Sets of numbers in different number systems and the Chomsky hierarchy

István Katsányi

It is a thoroughly studied subject within the discipline of formal languages and automata theory, that under which conditions will a set of numbers in m–ary notation be regular for a given $m \geq 1$. Cobham has solved the bases of this problem in [1]. His results were extended and generalized by many authors in many ways for example in the papers [2], [3], [4], [5], [6], [7]. Luca and Restivo suggested in their paper ([3]) to study the open problem of the context–free case. In this work, we examine the context–free, the context sensitive and the recursively enumerable classes in addition to the regular languages, hence the examination of the Chomsky–hierarchy in this regard becomes complete.

Let $\mathcal{N}$ denote the set of nonnegative integers, $\mathcal{REG}$, $\mathcal{CF}$, $\mathcal{CS}$, $\mathcal{RE}$ the classes of regular, context–free, context sensitive and recursively enumerable languages, respectively. For a set $A \subseteq \mathcal{N}$, and for an integer $a \geq 1$ let us denote by $L_a(A)$ the language that represents the set A in the base a number system.

One of the main results of our paper is the next theorem:

Theorem 1 *The following table is filled in correctly.*

	$a=1, b\geq 2$	$a\geq 2, b=1$	$a,b\geq 2$, $\exists n,m\geq 1:$ $a^n=b^m$	$a,b\geq 2$, $\nexists n,m\geq 1:$ $a^n=b^m$
$\mathcal{REG}$	$\mathcal{REG}$	$\mathcal{CS}$	$\mathcal{REG}$	$\mathcal{CS}^*$
$\mathcal{CF}$	$\mathcal{REG}$	$\mathcal{CS}$	$\mathcal{CF}$	$\mathcal{CS}^*$
$\mathcal{CS}$	$\mathcal{RE}^*$	$\mathcal{CS}$	$\mathcal{CS}$	$\mathcal{CS}$
$\mathcal{RE}$	$\mathcal{RE}$	$\mathcal{RE}$	$\mathcal{RE}$	$\mathcal{RE}$

Each element of the table determines the Chomsky–class, that for every set $A \subseteq \mathcal{N}$ the language $L_b(A)$ belongs to, provided that a and b have the property written in the heading of the column of the element, and $L_a(A)$ belongs to the class shown in the heading of the row of the element. With the exception of the elements marked with a $*$, the presented classes are the smallest ones in the Chomsky–hierarchy with this property.

The other main part of this paper is to study, that when do the arithmetical operations alter the Chomsky–class of the sets represented in a number system. First define some operation over sets of numbers:

Definition 1 *Let $A, B \subseteq \mathcal{N}$ be two sets of numbers, and let $c \geq 0$ be an integer. Let us define*

$$A+B=\{a+b \mid a\in A, b\in B\}, \qquad A\cdot B=\{ab \mid a\in A, b\in B\},$$
$$A^B=\{a^b \mid a\in A, b\in B\},$$
$$c+A=A+c=\{c\}+A, \qquad c\cdot A=A\cdot c=\{c\}\cdot A,$$
$$c^A=\{c\}^A, \qquad A^c=A^{\{c\}}.$$

Theorem 2 *For every base $a \geq 1$, the following table is filled in correctly.*

	$c+A$	$c\cdot A$	$A+B$	$A\cdot B$	A^B
$\mathcal{REG}$	$\mathcal{REG}$	$\mathcal{REG}$	$\mathcal{REG}$	$\mathcal{CS}^*$	$\mathcal{CS}^*$
$\mathcal{CF}$	$\mathcal{CF}$	$\mathcal{CF}$	$\mathcal{CS}^*$	$\mathcal{CS}^*$	$\mathcal{CS}^*$
$\mathcal{CS}$	$\mathcal{CS}$	$\mathcal{CS}$	$\mathcal{CS}$	$\mathcal{CS}$	$\mathcal{CS}$
$\mathcal{RE}$	$\mathcal{RE}$	$\mathcal{RE}$	$\mathcal{RE}$	$\mathcal{RE}$	$\mathcal{RE}$

Each element of the table determines the Chomsky–class, that the language $L_a(C)$ belongs to for every $c \geq 0$, $A, B \subseteq \mathcal{N}$, provided that C is the result of the operation written in the heading of the column of the element, and $L_a(A)$ (and $L_a(B)$, if appropriate) belongs to the class shown in the heading of the row of the element. With the exception of the elements marked with a $*$, the presented classes are the smallest ones in the Chomsky–hierarchy with this property.

As a corollary of theorem 2, we get results, which in some sense extend the theorem of Horváth about the ranges of polinomials published in [8]:

Theorem 3 *Let $a \geq 1$ be the base of our number system, $A_0, A_1 \subseteq \mathcal{N}$ be finite sets, $X \subseteq \mathcal{N}$ be a set for which $L_a(X) \in \mathcal{F}$, where $\mathcal{F}$ is one of the classes $\mathcal{RE}, \mathcal{CS}, \mathcal{CF}, \mathcal{REG}$. Then $L_a(A_1 \cdot X + A_0) \in \mathcal{F}$.*

Theorem 4 *Let $a \geq 1$ be the base of our number system, $A_0, A_1, \ldots A_n \subseteq \mathcal{N}$ be finite sets, $X \subseteq \mathcal{N}$ be a set for which $L_a(X) \in \mathcal{F}$, where $\mathcal{F}$ is one of the classes $\mathcal{RE}, \mathcal{CS}$. Then $L_a(A_n X^n + \cdots + A_1 \cdot X + A_0) \in \mathcal{F}$.*

Acknowledgement
The author wishes to thank László Hunyadvári for his useful suggestions.

References

[1] Alan Cobham. On the base-dependance of sets of numbers recognizable by finite automata. *Mathematical Systems Theory*, 3(2):186–192, 1969.

[2] Karel Culik II and Arto Salomaa. Ambiguity and decision problems concerning number systems. *Information and Control*, 56(3):139–153, 1983.

[3] Aldo de Luca and Antonio Restivo. Star-free sets of integers. *Theoretical Computer Science*, 43(2-3):265–275, 1986.

[4] Juha Honkala. Bases and ambiguity of number systems. *Theoretical Computer Science*, 31(1-2):61–71, May 1984.

[5] Juha Honkala. A decision method for the recognizability of sets defined by number systems. *RAIRO Inform. Théor. Appl.*, 20(4):395–403, 1986.

[6] Juha Honkala. A decision method for the unambiguity of sets defined by number systems. *The Journal of Universal Computer Science*, 1(9):648–653, September 1995.

[7] Christian Michaux and Roger Villemaire. Presburger arithmetic and recognizability of sets of natural numbers by automata: New proofs of Cobham's and Semenov's theorems. *Annals of Pure and Applied Logic*, 77(3):251–277, 19 February 1996.

[8] Horváth Sándor. Ranges of polinomials are non–context–free in any number system. To appear in: Pure Mathematics and Applications.