TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 354, Number 11, Pages 4597–4625
S 0002-9947(02)03048-9
Article electronically published on June 10, 2002

GROWTH AND ERGODICITY OF CONTEXT-FREE LANGUAGES

TULLIO CECCHERINI-SILBERSTEIN AND WOLFGANG WOESS

ABSTRACT. A language L over a finite alphabet $\mathbf{\Sigma}$ is called growth-sensitive if forbidding any set of subwords F yields a sublanguage L^F whose exponential growth rate is smaller than that of L. It is shown that every ergodic unambiguous, nonlinear context-free language is growth-sensitive. "Ergodic" means for a context-free grammar and language that its dependency di-graph is strongly connected. The same result as above holds for the larger class of essentially ergodic context-free languages, and if growth is considered with respect to the ambiguity degrees, then the assumption of unambiguity may be dropped. The methods combine a construction of grammars for 2-block languages with a generating function technique regarding systems of algebraic equations.

1. Introduction and main result

Let L be a language over the alphabet $\mathbf{\Sigma}$, that is, a subset of the free monoid $\mathbf{\Sigma}^*$ of all finite words over $\mathbf{\Sigma}$. Thus, $\mathbf{\Sigma}^*$ also contains the empty word ε. One writes $\mathbf{\Sigma}^+ = \mathbf{\Sigma}^* \setminus \{\varepsilon\}$. For a word $w \in \mathbf{\Sigma}^*$, write $|w|$ for its length (number of letters). The *growth* of L is the number

$$\gamma(L) = \limsup_{n\to\infty} \big|\{w \in L : |w| = n\}\big|^{1/n}.$$

We shall always suppose that L is infinite. Thus, $1 \le \gamma(L) \le |\mathbf{\Sigma}|$.

The general question addressed in this paper is the following: under which conditions is the growth of L sensitive to forbidding one (or more) subwords? More precisely, let $F \subset \mathbf{\Sigma}^*$ a nonempty set of nontrivial words that occur as subwords of elements of L, and define

$$L^F = \{w \in L : \text{no } v \in F \text{ is a subword of } w\}.$$

Then we ask under which conditions is it true that $\gamma(L^F) < \gamma(L)$ (strictly !).

Note that in principle, the question is of interest only when L has exponential growth, that is $\gamma(L) > 1$. Indeed, if $\gamma(L) = 1$, then either $\gamma(L^F) = 1$ or $\gamma(L^F) = 0$, in which case L^F is finite. Also note that without specific assumptions on L, one cannot expect growth-sensitivity. For example, if $\mathbf{\Sigma} = \{a, b, c, d\}$ and $L = \{a, b\}^* \cup \{c, d\}^*$, then $\gamma(L) = \gamma(L^a) = 2$.

In this paper, we shall prove the following.

Received by the editors December 19, 2001.
2000 *Mathematics Subject Classification.* Primary 68Q45; Secondary 05A16, 20F65, 68R15.
Key words and phrases. Context-free grammar, ambiguity, ergodicity, higher block languages, growth, Perron-Frobenius eigenvalue.
The first author was partially supported by TU Graz and by the Swiss National Science Foundation.

Main Result. *Every ergodic, unambiguous, non-linear context-free language is growth-sensitive with respect to forbidding arbitrary subwords.*

Of the notions appearing in this theorem, only "ergodic" is new, but for the convenience of the reader, and for setting our notation, we now explain all concepts involved.

We shall write $v \sqsubset w$ if v is a subword of w.

A *context-free grammar* is a quadruple $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$, where $\mathbf{V}$ is a finite set of *variables,* disjoint from the finite alphabet $\mathbf{\Sigma}$, the variable S is the *start symbol,* and $\mathbf{P} \subset \mathbf{V} \times (\mathbf{V} \cup \mathbf{\Sigma}^*)$ is a finite set of *production rules.* We write $T \vdash u$ or $(T \vdash u) \in \mathbf{P}$ if $(T, u) \in \mathbf{P}$. For $v, w \in (\mathbf{V} \cup \mathbf{\Sigma})^*$, we write $v \Longrightarrow w$ if $v = v_1 T v_2$ and $w = v_1 u v_2$, where $T \vdash u$, $v_1 \in (\mathbf{V} \cup \mathbf{\Sigma})^*$ and $v_2 \in \mathbf{\Sigma}^*$. A *rightmost derivation* is a sequence $v = w_0, w_1, \dots, w_k = w \in (\mathbf{V} \cup \mathbf{\Sigma})^*$ such that $w_{i-1} \Longrightarrow w_i$; we then write $v \overset{*}{\Longrightarrow} w$. For $T \in \mathbf{V}$, we consider the language $L_T = \{w \in \mathbf{\Sigma}^* : T \overset{*}{\Longrightarrow} w\}$. The *language generated by* $\mathcal{C}$ is $L(\mathcal{C}) = L_S$.

A context-free language is a language generated by a context-free grammar.

Basic Example (Dyck language)**.** Let $N \geq 2$ and $\mathbf{\Sigma} = \{a_1, \bar{a}_1, \dots, a_N, \bar{a}_N\}$. Take $\mathbf{V} = \{S\}$ and the productions $S \vdash \varepsilon$ and $S \vdash a_i S \bar{a}_i S$, $i = 1, \dots, N$.

Thinking of the a_i and $\bar{a}_i$ as N different "open" and "closed" parenthesis symbols, the language L generated by this grammar consists of all correctly nested parenthesis expressions over these symbols. For example,

$$S \vdash a_2 S \bar{a}_2 S \Longrightarrow a_2 S \bar{a}_2 a_1 S b_1 S \Longrightarrow a_2 S \bar{a}_2 a_1 S \bar{a}_1 \Longrightarrow a_2 S \bar{a}_2 a_1 a_2 S \bar{a}_2 S \bar{a}_1$$
$$\Longrightarrow a_2 S \bar{a}_2 a_1 a_2 S \bar{a}_2 \bar{a}_1 \Longrightarrow a_2 S \bar{a}_2 a_1 a_2 \bar{a}_2 \bar{a}_1 \Longrightarrow a_2 \bar{a}_2 a_1 a_2 \bar{a}_2 \bar{a}_1$$

is the – unique – rightmost derivation of $a_2 \bar{a}_2 a_1 a_2 \bar{a}_2 \bar{a}_1$. □

A grammar and the language generated by it are called *linear,* if every production rule in $\mathbf{P}$ is of the form $T \vdash v_1 U v_2$ or $T \vdash v$, where $v, v_1, v_2 \in \mathbf{\Sigma}^*$ and $T, U \in \mathbf{V}$. If furthermore in this situation one always has $v_2 = \varepsilon$ (the empty word), then grammar and language are called *right linear.* Analogously, it is called *left linear,* if instead one always has $v_1 = \varepsilon$. In both cases, language and grammar are also called *regular.* (It is well known that left and right linear languages are the same, i.e., every left linear language is also generated by a right linear grammar and conversely.)

Returning to a general context-free grammar $\mathcal{C}$, for a given variable $T \in \mathbf{V}$, we define the *ambiguity degree* $d_T(w)$ of a word $w \in \mathbf{\Sigma}^*$ as the number of all different rightmost derivations $T \overset{*}{\Longrightarrow} w$. We have $d_T(w) > 0$ if and only if $w \in L_T$. We shall assume that $d_T(w) < \infty$ always; indeed, we shall impose natural conditions guaranteeing this. The grammar is called *unambiguous,* if $d_S(w) = 1$ for all $w \in L$. A context-free language is called unambiguous if it is generated by some unambiguous grammar.

A context-free grammar $\mathcal{C}$ is called *reduced,* if there is no useless variable, i.e., each variable is used in some rightmost derivation of a word in $L(\mathcal{C})$. In particular, $L_T \neq \emptyset$ for each variable T. There is a simple algorithm that transforms any context-free grammar into a reduced one, see e.g. Harrison [19, Th. 3.2.1], and this algorithm preserves the ambiguity degrees $d_S(w)$, $w \in \mathbf{\Sigma}^*$. In the sequel, we shall consider only reduced grammars.

Next, we define the *dependency di-graph* $\mathcal{D} = \mathcal{D}(\mathcal{C})$; compare with Kuich [22]. This is an oriented graph with vertex set $\mathbf{V}$, with an edge from T to U (notation

$T \to U$) if there is a production $T \vdash u$ with $U \sqsubset u$. We write $T \xrightarrow{*} U$ if in $\mathcal{D}$ there is an oriented path of length ≥ 0 from T to U. A subset $\mathbf{W}$ of $\mathbf{V}$ is called *strongly connected* in $\mathcal{D}$, if for any pair of vertices T, U in $\mathbf{W}$ we have $T \xrightarrow{*} U$ and $U \xrightarrow{*} T$. Thus, $\mathbf{V}$ itself is strongly connected if and only if the adjacency matrix of $\mathcal{D}$ (with entries $a(T,U) = 1$ if $T \to U$, and $= 0$, otherwise) is irreducible as a non-negative matrix (i.e., for any U, V, there is n such that the (U,V)-entry of the n-th matrix power is positive).

Definition 1. We say that $\mathcal{C}$ is *ergodic* if the dependency di-graph of $\mathcal{C}$ is strongly connected and has at least one edge. We say that a context-free language L is ergodic, if L is generated by an ergodic, reduced context-free grammar. If this grammar is also unambiguous, we say that L is an ergodic, unambiguous context-free language.

The Dyck language (see above) is the basic example of this type, so that our Main Result applies to it. (Regarding the latter, at least to us it would not have been apparent before this work that forbidding, say, the subword $\bar{a}_2 a_1 a_2 \bar{a}_2$ leads to a language which is "exponentially smaller".)

The Main Result also holds for all ergodic, *regular* languages with the additional requirement in the definition of ergodicity that every terminal rule is of the form $T \vdash \varepsilon$. (The terminal rules are those of the form $T \vdash w$ with $w \in \mathbf{\Sigma}^*$. See §5 for variants; analogous modifications are also necessary in the linear case.) This is much simpler to prove and in principle known, see e.g. the remark of Grigorchuk and de la Harpe [14, §B]. A proof in the somewhat different context of symbolic dynamics and in some "disguise" is contained in the book of Lind and Marcus [26, Cor. 4.4.9]. See also Ceccherini-Silberstein, Fiorenzi and Scarabotti [4] and Ceccherini-Silberstein, Machì and Scarabotti [6]. In between these two theorems, on regular and on context-free languages, respectively, a small gap remains open, concerning *linear* languages that are not regular. Our approach does not work for these languages.

A few general words on growth are due here. Our own interest in growth arises from growth of *groups.* Given a finitely generated group G with a finite set of generators A, the associated growth sequence $V(n) = V_{G,A}(n)$ is obtained by counting the number of group elements that can be written as a product of no more than n elements of $A \cup A^{-1}$. The study of growth of groups was introduced in the context of Riemannian geometry by Efremovic [9], Schwarzc [34] and independently Milnor [27]. We refer to the survey by Grigorchuk and de la Harpe [14] and the book by de la Harpe [18] for (almost) all facts and references concerning growth.

Growth of groups can be interpreted in terms of growth of graphs (counting the number of elements of balls of radius n with respect to the graph metric), when one considers the Cayley graph of G with respect to A. But obviously, there is also an interpretation in terms of languages (with the irrelevant difference that what we are considering here corresponds to growth of spheres instead of balls). Indeed, the use of *regular languages* for studying the growth of a large class of groups – the *automatic* groups – has become a very popular topic in the 1990's; see the influential book by Epstein with coauthors [10].

A landmark in the study of growth of groups was Gromov's [16] classification of all groups with *polynomial growth,* i.e., where $V(n) \leq C\, n^d$, as the *virtually nilpotent* groups. Solvable groups that are not virtually nilpotent have *exponential growth,* i.e., $V(n) \geq \lambda^n$ with $\lambda > 1$; see Wolf [40] and Milnor [28]). Later, Grigorchuk [12],

[13] provided examples of groups with *intermediate growth,* i.e., faster than any polynomial but subexponential.

For context-free languages, Trofimov [36] has shown that the growth is either polynomial or exponential. This has also been proved independently by Incitti [21] and Bridson and Gilman [3]. The context-free languages with polynomial growth are characterized as the *bounded* languages, i.e., those that are contained in $w_1^* w_2^* \cdots w_k^*$ for finitely many words $w_1, \dots, w_k$ over the terminal alphabet. Trofimov [36] also gave an example of a *context-sensitive* language that has intermediate growth. Independently, Grigorchuk and Machì [15] have considered the same example, showing that it is even an *indexed* language (a specific form of being context-sensitive).

The problem of studying growth-sensitivity for context-free languages was suggested to the first author by R. I. Grigorchuk. Let us now come back to this question.

In general, if $\mathcal{L}$ is a given class of languages, a strategy to prove growth-sensitivity for all $L \in \mathcal{L}$ is the following subdivision into two steps.

Step 1. Consider the set $(\mathbf{\Sigma}^2)^*$ of all words over $\mathbf{\Sigma}^2$. Its letters are of the form (ab), where $a, b \in \mathbf{\Sigma}$. Define the mapping $\phi : \mathbf{\Sigma}^* \to (\mathbf{\Sigma}^2)^*$ by $\phi(w) = \varepsilon$ if $|w| \le 1$, and

$$\phi(a_1 \cdots a_n) = (a_1 a_2)(a_2 a_3) \cdots (a_{n-2} a_{n-1})(a_{n-1} a_n) \quad \text{if } n \ge 2\,.$$

For any language $L \subset \mathbf{\Sigma}^*$, consider the language $\phi(L)$ over the alphabet

$$\mathbf{\Sigma}_2 = \mathbf{\Sigma}_2(L) = \{(ab) : a, b \in \mathbf{\Sigma}\,,\ ab \sqsubset w \text{ for some } w \in L\}\,.$$

Note that $|\phi(w)| = |w| - 1$ when $|w| \ge 1$, whence $\gamma(L) = \gamma\big(\phi(L)\big)$.

Step 1 is the following. Show that $L \in \mathcal{L}$ implies $\phi(L) \in \mathcal{L}$.

Step 2. Show that each $L \in \mathcal{L}$ is growth-sensitive to forbidding one (or more) *letters,* i.e., elements of its alphabet $\mathbf{\Sigma}$.

Then each $L \in \mathcal{L}$ will be growth-sensitive to forbidding any nonempty $F \subset \mathbf{\Sigma}^* \setminus \{\varepsilon\}$. Indeed, it is clearly enough to prove this when $F = \{v_1\}$ consists of a single word v_1. If $m = |v_1|$, then after $m-1$ iterations, $v_m = \phi^{(m-1)}(v_1)$ is a *letter* in the alphabet of $\phi^{(m-1)}(L)$, and we have

$$\gamma(L) = \gamma\big(\phi^{(m-1)}(L)\big) \quad \text{and} \quad \gamma(L^{\{v_1\}}) = \gamma\Big(\big(\phi^{(m-1)}(L)\big)^{\{v_m\}}\Big)\,.$$

The structure of this paper is as follows. In §2, we extend the concept of ergodicity of a grammar by defining *essentially ergodic* grammars. Then we explain how to pass from a context-free grammar to a grammar that generates the associated 2-block language. The main difficulty is to show that this algorithm, which proceeds in several steps, preserves essential ergodicity (Theorem 1), i.e., we work out Step 1. In §3, we present a general method for characterizing the radius of convergence of power series with nonnegative coefficients that satisfy a system of algebraic equations with "nice" coefficient functions. This is then applied in §4 to elaborate Step 2. We prove that every essentially ergodic, unambiguous context-free language of *convergent type* is growth-sensitive with respect to forbidding letters. Together with Theorem 1, this yields an extended version of the Main Result. In §5, we present a further extension, where the ambiguity degrees are taken into account in the definition of the growth and the hypothesis of unambiguity is dropped (Theorem 3). We also explain how our proofs can be modified to obtain the Main

Result for regular languages. Finally, in §6, we give examples regarding groups and languages.

2. Normal forms and 2-block languages

We shall start with Step 1 of the proof-strategy outlined in the Introduction. Note that since we have defined ergodicity in terms of grammars, we have to use a grammar-oriented approach. In the sequel, it be will necessary to work with grammars that are standardized in some way, i.e., we consider normal forms. Two grammars are called *equivalent* if they generate the same language. Since the formal language literature does not deal with ergodicity, we shall now display the relevant algorithms for transforming a general context-free grammar into an equivalent one that has a specific normal form which is suitable for passing to a grammar for the associated 2-block language. We need to see if and how these algorithms preserve ambiguity degrees and ergodicity. We remark that for our purpose, it is irrelevant if we include ε in our language or not. Thus, we may define ergodicity in a slightly more general way by considering only $L \setminus \{\varepsilon\}$.

It will turn out that there are crucial steps that destroy ergodicity. For this reason, we shall have to extend the definition of ergodicity, and we shall see that the new property will be preserved in each step.

For any language $L \subset \mathbf{\Sigma}^*$, we define its *subword closure* as $\mathrm{SUB}(L) = \{v \in \Sigma^* : v \sqsubset w \text{ for some } w \in L\}$.

Given a context-free grammar $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$, we say that the start symbol is *isolated,* if it does not occur on the right-hand side of any production.

Let $\mathcal{D}(\mathcal{C})$ be the dependency-digraph associated with $\mathcal{C}$, and let $\mathbf{V} = \bigcup_{j=0}^{N} \mathbf{V}_j$ be the decomposition of its vertex set into *strong components.* The latter are the equivalence classes with respect to the relation where $T \sim U$ if $T \overset{*}{\to} U$ and $U \overset{*}{\to} T$. We assume that the class of S is $\mathbf{V}_0$. There is a partial order on the set of strong components; the component $\mathbf{V}_j$ precedes the component $\mathbf{V}_k$ (notation $\mathbf{V}_j \preccurlyeq \mathbf{V}_k$) if there is an oriented path from $T \in \mathbf{V}_j$ to U in $\mathbf{V}_k$ (this is independent of the choice of representatives). We now associate a grammar $\mathcal{C}_j$ with each strong component $\mathbf{V}_j$. To this end, we associate a "letter" c_T with each variable $T \in \mathbf{V}$, distinct from the elements of $\mathbf{\Sigma}$, and write $\mathbf{\Sigma}_j = \{c_T : T \in \mathbf{V} \setminus \mathbf{V}_j\}$. The alphabet of $\mathcal{C}_j$ is $\mathbf{\Sigma} \cup \mathbf{\Sigma}_j$, the set of variables is $\mathbf{V}_j$, and the set of productions $\mathbf{P}_j$ is obtained from $\mathbf{P}$ by taking only those rules which have an element of $\mathbf{V}_j$ on the left-hand side and replacing each occurring variable $T \in \mathbf{V} \setminus \mathbf{V}_j$ with c_T. As the start symbol, we choose and fix a representative S_j of $\mathbf{V}_j$ (the choice has no influence on our definition).

Definition 2. Let $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$ be a reduced context-free grammar with $|L(\mathcal{C})| = \infty$.

We say that a strong component $\mathbf{V}_j$ of $\mathcal{D}(\mathcal{C})$ is *regular,* if the associated grammar $\mathcal{C}_j$ is regular. In this case, every variable in $\mathbf{V}_j$ is also called regular. We write $\mathbf{V}_{\mathrm{reg}}$ for the set of all regular variables.

We say that $\mathcal{C}$ is *essentially ergodic* if

(i) there is precisely one strong component $\mathbf{V}_j = \mathbf{V}_{\mathrm{ess}}$ (the *essential variables*) such that neither the grammar $\mathcal{C}_j = \mathcal{C}_{\mathrm{ess}}$ nor the language generated by it are regular, and

(ii) for each $w \in L(\mathcal{C})$, there is $T \in \mathbf{V}_{\mathrm{ess}}$ such that $w \in \mathrm{SUB}(L_T)$.

This variant of Definition 1 is useful only when L is a context-free, non-regular language. In this case, there must be at least one non-regular strong component. Indeed, otherwise one could use the Substitution Theorem for regular languages (see Harrison [19, §3.4]), working backwards from the strong components that are maximal with respect to $\preccurlyeq$, to show that in $\mathcal{C}$, each language L_T ($T \in \mathbf{V}$) had to be regular. For non-regular grammars, Definition 2 clearly extends Definition 1.

We remark that left (resp. right) linearity of $\mathcal{C}_j$ means that in every production $T \vdash w$ in $\mathbf{P}$ to $T \in \mathbf{V}_j$, there is at most one element of $\mathbf{V}_j$ contained in w, which – if present – must be in the leftmost (resp. rightmost) position.

Regarding condition (ii), first observe that (strict) ergodicity in the sense of Definition 1 implies that $\mathrm{SUB}(L_T) = \mathrm{SUB}(L_U)$ for all $U, T \in \mathbf{V}$. In general, given condition (i), one sees that (ii) is equivalent to $L(\mathcal{C}) \subset \mathrm{SUB}(L_T)$ for *every* $T \in \mathbf{V}_{\mathrm{ess}}$. Later on in this section, we shall give examples that will clarify why we need this generalization of Definition 1. (This need is of course based on our approach.)

Theorem 1. *Let L be a context-free language. Then $L_2 = \phi(L) \setminus \{\varepsilon\}$, the 2-block language associated with L, is also context-free. If L is generated by a grammar that is unambiguous and/or essentially ergodic, then the same is true for L_2.*

The proof of this theorem needs several preparatory steps and will be completed at the end of this section.

A. Reduced grammar. We have already said in the Introduction that we always assume our grammars to be reduced (there are no superfluous variables).

B. ε-freeness. A context-free grammar is called ε-free if it contains no production of the form $T \to \varepsilon$. Suppose that L is generated by the reduced context-free grammar $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$. There is a simple algorithm that transforms $\mathcal{C}$ into a reduced ε-free grammar $\mathcal{C}'$ that generates $L \setminus \{\varepsilon\}$; see e.g. Harrison [19, §4.3]. Here, we make a small additional effort to ensure that the algorithm preserves being reduced; compare with Kuich [23, Thm. 5.4]. Let $\mathbf{V}_\varepsilon = \{T \in V : L_T \ni \varepsilon\}$ and $\mathbf{V}_\varepsilon^o = \{T \in \mathbf{V} : L_T = \{\varepsilon\}\}$. We set $\mathbf{V}' = \mathbf{V} \setminus \mathbf{V}_\varepsilon^o$. Regarding the productions, we first eliminate all $T \vdash \varepsilon$, where $T \in \mathbf{V}$, and all $T \vdash w$ with $w \in \mathbf{V}_\varepsilon^{o+}$. (In the latter case, also $T \in \mathbf{V}_\varepsilon^o$.) Next, if $T \vdash w$ is one of the remaining productions, then we replace it by all those new productions $T \vdash v$ where $v \in (\mathbf{\Sigma} \cup \mathbf{V}')^+$ can be obtained from w by deleting all occurring elements of $\mathbf{V}_\varepsilon^o$ and all, some or none occurring elements of $\mathbf{V}_\varepsilon \setminus \mathbf{V}_\varepsilon^o$.

This algorithm may decrease the ambiguity degrees. Namely, when $d_T(\varepsilon) \geq 2$ for some $T \in \mathbf{V}_\varepsilon$, then in the new productions this information is lost at the points where T is deleted on a right-hand side. But obviously, if $\mathcal{C}$ is unambiguous, then so is $\mathcal{C}'$.

Suppose that $\mathcal{C}$ is (strictly) ergodic (and, of course, $L(\mathcal{C}) \neq \{\varepsilon\}$). Then we must have $\mathbf{V}_\varepsilon^o = \emptyset$, since in $\mathcal{D}(\mathcal{C})$, there must be a path from any T to some U with $L_U \neq \{\varepsilon\}$. This simplifies the above algorithm a little, and one sees that $\mathcal{C}'$ is again ergodic.

After a moment's thought one also sees that when $\mathcal{C}$ is essentially ergodic then so is $\mathcal{C}'$.

C. Chain rules. A chain rule is a production of the form $T \vdash U$, where $T, U \in \mathbf{V}$. Chain rules can be unpleasant because they may cause large ambiguity degrees. For example, let $\mathcal{C} = (\{S, T\}, \{a\}, \{S \vdash T\,,\ T \vdash S\,,\ T \vdash a\}, S)$. Then $L(\mathcal{C}) = \{a\}$,

but $d_S(a) = \infty$, because one may go back and forth between S and T an arbitrary number of times before finally producing a.

Again, there is a simple algorithm that transforms a reduced grammar $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$ into an equivalent reduced grammar $\mathcal{C}' = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}', S)$ without chain rules; see e.g. Harrison [19, §4.3] or Kuich [23, Cor. 5.3]. The new productions are obtained from the old ones as follows: (a) if $(U \vdash w) \in \mathbf{P}$ with $w \notin \mathbf{V}$ and $T \stackrel{*}{\Longrightarrow} U$ by a sequence of chain rules in $\mathcal{C}$, then add the rule $T \vdash w$, and (b) delete all chain rules.

Clearly, if $\mathcal{C}$ is ε-free, then so is $\mathcal{C}'$, and the same is true for ergodicity as well as essential ergodicity. Ambiguity degrees are not necessarily preserved by this algorithm, but they can only decrease. In particular, if $\mathcal{C}$ is unambiguous, then so is $\mathcal{C}'$.

We remark that the algorithm of **B** that produces an ε-free grammar may generate chain rules, even if started with a grammar that was already chain-rule-free. However, the new chain rules will not be "harmful" in the sense of the above example.

In any case, if we start with a reduced, (essentially) ergodic context-free grammar $\mathcal{C}$, then we can first apply the algorithm of **B** and then eliminate chain rules to end up with a context-free grammar $\mathcal{C}'$ that generates $L(\mathcal{C}) \setminus \{\varepsilon\}$ and is reduced, (essentially) ergodic, ε-free and chain-rule-free; if $\mathcal{C}$ is unambiguous, then so is $\mathcal{C}'$.

D. Binary form. We say that a context-free grammar has *binary form (BF),* if each right-hand side of a production rule is contained in $\mathbf{\Sigma} \cup \mathbf{\Sigma}^2 \cup \mathbf{\Sigma V} \cup \mathbf{V\Sigma} \cup \mathbf{V}^2$. (Remark: Here, we confine ourselves to ε-free languages.) It is easy to pass from any reduced, chain-rule-free and ε-free grammar $\mathcal{C}$ to an equivalent one in BF. We have to consider all rules $T \vdash w$ in $\mathbf{P}$, where $w \in (\mathbf{V} \cup \mathbf{\Sigma})^+$ has length ≥ 3, and take some care in order to preserve right, resp. left, linearity. Our procedure is recursive, starting with a rule where $|w|$ on the right-hand side is maximal.

Case 1. We have $w = a_1 v$ with $a_1 \in \mathbf{\Sigma}$ and $v \in (\mathbf{V} \cup \mathbf{\Sigma})^+$ with $|v| \geq 2$. Then we introduce a new variable T', distinct from all other ones, and replace $T \vdash w$ with the two rules

$$T \vdash a_1 T' \quad \text{and} \quad T' \vdash v\,.$$

Case 2. We have $w = v a_k$ with $a_k \in \mathbf{\Sigma}$ and $v \in \mathbf{V}(\mathbf{V} \cup \mathbf{\Sigma})^+$. In this case, we introduce a new variable T' and replace $T \vdash w$ with the two rules

$$T \vdash T' a_k \quad \text{and} \quad T' \vdash v\,.$$

Case 3. We have $w = Uv$ with $U \in \mathbf{V}$ and $v \in (\mathbf{V} \cup \mathbf{\Sigma})^+ \mathbf{V}$. Again, we introduce a new variable T' and replace $T \vdash w$ with the two rules

$$T \vdash U T' \quad \text{and} \quad T' \vdash v\,.$$

We iterate this procedure until all right-hand sides of the production rules have length at most 2.

In the new grammar, the ambiguity degrees with respect to the variables in $\mathbf{V}$ are the same as the "old" ones. Let us discuss what this algorithm does to essential ergodicity.

Proposition 1. *The algorithm for passing to a grammar in binary form preserves essential ergodicity.*

Proof. We can use induction, a single step consisting in replacing a rule $T \vdash w$ by two new rules according to one of the three cases. For the moment, let us write $\mathcal{C}' = (\mathbf{V}', \mathbf{\Sigma}, \mathbf{P}', S)$ for the new grammar obtained after this replacement.

Write $\mathbf{V}_j$ for the strong component of T with respect to $\mathcal{C}$. In each of the three cases, we distinguish between two possibilities.

(a) The word v contains some variable $Z \in \mathbf{V}_j$. Then we have in $\mathcal{D}(\mathcal{C}')$ that $T \to T' \xrightarrow{*} Z$. Therefore $\mathbf{V}'_j = \mathbf{V}_j \cup \{T'\}$, while all the other strong components and their nature (regular or essential), as well as the partial order $\preccurlyeq$, remain unchanged.

Suppose that $\mathbf{V}_j$ is a regular component. If it is right linear, then Z must be the rightmost element in w. Thus, Case 2 cannot occur. If Case 1 occurs, then also $\mathbf{V}'_j$ is right linear. If Case 3 occurs, then we cannot have $U \in \mathbf{V}_j$ since this would contradict right linearity. Thus, $\mathbf{V}_j$ precedes the strong component of U strictly in $\mathcal{D}(\mathcal{C})$, so that $\mathbf{V}'_j$ is right linear in this case as well.

On the other hand, if $\mathbf{V}_j$ is left linear, then Z must be the leftmost element in w, and Case 1 is excluded as well as Case 3. In the remaining Case 2, one sees that $\mathbf{V}'_j$ is left linear again.

Finally, if $\mathbf{V}_j$ is the essential component in $\mathcal{D}(\mathcal{C})$, then clearly $\mathbf{V}'_j$ is also *the* essential component in $\mathcal{D}(\mathcal{C}')$, and condition (ii) of Definition 2 holds for $\mathcal{C}'$, too.

(b) The word v contains no element of $\mathbf{V}_j$. Then T' forms a one-point strong component in the graph $\mathcal{D}(\mathcal{C}')$, whence it is regular. The strong components of $\mathcal{D}(\mathcal{C})$ are also strong components of $\mathcal{D}(\mathcal{C}')$ and as such preserve their nature (regular or essential). Therefore $\mathcal{C}'$ is essentially ergodic. □

E. Operator normal form. A context-free grammar is said to have *operator normal form (ONF),* if no right-hand side of any production rule has a subword in $\mathbf{V}^2$.

Using a small variation of Theorem 5.9 of Kuich [23], we describe how to pass from a reduced grammar $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$ in BF to an equivalent grammar $\mathcal{C}' = (\mathbf{V}', \mathbf{\Sigma}, \mathbf{P}', S')$ that has ONF. Again, we take care to obtain a reduced grammar.

For any language $L \subset \mathbf{\Sigma}^*$ and $a \in \mathbf{\Sigma}$, we define $La^{-1} = \{w \in \mathbf{\Sigma}^* : wa \in L\}$.

Starting with $\mathcal{C}$ in BF and $T \in \mathbf{V}$, we first define $\mathbf{\Sigma}_T = \{a \in \mathbf{\Sigma} : L_T a^{-1} \neq \emptyset\} = \{a \in \mathbf{\Sigma} : T \stackrel{*}{\Longrightarrow} wa \text{ for some } w \in \mathbf{\Sigma}^*\}$. Since our grammar is reduced (basic assumption), this set is nonempty for every $T \in \mathbf{V}$. We now introduce the set of new variables

$$\mathbf{V}' = \{S'\} \cup \{[Ta] : T \in \mathbf{V}\,,\ a \in \mathbf{\Sigma}_T\}\,.$$

The new productions are

$$\begin{array}{rll} S' \vdash [Sa]a & & \text{for all } a \in \mathbf{\Sigma}_S\,, \\ [Ta] \vdash \varepsilon & & \text{whenever } T \vdash a \text{ in } \mathbf{P}\,, \\ [Ta] \vdash b & & \text{whenever } T \vdash ba \text{ in } \mathbf{P}\,, \\ [Ta] \vdash b[Va] & & \text{whenever } T \vdash bV \text{ in } \mathbf{P}\,,\ b \in \mathbf{\Sigma}\,,\ a \in \mathbf{\Sigma}_V\,, \\ [Ta] \vdash [Ub]b & & \text{whenever } T \vdash Ua \text{ in } \mathbf{P}\,,\ a \in \mathbf{\Sigma}\,,\ b \in \mathbf{\Sigma}_U\,, \quad \text{and} \\ [Ta] \vdash [Ub]b[Va] & & \text{whenever } T \vdash UV \text{ in } \mathbf{P}\,,\ a \in \mathbf{\Sigma}_V\,,\ b \in \mathbf{\Sigma}_U\,. \end{array} \tag{2.1}$$

We remark that when $a \in \mathbf{\Sigma}_V$ and $T \vdash UV$ or $T \vdash bV$ in $\mathbf{P}$, then also $a \in \mathbf{\Sigma}_T$, since L_U is non-empty, $\mathcal{C}$ being reduced.

Analogously to the cited Theorem 5.9 of Kuich [23], the ambiguity degrees d_S with respect to $\mathcal{C}$ and $d_{S'}$ with respect to $\mathcal{C}'$ are the same. Furthermore, $L_{[Ta]} =$

$L_T a^{-1}$, where L_T refers to $\mathcal{C}$ and $L_{[Ta]}$ refers to $\mathcal{C}'$. In particular, it may well happen that $[Ta]$ is regular even when $T \in \mathbf{V}_{\rm ess}$; see Example 1 further below.

We now turn to the question of whether passing from $\mathcal{C}$ to $\mathcal{C}'$ preserves essential ergodicity.

Proposition 2. *If $\mathcal{C}$ in BF is essentially ergodic, then so is $\mathcal{C}'$ in ONF as defined in (2.1).*

Proof. Passing from $\mathbf{V}$ to $\mathbf{V}'$ preserves the structure of strong components and the partial order $\preccurlyeq$ in the following sense.

If $[Ta] \xrightarrow{*} [Ub]$ in $\mathcal{D}(\mathcal{C}')$, then $T \xrightarrow{*} U$ in $\mathcal{D}(\mathcal{C})$.

In particular, if $[Ta] \sim [Ub]$ in $\mathcal{D}(\mathcal{C}')$, then $T \sim U$ in $\mathcal{D}(\mathcal{C})$. We see immediately that when the strong component of T in $\mathcal{D}(\mathcal{C})$ is right or left linear, then so is the strong component of every $[Ta]$ in $\mathcal{D}(\mathcal{C}')$. Also, the start symbol S' is isolated and hence forms a regular strong component by its own.

Consequently, we only have to check what happpens to the unique essential strong component $\mathbf{V}_{\rm ess}$.

Claim 1. If $T, U \in \mathbf{V}_{\rm ess}$ and $b \in \mathbf{\Sigma}_U$, then there is $a \in \mathbf{\Sigma}_T$ such that $[Ta] \xrightarrow{*} [Ub]$ in $\mathcal{D}(\mathcal{C}')$.

Proof. We use induction on the number n of steps in a $\mathcal{D}(\mathcal{C})$-path from T to U. If $n = 0$, then there is nothing to prove, $[Ub] \xrightarrow{*} [Ub]$. Suppose the statement holds for n. Then in $\mathcal{D}(\mathcal{C})$ we have $T \to \bar{T}$ and, in n steps, $\bar{T} \xrightarrow{*} U$, with $\bar{T} \in \mathbf{V}_{\rm ess}$. By assumption, there is $\bar{a} \in \mathbf{\Sigma}_{\bar{T}}$ such that $[\bar{T}\bar{a}] \xrightarrow{*} [Ub]$ in $\mathcal{D}(\mathcal{C}')$. The edge $T \to \bar{T}$ can arise from four types of productions in $\mathbf{P}$.

(1) If $T \vdash \bar{b}\bar{T}$ with $\bar{b} \in \mathbf{\Sigma}$, then $[T\bar{a}] \vdash \bar{b}[\bar{T}\bar{a}]$ in $\mathbf{P}'$, and we can choose $a = \bar{a}$.

(2) If $T \vdash \bar{T}\bar{b}$ with $\bar{b} \in \mathbf{\Sigma}$, then $[T\bar{b}] \vdash [\bar{T}\bar{a}]\bar{a}$ in $\mathbf{P}'$, and we can choose $a = \bar{b}$.

(3) If $T \vdash \bar{T}\bar{U}$ with $\bar{U} \in \mathbf{V}$, then $[T\bar{b}] \vdash [\bar{T}\bar{a}]\bar{a}[\bar{U}\bar{b}]$ in $\mathbf{P}'$ for any $\bar{b} \in \mathbf{\Sigma}_{\bar{U}}$, and we can choose $a = \bar{b}$.

(4) If $T \vdash \bar{U}\bar{T}$ with $\bar{U} \in \mathbf{V}$, then $[T\bar{a}] \vdash [\bar{U}\bar{b}]\bar{b}[\bar{T}\bar{a}]$ in $\mathbf{P}'$ for any $\bar{b} \in \mathbf{\Sigma}_{\bar{U}}$, and we can choose $a = \bar{a}$.

This proves Claim 1.

Claim 2. Let $T \in \mathbf{V}_{\rm ess}$ and $a \in \mathbf{\Sigma}_T$. Then there is the following alternative.

Either (I) the variable $[Ta]$ is regular in $\mathcal{C}'$, or (II) there are variables $\bar{T}, \bar{U} \in \mathbf{V}_{\rm ess}$ such that

$$[Ta] \xrightarrow{*} [\bar{T}a] \ \text{ in } \ \mathcal{D}(\mathcal{C}') \quad \text{and} \quad \begin{cases} \bar{T} \vdash \bar{U}a & \text{in } \mathbf{P} \text{ or} \\ \bar{T} \vdash \bar{U}\bar{V} & \text{in } \mathbf{P}\text{, where } \bar{V} \in \mathbf{V} \text{ and } a \in \mathbf{\Sigma}_{\bar{V}}\,. \end{cases}$$

Proof. Suppose that (II) does not hold for $[Ta]$. Let $\mathbf{V}'_j$ be the strong component of $[Ta]$ in $\mathcal{D}(\mathcal{C}')$, and consider the associated grammar $\mathcal{C}'_j$ with $[Ta]$ as the start symbol. The productions of $\mathcal{C}'_j$ can only arise from rules in $\mathbf{P}$ whose right hand side does *not* start with some $U \in \mathbf{V}_{\rm ess}$. This means that the productions of $\mathcal{C}'_j$ can only have one of the following five forms

$$[\bar{T}a] \vdash \varepsilon\,,\ [\bar{T}a] \vdash b\,,\ [\bar{T}a] \vdash b[\bar{V}a]\,,\ [\bar{T}a] \vdash cb\,,\ [\bar{T}a] \vdash cb[\bar{V}a]\,,$$

where $b \in \mathbf{\Sigma}$, $\bar{V} \in \mathbf{V}_{\rm ess}$ and $c = c_{[\bar{U}b]}$ is a "letter" associated with a variable $[\bar{U}b]$, where $U \in \mathbf{V} \setminus \mathbf{V}_{\rm ess}$. In other words, the grammar $\mathcal{C}'_j$ is right linear. This proves Claim 2.

Claim 3. Let $T, U \in \mathbf{V}_{\rm ess}$, $a \in \mathbf{\Sigma}_T$ and $b \in \mathbf{\Sigma}_U$. If $[Ta]$ is non-regular, then $[Ta] \xrightarrow{*} [Ub]$ in $\mathcal{D}(\mathcal{C}')$.

Proof. Let $\bar{T}$ and $\bar{U}$ be as in the statement of Claim 2. By Claim 1 there is $\bar{a} \in \mathbf{\Sigma}_{\bar{U}}$ such that $[\bar{U}\bar{a}] \xrightarrow{*} [Ub]$. But now, since $\bar{U}$ is the first element on the right-hand side of a production in $\mathbf{P}$ whose left-hand side is $\bar{T}$, we have $[\bar{T}a] \to [\bar{U}\bar{a}]$, which proves Claim 3.

Thus, all non-regular elements $[Ta] \in \mathbf{V}'$, where $T \in \mathbf{V}_{\mathrm{ess}}$ and $a \in \mathbf{\Sigma}_T$, form a single strong component in $\mathcal{D}(\mathcal{C}')$. This proves part (i) of the definition of essential ergodicity.

To prove part (ii), we use Claim 2 once more, which tells us that there are $[\bar{T}a] \in \mathbf{V}'_{\mathrm{ess}}$ and a rule in $\mathbf{P}$ of the form $\bar{T} \vdash \bar{U}a$ or $\bar{T} \vdash \bar{U}\bar{V}$ with $a \in \mathbf{\Sigma}_{\bar{V}}$, such that $\bar{U} \in \mathbf{V}_{\mathrm{ess}}$. This implies that every element of $L_{\bar{U}}$ (the language generated by $\bar{U}$ in $\mathcal{C}$) occurs as an initial subword of some element in $L_{[\bar{T}a]}$ (the language generated by $[\bar{T}a]$ in $\mathcal{C}'$). Therefore, since our language $L = L(\mathcal{C}) = L(\mathcal{C}')$ satisfies $L \subset \mathrm{SUB}(L_{\bar{U}})$, we also get $L \subset \mathrm{SUB}(L_{[\bar{T}a]})$.

We have not yet concluded our construction of a grammar in ONF, because $\mathcal{C}'$ has ε-rules. Therefore, we apply the algorithm of $\mathbf{B}$ to eliminate the latter. We end up with a grammar $\mathcal{C}'' = (\mathbf{V}', \mathbf{\Sigma}, \mathbf{P}'', S')$ where S' is isolated and the right-hand side of every production is in

$$\mathbf{\Sigma} \cup \mathbf{\Sigma}\mathbf{V}' \cup \mathbf{V}'\mathbf{\Sigma} \cup \mathbf{V}'\mathbf{\Sigma}\mathbf{V}' .$$

We conclude that if $\mathcal{C}$ is essentially ergodic and/or unambiguous, then so is $\mathcal{C}''$.

Example 1. Consider the grammar $\mathcal{C}$ with $\mathbf{\Sigma} = \{a, b, c\}$, $\mathbf{V} = \{S, U\}$ and productions

$$S \vdash a\,,\ S \vdash bS\,,\ S \vdash SU\,,\ U \vdash c\,,\ U \vdash SU\,.$$

In ONF and after the elimination of ε-rules, we get the following grammar: $\mathbf{V}' = \{S', [Sa], [Sc], [Uc]\}$, and the new productions are

$$\begin{gathered} S' \vdash a \ \mid\ [Sa]a \ \mid\ [Sc]c \\ [Sa] \vdash b \ \mid\ b[Sa]\,, \\ [Sc] \vdash b[Sc] \ \mid\ a \ \mid\ [Sa]a \ \mid\ a[Uc] \ \mid\ [Sa]a[Uc] \ \mid\ [Sc]c \ \mid\ [Sc]c[Uc]\,, \\ [Uc] \vdash a \ \mid\ [Sa]a \ \mid\ a[Uc] \ \mid\ [Sa]a[Uc] \ \mid\ [Sc]c \ \mid\ [Sc]c[Uc]\,. \end{gathered}$$

(Here, as commonly used, the symbols "|" serve to separate the different right-hand sides of rules having the same left-hand side.)

Although $\mathcal{C}$ was (strictly) ergodic, we get that $\mathbf{V}'_{\mathrm{ess}} = \{[Sc], [Uc]\} \subsetneq \mathbf{V}'$. The grammar associated with the strong component $\{[Sa]\}$ is clearly regular; it generates the language $\{b^n : n \geq 1\}$. This clarifies (at least in part) why we need condition (i) in Definition 2. □

Example 2. Take the grammar $\mathcal{C}$ of Example 1. We observe that $bc \notin \mathrm{SUB}\big(L(\mathcal{C})\big)$. Indeed, suppose the contrary. A letter b in some $w \in L(\mathcal{C})$ can only come from the production $S \vdash bS$. Hence the c of bc must be the first letter of a word v such that $S \overset{*}{\Longrightarrow} v$. But in the grammar $\mathcal{C}$, this is impossible.

Now construct a new grammar $\bar{\mathcal{C}}$ by adding a new start symbol $\bar{S}$ and the production rules

$$\bar{S} \vdash bc\,,\ \bar{S} \vdash a\,,\ \bar{S} \vdash bS\,,\ \bar{S} \vdash SU\,.$$

(The last three rules amount to first introducing only the chain rule $\bar{S} \vdash S$ and then applying the algorithm of $\mathbf{C}$.)

In $\bar{\mathcal{C}}$, the start symbol $\bar{S}$ is isolated, and all other variables are essential and form a strongly connected component in $\mathcal{D}(\mathcal{C}')$. However, $L(\bar{\mathcal{C}}) = L(\mathcal{C}) \cup \{bc\}$ is not growth-sensitive with respect to forbidding bc.

This clarifies why we need condition (ii) in Definition 2.

F. The 2-block language. We now want to present an algorithm for passing from a grammar $\mathcal{C}$ with $L(\mathcal{C}) = L$ to a grammar $\mathcal{C}'$ which generates the 2-block language $L_2 = \phi(L) \setminus \{\varepsilon\}$. By the above we can assume that $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$ is reduced, in ONF with isolated start symbol, and the production rules have their right-hand sides in $\mathbf{\Sigma} \cup \mathbf{\Sigma V} \cup \mathbf{V\Sigma} \cup \mathbf{V\Sigma V}$.

For $T \in \mathbf{V}$, a *T-sentential form* is any element w of $(\mathbf{\Sigma} \cup \mathbf{V})^*$ such that there is a rightmost derivation $T \stackrel{*}{\Longrightarrow} w$. A *sentential form* is a T-sentential form for some T. Since $\mathcal{C}$ is in ONF, a straightforward induction argument shows that a sentential form does not have any subword in $\mathbf{V}^2$. Thus, a sentential form looks as follows.

$$w = T_1 v_1 T_2 v_2 \cdots T_k v_k T_{k+1}\,, \tag{2.2}$$

where $v_i \in \mathbf{\Sigma}^+$, $T_i \in \mathbf{V}$ and possibly T_1 and/or T_{k+1} may be missing. Since $\mathcal{C}$ is reduced, every v_i is in $\mathrm{SUB}(L)$. Let a_i and b_i be the first and last letters of v_i, respectively.

We now transform each sentential form w as in (2.2) into a new expression $\Phi(w)$ by using ϕ and inserting brackets as follows.

$$\Phi(w) = [T_1 a_1]\phi(v_1)[b_1 T_2 a_1]\phi(v_2)\cdots[b_{k-1}T_k a_k]\phi(v_k)[b_k T_{k+1}]\,,$$

where $[T_1 a_1]$ and/or $[b_k T_{k+1}]$ will be missing when T_1 and/or T_{k+1} are missing in w, respectively. The resulting expressions in the square brackets will become our new variables. In principle, $[bTa]$ stands for "anything that be derived from T, with a leading b and a final a added", or more precisely its image under ϕ. The meaning of the variables $[bT]$ and $[Ta]$ is analogous. We write $\mathbf{V}'$ for the set consisting of $[S]$ and all expressions $[Ta]$, $[bT]$ and $[bTa]$ that appear in some $\Phi(w)$, where w is a sentential form of $\mathcal{C}$. (If necessary, it is easy to write down an algorithm for finding all of them.)

Slightly more generally, we can interpret Φ as a mapping from $\{w \in (\mathbf{V} \cup \mathbf{\Sigma})^* : w \text{ contains no element of } \mathbf{V}^2\}$ to $(\mathbf{V}' \cup \mathbf{\Sigma}^2)^*$, i.e., the argument w does not necessarily have to be a sentential form. It is clear that from $\Phi(w)$ one can reconstruct w. That is, the mapping Φ is one-to-one. Also, the restriction of Φ to $\mathbf{\Sigma}^*$ coincides with ϕ.

We now exhibit the new grammar $\mathcal{C}' = (\mathbf{V}', \mathbf{\Sigma}_2, \mathbf{P}', [S])$ for L_2. The next list displays the rules in $\mathbf{P}$ followed by the corresponding new rules in $\mathbf{P}'$.

$$\begin{aligned}
&\text{If } S \vdash bT &&: [S] \vdash [bT]\,;\\
&\text{if } S \vdash Ta &&: [S] \vdash [Ta]\,;\\
&\text{if } S \vdash TbU &&: [S] \vdash [Tb][bU]\,;\\
&\text{if } T \vdash c &&: [Ta] \vdash (ca)\,,\ [bT] \vdash (bc)\,,\ [bTa] \vdash (bc)(ca)\,;\\
&\text{if } T \vdash cU &&: [Ta] \vdash [cUa]\,,\ [bT] \vdash (bc)[cU]\,,\ [bTa] \vdash (bc)[cUa]\,;\\
&\text{if } T \vdash Uc &&: [Ta] \vdash [Uc](ca)\,,\ [bT] \vdash [bUc]\,,\ [bTa] \vdash [bUc](ca)\,;\\
&\text{if } T \vdash UcV &&: [Ta] \vdash [Uc][cVa]\,,\ [bT] \vdash [bUc][cV]\,,\ [bTa] \vdash [bUc][cVa]\,.
\end{aligned} \tag{2.3}$$

Here, $T, U, V \in \mathbf{V} \setminus \{S\}$ and $a, b, c \in \mathbf{\Sigma}$ have to be such that the occurring expressions in brackets belong to $\mathbf{V}'$. By the construction of $\mathcal{C}'$, for any sequence of

sentential forms $w_1, w_2, \dots, w_n$ with respect to $\mathcal{C}$, we have

$$S \Longrightarrow w_1 \Longrightarrow w_2 \Longrightarrow \dots \Longrightarrow w_n \quad \text{in } \mathcal{C}$$
$$\text{if and only if}$$
$$[S] \Longrightarrow \Phi(w_1) \Longrightarrow \Phi(w_2) \Longrightarrow \dots \Longrightarrow \Phi(w_n) \quad \text{in } \mathcal{C}'.$$

Therefore $\mathcal{C}'$ generates $\phi(L) \setminus \{\varepsilon\} = L_2$ as required, and the ambiguity degrees are preserved, that is,

$$d_{[S]}\big(\phi(w)\big) = d_S(w) \quad \text{for every } w \in L.$$

Before proceeding, let us note that $\mathcal{C}'$ has chain rules, which can be eliminated by the algorithm of **C** but are "harmless" anyway (they cannot be concatenated into an infinite loop – Harrison [19] uses "cycle-free" for such grammars). Thus, we continue to work with $\mathcal{C}'$.

We say that $[bTa]$ is an *interior* variable of $\mathcal{C}'$ if $T \in \mathbf{V}_{\mathrm{ess}}$ and the string bTa occurs in a sentential form of $\mathcal{C}$ that derives from some element of $\mathbf{V}_{\mathrm{ess}}$.

Proposition 3. *If $\mathcal{C}$ (in ONF as assumed above) is essentially ergodic, then so is $\mathcal{C}'$, and $\mathbf{V}'_{\mathrm{ess}}$ consists of the interior variables.*

Proof. First of all, the start symbol $[S]$ is once more isolated and consequently regular. Also, we observe once more that when $T, U \in \mathbf{V} \setminus \{S\}$ and there is a path in $\mathcal{D}(\mathcal{C}')$ from some variable of the form $[bT]$, $[Ta]$ or $[bTa]$ (the variables *induced* by T) to some variable induced by U, then there must be a path from T to U in $\mathcal{D}(\mathcal{C})$. In the same way as in the proof of Proposition 2 we conclude that the only variables in $\mathbf{V}'$ that may be non-regular are those induced by some element of $\mathbf{V}_{\mathrm{ess}}$.

Next observe that in the right-hand side of any production in $\mathbf{P}'$, a variable of the form $[Ta]$ can only occur as the first element. Thus, each $[Ta] \in \mathbf{V}'$ must be regular (its strong component is left linear). In the same way, each $[bT] \in \mathbf{V}'$ is regular (its strong component is right linear).

A look at (2.3) shows us that in the relation $\overset{*}{\to}$ of $\mathcal{D}(\mathcal{C}')$, a variable of the form $[bTa]$ can only precede variables of the same type. Thus, we are left with considering the variables $[bTa] \in \mathbf{V}'$, where $T \in \mathbf{V}_{\mathrm{ess}}$.

Claim 1. Let $[bTa], [dUc] \in \mathbf{V}'$ be such that $T, U \in \mathbf{V}_{\mathrm{ess}}$ and $[dUc]$ is interior. Then $[bTa] \overset{*}{\to} [dUc]$ in $\mathcal{D}(\mathcal{C}')$. In particular, the interior variables of $\mathcal{C}'$ form a strong component.

Proof. By assumption, there is $\bar{T} \in \mathbf{V}_{\mathrm{ess}}$ such that the string dUc appears in a $\bar{T}$-sentential form of the grammar $\mathcal{C}$. Furthermore, $T \overset{*}{\to} \bar{T}$ in $\mathcal{D}(\mathcal{C})$ which is the same as saying that $\bar{T}$ appears in a T-sentential form. Concatenating, we see that dUc is also part of a T-sentential form w. But then $\Phi(bwa)$ is a $[bTa]$-sentential form of $\mathcal{C}'$ and contains the variable $[dUc]$ of $\mathbf{V}'$, whence $[bTa] \overset{*}{\to} [dUc]$.

Claim 2. If $T_0 \in \mathbf{V}_{\mathrm{ess}}$ and $[b_0 T_0 a_0] \in \mathbf{V}'$ is non-interior, then it is regular.

Proof. If in $\mathbf{P}'$ we have a production of the form $[bTa] \vdash [bUc](ca)$, $[bTa] \vdash (bc)[cVa]$ or $[bTa] \vdash [bUc][cVa]$, respectively, then we call the edge $[bTa] \to [bUc]$ of $\mathcal{D}(\mathcal{C}')$ a *left edge* and the edge $[bTa] \to [cVa]$ a *right edge*. Now let $\mathcal{C}'_j$ be the strong component of $[b_0 T_0 a_0]$. We claim that this component either has only left edges (in which case it is left linear) or else has only right edges (in which case it is right linear).

Assume that there are edges of both types. Then there must be a vertex, say $[dUa]$, which is the endpoint of a right edge of the form $[bTa] \to [dUa]$ and the

initial point of a left edge of the form $[dUa] \to [dVc]$. Transferring this back to $\mathcal{C}$, we see that there must be productions of the form $T \vdash vdU$ and $U \vdash Vcw$ with $v, w \in \mathbf{V} \cup \{\varepsilon\}$. Thus, we find the T-sentential form $vdVcw$. It contains dVc, and $T, U, V \in \mathbf{V}_{\text{ess}}$. Therefore $\mathcal{C}'_j$ contains the interior variable $[dVc]$, which contradicts the result of Claim 1.

Claim 3. Every $T \in \mathbf{V}_{\text{ess}}$ occurs in some inner variable of $\mathcal{C}'$. In particular, the component formed by the interior variables in $\mathbf{V}'$ is non-regular.

Proof. This is similar to the proof of Claim 2. First of all, left and right edges can also be defined in $\mathcal{D}(\mathcal{C})$ for the productions in $\mathbf{P}$ whose left-hand side is in $\mathbf{V}_{\text{ess}}$. Since $\mathbf{V}_{\text{ess}}$ is non-regular by assumption, there are both left and right edges in $\mathbf{V}_{\text{ess}}$. Thus, for every $T \in \mathbf{V}_{\text{ess}}$ we can find a path within $\mathbf{V}_{\text{ess}}$ that ends at T and contains both left and right edges. Arguing as above, we conclude that there are $a, b \in \mathbf{\Sigma}$ such that $[bTa]$ is an inner variable of $\mathcal{C}'$. We have proved the first part of the claim. Since $\mathbf{V}_{\text{ess}}$ has edges of both types, the component formed by the inner variables must also have both left and right edges and cannot be regular.

We have now verified condition (i) of Definition 2. It is easy to see that (ii) holds as well: take an element of L_2. It is of the form $\phi(v)$, where $v \in L$, $|v| \geq 2$. By assumption, there are $T \in \mathbf{V}_{\text{ess}}$ and $w \in L_T$ such that $v \sqsubset w$. By Claim 3, there are $a, b \in \mathbf{\Sigma}$ such that $[bTa] \in \mathbf{V}'_{\text{ess}}$. We have $\phi(v) \sqsubset \phi(bwa) \in L_{[bTa]}$. □

Exercise. Transform the grammar of Example 1 into a grammar that generates L_2, and draw the dependency di-graph of the latter.

With the last proposition we have concluded the proof of Theorem 1.

We remark that from the above proofs one sees more generally that the algorithm for passing from a reduced context-free grammar to a grammar that generates the corresponding two-block language preserves the number of non-regular components. This observation may support the belief that our approach of defining regular and non-regular components is quite natural and may be useful in other circumstances, too.

3. Generating functions satisfying polynomial equations

In preparing Step 2 of our proof, we present a general result regarding the radius of convergence of certain generating functions. Different variants of it, at different levels of generality, have been used in different contexts; see e.g. Lalley [24], [25], Hersonsky and Hubbard [20], Drmota [8] or Nagnibeda and Woess [31]. We have learned this from the paper of Lalley [24], whose method (regarding random walks on free froups) is exposed in detail in Woess [39, pp. 210-211]. For the sake of completeness, and also because assumptions and setup vary somewhat in these references, we shall present the proof.

Let $f_i(z) = \sum_{n \geq 0} f_{i,n}\, z^n$, $i = 1, \dots, \nu$, be the generating functions of the non-negative sequences $(f_{i,n})_{n \geq 0}$, and let $r_i \leq \infty$ be the radius of convergence of $f_i(z)$. We suppose that $f_i(0) = 0$ and that $r = \min_i r_i > 0$. We assume that the $f_i(z)$ satisfy a system of equations

$$f_i(z) = \mathcal{Q}_i\big(z, f_1(z), \dots, f_\nu(z)\big)\,, \quad i = 1, \dots, \nu\,, \tag{3.1}$$

where

$$\mathcal{Q}_i(z, y_1, \dots, y_\nu) = \sum_{|\mathbf{n}| \leq N} a_{i,\mathbf{n}}(z) \mathbf{y}^{\mathbf{n}}\,, \quad z \in \mathbb{C}\,,\ i = 1, \dots, \nu\,,$$

are polynomials of degree at most N in the variables $y_1, \dots, y_\nu$. Here, $\mathbf{y} = (y_1, \dots, y_\nu)$, $\mathbf{n} = (n_1, \dots, n_\nu) \in \mathbb{N}_0^\nu$, $\mathbf{y}^{\mathbf{n}} = y_1^{n_1} \cdots y_\nu^{n_\nu}$ and $|\mathbf{n}| = n_1 + \cdots + n_\nu$. We further assume that the coefficient functions $a_{i,\mathbf{n}}(z)$ are also expressed as power series around $z = 0$ with non-negative coefficients and radii of convergence $R_{i,\mathbf{n}}$ such that $R = \min_i R_i > 0$, where $R_i = \min_{\mathbf{n}} R_{i,\mathbf{n}}$. We assume that at least one of the functions $a_{i,\mathbf{n}}(z)$ is non-constant.

From the system (3.1), we want to extract information about the radii of convergence r_i. Note that r_i has to be a singularity of $f_i(z)$ by Pringsheim's theorem.

The *dependency di-graph* $\mathcal{D}$ of our system (3.1) of equations has vertex set $\{1, \dots, \nu\}$, and there is an oriented edge from i to j (notation $i \to j$), if y_j appears in a non-zero term of $\mathcal{Q}_i(z, y_1, \dots, y_\nu)$. The following is straightforward.

Lemma 1. *If $i \to j$, then $r_i \le \min\{r_j, R_i\}$. In particular, if $\mathcal{D}$ is strongly connected, then all r_i coincide, $r_i = r \le R$, and either $f_i(r) < \infty$ for all i or $f_i(r) = \infty$ for all i.*

For $0 \le z \le r$, consider the Jacobian matrix of our system of equations:

$$\mathfrak{J}(z) = \left(\frac{\partial \mathcal{Q}_i}{\partial y_j}\big(z, f_1(z), \dots, f_\nu(z)\big) \right)_{i,j=1}^{\nu} .$$

This is a non-negative matrix whose entries are increasing in $z \ge 0$. Our assumption that $f_i(0) = 0$ for all i implies that $\mathfrak{J}(0)$ is the zero matrix. Furthermore, for $0 < z < r$, the (i,j)-entry of $\mathfrak{J}(z)$ is positive precisely when $i \to j$ in $\mathcal{D}$.

We now assume that $\mathcal{D}$ is strongly connected, and use the Perron-Frobenius theory of nonnegative matrices; see Seneta [35]: if $0 < z < r$, then $\mathfrak{J}(z)$ has a positive eigenvalue $\lambda(z)$, which has maximal absolute value among all eigenvalues of $\mathfrak{J}(z)$, algebraic and geometric multiplicity 1 and strictly positive left and right eigenvectors. We have $\lambda(0) = 0$. As all entries of $\mathfrak{J}(z)$ increase with z, so does $\lambda(z)$.

Proposition 4. *If $\mathcal{D}$ is strongly connected, then $r < R$ if and only if there is $z \in (0\,,\, R)$ such that $\lambda(z) = 1$.*

If this is the case, then $\mathfrak{J}(r)$ is finite and $r = \min\{z > 0 : \lambda(z) = 1\}$.

Proof. We start by supposing that $r < R$; the "if" will become clear from the proof.

Consider first the case when each $\mathcal{Q}_i$ is affine in $y_1, \dots, y_\nu$, that is, $a_{i,\mathbf{n}}(z) = 0$ when $|\mathbf{n}| > 1$. Then (3.1) is a system of linear equations in $f_1(z), \dots, f_\nu(z)$. Therefore $\mathfrak{J}(z) = \big(a_{i,\mathbf{e}(j)}(z)\big)_{i,j=1}^{\nu}$, where $\mathbf{e}(j)$ is the unit vector with a 1 in position j. In particular, $\mathfrak{J}(z)$ is finite for $0 < z < R$. If we denote $\mathbf{f}(z) = \big(f_i(z)\big)_{i=1}^{\nu}$ and $\mathbf{a}(z) = \big(a_{i,\mathbf{0}}(z)\big)_{i=1}^{\nu}$, two column vectors, then we can write the system (3.1) in the form $\mathbf{f}(z) = \mathfrak{J}(z)\mathbf{f}(z) + \mathbf{a}(z)$. We find

$$\mathbf{f}(z) = \big(I - \mathfrak{J}(z)\big)^{-1}\mathbf{a}(z)\,.$$

Since both $\mathbf{a}(z)$ and $\mathfrak{J}(z)$ are analytic in each $z \in (0\,,\, R)$, we find a singularity of $\mathbf{f}(z)$ in the latter interval precisely where $I - \mathfrak{J}(z)$ is non-invertible. Thus, r is the unique positive solution of $\det\big(I - \mathfrak{J}(z)\big) = 0$. This proves the proposition when the system (3.1) is linear.

Let us now turn to the case when the system (3.1) is not linear. Then there must be $i(0) \in \{1, \dots, \nu\}$ such that the polynomial $\mathcal{Q}_{i(0)}$ contains a term of the form $a(z)y_{i(1)} \cdots y_{i(\ell)}$, where $a(z)$ is one of the non-zero $a_{i,\mathbf{n}}(z)$ and $\ell \ge 2$. The

indices $i(0), \dots, i(\ell)$ do not have to be distinct. In particular, we get

$$f_{i(0)}(z) \geq a\, z^b\, f_{i(1)}(z) \cdots f_{i(\ell)}(z) \quad \text{for } 0 < z < r\,, \quad (a > 0\,,\ b \in \mathbb{N}_0)\,. \tag{3.2}$$

Here, $a\, z^b$ is a nonzero term in the expansion of $a(z)$. Now, since the dependency graph is strongly connected, there is a path $i(\ell) = j(0) \to j(1) \to \cdots \to j(m) = i(0)$ in $\mathcal{D}$. This means that for each $\kappa \in \{1, \dots, m\}$, the variable $y_{j(\kappa)}$ occurs in a nonzero term of the polynomial $\mathcal{Q}_{j(\kappa-1)}$. Therefore for $0 < z < r$

$$f_{j(\kappa-1)}(z) \geq a_\kappa\, z^{b_\kappa}\, f_{j(\kappa)}(z) \times \text{further terms} \quad (a_\kappa > 0\,,\ b_\kappa \in \mathbb{N}_0)\,,$$

where the "further terms" are a possibly empty product of some of the functions $f_i(z)$. Concatenating all these inequalities, we find in the end an inequality of the same form as in (3.2), with possibly different values $a > 0$, $b \geq 0$ and $\ell \geq 2$, such that $i(\ell) = i(0)$. Dividing by $f_{i(0)}(z)$ and letting $z \to r$ from the left, we obtain

$$1 \geq a\, r^b\, f_{i(1)}(r) \cdots f_{i(\ell-1)}(r)\,.$$

In particular, $f_{i(1)}(r) < \infty$, and consequently we have the following.

$$\textit{If (3.1) is non-linear, then } \ f_i(r) < \infty \ \textit{ for all } i\,. \tag{3.3}$$

Note that the proof of (3.3) did not require that $r < R$ strictly.

(3.3) proves that $\mathfrak{J}(r)$ is finite in each entry.

For $0 \leq z \leq r$, $\lambda(z)$ is a strictly increasing, continuous (real-analytic) function with $\lambda(0) = 0$. Suppose that $\lambda(r) > 1$. Let $s = \min\{z > 0 : \lambda(z) = 1\}$. Then $s < r$, and each $f_i(z)$ is analytic at $z = s$. Let $(x_1, \dots, x_\nu)$ be a left Perron-Frobenius eigenvector (unique up to normalization) of $\mathfrak{J}(s)$ with eigenvalue $\lambda(s) = 1$. We have $x_i > 0$ for all i. Expand the function

$$\mathcal{F}(z, y_1, \dots, y_\nu) = \sum_{i=1}^{\nu} x_i \Big(y_i - \mathcal{Q}_i(z, y_1, \dots, y_\nu) \Big)$$

in a Taylor series around the point $\mathfrak{z} = \big(s, f_1(s), \dots, f_\nu(s)\big)$. Then

$$\mathcal{F}(\mathfrak{z}) = 0\,, \quad \frac{\partial \mathcal{F}}{\partial z}(\mathfrak{z}) = -\beta < 0\,, \quad \frac{\partial \mathcal{F}}{\partial y_i}(\mathfrak{z}) = 0,$$

$$\frac{\partial^2 \mathcal{F}}{\partial z^2}(\mathfrak{z}) = -\gamma_0 \leq 0\,, \quad \frac{\partial^2 \mathcal{F}}{\partial z \partial y_i}(\mathfrak{z}) = -\gamma_i \leq 0\,, \quad \frac{\partial^2 \mathcal{F}}{\partial y_i \partial y_j}(\mathfrak{z}) = -\gamma_{i,j} \leq 0$$

($\beta > 0$ holds because some $a_{i,\mathbf{n}}(z)$ is non-constant), and by the assumption of non-linearity of the system (3.1), there are i, j such that $\gamma_{i,j} > 0$. Substituting $y_i = f_i(z)$ in the Taylor expansion, we obtain

$$\begin{aligned} \beta(s - z) = {} & \tfrac{1}{2}\gamma_0 (s - z)^2 + \tfrac{1}{2}(s - z) \sum_{i=1}^{\nu} \gamma_i \big(f_i(s) - f_i(z) \big) \\ & + \tfrac{1}{2} \sum_{i,j=1}^{\nu} \gamma_{i,j} \big(f_i(s) - f_i(z) \big) \big(f_j(s) - f_j(z) \big) + h.o.t.\,, \end{aligned}$$

where $h.o.t.$ stands for higher-order terms. Since each $f_i(z)$ is analytic at $z = s$, we can insert their Taylor expansions at s and find an identity of the form $\beta(s - z) = \frac{1}{2}\gamma(s - z)^2 + h.o.t$, where $\gamma > 0$, a contradiction.

Therefore $\lambda(r) \leq 1$. If $\lambda(r) < 1$, then $I - \mathfrak{J}(r)$ would be invertible, and as we have assumed $r < R$, the Implicit Function Theorem would imply that the $f_i(z)$ are analytic at $z = r$. Thus $\lambda(r) = 1$. □

Remark 1. When (3.1) is nonlinear, proceeding as in Lalley [24] (see also Woess [39, p. 211]), one can show that for complex z near r, except for $z \in (r\,,\,\infty)$, the $f_i(z)$ have a Puiseux expansion

$$f_i(z) = f_i(r) - b_i(r-z)^{1/2} + h.o.t.$$

This relies on the fact that $z = r$ is an algebraic singularity of $f_i(z)$, since $r < R$.

It can be used to derive a precise asymptotic estimate of the coefficients $f_{i,n}$ of the form $c_i\, r^{-n}\, n^{-3/2}$, when one knows that $f_i(z)$ has no singularities on the circle of convergence besides $z = r$. □

Our next goal is to generalize Proposition 4 to the situation when the dependency di-graph is not necessarily strongly connected. We decompose $\mathcal{D}$: consider once more the equivalence relation $\sim$ on $\{1, \dots, \nu\}$, where $i \sim j$ if $i = j$ or there are oriented paths in $\mathcal{D}$ from i to j and from j to i. The equivalence classes are the strong components of $\mathcal{D}$. There is a partial order on the set of strong components; the component $[i]$ of i precedes the component $[j]$ of j (notation $[i] \preccurlyeq [j]$) if there is an oriented path from i to j in $\mathcal{D}$.

From Lemma 1, we see that $r_i = r_j$ when $i \sim j$. We write $r_{[i]}$ for the common radius of convergence associated with the strong component of i. Also, we write $R_{[i]} = \min\{R, r_{[j]} : [i] \prec [j]\}$. With a given strong component $[i]$, we associate the restricted Jacobian matrix

$$\mathfrak{J}_{[i]}(z) = \left(\frac{\partial \mathcal{Q}_k}{\partial y_l}\big(z, f_1(z), \dots, f_\nu(z)\big) \right)_{k,l \in [i]}.$$

It is well defined for $0 \le z < r_{[i]}$, since the $f_j(z)$ that actually do occur in each $\mathcal{Q}_k$, $k \in [i]$, are such that $[i] \preccurlyeq [j]$, whence $r_{[i]} \le r_{[j]}$. We write $\lambda_{[i]}(z)$ for the Perron-Frobenius eigenvalue of $\mathfrak{J}_{[i]}(z)$, $0 < z < r_{[i]}$. The following is now an immediate consequence of Proposition 4.

Corollary 1. *We have $r_{[i]} < R_{[i]}$ if and only if there is $z \in (0\,,\,R_{[i]})$ such that $\lambda_{[i]}(z) = 1$.*

If this is the case, then $\mathfrak{J}_{[i]}(r_{[i]})$ is finite and $r_{[i]} = \min\{z > 0 : \lambda_{[i]}(z) = 1\}$.

Proof. In each polynomial $\mathcal{Q}_k$, $k \in [i]$, we replace each y_j for which $[i] \prec [j]$ with the function $f_j(z)$. Thus, we obtain a new system of equations over $[i]$ which satisfies all assumptions of Proposition 4. □

Combining Proposition 4 with Corollary 1, we get the following without the assumption of strong connectedness of the dependency di-graph.

Corollary 2. *Suppose that the generating functions $f_i(z)$, $i = 1, \dots, \nu$, satisfy a system of polynomial equations of the form (3.1); let $r = \min_i r_i$ be their minimal radius of convergence, R the minimal radius of convergence of the coefficient functions of the $\mathcal{Q}_i$, and $\mathfrak{J}(z)$ the Jacobian matrix of the system (as above). Then $r < R$ if and only if there are $z \in (0\,,\,R)$ and i such that $\lambda_{[i]}(z) = 1$.*

If this is the case, then $\mathfrak{J}(r)$ is finite and there is an $i \in \{1, \dots, \nu\}$ such that $r = \min\{z > 0 : \lambda_{[i]}(z) = 1\}$.

Proof. Let the strong component $[i]$ be maximal with respect to $\prec$ such that $r = r_{[i]}$. Then $r_{[i]} < R_{[i]}$, and we can apply Corollary 1. □

4. Growth-sensitivity with respect to letters

We finally embark on Step 2 of our proof-strategy. Let L be generated by the essentially ergodic, reduced context-free grammar $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$ without chain rules. For $T \in \mathbf{V}$, we can consider $d_T(\cdot)$ as a measure on $\mathbf{\Sigma}^*$: for $M \subset \mathbf{\Sigma}^*$ (typically finite), $d_T(M) = \sum_{w \in M} d_T(w)$. We then can define the corresponding growth number

$$\gamma(L, \mathcal{C}) = \limsup_{n \to \infty} d_S(\mathbf{\Sigma}^n)^{1/n} .$$

This is the inverse of the radius of convergence of the power series

$$f_S(z) = \sum_{n=0}^{\infty} d_S(\mathbf{\Sigma}^n)\, z^n , \quad z \in \mathbb{C} .$$

We say that the grammar $\mathcal{C}$ is of *convergent type,* if $f_S(1/\gamma) < \infty$, and of *divergent type,* if $f_S(1/\gamma) = \infty$, where $\gamma = \gamma(L, \mathcal{C})$.

If $\mathcal{C}$ is unambiguous, $\gamma(L, \mathcal{C})$ coincides with $\gamma(L)$.

Now let $F \subset \mathbf{\Sigma}$ be a set of forbidden *letters.* Then clearly L^F is generated by the grammar $\mathcal{C}^F = (\mathbf{V}, \mathbf{\Sigma} \setminus F, \mathbf{P}^F, S)$, where $\mathbf{P}^F$ is the set of all productions in $\mathbf{P}$ whose right-hand sides contain no element of F. Of course, $\mathcal{C}^F$ is not necessarily essentially ergodic. However, it is important to note that $\mathcal{C}^F$ is unambiguous when $\mathcal{C}$ is unambiguous. Step 2 will be accomplished by the following.

Theorem 2. *Suppose that L is generated by the essentially ergodic, reduced, context-free grammar $\mathcal{C}$ without chain and ε-rules. If $\mathcal{C}$ is of convergent type, then for any non-empty $F \subset \mathbf{\Sigma}$,*

$$\gamma(L^F, \mathcal{C}^F) < \gamma(L, \mathcal{C}) .$$

In particular, if $\mathcal{C}$ is also unambiguous, L is growth-sensitive.

The starting point is a famous theorem of Chomsky and Schützenberger [7] which transforms a context-free grammar into a system of algebraic equations for the associated formal power series. Let $\mathcal{C} = (\mathbf{V}, \mathbf{\Sigma}, \mathbf{P}, S)$ be any reduced, context-free grammar without chain and ε-rules. With each $T \in \mathbf{V}$ we associate the generating function (power series)

$$f_T(z) = \sum_{n=0}^{\infty} d_T(\mathbf{\Sigma}^n)\, z^n , \quad z \in \mathbb{C} .$$

Furthermore, consider complex variables y_T, $T \in \mathbf{V}$, in addition to the complex variable z. Define $\pi(a) = z$ for every $a \in \mathbf{\Sigma}$ and $\pi(T) = y_T$ for every $T \in \mathbf{V}$, and for $u = u_1 \cdots u_k \in (\mathbf{\Sigma} \cup \mathbf{V})^*$, let $\pi(u) = \pi(u_1) \cdots \pi(u_k)$, where the latter is a product of commuting complex variables. With $T \in \mathbf{V}$ we associate the polynomial

$$\mathcal{P}_T(z; y_U, U \in \mathbf{V}) = \sum_{T \vdash u} \pi(u) \tag{4.1}$$

in the complex variables z and y_U, $U \in \mathbf{V}$. Then the theory of Chomsky and Schützenberger [7] implies that the functions $f_T(z)$ satisfy the system of equations

$$f_T(z) = \mathcal{P}_T\big(z; f_U(z), U \in \mathbf{V}\big) , \quad T \in \mathbf{V} . \tag{4.2}$$

We remark that Chomsky and Schützenberger prove this in terms of formal power series in non-commuting variables; we have turned them into commuting complex variables via the homomorphism π.

Each $f_T(z)$ is a power series with non-negative coefficients. Therefore its radius of convergence r_T is the smallest positive singularity of $f_T(z)$. We have

$$\gamma(L,\mathcal{C}) = 1/r\,, \quad \text{where } r = r_S\,.$$

Furthermore, each of the polynomials $\mathcal{P}_T$ has non-negative coefficients.

The dependency-digraph of (4.2) is of course $\mathcal{D}(\mathcal{C})$. In particular, in analogy with Lemma 1, we have

Lemma 2. *If $T, U \in \mathbf{V}$ and $T \overset{*}{\to} U$ in $\mathcal{D}(\mathcal{C})$, then $r_T \le r_U$. In particular, the radius of convergence is the same for variables in the same strong component.*

The following lemma (as well as the last one) does not require essential ergodicity.

Lemma 3. *Suppose that $L(\mathcal{C})$ is infinite.*

(a) *If $\mathcal{C}$ is of convergent type, then it has non-regular variables, and there is a non-regular strong component $\mathbf{V}_j$ such that $r_T = r$ and $f_T(r) < \infty$ for every $T \in \mathbf{V}_j$.*

(b) *On the other hand, if all variables are regular, then L_T is a rational language and $f_T(z)$ is a rational function for each $T \in \mathbf{V}$.*

Proof. (a) If S is non-regular, then we have nothing to prove. Assume that $S \in \mathbf{V}_{\mathrm{reg}}$. Since $\mathcal{C}$ is reduced, we have $S \overset{*}{\to} T$ for every non-regular $T \in \mathbf{V}$, whence $r \le r_T$. Now let $\mathbf{V}_0$ be the strong component of S and $\mathcal{C}_0$ the associated grammar. It is left or right linear. Via the system (4.2), this implies that $f_S(z)$ is a *rational* function of z and of the functions $f_T(z)$, where $S \to T$ and $T \notin \mathbf{V}_0$. For any of these variables T that is regular, we can use the same argument and continue.

At the end of this recursive procedure, we find that $f_S(z)$ is a rational function of z and of the functions $f_T(z)$, where $T \in \mathbf{V} \setminus \mathbf{V}_{\mathrm{reg}}$, i.e., there are relatively prime polynomials $\mathcal{R}_1$ and $\mathcal{R}_2$ in the variables z and y_T, $T \in \mathbf{V} \setminus \mathbf{V}_{\mathrm{reg}}$, such that

$$f_S(z) = \frac{\mathcal{R}_1(z; f_T(z), T \in \mathbf{V} \setminus \mathbf{V}_{\mathrm{reg}})}{\mathcal{R}_2(z; f_T(z), T \in \mathbf{V} \setminus \mathbf{V}_{\mathrm{reg}})}\,.$$

Both numerator and denominator are analytic for positive $z < \tilde{r} = \min\{r_T : T \in \mathbf{V} \setminus \mathbf{V}_{\mathrm{reg}}\}$. Recall that r is the smallest positive singularity of $f_S(z)$. Thus $r < \tilde{r}$ would yield that this had to be a pole of $f_S(z)$, contradicting the fact that $\mathcal{C}$ is of convergent type. This means that $r = \tilde{r}$.

We cannot have $f_T(r) = \infty$ for all $T \in \mathbf{V} \setminus \mathbf{V}_{\mathrm{reg}}$, since otherwise, by Lemma 1, also $f_S(r) = \infty$, as $S \overset{*}{\to} T$ in $\mathcal{D}(\mathcal{C})$. This completes the proof of (a).

(b) Now suppose on the other hand that all variables are regular. We proceed inductively. Let $\mathbf{V}_j$ be a strong component of $\mathcal{D}(\mathcal{C}_j)$. If it is maximal with respect to $\preccurlyeq$, then clearly L_T is regular for each $T \in \mathbf{V}_j$. Next, supppose that $\mathbf{V}_j$ is such that L_U is regular for all $U \in \mathbf{V} \setminus \mathbf{V}_j$ with $T \overset{*}{\to} U$. Since the grammar $\mathcal{C}_j$ is regular by assumption, we can use the Substitution Theorem for regular languages; see Harrison [19, §3.4]: substituting L_U in the place of $c_U \in \mathbf{\Sigma}_j$ for each $U \in \mathbf{V}_k$ with $\mathbf{V}_j \prec \mathbf{V}_k$ implies regularity of L_T, $T \in \mathbf{V}_j$. The same inductive argument also shows rationality of all $f_T(z)$. □

Proof of Theorem 2. Let $\mathcal{C}$ be a reduced, ergodic context-free grammar that has no chain or ε-rules. Our first step is to restrict the system (4.2) to the variables of $\mathbf{V}_{\mathrm{ess}}$. We enumerate $\mathbf{V}_{\mathrm{ess}} = \{T_1, T_2, \dots, T_\nu\}$ and write $y_i = y_{T_i}$ and $f_i(z) = f_{T_i}(z)$ for the complex variables and generating functions associated with the T_i. For each

$i \in \{i, \dots, \nu\}$, we define a polynomial $\mathcal{Q}_i(z, y_1, \cdots, y_\nu)$ in the y_i by substituting in $\mathcal{P}_{T_i}$ for each appearing y_U with $U \notin \mathbf{V}_{\text{ess}}$ the corresponding function $f_U(z)$. The latter $U \in \mathbf{V}$ must be such that $T \to U$ in one step in $\mathcal{D}(\mathcal{C})$. In particular, by Lemma 3(b), $f_U(z)$ is rational and either $r_U = \infty$ or $f_U(r_U) = \infty$.

We have obtained a system that is precisely of the form (3.1), and the coefficient functions $a_{i,\mathbf{n}}(z)$ are generating functions of nonnegative sequences that are either polynomials or rational functions. Therefore either $R = \infty$ or else R is a pole of one of the coefficient functions. On the other hand, we are considering infinite languages, whence $r < \infty$, and by Lemmas 1 and 3, also $f_i(r) < \infty$ for all i. Thus, we may apply Proposition 4 and find that $r = \min\{z > 0 : \lambda(z) = 1\}$.

Next, given $F \subset \Sigma$, consider the grammar $\mathcal{C}^F$. It is reduced, but not necessarily ergodic. If necessary, we perform another reduction by eliminating from $\mathbf{V}$ all variables that cannot be reached from S in the dependency di-graph of $\mathcal{C}^F$, thus obtaining a set of variables $\mathbf{V}^F \subset \mathbf{V}$ and the corresponding subset $\mathbf{V}^F_{\text{ess}} = \{T_1, \dots, T_{\nu'}\} \subset \mathbf{V}_{\text{ess}}$ (we assume without loss of generality that the deleted variables are $T_{\nu'+1}, \dots, T_\nu$). Also, we eliminate from $\mathbf{P}^F$ the "superfluous" production rules (those that contain some variable in $\mathbf{V} \backslash \mathbf{V}^F$). For the sake of simplicity, we shall again write $\mathbf{P}^F$ for this new set of production rules, and $\mathcal{C}^F = (\mathbf{V}^F, \mathbf{\Sigma}, \mathbf{P}^F, S)$. For the associated polynomials, Jacobian matrix, Perron-Frobenius eigenvalue, etc., we shall always use the superscript F. Thus, writing $r^F = r^F_S$, we have $r^F = 1/\gamma(L^F, \mathcal{C}^F)$. Since the coefficients of $f^F_S(z)$ are $\leq$ those of $f_S(z)$, we have $r^F \geq r_S = r$.

Case 1. $\mathcal{C}^F$ is of divergent type. Since $f^F_S(r) \leq f_S(r) < \infty$, we cannot have $r^F = r$. Therefore $r^F > r$, and $\gamma(L^F, \mathcal{C}^F) < \gamma(L, \mathcal{C})$.

Case 2. $\mathcal{C}^F$ is of convergent type. Then the strong components of $\mathcal{D}(\mathcal{C}^F)$ cannot all be regular by Lemma 3(a). Observe that $\mathcal{D}(\mathcal{C}^F)$ is obtained from $\mathcal{D}(\mathcal{C})$ by deleting some edges and vertices. Therefore each strong component of $\mathcal{D}(\mathcal{C}^F)$ is – as a set of vertices – contained in some strong component of $\mathcal{D}(\mathcal{C})$. In particular, every non-regular strong component of $\mathcal{D}(\mathcal{C}^F)$ must be contained in $\mathbf{V}_{\text{ess}}$. Since $\mathcal{C}^F$ is reduced, each of these components can be reached from S in $\mathcal{D}(\mathcal{C}^F)$. Lemma 3 now shows that we must have

$$r^F = \min\{r^F_i : i = 1, \dots \nu'\},$$

where $r^F_i = r^F_{T_i}$.

Now let $U \in \mathbf{V}^F \setminus \mathbf{V}_{\text{ess}}$ be a variable that occurs in the right-hand side of some production $T \vdash u$ in $\mathbf{P}^F$, where $T \in \mathbf{V}_{\text{ess}}$. Since passing from $\mathcal{C}$ to $\mathcal{C}^F$ preserves regularity of the "surviving" variables, we see that U is a regular variable of $\mathcal{C}^F$. Lemma 3(b) shows that either $r^F_U = \infty$ or r^F_U is a pole of $f^F_U(z)$. In both cases, we must have $r^F < r^F_U$, since L^F is of convergent type.

For $i = 1, \dots, \nu'$, we can now construct the polynomials $\mathcal{Q}^F_i(z, y_1, \dots, y_{\nu'})$ in the variables $y_1, \dots, y_{\nu'}$ in the same way as we did above for $\mathcal{Q}_i(z, y_1, \dots, y_\nu)$. Their coefficient functions arise from the rational functions $f^F_U(z)$ that we have just discussed. In particular, denoting by R^F the minimum among the radii of convergence of these coefficient functions, we have $r^F < R^F$.

Thus, we can apply Corollary 2. Let $\mathcal{D}^F$ be the dependency-digraph of the $\mathcal{Q}^F_i$, $i = 1, \dots, \nu'$ (a subgraph of $\mathcal{D}(\mathcal{C}^F)$). Then there is a strong component $[k]$ of $\mathcal{D}^F$ such that $r^F = \min\{z > 0 : \lambda^F_{[k]}(z) = 1\}$. Consider the matrix $\mathfrak{J}^F_{[k]}(z)$. We extend

it to a matrix over $\{1,\dots,\nu\}$ by setting all elements outside of $[k]\times[k]$ equal to 0. We also write $\mathfrak{J}^F_{[k]}(z)$ for this extended matrix.

Consider the generating functions $f_i^F(z)$ and the "originial" $f_i(z)$ associated with $\mathcal{C}$, where $i\in\{1,\dots,\nu'\}$. Then clearly $r_i^F\geq r$. At this point, recall condition (ii) in the definition of essential ergodicity. It implies that $f_i^F(z)<f_i(z)$ strictly for $0<z\leq r$, since $L^F_{T_i}\subset L_{T_i}$ strictly, because the latter language contains words having a letter in F. This implies that for all $z\in(0\,,\,r]$, $i\in\{1,\dots,\nu'\}$,

$$\mathcal{Q}_i^F\big(z,f_1^F(z),\dots,f_{\nu'}^F(z)\big)<\mathcal{Q}_i\big(z,f_1(z),\dots,f_\nu(z)\big)\,.$$

We now appeal once more to the assumption that L^F is of convergent type. It implies that there must be some rule of the form $T\vdash u$ in $\mathbf{P}^F$ such that $T\in\mathbf{V}_{\mathrm{ess}}\cap\mathbf{V}^F$ and u contains at least two (not necessarily distinct) elements of $\mathbf{V}_{\mathrm{ess}}\cap\mathbf{V}^F$. Indeed, otherwise the system

$$f_i^F(z)=\mathcal{Q}_i^F\big(z,f_1^F(z),\dots,f_{\nu'}^F(z)\big)\,,\quad i=1,\dots,\nu'\,,$$

would be linear, its solutions rational functions in z, and their singularities would be poles.

Collecting all these facts, we get that

$$\mathfrak{J}^F_{[k]}(z)\leq\mathfrak{J}(z)\quad\text{and}\quad\mathfrak{J}^F_{[k]}(z)\neq\mathfrak{J}(z)\quad\text{for all } z\in(0\,,\,r]\,.$$

But then Thm. 1.6 in Seneta [35] implies that

$$\lambda^F_{[k]}(z)<\lambda(z)\leq\lambda(r)=1\quad\text{for all } z\in(0\,,\,r]\,.$$

This yields that $r^F>r$, thus concluding the proof. □

Remark 2. In the case of a (strictly) ergodic grammar $\mathcal{C}$, non-linearity implies that $\mathcal{C}$ is of convergent type, as follows from (3.3). Thus the Main Theorem stated in the introduction follows. □

5. An extension

The same method that we have used for the unambiguous case also applies for proving the following extension of the Main Theorem, resp. Theorem 2.

Theorem 3. *Suppose that L is generated by the essentially ergodic, reduced, context-free grammar $\mathcal{C}$ without chain and ε-rules. If $\mathcal{C}$ is of convergent type, then*

$$\gamma(L^F,\mathcal{C}^F)<\gamma(L,\mathcal{C})$$

for any nonempty $F\subset\mathrm{SUB}(L)$.

The point is that F may be arbitrary (not necessarily a subset of $\mathbf{\Sigma}$ as in Theorem 2). In particular, we get a version of the Main Theorem that allows ambiguity.

Proof (outline). We can use the same two steps as before, and Step 2 remains unchanged. Modifications are necessary in Step 1, because the initial steps before constructing a grammar for the 2-block language may decrease the ambiguity degrees, which had no importance in the unambiguous case.

A way to overcome this is to consider context-free grammars with *weighted productions.* That is, we have $\mathcal{C}=(\mathbf{V},\mathbf{\Sigma},\mathbf{P},d,S)$ where $\mathbf{V},\mathbf{\Sigma},\mathbf{P},S$ have maintained their meaning, and $d:\mathbf{P}\to\mathbb{N}$ assigns a positive integer *weight* $d(T\vdash v)$ to each production $T\vdash v$. The case when $d(T\vdash v)=1$ for every production corresponds to an "ordinary" context-free grammar. In general, when $d(T\vdash v)=k$, one should

think of this as k "parallel" productions of the same form, each of which can be chosen in a step of a rightmost derivation where T is replaced by v.

For a rightmost derivation $T \stackrel{*}{\Longrightarrow} w$, its weight is then the product of the weights of all production rules used in that derivation. Finally, for $w \in L_T$ we redefine its *ambiguity degree* $d_T(w)$ as the sum of the weights of all rightmost derivations $T \stackrel{*}{\Longrightarrow} w$.

It is then easy to take care of the weights of the productions in each of the steps described in §2 so that the ambiguity degrees for the respective start symbols remain the same.

For example, in §2.**B** (ε-freeness), suppose we have a rule $T \vdash w$ in $\mathbf{P}$ and transform it into a rule $T \vdash v$ in $\mathbf{P}'$ by deleting from w the variables $T_1, \ldots, T_k \in \mathbf{V}_\varepsilon$. Then we have to set the weight of $T \vdash v$ in $\mathbf{P}'$ as

$$d'(T \vdash v) = d(T \vdash w) \cdot d_{T_1}(\varepsilon) \cdots d_{T_k}(\varepsilon)\,.$$

We leave it as an exercise to work out the necessary modifications in the other steps described in §2.

Therefore, in the end one finds a grammar for L_2 where

$$d_{[S]}\big(\phi(w)\big) = d_S(w) \quad \text{for every } w \in L\,,\ |w| \geq 2\,.$$

The remaining parts of the proof are unchanged. □

We remark that the natural realm for considerations like those of the last proof is that of formal power series with integer coefficients, compare e.g. with Kuich [23].

This may also be the right place to explain why our approach fails for *linear* grammars. As mentioned in the introduction, in the definition of ergodicity for regular and linear grammars we need an additional requirement.

To see this, consider as an example the linear grammar with the only variable S, alphabet $\mathbf{\Sigma} = \{a, b, c\}$ and the production rules $S \vdash aSb$, $S \vdash c$ and $S \vdash ba$. It generates the language $\{a^n c b^n, a^n b a b^n : n \geq 0\}$ which is insensitive with respect to forbidding c as well as with respect to forbidding ba. The same is true for the right linear grammar obtained by exchanging a and S in the right-hand side of the first of the three rules.

A possible way to overcome this is to require that every terminal rule be an ε-rule. More generally, we can require that every right-hand side $v \in \mathbf{\Sigma}^*$ of a terminal rule $T \vdash v$ appear as a subword of some non-terminal sentential form, i.e., one in $\mathbf{\Sigma}^*\mathbf{V}\mathbf{\Sigma}^*$. The latter property is preserved by passing to the 2-block grammar.

With this modified definition of (strong) ergodicity, growth-sensitivity with respect to *letters* remains valid. However, Step 1 fails in the linear (non-regular) case, because the grammar for L_2 according to (2.3) may contain variables of all three types $[bT]$, $[Ta]$ and $[bTa]$ (while in the right linear case only the first type occurs). Thus, there will be several strong components in the di-graph corresponding to L_2, even when there was only one at the beginning. But then the problem is that linear languages are of divergent type, so that with an adapted version of essential ergodicity one cannot use the argument of Lemma 3 to show that r is the radius of convergence associated with the only essential component.

On the other hand, in the regular case, starting with a strictly ergodic grammar for L, the regular grammar corresponding to (2.3) (it is much simpler than that) will also be strictly ergodic with the only exception of the isolated start symbol $[S]$.

These last lines also explain how to use a (simpler) variant of the methods used here in order to prove growth-sensitivity of ergodic, regular languages.

6. Examples of languages associated with groups

It was the connection with finitely generated groups that originally motivated the present research. We refer to Gilman [11] and Rees [33] for surveys on groups and languages.

Let G be a finitely generated group, and let $\mathbf{\Sigma}$ be a finite alphabet. Suppose that we have a mapping $\pi : \mathbf{\Sigma} \to G$ (typically, but not necessarily one-to-one), and write π also for its canonical extension as a semigroup homomorphism from the free monoid $\mathbf{\Sigma}^*$ into G. We assume that this extension is onto, i.e., G is generated as a monoid by the set $A = \pi(\mathbf{\Sigma})$. The corresponding *Cayley graph* $X(G, \mathbf{\Sigma}, \pi)$ of G is the oriented, labelled multigraph with vertex set G, where for each $g \in G$ and $a \in \mathbf{\Sigma}$ there is an edge with label a from g to $g\pi(a)$. (The main reason why we distinguish between A and $\mathbf{\Sigma}$ is that in certain situations some care is necessary when dealing with idempotents in A.)

In this setting, there are various languages over $\mathbf{\Sigma}$ which are of interest in the study of the structure of G. Here, we shall consider two classes.

A. Geodesic normal forms, growth and relative growth. A *normal form* is a language $L = \mathrm{Nf}_G \subset \mathbf{\Sigma}^*$ such that $\pi|_L$ is one-to-one onto G. For $g \in G$, we write $w(g) = \pi|_L^{-1}(g)$. A normal form is called *geodesic* if the word length $|w(g)|$ is minimal for each $g \in G$. In this case, the growth of the language Nf_G is the growth of the group G with respect to A. Furthermore, if H is a subgroup of G, then the growth of the sublanguage $\mathrm{Nf}_G(H) = \{w(h) : h \in H\}$ of Nf_G is the *relative growth* of H in G with respect to A.

In the context of regular normal forms of finitely generated groups, the study of growth-sensitivity has been proposed by Grigorchuk and de la Harpe [14] as a tool for proving Hopfianity of a given group or class of groups; see also Ceccherini-Silberstein and Scarabotti [5]. A group is called *Hopfian* if it is not isomorphic with a proper quotient of itself. The basic example where this tool applies is the free group.

Example 1 (free groups)**.** Let $\mathbb{F}_m$ be the free group with free generators $a_1, \dots, a_m$. We set $A = \{a_1, a_1^{-1}, \dots, a_m, a_m^{-1}\}$ and $\mathbf{\Sigma} = \{\mathbf{a}_1, \mathbf{a}_{-1}, \dots, \mathbf{a}_m, \mathbf{a}_{-m}\}$. The mapping π is defined by $\pi(\mathbf{a}_{\pm i}) = a_i^{\pm 1}$. In this case there is of course a unique geodesic normal form $\mathrm{Nf}_{\mathbb{F}_m}$ which consists of all reduced words over $\mathbf{\Sigma}$, where *reduced* means that none of the strings $\mathbf{a}_i\mathbf{a}_{-i}$, $i \in I = \{\pm 1, \cdots, \pm m\}$, may appear as a subword. There is a well known unambiguous, right linear grammar generating this language. We set $\mathbf{V} = \{S, T_i : i \in I\}$; the production rules are

$$S \vdash \varepsilon\,, \quad S \vdash \mathbf{a}_i T_i \quad \forall\, i \in I\,, \quad T_i \vdash \varepsilon\,, \quad T_i \vdash \mathbf{a}_j T_j \quad \forall j \in I \setminus \{-i\}\,.$$

This grammar is ergodic with exception of the isolated start symbol, whence $\mathrm{Nf}_{\mathbf{F}_m}$ is growth-sensitive. □

In the 1990's, it has become very popular to study normal forms of groups and similarly associated languages that are regular; see the book by Epstein with coauthors [10]. However, there is no general theory regarding growth-sensitive automatic groups. Arzhantseva and Lysenok [1] have recently proved that every

non-elementary word-hyperbolic group has a growth-sensitive, regular geodesic normal form. For a specific example, where other related languages are also studied, see Bartholdi and Ceccherini-Silberstein [2].

The next question is whether there are groups with context-free geodesic normal forms to which our Main Theorem or Theorem 3 can be applied. The answer is "no" because of the following simple observation; see e.g. Parry [32, statement (0.1)].

Lemma 5. *If* Nf_G *is a geodesic normal form of a finitely generated group, then* Nf_G *is of divergent type.*

Proof. The number of words of length n in Nf_G is the same as s_n, the number of elements at distance n from id_G in the corresponding Cayley graph of the given group G. Now we have $s_{m+n} \le s_m s_n$ for all m, n, whence the multiplicative version of Kingman's Subadditive Lemma (compare with the supermultiplicative variant in Seneta [35, Lemma A.4]) implies that $s_n^{1/n} \to \gamma(\mathrm{Nf}_G)$ from above. In particular, $s_n \gamma(\mathrm{Nf}_G)^{-n} \ge 1$ for all n. □

We remark immediatley that the same argument does not apply when we consider relative growth of a subgroup, since submultiplicativity fails in this situation. We shall now give an example of a geodesic normal form Nf_G of a finitely generated group G that is context-free, nonlinear, unambiguous and essentially ergodic, but – in view of Lemma 5 – of course not of convergent type. We shall also exhibit a natural, infinitely generated subgroup G_0 such that $\mathrm{Nf}_G(G_0)$ has all the above properties and *is* of convergent type, so that Theorem 2 applies. These examples are based on Parry [32].

Example 2 (lamplighter groups)**.** This is a fancier name for what is usually called a *wreath product* $K \wr H$ of two groups H and K.

A *finite configuration* is a function $\eta : H \to K$ such that its *support* $\mathrm{supp}(\eta) = \{h \in H : \eta(h) \ne \mathrm{id}_K\}$ is finite. The product of two configurations is their pointwise product in K. Thus, the set of all finite configurations becomes a group G_0 whose identity element is the configuration ι_{id_K} with empty support. We can embed K into G_0 via $k \mapsto \iota_k$, where ι_k is the configuration with $\iota_k(\mathrm{id}_H) = k$ and $\iota_k(h) = \mathrm{id}_K$ if $h \ne \mathrm{id}_H$. The group H acts on G_0 via $(h, \eta) \mapsto T_h\eta$, where $T_h\eta(h') = \eta(h^{-1}h')$. Our lamplighter group $G = H \wr K$ is the resulting semidirect product $H \leftthreetimes G_0$. It consists of all pairs (h, η), where $h \in H$ and $\eta \in G_0$, and the group operation is

$$(h_1, \eta_1)(h_2, \eta_2) = (h_1 h_2, \eta_1 T_{h_1}\eta_2)\,.$$

Its unit element is $(\mathrm{id}_H, \iota_{\mathrm{id}_K})$. We have embeddings $H \hookrightarrow G$ via $h \mapsto (h, \iota_{\mathrm{id}_K})$ and $K \hookrightarrow G$ via $k \mapsto (\mathrm{id}_H, \iota_k)$. Also, G_0 embeds into G by $\eta \mapsto (\mathrm{id}_H, \eta)$. In the sequel, we shall identify H, K and G_0 with their respective embeddings into G. Thus, every element $g = (h, \eta) \in G$ can be written (non-uniquely!) as a product

$$g = h_0 k_1 h_1 k_2 \cdots h_{q-1} k_q h_q\,, \quad \text{where}$$
$$q \ge 0\,,\ h_0, h_q \in H\,,\ h_1, \ldots, h_{q-1} \in H \setminus \{\mathrm{id}_H\}\,,\ k_1, \ldots, k_q \in K \setminus \{\mathrm{id}_K\}\,.$$

If A_H and A_K are generating sets of H and K, respectively, then the union of their embeddings is a generating set of G. Here, we shall assume that A_H is *symmetric* ($A_H^{-1} = A_H$).

Now suppose that we have disjoint alphabets $\mathbf{\Sigma}_H$ and $\mathbf{\Sigma}_K$ with corresponding mappings π_H and π_K, respectively. We set $\mathbf{\Sigma} = \mathbf{\Sigma}_H \cup \mathbf{\Sigma}_K$ and write π for the

corresponding mapping $\mathbf{\Sigma}^* \to G$, where of course $\pi(a)$ is intended as the embedding into G of $\pi_H(a)$ or $\pi_K(a)$, respectively. Let Nf_H and Nf_K be geodesic normal forms of H and K associated with $(\mathbf{\Sigma}_H, \pi_H)$ and $(\mathbf{\Sigma}_K, \pi_K)$, respectively. Parry [32, Thm. 1.2] describes a method for obtaining from the latter a geodesic normal form Nf_G of G associated with $(\mathbf{\Sigma}, \pi)$.

Let $g = (h, \eta)$. In the Cayley graph $X(H, \mathbf{\Sigma}_H, \pi_H)$, select a shortest *tour* that starts in id_H, visits all points in $\mathrm{supp}(\eta)$ and ends up in h. That is, we enumerate $\mathrm{supp}(\eta) = \{h_1, \dots, h_q\}$ in such a way that, setting $h_0 = \mathrm{id}_H$ and $h_{q+1} = h$, the length $\sum_{i=1}^{q} \mathrm{dist}(h_{i-1}, h_i)$ of the tour is minimal. Here, dist denotes the graph distance in $X(H, \mathbf{\Sigma}_H, \pi_H)$. Set $k_i = \eta(h_i)$, $i = 1, \dots, q$. Then

$$g = h'_0 k_1 h'_1 k_2 \cdots h'_{q-1} k_q h'_q \quad \text{with} \quad h'_i = h_i^{-1} h_{i+1}\,,\ i = 0, \dots, q\,.$$

If we choose shortest representations of the h'_i and k_i, then this will be a shortest representation of G. In other words,

$$w_G(g) = w_H(h'_0) w_K(k_1) w_H(h'_1) w_K(k_2) \cdots w_H(h'_{q-1}) w_K(k_q) w_H(h'_q)\,,$$

where w_H and w_K correspond to the normal forms Nf_H and Nf_K, respectively, and $w_G(g)$ – which depends on the choice of the tour – is the representation of g in our normal form Nf_G of G.

If one wants to formalize Nf_G in terms of a grammar, then the main problem is to do this in terms of an algorithm that produces admissible tours without ambiguity. Parry [32] shows how this can be done when the Cayley graph of H is a (necessarily homogeneous) tree $\mathcal{T}$. Here, we explain this in the (not really restrictive) case when

$$H = \langle a_1, \dots, a_N \mid a_1^2 = \cdots = a_N^2 = \mathrm{id} \rangle$$

is the free product of $N \geq 2$ copies of the two-element-group. It is generated by the symmetric set $A_H = \{a_1, \dots, a_N\}$. Let $\mathbf{\Sigma} = \{\mathbf{a}_1, \dots, \mathbf{a}_N\}$ and $\pi_H(\mathbf{a}_i) = a_i$. Every element of H has a unique shortest representation

$$h = a_{i_1} \cdots a_{i_n}\,, \quad n \geq 0\,,\ a_{i_{j+1}} \neq a_{i_j}\,.$$

Obviously, for our geodesic normal form of H we choose $w_H(h) = \mathbf{a}_{i_1} \cdots \mathbf{a}_{i_n}$. In the (non-oriented) tree $\mathcal{T} = X(H, \mathbf{\Sigma}_H, \pi_H)$, two elements h, h' are neighbours if $h' = h a_i$ for some i. The label of that edge (in both orientations) is $\mathbf{a}_i$.

If we have $g = (h, \eta) \in G$, then a tour as above must visit every vertex of the finite subtree $\mathcal{T}_0(g)$ spanned by $\mathrm{supp}(\eta) \cup \{\mathrm{id}_H, h\}$. The vertex set of $\mathcal{T}_0$ consists of all elements h of H for which $w_H(h)$ is a (possibly trivial) prefix of some $w_H(h_i)$.

A tour is obtained by running along a "contour" of a plane realization of $\mathcal{T}_0(g)$ that starts in id_H and ends in h. To specify one given tour, we may agree to use a contour that follows the lexicographic ordering induced by $\mathbf{a}_1 < \cdots < \mathbf{a}_n$. In addition, when $h \neq \mathrm{id}_H$, this order has to be modified such that the branch containing h is the last one which the tour visits. Compare with the explanation given by Parry [32, statements (2.2)–(2.4)]. Thus $\mathcal{T}_0(g)$ and $h \in \mathcal{T}_0$ determine the tour uniquely. Conversely, given a finite subtree $\mathcal{T}_0 \neq \mathrm{id}_H$ and an element $h \in \mathcal{T}_0$, then for an element $g = (h, \eta) \in G$ we will have $\mathcal{T}_0 = \mathcal{T}_0(g)$ if and only if $\partial \mathcal{T}_0 \subset \mathrm{supp}(\eta) \cup \{h\}$. Here, $\partial \mathcal{T}_0$ denotes the set of *leaves* of $\mathcal{T}_0$ (the vertices of $\mathcal{T}_0 \setminus \{\mathrm{id}_H\}$ having only one neighbour in $\mathcal{T}_0$).

For example, if $\partial \mathcal{T}_0 = \{a_1, a_2 a_1, a_2 a_3, a_3 a_1, a_3 a_2\}$ and $h = a_2$, then the tour is

$$\begin{aligned} &\mathrm{id}_H \to a_1 \to \mathrm{id}_H \to a_3 \to a_3 a_1 \to a_3 \to a_3 a_2 \\ &\qquad \to a_3 \to \mathrm{id}_H \to a_2 \to a_2 a_1 \to a_2 \to a_2 a_3 \to a_2 = h\,. \end{aligned} \tag{6.1}$$

The successive labels read along this tour give the word

$$w = \mathbf{a}_1\mathbf{a}_1\mathbf{a}_3\mathbf{a}_1\mathbf{a}_1\mathbf{a}_2\mathbf{a}_2\mathbf{a}_3\mathbf{a}_2\mathbf{a}_1\mathbf{a}_1\mathbf{a}_3\mathbf{a}_3\,.$$

The leaves of $\mathcal{T}_0$ are recognizable from w: if one sees a substring $\mathbf{a}_i\mathbf{a}_i$, then one truncates w after the first of the two and takes π_H of the remaining prefix of w. In order to obtain an element (h,η) corresponding to $\mathcal{T}_0$, one *has* to insert an element of $K\setminus\{\mathrm{id}_K\}$ at each leaf (unless it is the endpoint of the tour), and one *may* insert one at every other vertex (but only once, e.g. at the first visit of the tour!). Thus, the elements of our language Nf_G that correspond to the tour described by w will be of the form

$$\tilde{w} = u_1\mathbf{a}_1v_1\mathbf{a}_1\mathbf{a}_3u_2\mathbf{a}_1v_2\mathbf{a}_1\mathbf{a}_2v_3\mathbf{a}_2\mathbf{a}_3\mathbf{a}_2u_3\mathbf{a}_1v_4\mathbf{a}_1\mathbf{a}_3v_5\mathbf{a}_3\,,$$

where $u_1,u_2,u_3\in\mathrm{Nf}_K$ and $v_1,\dots,v_5\in\mathrm{Nf}_K\setminus\{\varepsilon\}$. The resulting element $g=(h,\eta)=\pi(\tilde{w})\in G$ has

$$\eta(\mathrm{id}_H)=\pi_K(u_1)\,,\ \eta(a_1)=\pi_K(v_1)\,,\ \eta(a_3)=\pi_K(u_2)\,,\ \eta(a_3a_1)=\pi_K(v_2)\,,$$
$$\eta(a_3a_2)=\pi_K(v_3)\,,\ \eta(a_2)=\pi_K(u_3)\,,\ \eta(a_2a_1)=\pi_K(v_4)\,,\ \eta(a_2a_3)=\pi_K(v_5)\,.$$

With this example in mind, we can construct an unambiguous context-free grammar for Nf_G. We assume that we have already an unambiguous grammar $\mathcal{C}_K=(\mathbf{V}_K,\mathbf{\Sigma}_K,\mathbf{P}_K,S_K)$ that generates $\mathrm{Nf}_K\setminus\varepsilon$. We write S_K for its start symbol. Our grammar for Nf_G needs a new start symbol, S. We have $\mathbf{\Sigma}=\mathbf{\Sigma}_K\cup\{\mathbf{a}_1,\dots,\mathbf{a}_N\}$, the variables are $\mathbf{V}=\mathbf{V}_K\cup\{Z,T_1,U_1,\dots,T_N,U_N\}$, and $\mathbf{P}$ is the union of $\mathbf{P}_K$ with the following set of rules:

$$\begin{aligned}
&S\vdash ZT_{i_1}\cdots T_{i_r}\quad\text{for } 0\le r\le N\,,\ i_1<\cdots<i_r\,,\\
&S\vdash ZT_{i_1}\cdots T_{i_r}U_j\quad\text{for } 0\le r\le N-1\,,\ i_1<\cdots<i_r\,,\ j\ne i_l\ \forall\, l\,,\\
&T_i\vdash \mathbf{a}_iZT_{i_1}\cdots T_{i_r}\mathbf{a}_i\quad\text{for } 1\le r\le N-1\,,\ i_1<\cdots<i_r\,,\ i_l\ne i\ \forall\, l\,,\\
&T_i\vdash \mathbf{a}_iS_K\mathbf{a}_i\,,\\
&U_i\vdash \mathbf{a}_iZT_{i_1}\cdots T_{i_r}\quad\text{for } 0\le r\le N-1\,,\ i_1<\cdots<i_r\,,\ i_l\ne i\ \forall\, l\,,\\
&U_i\vdash \mathbf{a}_iZT_{i_1}\cdots T_{i_r}U_j\quad\text{for } 0\le r\le N-2\,,\ j\ne i\,,\ i_1<\cdots<i_r\,,\ i_l\ne i,j\ \forall\, l\,,\\
&Z\vdash\varepsilon\quad\text{and}\quad Z\vdash S_K\,.
\end{aligned}$$

We briefly explain the meaning of the variables. Of course, S stands for an arbitrary "lexicographic" tour with all possible insertions of elements of Nf_K, resp. $\mathrm{Nf}_K\setminus\{\varepsilon\}$. The variable Z stands for such an insertion. The variable T_i stands for a complete tour (loop) around a branch (of a tree $\mathcal{T}_0$) that starts with the edge labelled $\mathbf{a}_i$ which emanates from the current vertex of $\mathcal{T}$. The first and the last label (letter) of this tour will both be $\mathbf{a}_i$. If the other endpoint of that edge is a leaf, then we produce an element of $\mathrm{Nf}_K\setminus\{\varepsilon\}$ there before going back. Otherwise, we produce an element of Nf_K and have to insert one or more loops starting at that endpoint. The variable U_i stands for a tour that ends at a vertex h which is distinct from the current vertex of $\mathcal{T}$ and lies on the branch whose initial edges is labelled $\mathbf{a}_i$. This tour has $\mathbf{a}_i$ as its first label. After that, there are three possibilities: (1) we have reached h, which is a leaf, and we produce some element of Nf_K (at the terminal point it may be $=\varepsilon$), (2) we have reached h, produce an element of Nf_K and perform additional loops around some of the branches emanating from h, (3) we have not yet reached h, produce an element of Nf_K, perform loops around some of the branches that do not contain h and proceed further towards h.

Now suppose that Nf_K is regular and that $\mathcal{C}_K$ is an unambiguous, left or right linear grammar. In the dependency di-graph associated with our grammar $\mathcal{C}$ for Nf_G, we see that S is isolated, and the other strong components are $\mathbf{V}_T = \{T_1, \dots, T_N\}$, $\mathbf{V}_U = \{U_1, \dots, U_N\}$, $\{Z\}$ and the strong components of $\mathcal{D}(\mathcal{C}_K)$. The only non-regular component is $\mathbf{V}_T$. Therefore $\mathcal{C}$ is essentially ergodic. However, we know that Nf_G cannot be of convergent type.

From the grammar $\mathcal{C}$, we can recover the results of Parry [32]. If $N = 2$, then the grammar associated with the component $\mathbf{V}_T$ is linear. Therefore $f_L(z)$ is a rational function. If $N \geq 3$, then one sees from our grammar that $f_L(z)$ is algebraic, but not rational. Indeed, it follows from (3.3) that each of the languages L_{T_i}, consisting of all words over $\mathbf{\Sigma}$ that can be derived from T_i, is of convergent type.

Next, let us study the relative growth of the subgroup G_0 with respect to the set of generators $A_H \cup A_K$. It is easy to obtain a grammar for $\mathrm{Nf}_G(G_0)$. Take the above grammar $\mathcal{C}$ which generates Nf_G and delete from $\mathbf{P}$ all rules containing some U_i and from $\mathbf{V}$ all the variables U_i, $i = 1, \dots, N$. The resulting unambiguous grammar $\mathcal{C}_0$ generates $\mathrm{Nf}_G(G_0)$, it is essentially ergodic, and since S is isolated and each L_{T_i} is of convergent type, $\mathrm{Nf}_G(G_0)$ is also of convergent type. Therefore this language is growth-sensitive. Note that this holds independently of K and Nf_K, as long as the latter language is regular.

Remark 3. Rees [33] mentions that no example is known of a non-automatic group with a context-free *combing*. The precise definitions can be found in that reference. The above normal form of the lamplighter group is *not* a combing, since the *fellow traveller property* fails for pairs of group elements that are neighbours in the corresponding Cayley graph when one of the two lies in G_0 and the other does not. E.g., if in the above specific example (6.1) one replaces $h = a_2$ with $h = \mathrm{id}_H$, then the tour around $\mathcal{T}_0$ changes drastically. However, pairs of neighbours (η_1, h_1) and (η_2, h_2) do have the fellow traveller property when $h_1 = h_2$. □

B. Word problem. The *word problem* associated with G, $\mathbf{\Sigma}$ and π is the language $W(G) = W(G, \mathbf{\Sigma}, \pi) = \{w \in \mathbf{\Sigma}^* : \pi(w) = \mathrm{id}_G\}$. Its growth is the *spectral radius* of $X(G, \mathbf{\Sigma}, \pi)$. Contrary to geodesic normal forms, we have the following.

Proposition 5. *Unless G is finite, or contains a finite-index free abelian subgroup with one or two generators, $W(G, \mathbf{\Sigma}, \pi)$ is always of convergent type.*

This is in fact a deep theorem. It relies on Gromov's [16] classification of groups with polynomial growth and Varopoulos' [37] classification of groups carrying a recurrent random walk, in combination with a subtle theorem of Guivarc'h [17]; see Woess [39, Thm. 7.8], and, for an explanation regarding the link with language theory, Woess [38].

Example 3 (free groups)**.** In the setting of Example 1 (free groups), we find a context-free, unambiguous grammar for $W(\mathbb{F}_m)$ that has the same alphabet and variables and production rules

$$S \vdash \varepsilon\,, \quad S \vdash \mathbf{a}_i T_i \mathbf{a}_{-i} S \quad \forall\, i \in I\,,$$
$$T_i \vdash \varepsilon\,, \quad T_i \vdash \mathbf{a}_j T_j \mathbf{a}_{-j} T_i \quad \forall\, j \in I\,,\ j \neq -i\,.$$

If $|\mathbf{\Sigma}| \geq 2$, then it is essentially ergodic, $\mathbf{V}_{\mathrm{ess}} = \{T_i : i \in I\}$, and the variable S is regular. In view of Proposition 5, $W(\mathbb{F}_m)$ is growth-sensitive.

We remark that $W(\mathbb{F}_m)$ is a well-studied language, sometimes called the *extended Dyck language.* It is called the *Dyck set* by Harrison [19], who uses *semi-Dyck set* for what we have called the Dyck language in the Basic Example. □

More generally, Muller and Schupp [30] have shown that a finitely generated group G has a context-free word problem with respect to some ($\Longleftrightarrow$ every) finite set of generators if and only if G has a free subgroup with finite index, and for these groups, $W(G)$ is always unambiguous. The method of Lalley [24] seems to imply that at least for a large class of sets of generators of $\mathbb{F}_m$, the associated word problem is essentially ergodic, and this should extend to all virtually free groups. We intend to study this question in a future note.

Acknowledgements

The first author expresses his gratitude for the hospitality–during various phases of this work–of the Mathematics Institute at TU Graz, of the Séction de Mathématiques de l'Université de Genève, and (together with the second author) of the Erwin Schrödinger Institute at Vienna.

We thank Rostislav I. Grigorchuk for posing the problem, and we acknowledge fruitful discussions with and suggestions by Luc Boasson, Anna Erschler, Werner Kuich and Tatiana Smirnova–Nagnibeda.

References

[1] G. N. Arzhantseva and I. G. Lysenok, *Growth tightness for word hyperbolic groups*, preprint, Univ. Genève, `http://www.unige.ch/math/biblio/preprint/pp2001.html`, 2001.

[2] L. Bartholdi and T. G. Ceccherini-Silberstein, *Growth series and random walks on some hyperbolic graphs*, in print, Monatsh. Math (2002).

[3] M. R. Bridson and R. H. Gilman, *Context-free languages of sub-exponential growth*, preprint, `http://attila.stevens-tech.edu/~rgilman/`, 1999.

[4] T. Ceccherini-Silberstein, F. Fiorenzi and F. Scarabotti, *The Garden of Eden theorem for cellular automata and for symbolic dynamical systems*, to appear, Proceedings of the workshop "Random walks and geometry" held at Erwin Schroedinger Institute for Mathematical Physics, Vienna, Austria, 18.06-13.07 2001. Vadim A. Kaimanovich, Klaus Schmidt, Wolfgang Woess, editors. De Gruyter, Berlin, 2002.

[5] T. Ceccherini-Silberstein and F. Scarabotti, *Random walks, entropy and hopfianity of free groups*, to appear, Proceedings of the workshop "Random walks and geometry" held at Erwin Schroedinger Institute for Mathematical Physics, Vienna, Austria, 18.06-13.07 2002. Vadim A. Kaimanovich, Klaus Schmidt, Wolfgang Woess, editors. De Gruyter, Berlin, 2002.

[6] T. Ceccherini-Silberstein, A. Machì and F. Scarabotti, *On the Entropy of Regular Languages*, preprint, 2002.

[7] N. Chomsky and P.M. Schützenberger, *The algebraic theory of context-free languages*, Computer Programming and Formal systems, (P. Braffort abd D. Hirschberg, eds.) North-Holland, Amsterdam, 1963, pp. 118–161. MR **27:**2371

[8] M. Drmota, *Systems of functional equations*, Random Struct. Alg. **10** (1997), 103–124. MR **99e:**05011

[9] V. A. Efremovic, *The proximity geometry of Riemannian manifolds*, Uspekhi Math. Nauk. **8** (1953), 189.

[10] D. Epstein with J. Cannon, D. Holt, S. Levy, M. Paterson and W. Thurston, *Word Processing in Groups*, Jones and Bartlett Publishers, Boston MA, 1992. MR **93i:**20036

[11] R. H. Gilman, *Formal languages and infinite groups*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., Amer. Math. Soc. **25** (1996), 27–51. MR **97a:**20054

[12] R. I. Grigorchuk, *On the Milnor problem of group growth*, Dokl. Akad. Nauk SSSR **271** (1983), 30–33. MR **85g:**20042

[13] R. I. Grigorchuk, *Degrees of growth of finitely generated groups and the theory of invariant means*, Izv. Akad. Nauk SSSR Ser. Mat. **48** (1984), 939–985. MR **86h:**20041

[14] R. I. Grigorchuk and P. de la Harpe, *On problems related to growth, entropy, and spectrum in group theory*, J. Dynam. Control Systems **3** (1997), 51–89. MR **98d:**20039
[15] R. I. Grigorchuk and A. Machì *An example of an indexed language of intermediate growth*, Theoret. Comput. Sci. **215** (1999), 325–327. MR **99k:**68092
[16] M. Gromov, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. **53** (1981), 53–73. MR **83b:**53041
[17] Y. Guivarc'h, *Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire*, Astérisque **74** (1980), 15–28. MR **82g:**60016
[18] P. de la Harpe, *Topics in Geometric Group Theory. Chicago Lectures in Mathematics*, University of Chicago Press, Chicago, IL, 2000. MR **2001i:**20081
[19] M. Harrison, *Introduction to formal language theory*, Addison-Wesley Publishing Co., Reading, Mass, 1978. MR **80h:**68060
[20] S. Hersonsky and J. Hubbard, *Groups of automorphisms of trees and their limit sets*, Ergodic Theory Dyn. Syst. **17** (1997), 869-884. MR **98k:**57005
[21] R. Incitti The growth function of context-free languages, Theoret. Comput. Sci. **255** (1999), 601–605. MR **2001m:**68098
[22] W. Kuich, *On the entropy of context-free languages*, Information and Control **16** (1970), 173–200. MR **42:**4343
[23] W. Kuich, *Semirings and formal power series: their relevance to formal languages and automata*, Handbook of Formal Languages, vol. 1, Springer, Berlin, 1997, pp. 609–677. MR **98m:**68152
[24] St. P. Lalley, *Finite range random walk on free groups and homogeneous trees*, Ann. Probab. **21** (1993), 2087–2130. MR **94m:**60051
[25] St. P. Lalley, *Random walks on regular languages and algebraic systems of generating functions*, Contemp. Math. 287, Amer. Math. Soc., Providence, RI, 2001, pp. 201–230.
[26] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, 1995. MR **97a:**58050
[27] J. Milnor, *A note on curvature and fundamental group*, J. Differential Geometry **2** (1968), 1–7. MR **38:**636
[28] J. Milnor, *Growth of finitely generated solvable groups*, J. Differential Geometry **2** (1968), 447–449. MR **39:**6212
[29] D. E. Muller and P. E. Schupp, *Groups, the theory of ends, and context-free languages*, J. Comput. Syst. Sci. **26** (1983), 295–310. MR **84k:**20016
[30] D. E. Muller and P. E. Schupp, *The theory of ends, pushdown automata, and second order logic*, Theoret. Comput. Sci. **37** (1985), 51-75. MR **87h:**03014
[31] T. Nagnibeda and W. Woess, *Random walks on trees with finitely many cone types*, J. Theoret. Probab. **15** (2002), 399-438.
[32] W. Parry, *Growth series of some wreath products*, Trans. Amer. Math. Soc. **331** (1992), 751–759. MR **92h:**20061
[33] S. Rees, *Hairdressing in groups: a survey of combings and formal languages*, The Epstein birthday schrift, Geom. Topol. Monogr. **1**, 1998, pp. 493–509. MR **99m:**20069
[34] A. S. Schwarzc, *Volume invariants of coverings*, Dokl. Ak. Nauk. **105** (1955), 32-34.
[35] E. Seneta, *Non-negative Matrices and Markov Chains. 2nd edition*, Springer, New York, 1981. MR **85i:**60058
[36] V. I. Trofimov, *Growth functions of some classes of languages*, Cybernetics **17** (1981), 727–731. MR **84d:**68088
[37] N. Th. Varopoulos, *Théorie du potentiel sur des groupes et des variétés*, C. R. Acad. Sci. Paris, Série I **302** (1986), 865–868. MR **87c:**22020
[38] W. Woess, *Context-free languages and random walks on groups*, Discrete Math. **67** (1987), 81–87. MR **88m:**60020
[39] W. Woess, *Random Walks on Infinite Graphs and Groups*, Cambridge Univ. Press, Cambridge, 2000. MR **2001k:**60006
[40] J. A. Wolf, *Growth of finitely generated solvable groups and curvature of Riemanniann manifolds*, J. Differential Geometry **2** (1968), 421–446. MR **40:**1939

Dipartimento di Ingegneria, Università del Sannio, Corso Garibaldi 107, I-82100 Benevento, Italy
E-mail address: tceccher@mat.uniroma1.it

Institut für Mathematik, Technische Universität Graz, Steyrergasse 30, A-8010 Graz, Austria
E-mail address: woess@weyl.math.tu-graz.ac.at