Theoretical Computer Science 215 (1999) 325–327

Theoretical Computer Science

Note

An example of an indexed language of intermediate growth

R.I. Grigorchuk[a], A. Machì[b,*]

[a] Steklov Mathematical Institute, Gubkina Str. 8, Moscow 117 966, Russia
[b] Dipartimento di Matematica, Università di Roma "La Sapienza", Piazzale Aldo Moro 2, 00185 Roma, Italy

Received January 1998; received in revised form March 1998
Communicated by D. Perrin

Abstract

We give an example of a language belonging to the class of indexed languages whose growth is intermediate. In particular, the growth function of this language is transcendental.

1. An important feature of a formal language L over a finite alphabet is its growth function $\gamma_L(z)$, that is, the function determined by the series $\sum_{n\geqslant 0}\gamma(n)z^n$, where $\gamma(n)$ is the number of words of L of length at most n. It is known that if L is regular than $\gamma_L(z)$ is rational, and that if L is a context-free unambiguous language then $\gamma_L(z)$ is algebraic [2]. It is also known that there exist context-free languages for which this function is transcendental [3]. In [4] a formula is given for the growth series of a language defined by a set of forbidden words. The rate of growth of the sequence $\{\gamma(n)\}$ is called the *growth of the language* L; it can be polynomial, exponential or between the two, the so-called *intermediate* growth. If $\gamma_L(z)$ is algebraic, then only the first two cases are possible. In [3, p. 307] the question is asked as to whether there exist context-free languages of intermediate growth. We were unable to answer this question, and indeed we believe that such languages do not exist. In this note, we produce an example of an indexed language of intermediate growth belonging to a class which is as close as possible to that of context-free languages.

Theorem 1. *The language* $L\subseteq\{a,b\}^*$ *of the words*

$$ab^iab^j\cdots ab^k,$$

* Corresponding author. E-mail: machi@mat.uniroma1.it.

0304-3975/99/$ – see front matter
PII: S0304-3975(98)00161-3

where $0 \leqslant i \leqslant j \leqslant \cdots \leqslant k$ *are integers, is an indexed language of intermediate growth recognizable by a one-way deterministic non-erasing stack automaton* (1DNESE). *Its growth series is given by*

$$\gamma_L(z) = \prod_{n \geqslant 1} (1 - z^n)^{-1}.$$

Therefore,

$$\gamma_L(n) = \pi(n),$$

where π *is the partition function. Thus, asymptotically*:

$$\gamma_L(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi\sqrt{\frac{2}{3}n}\right).$$

Proof. 1. The automaton works as follows. If a word w is a string of a's, then w is accepted. If w begins with a string of a's followed by a string s_1 of b's, then we push all the b's of s_1 in the stack until a new a is met (if no new a is met the word is accepted). Using the head of the stack we compare the number of b's in the stack with that of the next string s_2 of b's (for each b of s_2 the head starting from the top of the stack makes one step towards the bottom). If the length of s_2 is shorter than the number of b's in the stack, w is rejected. Otherwise the head is taken back to the top, and the b's of s_2 that are left are pushed in the stack. Thus, when starting reading string s_i, the number of b's in the stack equals that of the string s_{i-1}. Therefore, w is accepted if and only if the length of each string s_i of b's is greater than or equal to that of s_{i-1}, i.e. if and only if w belongs to L.

2. Let us introduce the function of two variables

$$\Gamma_L(u,z) = \sum_{w \in L} z^{|w|_b} u^{|w|_a},$$

where $|w|_x$ denotes the number of occurrences of the letter x in the word w. We have

$$\Gamma_L(u,z) = \prod_{n \geqslant 0} (1 - uz^n)^{-1}.$$

Indeed,

$$\prod_{n \geqslant 0} (1 - uz^n)^{-1} = \prod_{n \geqslant 0} (1 + uz^n + \cdots + u^k z^{kn} + \cdots)$$

$$= \sum_{n \geqslant 0} \sum_{0 \cdot k_0 + 1 \cdot k_1 + \cdots + i \cdot k_i = n} (uz^0)^{k_0} (uz^1)^{k_1} \cdots (uz^i)^{k_i}$$

and the terms in the last sum are in a one-to-one correspondence with the words of L. Setting $\Gamma_L(z) = \Gamma_L(u,z)$ we obtain

$$\sum_{n \geqslant 0} \gamma_L(n) z^n = \Gamma_L(z) = \prod_{n \geqslant 0} (1 - z^{n+1})^{-1} = \prod_{n \geqslant 1} (1 - z^n)^{-1}$$

$$= \sum_{n \geqslant 0} \pi(n) z^n,$$

where the last equality, as well as the asymptotic behaviour of $\pi(n)$ are well known in the theory of partition functions [1, Theorem 6.2]. $\square$

2. The following question arises: does the group of intermediate growth constructed in [5] have a geodesic normal form of the elements such that the corresponding language is an indexed language? The result of the present paper suggests that the answer could be in the affirmative.

Acknowledgements

We wish to thank J.-P. Allouche for calling our attention on the problem dealt with in this paper.

References

[1] G.E. Andrews, The Theory of Partitions, Encyclopedia of Mathematics and its Applications, Addison-Wesley, London, 1976.
[2] N. Chomsky, M.P. Schützenberger, The algebraic theory of context-free languages, in: P. Braffort, D. Hirschberg (Eds.), Computer Programming and Formal Systems, North-Holland, Amsterdam, 1963, pp. 118–161.
[3] P. Flajolet, Analytic models and ambiguity of context-free languages, Theoret. Comput. Sci. 49 (1987) 283–309.
[4] R.I. Grigorchuk, Growth functions, rewriting systems and Euler characteristic, Mat. Zametki 58 (5) (1995) 653–668 (Russian).
[5] R.I. Grigorchuk, The growth degrees of finitely generated groups and the theory of invariant means, Math. USSR Izv. 25 (1984) 259–300.