Journal of Computer and System Sciences **64**, 308–310 (2002)
doi:10.1006/jcss.2001.1804, available online at http://www.idealibrary.com on IDEAL®

Context-Free Languages of Sub-exponential Growth

Martin R. Bridson[1]

Mathematical Institute, 24–29 St. Giles, Oxford OX1 3LB, United Kingdom
E-mail: bridson@maths.ox.ac.uk

and

Robert H. Gilman

Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, New Jersey 07030
E-mail: rgilman@stevens-tech.edu

Received January 24, 2000; revised October 3, 2001

There do not exist context-free languages of intermediate growth.

Key Words: context-free languages; growth functions.

1. INTRODUCTION

The function γ whose value at each non-negative integer n is the number of words on length n in a fixed formal language L is called the growth function of L. Flajolet [3] asked if there are context-free languages of *intermediate growth*, that is, such that γ is not bounded above by a polynomial, but $\limsup \gamma(n)/r^n = 0$ for all $r > 1$. The answer to this question is a corollary to the following theorem.

THEOREM 1.1. *If L is a context-free language with growth function γ, then either there is a number $r > 1$ and integer n_0 such that $\gamma(n) \geqslant r^n$ for all $n \geqslant n_0$, or else L is a bounded language.*

A *bounded language* is one which is a subset of $w_1^* \cdots w_n^*$ for some words $\{w_1, \ldots, w_n\}$. Since it is clear that the growth of a bounded language is bounded above by a polynomial, we have the desired corollary.

COROLLARY 1.1. *There do not exist context-free languages of intermediate growth.*

[1] Supported by an EPSRC Advanced Fellowship.

0022-0000/02 $35.00

We note that by a recent result of Grigorchuk and Machì there are indexed languages of intermediate growth [4].

Corollary 1.1 was obtained independently by Incitti [6]. Theorem 1.1 occurs in our previous work [1] as a remark that the proof given there of the weaker result [1, Proposition 1.3] suffices for Theorem 1.1. In this note we give a quicker proof of Theorem 1.1 based on work of Ginsburg and Spanier [5], who also obtain a corresponding decidability result:

THEOREM 1.2 [5, Theorem 5.2]. *It is decidable whether or not the language L generated by a given context-free grammar is bounded; and if L is bounded, one can effectively find words $\{w_1, ..., w_n\}$ such that $L \subset w_1^* \cdots w_n^*$.*

2. PROOF OF THEOREM 1.1

Suppose the language L over the finite alphabet Σ is generated by a context-free grammar G with start symbol S. Without loss of generality we may assume that each non-terminal A of G participates in a derivation of some word in L. Write $A \stackrel{*}{\Rightarrow} \alpha$ to indicate that A derives the sentential form α. Following [5] we define for each non-terminal A

$$Y_A = \{u \mid A \stackrel{*}{\Rightarrow} uAv \text{ for some } v \in \Sigma^*\}$$

$$Z_A = \{v \mid A \stackrel{*}{\Rightarrow} uAv \text{ for some } w \in \Sigma^*\}.$$

THEOREM 2.1 [5, Theorem 5.1]. *A necessary and sufficient condition that the non-empty language L generated by a context-free grammar G be bounded is that the monoids Y_A and Z_A both be commutative for every non-terminal A of G.*

To complete the proof of Theorem 1.1 it suffices to show that if some Y_A or Z_A is not commutative, then L_A, the language of all words derivable from A, has growth function bounded below by an exponential. Indeed since A occurs in the derivation of at least one word in L, there is a derivation $S \stackrel{*}{\Rightarrow} uAv$ for some words $u, v \in \Sigma^*$, and it follows easily that the growth function of L is bounded below by an exponential once the growth function for L_A is.

Suppose a particular Y_A is not commutative (the argument is similar for Z_A) and pick two derivations

$$A \stackrel{*}{\Rightarrow} u_1Av_1 \qquad A \stackrel{*}{\Rightarrow} u_2Av_2, \qquad u_1u_2 \neq u_2u_1.$$

Choose an integer m with $m \geqslant |u_i|$, $m \geqslant |v_i|$, and $m \geqslant |w|$ for some $w \in L_A$, and then choose two more integers d, e such that u_1^d and u_2^e have the same length. Each derivation may be used to expand the non-terminal A occurring in u_1Av_1 and in u_2Av_2. By iterating these expansions, one sees that for every word $W = W(x_1, x_2)$ in the free monoid $\{x_1, x_2\}^*$, L_A contains $W(u_1^d, u_2^e)\, w\bar{W}(v_1^d, v_2^e)$, where $\bar{W}$ is W written backwards.

If two distinct words W and W' of the same length k yield the same element of L_A, then one of $W(u_1^d, u_2^e)$, $W'(u_1^d, u_2^e)$ must be a prefix of the other. Because u_1^d and

u_2^e have the same length, $W(u_1^d, u_2^e) = W'(u_1^d, u_2^e)$ in this case. As W and W' are distinct, we must have $u_1^d = u_2^e$. According to [2, Corollary 4.1, 5, Lemma 5.1], this implies that u_1 and u_2 commute. Since u_1 and u_2 do not commute, we conclude that the 2^k words W of length k yield 2^k distinct words $W(u_1^d, u_2^e)\, w\bar{W}(v_1^d, v_2^e)$ of length at most $(2k+1)\, mf$ in L_A. Here f is the maximum of $\{d, e\}$. Hence the growth function γ_A of L_A satisfies $\gamma_A((2k+1)\, mf) \geqslant 2^k$ if $k \geqslant 1$.

Suppose $n \geqslant 6mf$. Dividing n by mf we obtain $n = (2k+1)\, mf + r$ for some $k \geqslant 1$ and r with $0 \leqslant r < 2mf$. Thus $\gamma_A(n) \geqslant 2^k$ and $k \geqslant (n - 3mf)/(2mf) \geqslant n/(4mf)$. Hence $n \geqslant 6mf$ implies $\gamma_A(n) \geqslant r^n$ for $r = 2^{-4mf}$.

REFERENCES

1. M. Bridson and R. Gilman, Formal language theory and the geometry of 3-manifolds, *Comment. Math. Helv.* **71** (1996), 525–555.
2. C. Choffrut and J. Karhumäki, Combinatorics of words, *in* “Handbook of Formal Languages” Vol. 1, Springer-Verlag, New York/Berlin, 1997.
3. P. Flajolet, Analytic models and ambiguity of context-free languages, *Theoret. Comput. Sci.* **49** (1987), 283–309.
4. R. I. Grigorchuk and A. Machì, An example of an indexed language of intermediate growth, *Theoret. Comput. Sci.* **215** (1999), 325–327.
5. S. Ginsburg and E. Spanier, Bounded ALGOL-like languages, *Trans. Amer. Math. Soc.* **113** (1964), 333–368.
6. R. Incitti, The growth function of context-free languages, *Theoret. Comput. Sci.* **255** (2001), 601–605.