## Werk

**Label:** Article

**Jahr:** 1987

**PURL:** https://resolver.sub.uni-goettingen.de/purl?312901348_48-49|log35

UNIVERSITAS COMENIANA
ACTA MATHEMATICA UNIVERSITATIS COMENIANAE
XLVIII—XLIX — 1986

# INFINITE SETS OF PRIMES AND OF POWERS OF AN INTEGER $a$ ARE NOT CONTEXT-FREE

ELENA KALAŠOVÁ, Bratislava

## 1. Introduction

A simple algebraic condition is derived which shows that neither the set of $q$-ary representations of powers of an integer $a > 1$ nor the set of $q$-ary representations of primes nor any of their infinite subsets is a context-free language.

Our second result differs from that of [1] in that arbitrary base $q > 1$ is considered instead of $q = 2$.

We consider languages over the alphabet $A$ which consists of $q$ digits ($q > 1$), the digit zero included. Strings over the alphabet $A$ will be understood as $q$-ary representations of integers. The $q$-ary strings over $A$ will be denoted as $v$, $w$, $x$, $u_1$, $u_2$, ..., and the integers represented by these strings as $\boldsymbol{v}$, $\boldsymbol{w}$, $\boldsymbol{x}$, $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, ...; $l(x)$ denotes the length (number of digits) of the string $x$. The empty word $\varepsilon$ will have the length 0. The string obtained by concatenating $n$ times, $n > 0$, the string $x$ will be denoted as $x^n$. The notation $a|b$ will mean that $a$ divides $b$.

In the following proofs we make use of the Fermat's Theorem which states:

If $p$ is a prime and $p \nmid c$, then $c^{p-1} \equiv 1 \pmod{p}$; and of the Pumping Lemma:

For every context-free language $L$ there exists an $n$ such that $w$ in $L$ and $l(w) \geqslant n$ implies that $w = w_1w_2w_3w_4w_5$ with $w_2 \neq \varepsilon$ or $w_4 \neq \varepsilon$ and $w_1w_2^kw_3w_4^kw_5$ is in $L$, for $k = 0, 1, 2, \ldots$.

## 2. Derivation of Algebraic Result

**Theorem 1.** Let $a > 1$, $q$ be natural numbers. Suppose that there exists a prime $p$ such that $p \mid q$ and $p \nmid a$. Then neither the set $M$ of $q$-ary representations of powers of the integer $a$ nor any of the infinite subset of $M$ is a context-free language.

**Proof.** Assume the contrary, i.e. there exists an infinite context-free language $K \subseteq M$. From Pumping Lemma it follows that there exists $z$ in $K$ such that $z = uvwxy$. We can assume $x \neq \varepsilon$. (The case $x = \varepsilon$ and $v \neq \varepsilon$ is handled as follows: $U = u$, $V = W = \varepsilon$, $X = v$, $Y = wxy$). Thus $uv^iwx^iy$ is in $K$ for $i = 1, 2, \ldots$.

Denote this infinite increasing sequence of the powers of the integer $a$ by $(b_0, b_1, b_2, \ldots)$, i.e. $b_i = a^{l_i}$ for $i = 0, 1, \ldots$ and the sequence of lenghts of these strings by $(m_0, m_1, m_2, \ldots)$, i.e. $m_i = l(a^{l_i}) = l(uv^iwx^iy)$. The Pumping Lemma also implies that the lengths $m_i$ of the strings $a^{l_i}$ increase linearly. This holds because

$$1 \leqslant |m_i - m_{i-1}| = m_i - m_{i-1} \leqslant l(v) + l(x) \tag{1}$$

Obviously,

$$q^{m_i - 1} \leqslant a^{l_i} \leqslant q^{m_i}$$

$$q^{m_{i+1} - 1} \leqslant a^{l_{i+1}} \leqslant q^{m_{i+1}} \quad \text{for } i \geqslant 0.$$

We take the logarithm to the base $q$ of these inequalities and we obtain

$$m_i - 1 \leqslant l_i \log_q a \leqslant m_i \tag{2}$$

$$m_{i+1} - 1 \leqslant l_{i+1} \log_q a \leqslant m_{i+1} \tag{3}$$

(1), (2) and (3) imply that

$$0 \leqslant l_{i+1} - l_i \leqslant \log_a q(l(x) + l(v)) = \text{const.},$$

thus the sequence $\{(l_{i+1} - l_i)\}_{i=0}^{\infty}$ is bounded. As far as the sequence $\{l_i\}_{i=0}^{\infty}$ is increasing and $\{(l_{i+1} - l_i)\}_{i=0}^{\infty}$ is bounded, there must be a constant $m$ such that $m$ is the difference $m = (l_{i+1} - l_i)$ for infinitely many $i$. Consider all these pairs $a^l$, $a^{l+m}$ (for infinitely many $l$). Obviously there exist their representations

$$a^l = uv^iwx^iy$$

$$a^{l+m} = uv^jwx^jy, \quad j > i.$$

Then the difference $a^{l+m} - a^l = a^l(a^m - 1)$ is an integer which contains $k$ zeros in its representation and $k \geqslant l(y) + il(x)$. Then, necessarily

$$q^k \mid a^l(a^m - 1) \quad \text{which implies}$$

$$p^k \mid a^l(a^m - 1).$$

Moreover, the integer $k$ increases to infinity whenever $l$ increases to infinity. As far as we suppose that there exists the prime $p$ such that $p \mid q$ and $p \nmid a$, then $p^k \nmid a^l$ for any $k$, $l$.

Therefore, necessarily $p^k \mid a^m - 1$ for infinitely many $k$ and constant $m$ and we obtain a contradiction.

Thus neither the set $M$ of $q$-ary representations of powers of an integer $a$ nor any of its infinite subsets is a context-free language, as was to be shown.

**Example.** For $q = 10$, $a = 2$, $p = 5$ we obtain $10^k \mid 2^l(2^m - 1)$, $5^k \mid 2^m - 1$ for infinitely many $k$ and constant $m$; which is a contradiction.

**Remark.** A stronger assertion could be expected. It probably suffices that $a$, $q$ are not powers of the same integer. However, the assumption cannot be omitted completely because e.g. the powers of 4 in the base 8 form the language

$$L = \{20(00)^*\} \cup \{1(00)^*\} \cup \{4(00)^*\}$$

which is regular.

In this part we derive a result about the set of $q$-ary representations of primes and its infinite subsets. It will be convenient to have the following result:

**Lemma.** Let $p$ be a prime, $p > q$, let $u_1$, $u_2$, $u_3$ be $q$-ary strings ($u_1$ is not starting with zero), such that

$$q^{l(u_2)} \not\equiv 1 \pmod{p}.$$

Then

$$\boldsymbol{u}_1\boldsymbol{u}_2^{p-1}\boldsymbol{u}_3 \equiv \boldsymbol{u}_1\boldsymbol{u}_3 \pmod{p}.$$

**Proof.** Observe that

$$\boldsymbol{u}_1\boldsymbol{u}_2^{p-1}\boldsymbol{u}_3 + q^{l(u_3)}\cdot\boldsymbol{u}_2^{p-1} + q^{l(u_3)+(p-1)l(u_2)}\boldsymbol{u}_1 =$$

$$= \boldsymbol{u}_3 + q^{l(u_3)}\cdot(\boldsymbol{u}_2 + q^{l(u_2)}\cdot\boldsymbol{u}_2 + \ldots + q^{(p-2)l(u_2)}\cdot\boldsymbol{u}_2) + q^{l(u_3)}\cdot(q^{(p-1)})^{l(u_2)}\cdot\boldsymbol{u}_1 =$$

$$= \boldsymbol{u}_3 + q^{l(u_3)}\cdot\boldsymbol{u}_2\,\frac{q^{(p-1)l(u_2)} - 1}{q^{l(u_2)} - 1} + q^{l(u_3)}\cdot\boldsymbol{u}_1 = \boldsymbol{u}_3 + q^{l(u_3)}\cdot\boldsymbol{u}_1 = \boldsymbol{u}_1\boldsymbol{u}_3.$$

By Fermat's Theorem we obtain that $q^{l(u_2)(p-1)} \equiv 1 \pmod{p}$ and therefore

$$\frac{q^{(p-1)l(u_2)} - 1}{q^{l(u_2)} - 1} \equiv 0 \pmod{p}.$$

Thus $\boldsymbol{u}_1\boldsymbol{u}_2^{p-1}\boldsymbol{u}_3 \equiv \boldsymbol{u}_1\boldsymbol{u}_3 \pmod{p}$, as was to be shown.

**Corollary 1.** Let $p$ be a prime, $p > q$, let $u_1$, $u_2$, $u_3$, $u_4$, $u_5$ be $q$-ary strings such that

$$q^{l(u_2)} \not\equiv 1 \pmod{p} \text{ and}$$
$$q^{l(u_4)} \not\equiv 1 \pmod{p}.$$

Then

$$\boldsymbol{u}_1\boldsymbol{u}_2^{p-1}\boldsymbol{u}_3\boldsymbol{u}_4^{p-1}\boldsymbol{u}_5 \equiv \boldsymbol{u}_1\boldsymbol{u}_3\boldsymbol{u}_5 \pmod{p}.$$

**Corollary 2.** If $p = \boldsymbol{u}_1\boldsymbol{u}_2^r\boldsymbol{u}_3$ is a prime larger than $q$ and $q^{l(u_2)} \not\equiv 1 \pmod{p}$, then

$$\boldsymbol{u}_1\boldsymbol{u}_2^r\boldsymbol{u}_2^{p-1}\boldsymbol{u}_3 \equiv \boldsymbol{u}_1\boldsymbol{u}_2^r\boldsymbol{u}_3 \equiv 0 \pmod{p}.$$

**Corollary 3.** If $p = u_1 u_2^k u_3 u_4^k u_5$ is a prime larger than $q$ and $q^{l(u_2)} \not\equiv 1 \pmod{p}$ and $q^{l(u_4)} \not\equiv 1 \pmod{p}$, then

$$u_1 u_2^k u_2^{p-1} u_3 u_4^k u_4^{p-1} u_5 \equiv u_1 u_2^k u_3 u_4^k u_5 \equiv 0 \pmod{p}.$$

Now we combine these corollaries and the Pumping Lemma to show that no infinite set of primes is a context-free language.

**Theorem 3.** If $P$ is an infinite subset of the set of $q$-ary representations of primes, then $P$ is not a context-free language.

**Proof.** Let $B \subseteq (A - \{0\})A^*$ be an infinite context-free language. By the Pumping Lemma there exists $w \in B$ such that $w = w_1 w_2 w_3 w_4 w_5$, with $w_2 w_4 \neq \varepsilon$ and $w_1 w_2^k w_3 w_4^k w_5$ is in $B$ for $k = 1, 2, \ldots$. We assume that $w_2 \neq \varepsilon$ and $w_4 \neq \varepsilon$, in other cases we use Corollary 2 instead of Corollary 3.

Let $k$ be an integer such that

$$q^{l(w_2)} < w_1 w_2^k w_3 w_4^k w_5 \text{ and}$$
$$q^{l(w_4)} < w_1 w_2^k w_3 w_4^k w_5.$$

Then

$$q^{l(w_2)} \not\equiv 1 \pmod{w_1 w_2^k w_3 w_4^k w_5} \text{ and}$$
$$q^{l(w_4)} \not\equiv 1 \pmod{w_1 w_2^k w_3 w_4^k w_5}.$$

If $p$ is a prime, from Corollary 3 it follows that

$$s = w_1 w_2^{k+p-1} w_3 w_4^{k+p-1} w_5 \equiv 0 \pmod{p}.$$

Therefore $s$ is divisible by $p$ (i.e. $s$ is not a prime) and since $s$ and $p$ are in $B$, we see that $B$ cannot be an infinite subset of the set of primes, as was to be shown.

**Acknowledgment.** The author would like to thank Dr. I. Korec for many helpful discussions and comments.

## REFERENCES

[1] Hartmanis, J. Shank, H.: Of the Recognition of Primes by Automata, J. of the ACM, vol. 15, No 3, p. 382–389, 1968.

[2] Hopcroft, J. E.—Ullman, J. D.: Formal Languages and Their Relation to Automata, Adison-Wesley, 1969.

[3] Niven, I.—Zuckerman, H. S.: An Introduction to the Theory of Numbers, J. Wiley, New York, 1966.

*Author's address:*
RNDr. Elena Kalašová
Katedra algebry a teórie čísel
Matematicko-fyzikálna fakulta UK
842 15 Bratislava
ČSSR

Received: 8. 2. 1984

SÚHRN

NEKONEČNÉ MNOŽINY PRVOČÍSEL A MOCNÍN PRIRODZENÉHO ČÍSLA $a$ NIE SÚ BEZKONTEXTOVÉ

Elena Kalašová, Bratislava

V práci sa dokazuje, že množiny $q$-adických zápisov všetkých mocnín čísla $a$ (za istých obmedzení na $q$, $a$) a všetkých prvočísel neobsahujú nekonečný bezkontextový jazyk.

РЕЗЮМЕ

БЕСКОНЕЧНЫЕ МНОЖЕСТВА ПРОСТЫХ ЧИСЕЛ И СТЕПЕНЕЙ НАТУРАЛЬНЫХ ЧИСЕЛ $a$ НЕ ЯВЛЯЮТСЯ КОНТЕКСТНО-СВОБОДНЫМИ

Елена Калашова, Братислава

В работе доказано, что множества $q$-адичных предложений всех степеней натурального числа $a$ (при некоторых ограничениях) и всех простых чисел не содержают бесконечный контекстно-свободный язык.