$ is the word whose underlying linear order is $\sum_{i \in I} P_i$ and whose labeling function agrees with the labeling function of P_i on the elements of each P_i. In particular, when $u_0, u_1, \ldots, u_n, \ldots$ are words over Σ and I is the linear

[1] In [26], a singleton linear order is also called dense.

order ω or its reverse, then $\prod_{i\in I} u_i$ is the word $u_0u_1 \dots u_n \dots$ or $\dots u_n \dots u_1u_0$, respectively. When $u_i = u$ for each i, these words are denoted u^ω and $u^{-\omega}$, respectively.

In the following, we will sometimes make use of the substitution operation on words. Suppose that u is a word over Σ and for each letter $a \in \Sigma$, u_a is a word over Δ. Then the word $u[a \leftarrow u_a]_{a\in\Sigma}$ obtained by substituting u_a for each occurrence of a letter a in u (or replacing each occurrence of a letter a with u_a) is formally defined as follows. Let $u = (P, \le, \lambda)$ and $u_a = (P_a, \le_a, \lambda_a)$ for each $a \in \Sigma$. Then for each $i \in P$ let $u_i = (P_i, \le_i, \lambda_i)$ be an isomorphic copy of $P_{\lambda(i)}$. We define

$$u[a \leftarrow u_a]_{a\in\Sigma} = \prod_{i\in P} u_i.$$

Note that when $u = a^\omega$, then $u[a \leftarrow v]$ is v^ω, and similarly for $v^{-\omega}$. For any words $u_1, \dots, u_n$ over an alphabet Σ, we define

$$(u_1, \dots, u_n)^\eta = (a_1, \dots, a_n)^\eta[a_1 \leftarrow u_1, \dots, a_n \leftarrow u_n].$$

We call a word over an alphabet Σ well-ordered, scattered, dense, or quasi-dense if its underlying linear order has the appropriate property. For example, the words a^ω, $a^\omega b^\omega a$, $(a^\omega)^\omega$ over the alphabet $\{a, b\}$ are well-ordered, the words $a^\omega a^{-\omega}$, $a^{-\omega}a^\omega$ are scattered, the words a^η, $a^\eta ba^\eta$, $(a, b)^\eta$ are dense, and the words $(ab)^\eta$, $(a^\omega)^\eta$, $(a^\eta b)^\omega$ are quasi-dense.

From Lemma 2.1 we immediately have:

Lemma 2.2. *Any scattered generalized product of scattered words is scattered. Any well-ordered generalized product of well-ordered words is well-ordered. Moreover, every quasi-dense word is a dense product of (nonempty) scattered words.*

As already mentioned, we will usually identify isomorphic words, so that a word is an isomorphism type (or isomorphism class) of a labeled linear order. When Σ is an alphabet, we let Σ^*, Σ^ω and Σ^∞ respectively denote the set of all finite words, ω-words, and countable words over Σ. Σ^+ is the set of all finite nonempty words, and $\Sigma^{+\infty} = \Sigma^\infty\Sigma\Sigma^\infty$ is the set of all countable nonempty words over Σ. The length of a finite word w will be denoted $|w|$.

A language over Σ is any subset L of Σ^∞. When $L \subseteq \Sigma^*$ or $L \subseteq \Sigma^\omega$, we sometimes call L a language of finite words or ω-words, or an ω-language.

Languages are equipped with several operations. First of all, they are equipped with the usual set theoretic operations. We now define the generic operation of language substitution.

Suppose that $u \in \Sigma^\infty$ and for each $a \in \Sigma$, $L_a \subseteq \Delta^\infty$. Then the words in the language $u[a \leftarrow L_a]_{a\in\Sigma} \subseteq \Delta^\infty$ are obtained from u by substituting in all possible ways a word in L_a for each occurrence of each letter $a \in \Sigma$. Different occurrences of the same letter a may be replaced by different words in L_a.

Formally, suppose that $u = (P, \le, \lambda)$. For each $x \in P$ with $\lambda(x) = a$, let us choose a word $u_x = (P_x, \le_x, \lambda_x)$ which is isomorphic to some word in L_a. Then the language $u[a \leftarrow L_a]_{a\in\Sigma}$ consists of all words $\prod_{x\in P} u_x$.

Suppose now that $L \subseteq \Sigma^\infty$ and for each $a \in \Sigma$, $L_a \subseteq \Delta^\infty$. Then

$$L[a \leftarrow L_a]_{a\in\Sigma} = \bigcup_{u\in L} u[a \leftarrow L_a]_{a\in\Sigma}.$$

We call $L[a \leftarrow L_a]_{a\in\Sigma}$ the language obtained from L by substituting the language L_a for each $a \in \Sigma$.

As mentioned above, set theoretic operations on languages in Σ^∞ have their standard meaning. Below we define some other operations.

Let $L, L_1, L_2, \dots, L_m \subseteq \Sigma^\infty$. Then we define:

1. $L_1L_2 = \{ab\}[a \leftarrow L_1, b \leftarrow L_2] = \{uv : u \in L_1, v \in L_2\}$.
2. $L^* = \{a\}^*[a \leftarrow L] = \{u_1 \dots u_n : n < \omega,\ u_i \in L\}$.
3. $L^\omega = \{a^\omega\}[a \leftarrow L] = \{u_0u_1 \dots u_n \dots : u_i \in L\}$.
4. $L^{-\omega} = \{a^{-\omega}\}[a \leftarrow L] = \{\dots u_n \dots u_1u_0 : u_i \in L\}$.
5. $(L_1, \dots, L_m)^\eta = \{(a_1, \dots, a_m)^\eta\}[a_1 \leftarrow L_1, \dots, a_m \leftarrow L_m]$.
6. $L^\infty = \{a\}^\infty[a \leftarrow L]$.

The above operations are respectively called concatenation, star, $^\omega$-power, $^{-\omega}$-power, $^\eta$-power, and $^\infty$-power.

Some more operations. The reverse L^r of a language $L \subseteq \Sigma^\infty$ is defined as $L^r = \{u^r : u \in L\}$. The prefix language $\mathrm{Pre}(L)$ is given by $\mathrm{Pre}(L) = \{u : \exists v\ uv \in L\}$ and the suffix language $\mathrm{Suf}(L)$ is defined symmetrically. The infix language $\mathrm{In}(L)$ is $\{u : \exists v, w\ vuw \in L\}$, and the language $\mathrm{Sub}(L)$ of subwords of L is the collection of all words u such that there is an embedding $u \hookrightarrow v$ for some $v \in L$.

3. Büchi context-free languages

Recall from [27] that an (ordinary) context-free grammar (CFG) is a system $G = (N, \Sigma, P, S)$ where N and Σ are the disjoint alphabets of nonterminals and terminal symbols (or letters), P is a finite set of productions of the form $A \to p$ where $A \in N$ and $p \in (N \cup \Sigma)^*$, and $S \in N$ is the start symbol. Each context-free grammar $G = (N, \Sigma, P, S)$ generates a context-free language $L(G) \subseteq \Sigma^*$ which can be defined either by using the derivation relation $\Rightarrow^*$ or by using the concept of derivation trees.

We recall that for finite words $p, q \in (N \cup \Sigma)^*$ it holds that $p \Rightarrow q$ if p and q can be written as $p = p_1 A p_2$, $q = p_1 r p_2$ such that $A \to r$ is in P. The relations $\Rightarrow^+$ and $\Rightarrow^*$ are respectively the transitive closure and the reflexive-transitive closure of the direct derivation relation $\Rightarrow$. The context-free language generated by G is $L(G) = \{u \in \Sigma^* : S \Rightarrow^* u\}$. Two context-free grammars G and G' having the same terminal alphabet are called equivalent if $L(G) = L(G')$. We let CFL denote the class of all context-free languages.

A derivation tree is a partial mapping $t : \mathbb{N}^* \to N \cup \Sigma \cup \{\varepsilon\}$ whose domain $\mathrm{dom}(t)$ is finite, nonempty and prefix closed (i.e., $uv \in \mathrm{dom}(t) \Rightarrow u \in \mathrm{dom}(t)$). The elements of $\mathrm{dom}(t)$ are the vertices of t, and for any vertex v, $t(v)$ is the label of v. The empty word ε is the root of t, and $t(\varepsilon)$ is the root symbol. The vertices in $\mathrm{dom}(t)$ are equipped with both the lexicographic order and the prefix order. Let $x, y \in \mathrm{dom}(t)$. We say that $x \le y$ in the prefix order if $y = xz$ for some $z \in \mathbb{N}^*$. Moreover, we say that $x < y$ in the lexicographic order if $x = uiz$ and $y = ujz'$ for some $u, z, z' \in \mathbb{N}^*$ and $i, j \in \mathbb{N}$ with $i < j$. The leaves of t are the maximal elements of $\mathrm{dom}(t)$ with respect to the prefix order. When $x, y \in \mathrm{dom}(t)$ and $y = xi$ for some $i \in \mathbb{N}$, then we say that y is the ith successor of x and x is the predecessor of y. The function t is required to satisfy the local consistency condition that whenever $t(u) = A$ with $A \in N$ and u is not a leaf, then either $A \to \varepsilon \in P$ and $t(u1) = \varepsilon$ and $t(ui)$ is not defined for any $i \in \mathbb{N}$ with $i > 1$, or there is a production $A \to p$ such that $|p| = n$ with $n > 0$ and $t(ui)$ is defined for some $i \in \mathbb{N}$ iff $i \le n$, moreover, $t(ui)$ is the ith letter of p for each $i \le n$. When $t(u)$ is in $\Sigma \cup \{\varepsilon\}$, then u must be a leaf. The frontier of t is the linearly ordered set of leaves whose order is the lexicographic order. The frontier determines a word in $(N \cup \Sigma)^*$ whose underlying linear order is obtained from the frontier of t by removing all those vertices whose label is ε. The labeling function is the restriction of the function t to the remaining vertices. This word is sometimes called the frontier word of t. It is well-known that a word u in Σ^* belongs to $L(G)$ iff there is a derivation tree whose root is labeled S and whose frontier word is u.

We now define context-free grammars generating countable words.

Definition 3.1. A *context-free grammar with Büchi acceptance condition*, or BCFG, is a system $G = (N, \Sigma, P, S, F)$ where N, Σ, P, S are the same as above, and $F \subseteq N$ is the set of *designated nonterminals*.

Note that each BCFG has an underlying CFG. Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG. A derivation tree t is defined as above except that $\mathrm{dom}(t)$ may now be infinite. However, we require that at least one designated nonterminal occurs infinitely often along each infinite path. When the root symbol of t is A and the frontier word of t is p, we also write $A \Rightarrow^\infty p$. (Here, it is allowed that A is a terminal in which case $A = p$.) The language (of countable words) generated by G is $L^\infty(G) = \{u \in \Sigma^\infty : S \Rightarrow^\infty u\}$. When G and G' are BCFG's with the same terminal alphabet Σ generating the same language, then we say that G and G' are equivalent.

Definition 3.2. We call a set $L \subseteq \Sigma^\infty$ a Büchi context-free language, or a BCFL, if it can be generated by some BCFG, i.e., when $L = L^\infty(G)$ for some *BCFG* $G = (N, \Sigma, P, S, F)$.

Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG with underlying CFG $G' = (N, \Sigma, P, S)$. Then we define $L^*(G)$ as the CFL $L(G')$. Note that in general it does not hold that $L^*(G) = L^\infty(G) \cap \Sigma^*$. For an example, consider the BCFG G with productions $S \to aS$ and $S \to S$ where $N = F = \{S\}$ and a is the single terminal letter. Then $L^*(G)$ is empty, and $L^\infty(G) \cap \{a\}^* = \{a\}^*$. Later we will prove that for every BCFG $G = (N, \Sigma, P, S, F)$ it holds that $L^\infty(G) \cap \Sigma^*$ is a CFL. It is clear that CFL $\subseteq$ BCFL, for if $G = (N, \Sigma, P, S, F)$ is a BCFG with $F = \emptyset$, then $L^\infty(G) = L^*(G)$.

Example 3.3. Consider the sequence $(w_n)_{n<\omega}$ of words over $\{a\}$ defined inductively by $w_0 = a$, and for each $n < \omega$, $w_{n+1} = w_n^\omega$. Note that the order type of w_n is ω^n, for each n.

For each n, the BCFG $G_n = (N, \{a\}, P, S_n, N)$ with

$$N = \{S_0, \ldots, S_n\} \text{ and}$$
$$P = \{S_0 \to a\} \cup \{S_i \to S_{i-1} S_i : 1 \le i \le n\}$$

generates the singleton language $\{w_n\}$. Using this, it follows that the BCFG

$$G_n' = (N \cup \{S\}, \{a\}, P \cup \{S \to S_i : 0 \le i \le n\}, S, N)$$

generates the set $\{w_i : 0 \le i \le n\}$.

Example 3.4. Let Σ be an alphabet and let $a_1, \ldots, a_n \in \Sigma$ be letters in Σ. The singleton language containing the word $(a_1, \ldots, a_n)^\eta$ is a BCFL generated by $G = (\{S\}, \Sigma, \{S \to S a_1 S a_2 \ldots S a_n S\}, S, \{S\})$.

Example 3.5. Consider the language L over the 1-letter alphabet $\{a\}$ consisting of all words in $\{a\}^\infty$ whose domain is well-ordered of order type $<\omega^n$. Then L is generated by the BCFG $G = (N, \{a\}, P, S_n, N - \{S_n\})$ with

$$N = \{S_n, \ldots, S_0\} \quad \text{and}$$
$$P = \{S_i \to \varepsilon : 0 \le i \le n\} \cup \{S_0 \to a\} \cup \{S_i \to S_{i-1} S_i : 1 \le i \le n\}.$$

Let L' be the subset of L consisting of those words in L whose domain is a limit ordinal.[2] Then L' is the set of all finite concatenations of the words w_i, $1 \leq i < n$ of Example 3.3. L' is generated by the BCFG $G = (N, \{a\}, P, S, N - \{S\})$ with

$$\begin{aligned} N &= \{S, S_0, \ldots, S_{n-1}\} \quad \text{and} \\ P &= \{S \to S_i S : 1 \leq i < n\} \cup \{S \to \varepsilon\} \cup \\ &\quad \{S_0 \to a\} \cup \{S_i \to S_{i-1} S_i : 1 \leq i < n\}. \end{aligned}$$

Example 3.6. The language $\{a^\omega b^{-\omega}\}^* \cup \{a^\omega b^{-\omega}\}^\omega$ is a BCFL generated by $G = (N, \{a, b\}, P, S, N)$ with $N = \{S, X\}$ and $P = \{S \to XS, S \to \varepsilon, X \to aXb\}$.

Example 3.7. Using the fact (see e.g., Theorem 2.5 in [26]) that any countable linear order can be embedded into $\mathbb{Q}$, we get that Σ^∞ is a BCFL for any alphabet Σ, generated by the BCFG

$$G = (\{S\}, \Sigma, \{S \to SS, S \to \varepsilon\} \cup \{S \to a : a \in \Sigma\}, S, \{S\}).$$

Some more notation. Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG. We may define the direct derivation relation $\Rightarrow$ over words in $(N \cup \Sigma)^\infty$ in the same way as in the case of ordinary context-free grammars. We again let $\Rightarrow^+$ and $\Rightarrow^*$ denote the transitive closure and the reflexive-transitive closure of the relation $\Rightarrow$. When $p \Rightarrow^* q$ holds for some $p, q \in (N \cup \Sigma)^\infty$, we also say that there is a finite derivation of q from p.

We may also extend the relation $\Rightarrow^\infty$ to words in $(N \cup \Sigma)^\infty$. Suppose that γ is a countable word formed of possibly infinite derivation trees, say $\gamma = (P, \leq, \lambda)$, where $\lambda(x)$ is a derivation tree for each $x \in P$. Then the root symbols determine a word p in $(N \cup \Sigma)^\infty$: $p = (P, \leq, \lambda')$ where for each $x \in P$, $\lambda'(x)$ is the root symbol of $\lambda(x)$. Also, the frontier words of the trees $\lambda(x)$ determine a word q in $(N \cup \Sigma)^\infty$ and we write $p \Rightarrow^\infty q$.

Let t be a derivation tree. A finite path in t may be identified with a word in $\mathbb{N}^*$, its endpoint, and an infinite path may be identified with a word in $\mathbb{N}^\omega$. We extend the lexicographic order to paths. The lexicographically least (greatest, respectively) complete path of t will be called the leftmost (rightmost, respectively) complete path. (A path is complete if it is either infinite or its endpoint is a leaf.)

4. Normal forms

In this section we show that each BCFG can be transformed in polynomial time into an equivalent BCFG which is "weakly ε-free" and does not contain useless nonterminals nor any chain productions. Moreover, we show that each BCFG can be transformed in polynomial time into an equivalent "ε-free" BCFG having no useless nonterminals.

Definition 4.1. Let $G = (N, \Sigma, P, S, F)$ be a BCFG. We say that a nonterminal A is useful if there exist words $p, q \in (N \cup \Sigma)^*$ and $u \in \Sigma^\infty$ such that $S \Rightarrow^* pAq$ and $A \Rightarrow^\infty u$. We say that G contains no useless nonterminals if either $N = \{S\}$, $P = \emptyset$ and $F = \emptyset$, or each nonterminal is useful.

Note that when $G = (N, \Sigma, P, S, F)$ contains no useless nonterminals, then $L^\infty(G)$ is empty iff $N = \{S\}$, $P = \emptyset$ and $F = \emptyset$. Moreover, if $L^\infty(G)$ is not empty, then for each $A \in N$ there are words $u, v \in \Sigma^\infty$ with $S \Rightarrow^\infty uAv$.

Proposition 4.2. *For each BCFG $G = (N, \Sigma, P, S, F)$ one can construct in polynomial time an equivalent BCFG $G' = (N', \Sigma, P', S, F')$ which contains no useless nonterminals and satisfies $N' \subseteq N$, $F' \subseteq F$ and $P' \subseteq P$.*

Proof. In polynomial time, we find the largest set $Y \subseteq F$ such that for each $A \in Y$ there is a finite word $p \in (Y \cup \Sigma)^*$ with $A \Rightarrow^+ p$.

We claim that for any $B \in F$, $B \in Y$ iff $B \Rightarrow^\infty u$ holds for some $u \in \Sigma^\infty$. The necessity is obvious. Suppose now that $B \Rightarrow^\infty u$ for some $u \in \Sigma^\infty$. Then there exists a derivation tree t with root symbol B and frontier word u. Let Y_0 stand for the set of those nonterminals in F that label a vertex in t. Let $C \in Y_0$ be arbitrary. Then there exists a vertex x of t labeled C. For each maximal path starting at x, consider the first vertex different from x labeled in the set $\Sigma \cup \{\varepsilon\} \cup F$. By König's lemma, the set X of all such vertices is finite. By definition of Y_0, when some vertex $x \in X$ is labeled in F, then it is labeled in Y_0. Hence, for any $C \in Y_0$ there exists a word $p \in (Y_0 \cup \Sigma)^*$ with $C \Rightarrow^+ p$. Since Y is the greatest set of nonterminals with this property, this implies $Y_0 \subseteq Y$, hence from $B \in Y_0$ we get $B \in Y$, proving the claim.

Now in polynomial time we construct the set Y' of all nonterminals A such that $A \Rightarrow^* p$ for some $p \in (Y \cup \Sigma)^*$. If $S \notin Y'$ then $L^\infty(G)$ is empty. Otherwise the grammar $G' = (N', \Sigma, P', S, F')$ with $N' = Y'$, $P' = \{A \to p \in P : A \in Y', p \in (Y' \cup \Sigma)^*\}$, $F' = F \cap Y'$ satisfies the conditions. $\square$

Definition 4.3. Let $G = (N, \Sigma, P, S, F)$ be a BCFG. We call G weakly ε-free if either $L^\infty(G) = \emptyset$, or for each nonterminal A there is a nonempty word $u \in \Sigma^\infty$ with $A \Rightarrow^\infty u$, or $S \to \varepsilon$ is the only production.

Proposition 4.4. *For each $G = (N, \Sigma, P, S, F)$ one can construct in polynomial time an equivalent weakly ε-free BCFG containing no useless nonterminals.*

[2] By convention, 0 is considered as a limit ordinal here.

Proof. By Proposition 4.2, without loss of generality we may assume that G contains no useless nonterminals. When $P = \emptyset$ our claim is clear, so assume from now on that $P \neq \emptyset$. In polynomial time, we construct the set $N_1 \subseteq N$ of all nonterminals A such that there is a word $p \in (N \cup \Sigma)^*$ containing at least one occurrence of a letter in Σ with $A \Rightarrow^* p$. Then we have that $A \notin N_1$ iff the only word $u \in \Sigma^\infty$ with $A \Rightarrow^\infty u$ is the word $u = \varepsilon$. If $S \in N_1$ then we let $G' = (N', \Sigma, P', S, F')$ with $N' = N_1$, $F' = F \cap N'$, where P' consists of all productions of the form $A \to p$ with $A \in N'$ and $p \in (N' \cup \Sigma)^*$ such that p can be written as $p_1 \dots p_n$ with $A \to q_0 p_1 q_1 \dots p_n q_n \in P$ for some $q_0, \dots, q_n \in (N - N')^*$. If $S \notin N_1$, then we let $N' = \{S\}$, $P' = \{S \to \varepsilon\}$, and $F' = \emptyset$. In either case, G' is equivalent to G. □

As usual, a chain production is of the form $A \to B$, where A, B are nonterminals.

Proposition 4.5. *For each BCFG G one can construct in polynomial time an equivalent weakly ε-free BCFG G' without any chain productions.*

Proof. Without loss of generality we may assume that G is weakly ε-free. Suppose that $G = (N, \Sigma, P, S, F)$. Let $\overline{N}$ be a disjoint copy of N. For each $A \in N$ let Y_A be the set of all nonterminals B such that B can be derived from A by using only chain productions. We define $G' = (N', \Sigma, P', S, F')$ where $N' = N \cup \overline{N}$, $F' = F \cup \overline{N}$, and P' consists of the following productions. Let $B \in Y_A$. Assume that there is a derivation $A = A_0 \Rightarrow A_1 \Rightarrow \cdots \Rightarrow A_n = B$ such that $\{A_1, \dots, A_n\} \cap F \neq \emptyset$. Then for each production $B \to p$ where p is not a single nonterminal we add the productions $A \to q$ and $\overline{A} \to q$ to P', where to obtain q we replace each occurrence of a nonterminal C in p with $\overline{C}$. If there is an infinite derivation $A = A_0 \Rightarrow A_1 \Rightarrow \cdots$ such that $A_i \in F$ holds infinitely often, then we also add the productions $A \to \varepsilon$ and $\overline{A} \to \varepsilon$ to P. Finally, we keep all productions $A \to p$ where p is not a single nonterminal. □

If needed, we may remove useless nonterminals.

Corollary 4.6. *For each BCFG G one can construct in polynomial time an equivalent weakly ε-free BCFG G' without any chain productions which does not contain any useless nonterminals.*

We now define another stronger version of ε-freeness.

Definition 4.7. Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG. We say that G is ε-free if the following conditions hold:

1. G is weakly ε-free.
2. Except possibly for the production $S \to \varepsilon$, the right side of any other production is a nonempty word. Moreover, if $S \to \varepsilon$ is a production, then S does not occur on the right side of any other production.
3. If $\varepsilon \in L^\infty(G)$ then $S \to \varepsilon$ is a production.
4. For each derivation tree t whose frontier determines a nonempty word in Σ^∞ there is a derivation tree t' with the same root symbol and the same frontier word which is locally finite in the following strict sense: For each vertex $x \in \mathrm{dom}(t')$, the subtree $t'|_x$ of t' rooted at x has at least one leaf labeled in Σ.

Proposition 4.8. *For each BCFG G one can construct in polynomial time an equivalent ε-free grammar without useless nonterminals.*

Proof. Without loss of generality we may assume that G is ε-free and contains no useless nonterminals. If $N = \{S\}$ and $P = \emptyset$ or $P = \{S \to \varepsilon\}$ then $L^\infty(G)$ is either empty or $L^\infty(G) = \{\varepsilon\}$, and G is ε-free. Otherwise $L^\infty(G)$ contains a nonempty word and we construct (in polynomial time) the set N_ε of all nonterminals A such that $A \Rightarrow^\infty \varepsilon$: N_ε is the largest subset Y of N such that for all nonterminals A if there is a nontrivial finite derivation tree with root symbol A and frontier word in Y^* and such that each path from the root to a leaf labeled in Y contains some nonterminal in F, then $A \in Y$. Suppose that we remove all productions whose right side is ε and replace any other production $A \to p$ with all productions $A \to q$ such that $q \neq \varepsilon$ and q can be obtained from p by removing some occurrences of letters in N_ε. Let P' denote this set of productions. If $S \notin N_\varepsilon$ then we could define $G' = (N, \Sigma, P', S, F)$. In the opposite case we could introduce a new nonterminal S_0 and define $G' = (N \cup \{S_0\}, \Sigma, P' \cup \{S_0 \to S, S_0 \to \varepsilon\}, S_0, F)$. In either case, G' would be an ε-free BCFG without useless nonterminals which is equivalent to G. However, in worst case, the size of G' would be exponentially larger than the size of G. So we refine the above construction as follows.

Let us define the set

$$N' = N \cup \{[s] : A \to ps \in P \text{ for some } p \in (N \cup \Sigma)^*, s \in (N \cup \Sigma)^+\}$$

and the set P' consisting of the following rules:

1. For each $A \to p \in P$ with $p \neq \varepsilon$ let $A \to [p] \in P'$.
2. For each $[X_1 \dots X_n] \in N'$ with $n > 1$ let $[X_1 \dots X_n] \to X_1[X_2 \dots X_n] \in P'$.
3. For each $[X_1 \dots X_n] \in N'$ with $n > 1$ and $X_1 \in N_\varepsilon$ let $[X_1 \dots X_n] \to [X_2 \dots X_n] \in P'$.
4. For each $[X_1 \dots X_n] \in N'$ with $n \geq 1$ and $X_2, \dots, X_n \in N_\varepsilon$ let $[X_1 \dots X_n] \to X_1 \in P'$.

Finally, let $G' = (N', \Sigma, P', S, F)$ if $S \notin N_\varepsilon$, otherwise let $G' = (N' \cup \{S_0\}, \Sigma, P' \cup \{S_0 \to S, S_0 \to \varepsilon\}, S_0, F)$ for some fresh symbol S_0. Then G' is an ε-free BCFG constructed in polynomial time which is equivalent to G. □

If needed, we can eliminate chain productions as in Proposition 4.5.

Corollary 4.9. *For each BCFG G one can construct in polynomial time an equivalent ε-free BCFG G' without any chain productions which does not contain any useless nonterminals.*

As an application of ε-free BCFG's, we will show in the rest of this section that an ordinary language of finite words is context-free iff it is a BCFL. To this end, we will use the following fact.

Proposition 4.10. *Suppose that $G = (N, \Sigma, P, S, F)$ is an ε-free BCFG. Then $L^\infty(G) \cap \Sigma^* = L^*(G)$.*

Proof. We have $\varepsilon \in L^\infty(G)$ iff $S \to \varepsilon$ is a production. By ε-freeness, for each derivation tree whose frontier word is in Σ^+ there is a derivation tree with the same root symbol and the same frontier which is strictly locally finite. But any such tree t is necessarily finite. Indeed, if t is infinite then it has an infinite path π. Since t is strictly locally finite, there must be an infinite number of vertices that do not lie on π but have their predecessor on π. Since each such vertex is the root of a strictly locally finite tree, it follows that the frontier of t is infinite, a contradiction. □

Corollary 4.11. *A language $L \subseteq \Sigma^*$ is in BCFL iff L is in CFL.*

Corollary 4.12. *There is a polynomial time algorithm to decide for a BCFG $G = (N, \Sigma, P, S, F)$ and a finite word $w \in \Sigma^*$ whether $w \in L^\infty(G)$.*

Remark 4.13. Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG with $F = N$. By an argument similar to the proof of the well-known pumping lemma for ordinary context-free languages we show that if $L^\infty(G) \cap \Sigma^*$ is infinite, then $L^\infty(G)$ contains an infinite word. Indeed, by the above proofs, without loss of generality we may assume that G is ε-free without chain productions. Since $L^\infty(G) \cap \Sigma^*$ is infinite, there is a word $w \in L^\infty(G) \cap \Sigma^+$ with a strictly locally finite derivation tree rooted S such that at least one nonterminal is repeated along some path. This implies that w can be written as $xyuvz$ such that $yv \neq \varepsilon$ and for some nonterminal A we have $S \Rightarrow^* xAz$, $A \Rightarrow^* yAv$ and $A \Rightarrow^* u$. Since $F = N$ we have $A \in F$. Thus, $S \Rightarrow^\infty xy^\omega v^{-\omega} z$, showing that $L^\infty(G)$ contains the infinite word $xy^\omega v^{-\omega} z$.

5. Closure properties

In this section we establish the fact that BCFL's are effectively closed under substitution and use this result to derive the closure of BCFL's under the operations of union, concatenation, $^\omega$-power, $^{-\omega}$-power, $^\eta$-power and $^\infty$-power. Recall the definition of language substitution from Section 2.

Theorem 5.1. *If the languages L, L_a, $a \in \Sigma$ are BCFL's then so is $L' = L[a \leftarrow L_a]_{a \in \Sigma}$. Moreover, given BCFG's generating the languages L, L_a, $a \in \Sigma$, one can effectively construct a BCFG generating L'.*

Proof. Let $L = L^\infty(G)$ where $G = (N, \Sigma, P, S, F)$, and for each $a \in \Sigma$, let $L_a = L^\infty(G_a)$, where $G_a = (N_a, \Delta, P_a, S_a, F_a)$. Without loss of generality we may assume that the sets $N, N_a,\ a \in \Sigma$ are pairwise disjoint. Now let P' be the set of productions obtained from the productions in P by replacing each occurrence of each letter $a \in \Sigma$ with S_a. Then let

$$\overline{N} = N \cup \bigcup_{a \in \Sigma} N_a$$
$$\overline{F} = F \cup \bigcup_{a \in \Sigma} F_a$$
$$\overline{P} = P' \cup \bigcup_{a \in \Sigma} P_a$$

and $\overline{G} = (\overline{N}, \Delta, \overline{P}, S, \overline{F})$. The BCFG $\overline{G}$ generates the language $L[a \leftarrow L_a]_{a \in \Sigma}$. □

Corollary 5.2. *The class BCFL is effectively closed under binary set union, concatenation, $^\omega$-power, $^{-\omega}$-power, $^\eta$-power and $^\infty$-power.*

Thus, for example, given a BCFG generating L, one can effectively construct a BCFG generating L^η.

Example 5.3. For any ordinary context-free language $L \subseteq \Sigma^*$, $L^\omega, L^{-\omega}, L^\eta, L^\infty$ are BCFL's.

Theorem 5.4. *Suppose that L is a Büchi context-free language. Then L^r,* Pre(L), Suf(L), In(L) *and* Sub(L) *are all effectively Büchi context-free languages.*

Proof. Suppose that L is a BCFL generated by the BCFG $G = (N, \Sigma, P, S, F)$. It is clear that L^r is generated by the BCFG $G^r = (N, \Sigma, P^r, S, F)$ where $P^r = \{X \to p^r : X \to p \in P\}$.

Regarding Pre(L), let $\overline{N} = \{\overline{X} : X \in N\}$ and $\overline{F} = \{\overline{X} : X \in F\}$. Then consider the grammar $\mathrm{Pre}(G) = (N \cup \overline{N}, \Sigma, P \cup \overline{P}, \overline{S}, F \cup \overline{F})$,

$$\begin{aligned} \overline{P} = &\{\overline{X} \to p\overline{Y} : X, Y \in N, \exists q\, X \to pYq \in P\} \\ &\cup \{\overline{X} \to pa : X \in N, a \in \Sigma, \exists q\, X \to paq \in P\} \\ &\cup \{\overline{S} \to \varepsilon\}. \end{aligned}$$

If G contains no useless nonterminals and $L^\infty(G)$ is not empty, then $G' = \text{Pre}(G)$ generates the language $\text{Pre}(L)$. To see this, consider a derivation tree t over the grammar G whose root symbol is S and whose frontier word u is in Σ^∞. If v is a prefix of u then we can partition the set of leaves into two disjoint sets K and R such that K is closed below and R is closed above with respect to the lexicographic order and such that v is isomorphic to the word determined by K. If K is empty then v is the empty word and since $\overline{S} \to \varepsilon$ is a production of G' we have $v = \varepsilon \in L^\infty(G')$. Assume now that K is not empty. Using t, we will construct a derivation tree for v over the grammar G'. To this end, let us relabel the root by $\overline{S}$. Then suppose that we have relabeled a vertex x originally labeled $X \in N$ by $\overline{X}$ such that every vertex in K is either lexicographically less than x or belongs to $t|_x$, moreover, $t|_x$ contains at least one leaf in K. Consider the successors $x_1, \ldots, x_m$ of x. There is a largest integer i such that the subtree $t|_{x_i}$ rooted at x_i contains a leaf in K. If x_i is labeled in Σ, or $i = 1$ and x_i is labeled ε, then x_i is the lexicographically greatest element of K. A derivation tree may be obtained from the relabeled tree by removing all vertices lexicographically greater than x_i. If x_i is labeled by a nonterminal Y then we relabel it $\overline{Y}$ and continue the process. If the process does not stop, then the vertices which are relabeled form a complete path π and a leaf belongs to K iff it is lexicographically less than π. A derivation tree of v over G' can be obtained by removing all vertices lexicographically greater than π.

Suppose now that t is a derivation tree over G' with root symbol $\overline{S}$ and frontier word v in Σ^∞. Then the inner vertices of t on the rightmost complete path π are labeled by nonterminals in $\overline{N}$ and all other inner vertices are labeled in N. Suppose that x is an inner vertex lying on π labeled $\overline{X}$. Let $x_1, \ldots, x_i$ denote the successors of x labeled $p_1, \ldots, p_i$, respectively. If $p_i = \overline{Y}$ is in $\overline{N}$ then there are some $q_1, \ldots, q_j \in N \cup \Sigma$ such that $X \to p_1 \ldots p_{i-1} Y q_1 \ldots q_j$ is a production of G. In this case let us add j new successors of x to the tree, labeled $q_1, \ldots, q_j$, respectively. If p_i is a terminal or $i = 1$ and $p_i = \varepsilon$, then x_i is the last vertex of π. Moreover, there exist $q_1, \ldots, q_j \in N \cup \Sigma$ such that $X \to p_1 \ldots p_i q_1 \ldots q_j$ is in P. We add j new successors of x to the tree, labeled $q_1, \ldots, q_j$, respectively. Replacing each vertex label $\overline{X}$ with X, the tree constructed in this way is a derivation tree over G whose frontier word is of the form vq for some $q \in (N \cup \Sigma)^\infty$. Since G contains no useless nonterminals, this tree can be completed to a derivation tree whose frontier word u is in Σ^∞. It is clear that v is a prefix of u.

Since $\text{Suf}(L) = (\text{Pre}(L^r))^r$ and $\text{In}(L) = \text{Suf}(\text{Pre}(L))$, it follows now that $\text{Suf}(L)$ and $\text{In}(L)$ are also BCFL's.

Last, we prove that $\text{Sub}(L)$ is a BCFL. For this reason, without loss of generality we may assume that whenever a terminal letter a occurs on the right side of a production, then the production is of the form $X \to a$. If G satisfies this condition, then a grammar generating $\text{Sub}(L)$ is obtained by adding all productions $X \to \varepsilon$ to the set P whenever $X \to a$ is in P for some $a \in \Sigma$. $\square$

Proposition 5.5. *For every alphabet Σ, the set of all dense words in Σ^∞ and the set of all quasi-dense words in Σ^∞ are BCFL's.*

Proof. We use Lemma 2.2. The set of all dense words in Σ^∞ can be given by the expression

$$(\Sigma \cup \{\varepsilon\})\Sigma^\eta(\Sigma \cup \{\varepsilon\}).$$

And a word in Σ^∞ is quasi-dense iff it is in the set

$$(L \cup \{\varepsilon\})L^\eta(L \cup \{\varepsilon\}),$$

where L is the set of all nonempty scattered words in Σ^∞. But even if L is the set of all nonempty words in Σ^∞, all words in the above language are quasi-dense. Since the set of all nonempty words $\Sigma^{+\infty} = \Sigma^\infty \Sigma \Sigma^\infty$ is a BCFL, the result follows. $\square$

Later we will prove that neither the set of all scattered words in Σ^∞ nor the set of all well-ordered words in Σ^∞ is a BCFL.

Remark 5.6. Since a language of finite words $L \subseteq \Sigma^*$ is a BCFL iff it is a CFL, and since CFL's are not closed under intersection, it follows that BCFL's are not closed under complementation and intersection either.

6. Some decidable properties of Büchi context-free languages

In this section our aim is to prove that the following properties are decidable for a Büchi context-free language L given by a BCFG.

- L is empty.
- L consists of finite words.
- L consists of infinite words.
- L consists of ω-words.
- L consists of well-ordered words.
- L consists of scattered words.
- L consists of dense words.

In each case, a polynomial time algorithm is obtained. We also establish a limitedness property of BCFL's.

Let $G = (N, \Sigma, P, S, F)$ be a BCFG. We define a directed graph Γ_G whose set of vertices is N. There is an edge $A \to B$ exactly when B occurs on the right side of a production whose left side is A. We partition N into strongly connected components. As usual, the strongly connected components can be partially ordered by $\mathcal{S} \leq \mathcal{S}'$ iff there is a sequence of nonterminals $A_0, \ldots, A_m$ such that $A_0 \in \mathcal{S}'$, $A_m \in \mathcal{S}$ and for each $i < m$ there is an edge from A_i to A_{i+1}.

Our first decidability result is immediate from the following theorem.

Theorem 6.1. *It is decidable in polynomial time whether a BCFG generates an empty language.*

Proof. Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG. By Proposition 4.2, we can construct in polynomial time a BCFG $G' = (N', \Sigma, P', S', F')$ which is equivalent to G and contains no useless nonterminals. Now $L^\infty(G)$ is empty iff P' is empty. □

Next we show that it is decidable whether a BCFL contains an infinite word.

Theorem 6.2. *Let $G = (N, \Sigma, P, S, F)$ be a weakly ε-free BCFG having no useless nonterminal. Then $L^\infty(G)$ contains an infinite word iff there is a strongly connected component $\mathcal{S}$ of Γ_G which contains a nonterminal in F, and there is a production $A \to p$ with $A \in \mathcal{S}$ such that $|p| \geq 2$ and at least one nonterminal in $\mathcal{S}$ occurs in p.*

Proof. Suppose first that there is a strongly connected component $\mathcal{S}$ satisfying the above condition. Let $B \in \mathcal{S} \cap F$. Since for some $A \in \mathcal{S}$ there is a production $A \to p$ such that $|p| \geq 2$ and p contains a nonterminal in $\mathcal{S}$, there exist words $q, r \in (N \cup \Sigma)^*$ with $qr \neq \varepsilon$ and $B \Rightarrow^* qBr$. Since G is weakly ε-free and contains no useless nonterminals, $B \Rightarrow^\infty uBv$ holds for some words $u, v \in \Sigma^\infty$ with $uv \neq \varepsilon$. Thus $B \Rightarrow^\infty u^\omega v^{-\omega}$, and since there are no useless nonterminals, it follows that $L^\infty(G)$ contains an infinite word of the form $u_0 u^\omega v^{-\omega} v_0$.

Suppose now that $L^\infty(G)$ contains an infinite word w. Consider a derivation tree t for w with root label S. Starting from the root, we can construct an infinite path π such that for each vertex x along π the frontier word of the subtree $t|_x$ is infinite. It is clear that there are infinitely many vertices along π that have at least two successors. We can thus find vertices x, y, z on π with $x < y < z$ in the prefix order such that x, z are labeled by the same nonterminal $B \in F$ and y has at least two successors. Now let $\mathcal{S}$ be the strongly connected component of B and let p be the right side of the production used to rewrite the nonterminal A at vertex y. □

Corollary 6.3. *It is decidable in polynomial time whether the language $L^\infty(G)$ generated by a given BCFG G consists of finite words.*

Our next task is to show that there exists a polynomial time algorithm to decide whether a BCFL specified by a BCFG contains only infinite words.

Theorem 6.4. *It is decidable in polynomial time whether the language $L^\infty(G)$ generated by a given BCFG $G = (N, \Sigma, P, S, F)$ contains only infinite words.*

Proof. In polynomial time, we compute as in Proposition 4.8 the set Y of all nonterminals A such that $A \Rightarrow^\infty \varepsilon$. Then we add all productions $A \to \varepsilon$ to P where $A \in Y$. Let G' denote the resulting grammar (N, Σ, P', S, F). We have that $L^\infty(G) \cap \Sigma^* = L^\infty(G') \cap \Sigma^* = L^*(G')$. Thus, $L^\infty(G)$ contains only infinite words iff $L^*(G')$ is empty. This latter problem is decidable in polynomial time since it is decidable in polynomial time for an ordinary CFG whether it generates the empty language. □

In our proofs below, we will make use of the notion of the *rank* of a scattered countable word, related to the Hausdorff rank of countable linear orders, cf. [26]. Let Σ be an alphabet. We define the sequence $(V_\alpha^\Sigma)_\alpha$ of subsets of Σ^∞, where α ranges over all countable ordinals. Let $V_0^\Sigma = \Sigma^*$. Then for any countable ordinal $\alpha > 0$, let V_α^Σ be the least set of words closed under finite concatenation which contains $\bigcup_{\beta<\alpha} V_\beta^\Sigma$ together with all words of the form $u_0 u_1 \ldots u_i \ldots$ and $\ldots u_i \ldots u_1 u_0$, where each u_i, $i < \omega$ is in $V_{\beta_i}^\Sigma$ for some β_i with $\beta_i < \alpha$.

The following fact is immediate from Hausdorff's theorem [26].

Proposition 6.5. *A word in Σ^∞ is scattered iff it belongs to V_α^Σ for some countable ordinal α.*

Definition 6.6. The *rank* of a scattered word w in Σ^∞ is the least ordinal α such that w is in V_α^Σ. If this ordinal is finite we say that w is of finite rank.

Example 6.7. Consider the following languages over the singleton alphabet. Let $L_0 = \{a\}$ and $L_{n+1} = \{w^\omega, w^{-\omega} : w \in L_n\}$, for all $n < \omega$. Then for each n and for each word $w \in L_n$, we have that w is scattered of rank n. In particular, let $w_0 = a$ and $w_{n+1} = w_n^\omega$, for all $n < \omega$. Then each w_n is scattered of rank n.

For later use we establish two properties of the rank.

Lemma 6.8. *Suppose that $u \in \Sigma^\infty$ is scattered of rank α. If v can be embedded in u then v is scattered of rank at most α.*

Proof. We argue by induction on α. When $\alpha = 0$ our claim is clear. So suppose that $\alpha > 0$. Then u is a finite concatenation of words of the form $u_0 u_1 \ldots$ or $\ldots u_1 u_0$, where each u_i, $i < \omega$ is scattered of rank $< \alpha$. Thus, if v embeds in u, then v is a finite concatenation of words of the form $v_0 v_1 \ldots$ or $\ldots v_1 v_0$, where each v_i embeds in u_i. Thus, by the induction hypothesis, each v_i is scattered of rank $< \alpha$ and thus v is also scattered of rank at most α. □

By Lemma 2.2, if a word w can be constructed from a scattered word u by replacing each occurrence of a letter of u with a scattered word, then w is also scattered.

Lemma 6.9. *Suppose that the word $u \in \Delta^\infty$ is scattered of rank at most 1. Moreover, suppose that the word $w \in \Sigma^\infty$ can be constructed from u by replacing each letter with a scattered word of rank at most n. Then w is also scattered and its rank is at most $n+1$. Moreover, if $u_0, u_1, \ldots$ are scattered words of finite rank at most n, and if there exist infinitely many i such that u_i is nonempty of rank n, then both $u_0u_1\ldots$ and $\ldots u_1u_0$ are scattered words of rank $n+1$.*

Proof. Since both claims formulate properties of the underlying linear order of the words, it suffices to prove our Lemma in the case when Σ and Δ are the unary alphabet $\{a\}$.

To prove the first claim, suppose that u is scattered of rank at most 1. Moreover, suppose that the word w can be constructed from u by replacing each letter with a scattered word of rank at most n. When u is finite then w is scattered of rank at most n. If $u = a^\omega$ or $u = a^{-\omega}$ then w is scattered of rank at most $n+1$. Otherwise u is a finite concatenation of finite words and the words a^ω and $a^{-\omega}$. Since any finite concatenation of scattered words of rank at most $n+1$ is scattered of rank at most $n+1$, it follows that w is scattered of rank at most $n+1$.

For each $n < \omega$, define L_n as follows. Let $L_0 = \{a\}$ and $L_{n+1} = \{w^\omega, w^{-\omega} : w \in L_n\}$, for all $n < \omega$. (See Example 6.7.) Assume that $u_0, u_1, \ldots$ are scattered words of finite rank at most n such that there exist infinitely many i such that u_i is nonempty of rank n. We show that there is a word in L_{n+1} which can be embedded in $u_0u_1\ldots$, and similarly for $\ldots u_1u_0$. Suppose that $n = 0$. Then $u_0u_1\ldots = a^\omega$ and $\ldots u_1u_0 = a^{-\omega}$ and our claim holds obviously. We proceed by induction on n. So suppose that $n > 0$ and that our claim holds for all $m < n$. Consider the word $u_0u_1\ldots$, the argument is similar for $\ldots u_1u_0$. Let I denote the infinite set of all $i < \omega$ such that u_i has rank n. By the induction assumption, for each $i \in I$ there is a word in L_n that can be embedded in u_i. But since L_n is finite, there is an infinite set $J \subseteq I$ and a word $w \in L_n$ that can be embedded in each u_j with $j \in J$. This implies that $w^\omega \in L_{n+1}$ can be embedded in $u_0u_1\ldots$.

By Example 6.7, each word in each L_n is scattered of rank n. Thus, by Lemma 6.8, $u_0u_1\ldots$ and $\ldots u_1u_0$ are scattered of rank at least $n+1$. On the other hand, both words are of rank at most $n+1$, so that they have rank $n+1$. □

Below we will also make use of the following simple fact.

Lemma 6.10. *The single word z_0 generated by the grammar*

$$(\{S\}, \{a, b, c\}, S, \{S \to aSbSc\}, \{S\})$$

is quasi-dense. In fact, the word b^η can be embedded in z_0.

The following result will be the key to prove that it is decidable for a BCFL whether it contains only scattered words.

Theorem 6.11. *Let $G = (N, \Sigma, P, S, F)$ be a weakly ε-free BCFG with no useless nonterminals. Then $L^\infty(G)$ consists of scattered words iff for each strongly connected component $\mathcal{S}$ of Γ_G with $\mathcal{S} \cap F \neq \emptyset$ and for each production $A \to p$ with $A \in \mathcal{S}$, the word p contains at most one occurrence of a nonterminal in $\mathcal{S}$.*

Proof. For one direction, suppose that $\mathcal{S}$ is a strongly connected component with $\mathcal{S} \cap F \neq \emptyset$ and for some $A \in \mathcal{S}$ there is a rule $A \to s$ such that s contains two or more occurrences of nonterminals in $\mathcal{S}$. This means that s has a decomposition $s = p_0A_1p_1A_2p_2$ with $A_1, A_2 \in \mathcal{S}$. Let B denote a nonterminal in $\mathcal{S} \cap F$. Since B is in $\mathcal{S}$, there exist finite derivations $B \Rightarrow^* q_0Aq_1$ and $A_i \Rightarrow^* r_{i,0}Br_{i,1}$, $i = 1, 2$. Thus, we have

$$B \Rightarrow^* q_0p_0r_{1,0}Br_{1,1}p_1r_{2,0}Br_{2,1}p_2q_1 = pBqBr.$$

Thus, also

$$B \Rightarrow^* ppBqBrqBr = p'Bq'Br'$$

where $p' = pp$, $q' = qBrq$, $r' = r$. Since G contains no useless nonterminals, there exist words $u, w \in \Sigma^\infty$ with $p' \Rightarrow^\infty u$ and $r' \Rightarrow^\infty w$. Since G contains no useless nonterminals and since G is weakly ε-free, there exists a nonempty word $v \in \Sigma^\infty$ with $q' \Rightarrow^\infty v$. Summing up,

$$B \Rightarrow^\infty uBvBw,$$

where v is nonempty. This implies that the word $z = z_0[a \leftarrow u, b \leftarrow v, c \leftarrow w]$ obtained by substituting u, v, w respectively for the letters a, b, c in the word z_0 of Lemma 6.10 can be generated from B. Since v^η embeds in z and is quasi-dense, so is z. Using the fact that there exists a derivation $S \Rightarrow^\infty u_0Av_0$, for some $u_0, v_0 \in \Sigma^\infty$, it follows that $L^\infty(G)$ contains the quasi-dense word u_0zv_0.

For the other direction, suppose that the condition stated in the theorem holds for each strongly connected component of Γ_G. Let n denote the number of nonterminals. We will argue by induction on n to show that each word in $L^\infty(G)$ is scattered of rank at most n. When S is the single nonterminal then there are two cases. If F is empty, then $L \subseteq \Sigma^*$. If $F = \{S\}$, then the right side of each production contains at most one occurrence of S. Thus every infinite derivation tree has a single infinite branch and thus by Lemma 6.9, any infinite word in $L^\infty(G)$ is scattered of rank 1.

In the induction step, suppose that $n > 1$. Consider a derivation tree t whose root symbol is S and whose frontier word w is in Σ^∞. If w is finite then it is scattered of rank 0, so assume that w is infinite. Then t is also infinite. Along each infinite path of t there is a first occurrence of a vertex labeled in F. Let X denote the set of all such vertices. Let us remove all vertices strictly below a vertex in X. Then the resulting derivation tree t_0 is finite, for if it was infinite, then by König's lemma it would contain an infinite path. Thus X is also finite. Note also that each vertex of the frontier of t_0 is either in X or a leaf labeled in $\Sigma \cup \{\varepsilon\}$.

Consider a vertex $x \in X$ labeled $A \in F$. Suppose that A occurs an infinite number of times in the tree $t|_x$. Then by assumption, all occurrences of A in $t|_x$ lie on the same infinite path π. Let Y denote the set of those vertices which do not lie on π but whose predecessor lies on π, and let u denote the word in $(N \cup \Sigma)^\infty$ formed by the labels of the vertices in Y. Then u is a scattered word of rank at most 1. Now if some $y \in Y$ is labeled in N, then $t|_y$ has no vertex labeled A. So by induction, the frontier of $t|_y$ determines a scattered word in Σ^∞ of rank at most $n-1$. If y is labeled in Σ, then the frontier of $t|_y$ determines a 1-letter word which is scattered of rank 0. Since u is scattered of rank at most 1 and the word determined by the frontier of $t|_x$ is obtained by replacing each letter of u by a scattered word of rank at most $n-1$, the frontier word of $t|_x$ is scattered of rank at most n.

Suppose now that A occurs a finite number of times as the label of a vertex of $t|_x$. In this case consider the finite tree whose vertices are those lying on a path from x to a vertex labeled A together with all successors of these vertices. For every leaf y of this tree, the frontier of $t|_y$ determines a scattered word of rank at most $n-1$. Since the frontier word of $t|_x$ is a finite concatenation of these words, it is scattered of rank at most $n-1$.

We have thus proved that for every $x \in X$, the frontier word of $t|_x$ is scattered of rank at most n. Let u_0 denote the frontier of t_0. Since w can be obtained from u_0 by replacing each occurrence of a letter which is a nonterminal in F by a scattered word of rank at most n, and since u_0 is finite, w is scattered of rank at most n. □

Example 6.12. Let Σ be any alphabet. Then for each $n < \omega$, the set L_n of all scattered words in Σ^∞ of rank at most n is a BCFL. Indeed, $L_0 = \Sigma^*$, and $L_{n+1} = (L_n^\omega \cup L_n^{-\omega})^*$.

Corollary 6.13. *It is decidable in polynomial time whether the language $L^\infty(G)$ generated by a given BCFG G contains only scattered words.*

The above corollary generalizes a result from [6].

Corollary 6.14. *Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG such that $L^\infty(G)$ contains only scattered words. Then the rank of each word in $L^\infty(G)$ is at most the number of nonterminals in N.*

Corollary 6.15. *Let $w_0 = a$ and $w_{n+1} = (w_n)^\omega$ for all $n < \omega$. There exists no BCFL consisting only of scattered words containing all words w_n, for all $n < \omega$. In particular, for any alphabet Σ, the set of all scattered words in Σ^∞ is not a BCFL. Similarly, the set of all well-ordered words in Σ^∞ is not a BCFL.*

Corollary 6.16. *For every alphabet Σ, including the singleton alphabet, the set of all BCFL's in Σ^∞ is not closed under complementation.*

Proof. The language of all quasi-dense words in Σ^∞ is a BCFL, while its complement, the language of all scattered words in Σ^∞ is not. □

By Corollary 6.14, it is natural to ask the following question. Suppose that G is a BCFG such that $L^\infty(G)$ contains only scattered words. Compute the maximal number n such that $L^\infty(G)$ contains a scattered word of rank n. This question motivates the following definition.

Definition 6.17. Suppose that $L \subseteq \Sigma^\infty$ consists of scattered words of finite rank bounded by some $n < \omega$. Then we define the rank of L as the maximum rank of a word in L.

By Theorem 6.11, the above rank is finite for BCFLs.

Theorem 6.18. *There is a polynomial time algorithm to compute the rank of a BCFL of scattered words generated by a BCFG.*

Proof. Let $G = (N, \Sigma, P, S, F)$ be a BCFG. Without loss of generality we may assume that G is weakly ε-free and contains no useless nonterminals. Thus, Γ_G satisfies the condition stated in Theorem 6.11. By Lemma 6.8, whenever A and B belong to the same strongly connected component and if n is the maximum rank of a scattered terminal word derivable from A, then n is also the maximum rank of a scattered terminal word derivable from B.

For each strongly connected component $\mathcal{S}$ of Γ_G, we define $\#(\mathcal{S})$ as follows. Suppose that $\mathcal{S}$ is minimal with respect to the ordering of the strongly connected components. Assume that $\mathcal{S}$ contains a nonterminal in F. Then for each production $A \to p$ with A in $\mathcal{S}$ either p contains a single occurrence of a nonterminal (which is then in $\mathcal{S}$), or p is a terminal word. If there is a production $A \to p$ with $A \in \mathcal{S}$ such that p contains a nonterminal but is not a single nonterminal, then we define $\#(\mathcal{S}) = 1$, since a scattered word in Σ^∞ of rank 1 can be derived from each nonterminal in $\mathcal{S}$, and no scattered word of rank >1 is derivable. Otherwise, if there is no such production, then we define $\#(\mathcal{S}) = 0$, since only finite terminal words are derivable from any nonterminal in $\mathcal{S}$. In the same way, if $\mathcal{S}$ contains no nonterminal in F, then we define $\#(\mathcal{S}) = 0$.

Suppose now that $\mathcal{S}$ is not a minimal strongly connected component and that $\#(\mathcal{S}')$ has already been defined for all strongly connected components strictly below $\mathcal{S}$. Moreover, suppose that for each such strongly connected component $\mathcal{S}'$ strictly below $\mathcal{S}$, $\#(\mathcal{S}') = n$ iff for each $A \in \mathcal{S}'$ there is a scattered terminal word of rank n derivable from A but no scattered word of rank $>n$ is derivable. There are two cases.

Case 1. Assume that $\mathcal{S}$ contains at least one nonterminal in F and there is a production whose left side is in $\mathcal{S}$ and whose right side is of length at least 2 and contains a nonterminal in $\mathcal{S}$. In this case we compute two values and define $\#(\mathcal{S})$ as the maximum of the two. The first value is $k+1$, where k is the maximum of all numbers $\#(S')$ such that $\mathcal{S}' \neq \mathcal{S}$ and there is a production whose left side is in $\mathcal{S}$ and whose right side contains both a nonterminal in $\mathcal{S}$ and a nonterminal in $\mathcal{S}'$. (We agree that the maximum of the empty set is 0.) The second value is ℓ, where ℓ is the maximum of the numbers $\#(\mathcal{S}')$ such that there is a production whose left side is in $\mathcal{S}$ and whose right side contains a nonterminal in $\mathcal{S}'$ but no nonterminal in $\mathcal{S}$.

Case 2. If the previous condition does not hold, then consider all values $\#(\mathcal{S}')$ such that $\mathcal{S}'$ is strictly below $\mathcal{S}$ and there is a production whose left side is in $\mathcal{S}$ and whose right side contains a nonterminal in $\mathcal{S}'$. We define $\mathcal{S}$ as the maximum of these numbers.

In either case, it is clear by Lemma 6.9 that $\#(\mathcal{S}) = n$ iff for each nonterminal A in $\mathcal{S}$ there is a derivation of a scattered terminal word of rank n from A, but no scattered terminal word of rank $>n$ is derivable. $\square$

Our next aim is to show that it is decidable in polynomial time whether a BCFG generates a language of well-ordered words.

Theorem 6.19. *Let $G = (N, \Sigma, P, S, F)$ be a weakly ε-free BCFG with no useless nonterminals. Then $L^\infty(G)$ contains only well-ordered words iff for each strongly connected component $\mathcal{S}$ of Γ_G which contains a nonterminal in F and for each production $A \to p$ with $A \in \mathcal{S}$, if p contains a letter which is a nonterminal in $\mathcal{S}$ then it contains a single such letter, and moreover, this letter is the rightmost letter of p.*

Proof. Suppose first that there is a strongly connected component $\mathcal{S}$ which contains a nonterminal in F but violates the above condition. It is easy to see that there is a nonterminal $A \in \mathcal{S} \cap F$ and words $p, q \in (N \cup \Sigma)^*$ such that q is nonempty and $A \Rightarrow^* pAq$. Since G does not contain useless nonterminals and since G is weakly ε-free, there exist words $u, v \in \Sigma^\infty$ with $v \neq \varepsilon$ such that $A \Rightarrow^\infty uAv$. Thus, $A \Rightarrow^\infty u^\omega v^{-\omega}$. Using again the assumption that G has no useless nonterminals, we have that $S \Rightarrow^\infty u_0 u^\omega v^{-\omega} v_0$ for some $u_0, v_0 \in \Sigma^\infty$. Since v is nonempty, this word is not well-ordered.

Suppose now that the condition of the theorem holds for each strongly connected component. Let n denote the number of nonterminals. We show by induction on n that each word in $L^\infty(G)$ is well-ordered. This is clear when S is the only nonterminal. So assume that $n > 1$ and consider a derivation tree t whose root is labeled S and whose frontier word w is in Σ^∞. Let us define X and t_0 in the same way as in the proof of Theorem 6.11.

Consider a vertex $x \in X$ labeled $A \in F$. Suppose that A occurs an infinite number of times in the tree $t|_x$. Then by assumption, all occurrences of A in $t|_x$ lie on the same infinite path π. Let Y denote the set of those vertices which do not lie on π but whose predecessors belong to π, and let u denote the word in $(N \cup \Sigma)^\infty$ formed by the labels of the vertices in Y. Then u is either finite or an ω-word. If some $y \in Y$ is labeled in N, then $t|_y$ has no vertex labeled A. So by induction, the frontier of $t|_y$ determines a well-ordered word in Σ^∞. If y is labeled in Σ, then the frontier of $t|_y$ determines a 1-letter word. Since u is well-ordered and the word w determined by the frontier of $t|_x$ is obtained by replacing each letter of u by a well-ordered word, and since any well-ordered generalized product of well-ordered words is well-ordered (Lemma 2.2), the frontier word of $t|_x$ is well-ordered.

Suppose now that A occurs a finite number of times as the label of a vertex in $t|_x$. In this case consider the finite tree whose vertices are those lying on a path from x to a vertex labeled A together with all successors of these vertices. For every leaf y of this tree, the frontier of $t|_y$ determines a well-ordered word. Since the frontier word of $t|_x$ is a finite concatenation of these words, it is also well-ordered.

We have thus proved that for every $x \in X$, the frontier word of $t|_x$ is well-ordered. Let u_0 denote the frontier of t_0. Since w can be obtained from u_0 by replacing each occurrence of a letter which is a nonterminal in F by a well-ordered word, and since u_0 is finite, w is well-ordered. $\square$

Corollary 6.20. *It is decidable in polynomial time whether the language $L^\infty(G)$ generated by a given BCFG G contains only well-ordered words.*

The next result answers the following question: When does it hold that a weakly ε-free BCFG generates a language of well-ordered words of order type at most ω?

Theorem 6.21. *Suppose that $G = (N, \Sigma, P, S, F)$ is a weakly ε-free BCFG without useless nonterminals and chain productions. Then $L^\infty(G)$ consists of finite and ω-words iff the following holds: Whenever $\mathcal{S}$ is a strongly connected component of Γ_G containing a nonterminal in F such that for at least one production whose left side is in $\mathcal{S}$, the right side of the production contains a nonterminal in $\mathcal{S}$, and whenever $A \in \mathcal{S}$, then there is no finite derivation $S \Rightarrow^* pAp'$ for any words $p, p' \in (N \cup \Sigma)^*$ such that $p' \neq \varepsilon$.*

Proof. Suppose that $L^\infty(G)$ consists of finite and ω-words. Suppose that $\mathcal{S}$ is a strongly connected component containing a nonterminal in F such that for at least one production whose left side is in $\mathcal{S}$, the right side of the production contains a nonterminal in $\mathcal{S}$. Assume to the contrary that $S \Rightarrow^* pAp'$ holds for some $A \in \mathcal{S}$ and words $p, p' \in (N \cup \Sigma)^*$ such that $p' \neq \varepsilon$. Since $\mathcal{S} \cap F \neq \emptyset$, without loss of generality we may assume that $A \in F$. By the assumption on $\mathcal{S}$ and since G is free of chain productions, there exist words $q, r \in (N \cup \Sigma)^*$ with $qr \neq \varepsilon$ and $A \Rightarrow^* qAr$. Since G is weakly ε-free and contains no useless nonterminals, there exist words $u, u', v, w \in \Sigma^\infty$ such that u' and vw are nonempty and the following derivations exist: $p \Rightarrow^\infty u$, $p' \Rightarrow^\infty u'$, $q \Rightarrow^\infty v$, $r \Rightarrow^\infty w$. Thus, $S \Rightarrow^\infty uv^\omega w^{-\omega} u'$. If $w \neq \varepsilon$, this word is not well-ordered. Thus $w = \varepsilon$. But then by $vw \neq \varepsilon$ we have that $v \neq \varepsilon$, and since $u' \neq \varepsilon$, $uv^\omega w^{-\omega} u'$ is not an ω-word.

Suppose now that Γ_G satisfies the condition in our Theorem. Then consider any derivation tree whose root is labeled S and whose frontier word is infinite. Then clearly, for every infinite path π of t and for every vertex x along π, the last successor of x is on the path π. This also implies that there is a single infinite path originating at the root, and that any other complete path is strictly less than this infinite path with respect to the lexicographic order. This clearly implies that w is an ω-word. $\square$

Corollary 6.22. *It can be decided in polynomial time whether the language generated by a BCFG contains only finite or ω-words.*

Corollary 6.23. *It can be decided in polynomial time whether the language generated by a BCFG contains only ω-words.*

The last property we are going to deal with in this section is that of denseness. We will show that it is decidable for a BCFG whether it generates a language whose words are all dense. We need a preliminary fact.

Proposition 6.24. *The following are decidable in polynomial time for a BCFG G:*

1. *Does $L^\infty(G)$ contain a word having a first letter?*
2. *Does $L^\infty(G)$ contain a word having a last letter?*
3. *Does $L^\infty(G)$ contain a word having two consecutive letters?*

Proof. Without loss of generality we may assume that G contains no useless nonterminals. Moreover, we may assume that whenever $A \Rightarrow^\infty \varepsilon$ holds for some $A \in N$, then $A \to \varepsilon$ is in P (see the proof of Proposition 4.8). It is well-known (see e.g., [1]) that the following relations can be computed in polynomial time. For any $A, B \in N \cup \Sigma$,

1. A *FIRST* B iff there is a finite derivation $A \Rightarrow^* Bp$ for some $p \in (N \cup \Sigma)^*$.
2. A *LAST* B iff there is a finite derivation $A \Rightarrow^* pB$ for some $p \in (N \cup \Sigma)^*$.
3. A *FOLLOW* B if there is a finite derivation $S \Rightarrow^* pABq$ for some $p, q \in (N \cup \Sigma)^*$.

Now $L^\infty(G)$ contains a word having a first letter iff S *FIRST* a holds for some $a \in \Sigma$, and symmetrically, $L^\infty(G)$ contains a word having a last letter iff S *LAST* a holds for some $a \in \Sigma$. Finally, $L^\infty(G)$ contains a word having two consecutive letters iff a *FOLLOW* b holds for some $a, b \in \Sigma$. $\square$

Theorem 6.25. *It is decidable in polynomial time for a BCFG G whether each word in $L^\infty(G)$ is dense.*

Proof. $L^\infty(G)$ contains only dense words iff it contains only infinite words, moreover, it does not contain any word having two consecutive letters. Both conditions are decidable in polynomial time. $\square$

7. A comparison

An ω-language is a subset of Σ^ω, where Σ is any alphabet. In this section, we compare the class of regular ω-languages [24,28] and the class of context-free ω-languages as defined by Cohen and Gold [15] with the class of those ω-languages that are BCFL's.

Recall from [24] that a Büchi automaton is a system $\mathbf{A} = (Q, \Sigma, \delta, q_0, F)$ which consists of an alphabet Q of states, an alphabet Σ of letters, a transition relation $\delta \subseteq Q \times \Sigma \times Q$, an initial state $q_0 \in Q$ and a set F of designated states. A run of the automaton $\mathbf{A}$ on a word $w = a_0a_1 \ldots \in \Sigma^\omega$ is a sequence of states $q_0, q_1, \ldots$ where q_0 is the initial state and $(q_i, a_i, q_{i+1}) \in \delta$ holds for all i. Moreover, it is required that at least one state in F occurs infinitely often in the run. The language $L(\mathbf{A})$ accepted by $\mathbf{A}$ is the set of all words in Σ^ω that have at least one run.

Definition 7.1. A language $L \subseteq \Sigma^\omega$ is regular if there is a Büchi automaton accepting L.

Proposition 7.2. *Let $G = (N, \Sigma, P, S, F)$ be a BCFG. Consider the collection of all derivation trees whose root symbol is labeled S and whose rightmost (complete) path is infinite, and let $K \subseteq N^\infty$ denote the set of all ω-words over N that are label sequences of such rightmost paths. Then K is a regular language.*

Proof. Consider the Büchi automaton $\mathbf{A} = (\{q_A : A \in N\}, N, \delta, q_S, \{q_A : A \in F\})$ where

$$\delta = \{(q_A, A, q_B) : \exists p\ A \to pB \in P\}.$$

Then $L(\mathbf{A})$ is the ω-language described in the proposition. $\square$

Proposition 7.3. *Every regular language $L \subseteq \Sigma^\omega$ is a BCFL.*

Proof. When $L = L(\mathbf{A})$ where $\mathbf{A} = (Q, \Sigma, \delta, q_0, F)$, define $G = (Q, \Sigma, P, q_0, F)$ with $P = \{q \to aq' : (q, a, q') \in \delta\}$. $\square$

A notion of context-free languages of ω-words was introduced in [15]. The following result was shown in [23], see also [9].

Theorem 7.4. *A language $L \subseteq \Sigma^\omega$ is a context-free language in the sense of Cohen and Gold [15] iff there is a regular language $K \subseteq \Delta^\omega$ and ordinary context-free languages $K_b \subseteq \Sigma^*$, $b \in \Delta$ such that $L = K[b \leftarrow K_b]_{b \in \Delta}$.*

Using this result we now prove:

Theorem 7.5. *The following two conditions are equivalent for a language $L \subseteq \Sigma^\omega$.*

1. *L is a BCFL.*
2. *L is context-free in the sense of Cohen and Gold [15].*

Proof. Suppose that $L = L^{\infty}(G)$, where G is a BCFG (N, Σ, P, S, F) which can be assumed to be weakly ε-free with no chain productions and no useless nonterminals. By the proof of Theorem 6.21, every infinite derivation tree rooted S whose frontier is an infinite word in Σ^{∞} has a single infinite path, the rightmost complete path. Let us relabel each vertex on this rightmost path by the finite word obtained by concatenating the labels of those successors of the vertex that do not lie on the rightmost path. Similarly to the proof of Proposition 7.2, the ω-words in $(N \cup \Sigma)^{\omega}$ that appear as these modified label sequences of such rightmost paths form a regular language K. For each $A \in N$, let $K_A \subseteq \Sigma^*$ be the language $L(G_A)$, where G_A is the ordinary context-free grammar (N, Σ, P, A). Moreover, for each letter $a \in \Sigma$ let $K_a = \{a\}$. Clearly, $L = K[A \leftarrow K_A, a \leftarrow K_a]_{A \in N, a \in \Sigma}$. Thus L is a context-free language in the sense of Cohen and Gold.

Suppose now that L is context-free in the sense of Cohen and Gold. Then there is a regular language $K \subseteq \Delta^{\omega}$ and context-free languages $K_b \subseteq \Sigma^*$, $b \in \Delta$ such that $L = K[b \leftarrow K_b]_{b \in B}$. Since K is regular, it is a BCFL by Proposition 7.3. Also, each K_b is a BCFL, so that L is a BCFL. $\square$

Remark 7.6. The papers [10,4] define finite automata acting on infinite words and using this automaton model, provide a definition of recognizable languages of both countable words and all words with no upper bound on the cardinality of the word. Here we briefly compare BCFL's with the class REC of recognizable languages of countable words. On one hand, for any alphabet Σ, both the set of all well-ordered words and the set of all scattered words in Σ^{∞} are in REC but not in BCFL. On the other hand, any nonregular context-free language in Σ^* is a BCFL which is not in REC. Thus, the two classes REC and BCFL are incomparable.

8. An undecidable property

In this section we show that for any fixed alphabet Σ, it is undecidable whether a BCFL $L = L^{\infty}(G)$ generated by a grammar $G = (N, \Sigma, P, S, F)$ is the universal language Σ^{∞}. In our proof we make use of a variant of a corresponding undecidability result for ordinary context-free languages of finite words.

First we note that the language $\Sigma^{+\infty}$ of all nonempty words in Σ^{∞} is a BCFL. Next, the set of all words in Σ^{∞} with no first letter is also a BCFL since it can be given as $(\Sigma^{+\infty})^{-\omega} \cup \{\varepsilon\}$. Consider now the set of all words in Σ^{∞} having a first letter. This set can be subdivided into two sets:

1. All words starting with an ω-word which is a BCFL given by $\Sigma^{\omega}\Sigma^{\infty}$.
2. All words starting with a nonempty finite word followed by a word that does not have a first letter. This is again a BCFL given by the expression $\Sigma^{+}((\Sigma^{+\infty})^{-\omega} \cup \{\varepsilon\})$.

Suppose now that $G = (N, \Sigma, P, S)$ is an ordinary CFG with no ε-productions generating the language of finite words $L = L(G) \subseteq \Sigma^{+}$. Then consider the following language $L' \subseteq \Sigma^{\infty}$. L' consists of all words in Σ^{∞} not having a first letter together with all words that start with an ω-word as well as those words starting with a finite word in L followed by a word not having a first letter. An expression for this language is

$$((\Sigma^{+\infty})^{-\omega} \cup \{\varepsilon\}) \cup \Sigma^{\omega}\Sigma^{\infty} \cup L((\Sigma^{+\infty})^{-\omega} \cup \{\varepsilon\}),$$

showing that L' is a BCFL. The following fact is clear.

Lemma 8.1. *$L' = \Sigma^{\infty}$ iff $L = \Sigma^{+}$.*

Since it is undecidable for an ordinary context-free grammar without ε-productions over a fixed alphabet of size at least two whether it generates the language of all finite nonempty words, and since BCFL's are effectively closed under the operations that appear in the above expressions, we immediately have that the universality problem is undecidable for BCFL's.

Proposition 8.2. *Let Σ be an alphabet of size at least two. Then it is undecidable for a BCFG $G = (N, \Sigma, P, S, F)$ whether $L^{\infty}(G) = \Sigma^{\infty}$.*

In the rest of this section we show that the above universality problem is undecidable even for the unary alphabet. In our argument, we will make use of Lemma 8.3. For a word w we let Θ_w denote the equivalence relation on $\mathrm{dom}(w)$ defined by $x\,\Theta_w\,y$ iff $x \leq y$ and the interval $[x, y]$ is finite, or symmetrically, $y \leq x$ and $[y, x]$ is finite.

Lemma 8.3. *Let $L_0 \subseteq \{a\}^{\infty}$ be the language $\{a, b\}^{\infty}[a \leftarrow a^{\omega}, b \leftarrow a^{-\omega}]$. Then a word $w \in \{a\}^{\infty}$ is in L_0 iff each Θ_w equivalence class is infinite. Moreover, both L_0 and its complement $\bar{L}_0$ are BCFL's.*

Proof. The only if part of the first claim is obvious. Suppose now that $w \in \{a\}^{\infty}$ is such that each Θ_w class is infinite. Then each factor determined by a Θ_w class is either a^{ω} or $a^{-\omega}$, or $a^{-\omega}a^{\omega}$. Let $u \in \{a, b, c\}^{\infty}$ be the word whose underlying linear order is the set of all Θ_w equivalence classes ordered by $C \leq C'$ iff $x \leq y$ for some $x \in C$ and $y \in C'$, where a class C is labeled a, b, or c depending on whether the factor determined by C is a^{ω}, $a^{-\omega}$ or $a^{-\omega}a^{\omega}$. Then let $v = u[a \leftarrow a, b \leftarrow b, c \leftarrow ba] \in \{a, b\}^{\infty}$. We clearly have that $w = v[a \leftarrow a^{\omega}, b \leftarrow a^{-\omega}]$.

It is clear that L_0 is a BCFL. Regarding $\bar{L}_0$, note that $\bar{L}_0 = L_1\{a\}^{+}L_2$, where L_1 is the set of all words in $\{a\}^{\infty}$ with no last letter, and L_2 is the set of all words in $\{a\}^{\infty}$ with no first letter, i.e., $L_1 = (\{a\}^{+\infty})^{\omega} \cup \{\varepsilon\}$ and $L_2 = (\{a\}^{+\infty})^{-\omega} \cup \{\varepsilon\}$. $\square$

Theorem 8.4. *It is undecidable for a BCFG G over the unary alphabet $\{a\}$ whether $L^{\infty}(G) = \{a\}^{\infty}$.*

Proof. Consider a BCFG $G' = (N, \{a, b\}, P, S, F)$ generating the language $L' = L^{\infty}(G') \subseteq \{a, b\}^{\infty}$. Let $L = \bar{L}_0 \cup L'[a \leftarrow a^{\omega}, b \leftarrow a^{-\omega}]$, where L_0 is the language of Lemma 8.3. Then we have that $L = \{a\}^{\infty}$ iff $L' = \{a, b\}^{\infty}$. Since one can effectively construct a BCFG G with $L^{\infty}(G) = L$, the result follows from Proposition 8.2. $\square$

9. Müller context-free languages

In this section we define context-free grammars with Müller acceptance condition and show that their generative power strictly exceeds the generating power of context-free grammars with Büchi acceptance condition. The Müller acceptance condition originates from [22].

Definition 9.1. A context-free grammar with Müller acceptance condition, or MCFG is a system $G = (N, \Sigma, P, S, \mathcal{F})$ where (N, Σ, P, S) is an (ordinary) CFG and $\mathcal{F}$ is a set of subsets of N.

Suppose that $G = (N, \Sigma, P, S, \mathcal{F})$ is an MCFG. A derivation tree t over G is defined as for BCFG's except that we now require that for every infinite path π of t, the set of nonterminals that occur infinitely often as a label of a vertex along π belongs to $\mathcal{F}$. When t is a derivation tree with root symbol X and frontier word p, we also write $X \Rightarrow^\infty p$.

Definition 9.2. Let $G = (N, \Sigma, P, S, \mathcal{F})$ be an MCFG. The language $L^\infty(G)$ generated by G is the collection of all words $u \in \Sigma^\infty$ that are frontier words of some derivation tree whose root symbol is S. A language $L \subseteq \Sigma^\infty$ is called a Müller context-free language, or an MCFL, if L is generated by some MCFG.

Theorem 9.3. *BCFL is strictly included in MCFL.*

Proof. Suppose that $G = (N, \Sigma, P, S, F)$ is a BCFG. Then let G' be the MCFG $(N, \Sigma, P, S, \mathcal{F})$ with $\mathcal{F} = \{Y \subseteq N : Y \cap F \neq \emptyset\}$. Clearly, $L^\infty(G) = L^\infty(G')$. Thus, BCFL $\subseteq$ MCFL.

It remains to show that the inclusion is strict. Consider the Müller context-free grammar $G = (\{S, X\}, \{a\}, P, S, \{\{X\}\})$ where P consists of the following productions:

$$S \to a|X$$
$$X \to SX.$$

Let $w_0 = a$ and $w_{n+1} = w_n^\omega$, for all $n < \omega$. Since $S \Rightarrow^* a$ and $S \Rightarrow^\infty S^\omega$, it follows by induction that $S \Rightarrow^\infty w_n$ for all $n < \omega$.

We show that $L^\infty(G)$ consists only of well-ordered (and thus scattered) words. Let t be a derivation tree with root symbol S and frontier word w in $\{a\}^\infty$. The Müller condition $\mathcal{F}$ ensures that each infinite path π of t (viewed as an ω-word) is contained in the language $\{1, 2\}^*\{2\}^\omega$.

Now suppose w is not well-ordered and let $x_1 > x_2 > \cdots$ be a decreasing chain of leaves of t in the lexicographic order. Define the infinite path $\pi = y_0, y_1, \ldots$ of t as follows: let $y_0 = \varepsilon$ and for each $k \geq 0$, let

$$y_{k+1} = \begin{cases} y_k \cdot 1 & \text{if } t|_{y_k \cdot 1} \text{ contains } x_i \text{ for some } i; \\ y_k \cdot 2 & \text{otherwise.} \end{cases}$$

Thus, for each k there exists some j such that $t|_{y_k}$ contains the leaves $x_j, x_{j+1}, \ldots$.

Since π is an infinite path of t, $\pi = u \cdot 2^\omega$ for some $u \in \{1, 2\}^*$. By the definition of π, none of the trees $t|_{u2^k1}$ contains any of the leaves $\{x_1, x_2, \ldots\}$. But then $t|_u$ cannot contain any of these leaves either, a contradiction.

Since $L^\infty(G)$ consists of well-ordered words and contains all words w_n, $n < \omega$, it is not a BCFL by Corollary 6.15. □

Remark 9.4. Let us add $S \to \varepsilon$ to the productions of the above grammar. Then the language generated by the grammar is the set of all words in $\{a\}^\infty$ whose underlying linear order is a well-order of some countable order type.

10. Conclusion and further research topics

We have defined two types of context-free grammars generating languages of countable words, BCFG's and MCFG's, corresponding to the Büchi- and Müller-type acceptance conditions of automata on ω-words and automata on infinite trees. We showed that BCFG's can be transformed into equivalent BCFG's that are (weakly) ε-free and do not have chain productions or useless nonterminals. We established several closure properties of the class BCFL of languages that can be generated by BCFG's. We proved that many properties, including several order theoretic properties of BCFL's, are decidable in polynomial time, whereas the universality problem is undecidable even for the single letter alphabet. We showed that the BCFL's of finite words are exactly the usual CFL's, and that the ω-languages that are BCFL's are exactly the context-free ω-languages of Cohen and Gold [15]. We showed that every BCFL of scattered words consists of words of finite bounded rank. Finally we showed that there is a language that can be generated by an MCFG which is not a BCFL.

It follows from the proof of Theorem 6.4 that it is decidable in polynomial time whether a finite word belongs to the language generated by a BCFG. The same question for regular words seems very interesting, where a regular word may be defined as a word generated by a BCFG which contains exactly one production for each nonterminal. (See [8] for other equivalent definitions.)

It is known, see e.g. [24], that if a Müller automaton on infinite trees accepts a nonempty language, then it accepts a regular tree, i.e., a tree which has, up to isomorphism, a finite number of subtrees. It follows from this result that if an MCFG generates a nonempty language, then this language contains at least one regular word.

The present paper focuses on BCFG's and BCFL's. It would be interesting to see how much differently MCFG's behave. We have seen that they have a strictly larger generative power. It would also be interesting to develop a suitable pushdown automaton model.

References

[1] A.V. Aho, J.D. Ullman, The Theory of Parsing, Translation, and Compiling, in: Prentice-Hall Series in Automatic Computation, vol. I, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.
[2] N. Bedon, Finite automata and ordinals, Theoret. Comput. Sci. 156 (1996) 119–144.
[3] N. Bedon, A. Bès, O. Carton, C. Rispal, Logic and rational languages of words indexed by linear orderings, in: CSR 2008, in: LNCS, vol. 5010, Springer, 2008, pp. 76–85.
[4] A. Bès, O. Carton, A Kleene theorem for languages of words indexed by linear orderings, in: DLT'2005, in: LNCS, vol. 3572, Springer, 2005, pp. 158–167.
[5] S.L. Bloom, Ch. Choffrut, Long words: the theory of concatenation and ω-power, Theoret. Comput. Sci. 259 (2001) 533–548.
[6] S.L. Bloom, Z. Ésik, Deciding whether the frontier of a regular tree is scattered, Fund. Inform. 55 (2003) 1–21.
[7] S.L. Bloom, Z. Ésik, Axiomating omega and omega-op powers of words, Theor. Inform. Appl. 38 (2004) 3–17.
[8] S.L. Bloom, Z. Ésik, The equational theory of regular words, Inform. and Comput. 197((2005) 55–89.
[9] L. Boasson, Context-free sets of infinite words, in: Theoretical Computer Science (Fourth GI Conf., Aachen, 1979), in: LNCS, vol. 67, Springer, 1979, pp. 1–9.
[10] V. Bruyère, O. Carton, Automata on linear orderings, J. Comput. System Sci. 73 (2007) 1–24.
[11] J.R. Büchi, Weak second-order arithmetic and finite automata, Z. Math. Log. Grund. Math. 6 (1960) 66–92.
[12] J.R. Büchi, On a decision method in restricted second order arithmetic, in: Logic, Methodology and Philosophy of Science (Proc. 1960 Internatational Congress), Stanford Univ. Press, 1962, pp. 1–11.
[13] J.R. Büchi, The monadic second order theory of ω_1, in: Decidable Theories, II, in: Lecture Notes in Math., vol. 328, Springer, 1973, pp. 1–127.
[14] Y. Choueka, Finite automata, definable sets, and regular expressions over ω^n-tapes, J. Comput. System Sci. 17 (1) (1978) 81–97.
[15] R.S. Cohen, A.Y. Gold, Theory of ω-languages, parts one and two, J. Comput. System Sci. 15 (1977) 169–208.
[16] B. Courcelle, Frontiers of infinite trees, RAIRO Theor. Inform. 12 (1978) 319–337.
[17] J. Doner, Tree acceptors and some of their applications, J. Comput. System Sci. 4 (1970) 406–451.
[18] Z. Ésik, M. Ito, W. Kuich, Linear languages of finite and infinite words, in: Automata, Formal Languages and Algebraic Systems, World Scientific, 2010, pp. 33–46.
[19] Z. Ésik, Sz. Iván, Context-free languages of countable words, in: ICTAC 2009, Proceedings, in: LNCS, vol. 5684, Springer, 2009, pp. 185–199.
[20] F. Gécseg, M. Steinby, Tree automata, Akadémiai Kiadó, Budapest, 1984.
[21] S. Heilbrunner, An algorithm for the solution of fixed-point equations for infinite words, RAIRO Theor. Inform. 14 (1980) 131–141.
[22] D. E. Müller, Infinite sequences and finite machines, in: Proceedings of the 4th IEEE Symp. on Switching Circuit Theory and Logical Design, 1963, pp. 3–16.
[23] M. Nivat, Sur les ensembles de mots infinis engendrés par une grammaire algébrique, RAIRO Inform. Théor. 12 (3) (1978) 259–278 (in French).
[24] D. Perrin, J.-E. Pin, Infinite Words, Elsevier, 2004.
[25] M.O. Rabin, Decidability of second-order theories and automata on infinite trees, Trans. Amer. Math. Soc. 141 (1969) 1–35.
[26] J.G. Rosenstein, Linear Orderings, Academic Press, 1982.
[27] A. Salomaa, Formal Languages, Academic Press, 1973.
[28] L. Staiger, ω-languages, in: Handbook of Formal Languages, vol. 3, Springer, Berlin, 1997, pp. 339–387.
[29] J.W. Thatcher, J.B. Wright, Generalized finite automata theory with an application to a decision problem of second-order logic, Math. Systems Theory 2 (1968) 57–81.
[30] W. Thomas, On frontiers of regular sets, RAIRO Theor. Inform. 20 (1986) 371–381.
[31] W. Thomas, Automata on infinite objects, in: Handbook of Theoretical Computer Science, vol. B, Elsevier, 1990, pp. 135–191.
[32] W. Thomas, Languages, automata, and logic, in: Handbook of Formal Languages, vol. 3, Springer, 1997, pp. 389–455.
[33] J. Wojciechowski, Classes of transfinite sequences accepted by finite automata, Fund. Inform. 7 (1984) 191–223.
[34] J. Wojciechowski, Finite automata on transfinite sequences and regular expressions, Fund. Inform. 8 (1985) 379–396.