On Primitive Multisets *

Sándor Horváth
Department of Computer Science, Eötvös Loránd University,
Budapest, Hungary
email : horvath@cs.elte.hu

Manfred Kudlek
Fachbereich Informatik, Universität Hamburg,
Hamburg, Germany
email : kudlek@informatik.uni-hamburg.de

Abstract

We consider multisets and present another proof that the set of primitive multisets is not algebraic. We also present a large class of word languages with semilinear Parikh image, containing the language $COPY$. Finally, we show that the class of multiset languages fulfilling the iteration lemmata for rational multisets has the cardinality of the continuum.

Kewords : Multisets, primitive multisets, iteration lemmata.

1 Introduction

The *conjecture* from 1991 (of Dömösi, Horváth, Ito [2]) that for every $k \geq 2$, the set Q_k of primitive words over some k-letter alphabet is *not* context-free, is still unsettled. However, the related languages L_k, for each $k \geq 2$, consisting only of those primitive words all permutations of which are also primitive, has been shown not to be context-free [21]. That result also implies that the sets mQ_k of primitive multisets over some k-letter alphabet is not semilinear. Semilinear sets correspond to algebraic (context-free) languages of multisets. Since the underlying operation $+$ is commutative in contrast to catenation $\cdot$ the classes of rational, linear and algebraic multiset languages coincide, and are also identical with the corresponding classes of Parikh sets of regular, linear and context-free word languages.

Using the iteration lemma for rational multiset languages we present another proof that mQ_k is not algebraic (semilinear) for $k \geq 2$. But mQ_k is in **PsDCS**, the class of Parikh images of deterministic context-sensitive languages.

We also introduce a large class of word languages with semilinear Parikh image, closed under homomorphism and containing the language $COPY$, but not the languages L_k.

Finally, we show that the class of multiset languages fulfilling the iteration lemma for rational multiset languages has the cardinality of the continuum. This implies and gives an alternative proof of the fact that the class of word languages fulfilling the iteration lemma for regular languages also has the cardinality of the continuum [1].

*Research supported by the German-Hungarian project WTZ HUN 00/040 / D-39/2000

2 Definitions

In the sequel, a *multiset* is defined just as an element $\alpha = (a_1, \cdots, a_k) \in \mathbb{N}^k$. Operations on multisets are defined by

sum ; $\alpha + \beta = (a_1 + b_1, \cdots, a_k + b_k)$

order : $\alpha \sqsubseteq \beta \Leftrightarrow a_j \leq b_j$ ($1 \leq j \leq k$)

difference : $\beta - \alpha = (b_1 - a_1, \cdots, b_k - a_k)$ if $\alpha \sqsubseteq \beta$

The *norm* of a multiset $\alpha = (a_1, \cdots, a_k) \in \mathbb{N}^k$ is defined by $|\alpha| = \sum_{i=1}^k a_i$.

If $\Sigma = \{s_1, \cdots, s_k\}$ is an alphabet then a multiset $\alpha = (a_1, \cdots, a_k) \in \mathbb{N}^k$ represents the *multiplicities* of symbols s_i in a word $w \in \Sigma^*$, especially the word $\nu(\alpha) = s_1^{a_1} \cdots s_k^{a_k}$. Let $\Sigma^{\oplus}$ denote the set of all multisets on *carrier* Σ. $\Sigma^{\oplus}$ can be identified with $\mathbb{N}^k$.

A multiset grammar is a quadruple $G = (V_N, V_T, A, P)$, where V_N, V_T are disjoint alphabets, the *nonterminal* and the *terminal* one, respectively, $A \subseteq V^{\oplus}$ is a finite set of multisets (its elements are called *axioms*), and P is a finite set of *multiset rewriting rules or productions* (in short, *rules*) of the form $\mu_1 \to \mu_2$, where μ_1, μ_2 are multisets over $V = V_N \cup V_T$ and $|\mu_1|_{V_N} \geq 1$.

For two multisets α_1, α_2 over V, we write $\alpha_1 \overset{r}{\Longrightarrow} \alpha_2$ for some $r : \mu_1 \to \mu_2 \in P$ if $\mu_1 \sqsubseteq \alpha_1$ and $\alpha_2 = (\alpha_1 - \mu_1) + \mu_2$. If r is understood, then we write $\Longrightarrow$ instead of $\overset{r}{\Longrightarrow}$. We denote by $\overset{*}{\Longrightarrow}$ the reflexive and transitive closure of the relation $\Longrightarrow$. The set of multisets (*language*) *generated* by G is defined by

$$M(G) = \{\beta \in V_T^{\oplus} \mid \alpha \overset{*}{\Longrightarrow} \beta, \text{ for some } \alpha \in A\}.$$

We classify such grammars in a Chomsky-like way as follows [19]:

1. Grammars G as above are said to be *arbitrary.*
2. If $|\mu_1| \leq |\mu_2|$ for all rules $\mu_1 \to \mu_2$ in P, then G is said to be *monotone.*
3. If $|\mu_1| = 1$ for all rules $\mu_1 \to \mu_2$ in P, then G is said to be *context-free.*
4. If $|\mu_1| = 1$ and $|\mu_2|_{V_N} \leq 1$ for all rules $\mu_1 \to \mu_2$ in P, then G is said to be *linear.*
5. If $|\mu_1| = 1, |\mu_2| \leq 2$, and $|\mu_2|_{V_N} \leq 1$ for all rules $\mu_1 \to \mu_2$ in P, then G is said to be *regular.*

We denote by **mARB, mMON, mCF, mLIN, mREG** the families of multiset languages generated by arbitrary, monotone, context-free, linear, and regular multiset grammars, respectively. By **FIN, REG, LIN, CF, CS, RE** we denote the families of finite, regular, linear, context-free, context-sensitive, and recursively enumerable languages, respectively. For a family **F** of languages we denote by **PsF** the family of Parikh sets of vectors associated with languages in **F**. The family of all semilinear languages is denoted by **SLin**.

3 Primitive Multisets

We first cite

Proposition 3.1 : *([19])*
mREG = mCF = PsREG = PsCF = SLin. □

Definition 3.1 : *(Primitive multiset)*
A multiset $\alpha \in \mathbb{N}^k$ with $\alpha \neq \mathbf{0}$ is called primitive iff there is no $m \in \mathbb{N}$ with $m > 1$ and no $\beta \in \mathbb{N}^k$ such that $\alpha = m \cdot \beta$, where $\mathbf{0} = (0, \cdots, 0)$.

Let mQ_k denote the set of all primitive multisets over $\mathbb{N}^k$. □

Let $\Sigma = \{s_1, \cdots, s_k\}$. For any $w \in \Sigma$ let $\pi(w) = \{u \mid \psi(u) = \psi(w)\}$, i.e. $\pi(w)$ is the set of all permutations of w, where $\psi(w)$ denotes the Parikh vector of w.

To mQ_k are associated two word languages :

$L_k = \{w \mid w \in \Sigma^*, \pi(w) \subseteq Q_k\}$, and $L'_k = \nu(mQ_k)$.

Trivially, $L_k = \bigcup_{w \in \Sigma^+, \pi(w) \subseteq Q_k} \pi(w)$.

Note that $\pi(abbaba) \not\subseteq Q_k$ since $ababab \in \pi(abbaba)$ but $ababab \notin Q_k$.

Clearly, $\psi(L_k) = \psi(L'_k) = mQ_k$.

Note that $\psi(Q_k) = I\!N^k \setminus (\bigcup_{i=1}^{k} \{n_i \cdot e_i \mid n_i > 1\} \cup \{\mathbf{0}\})$ is semilinear, with *unit vectors* $e_i = (0, \cdots 0, 1, 0, \cdots, 0)$. Thus $\psi(Q_k) \in \mathbf{PsCF} = \mathbf{SLin}$.

In contrast to the word case in which the set Q_k of primitive words fulfills all known iteration lemmata for **CF** and the conjecture is that $Q_k \notin \mathbf{CF}$, we shall show that mQ_k does not fulfill the iteration lemma for **PsCF**, giving another proof that $mQ_k \notin \mathbf{PsCF}$ ([21]).

For $\alpha = (a_1, \cdots, a_k) \in I\!N^k$ let $\Gamma(\alpha) = gcd(a_1, \cdots, a_k)$ if $\sum_{j=1}^{k} a_j \neq 0$ and $\Gamma(\alpha) = 0$ if $\sum_{j=1}^{k} a_j = 0$ (i.e. if $\alpha = \mathbf{0} \Leftrightarrow gcd(\alpha) = 0$).

Note that $gcd(0,3,5) = 1$, $gcd(2,0,6) = 2$ but $gcd(0,0,0) = 0$.

Lemma 3.1 :

$\alpha \in mQ_k \Leftrightarrow \Gamma(\alpha) = 1$

Proof : $\alpha = m \cdot \beta$ implies $\Gamma(\alpha) \geq m > 1$. $\Gamma(\alpha) = d > 1$ implies $\alpha = d \cdot \beta$.

□

Proposition 3.2 : *(Iteration lemma)[18]*

For any multiset language $M \in$ **PsCF** *there exists a $N = N(M) \in I\!N$ with $N > 0$ such that for any $\alpha \in M$ with $|\alpha| > N$ there exists $\beta \in I\!N^k$ with $|\beta| \leq N$ and $\forall m \geq -1 : \alpha + m \cdot \beta \in M$.*□

Theorem 3.1 :

mQ_k does not fulfill the iteration lemma for algebraic multiset languages **PsCF**.

Proof : It suffices to prove this for the case $k = 2$ since for $k > 2$ we can consider the subset $mQ_k \cap C_2 \subseteq mQ_k$ with $C_2 = \{(a_1, a_2, 0, \cdots, 0) \mid a_1, a_2 \in I\!N\}$.

We have to show that for arbitrary $N \in I\!N$ with $N > 0$ there exists $(p,q) \in mQ_2$ with $p + q > N$ such that for all $(x,y) \in I\!N^2$ with $0 < x + y < N + 1$ there exists an $r \in I\!N \cup \{-1\}$ such that $\Gamma(p + rx, q + ry) = gcd(p + rx, q + ry) > 1$.

Actually, we shall show that there exist infinitely many such r.

Let $N > 0$, $p = (N+1)!$, and $q = t \cdot (N+1)! + 1$ with $t > N$.

$(p,q) \in mQ_2$ since $gcd((N+1)!, t \cdot (N+1)! + 1) = gcd((N+1)!, 1) = 1$.

Case 1 : $0 < x < N + 1$, $y = 0$.

Since $p = hx$ it follows that $p + rx = (h + r)x$ for all $r > 0$. If $q = de$ with $d > 1$ (e.g. $d = q$) then for infinitely many $r > 0$ we have $h + r = sd$, and therefore it follows that$gcd(p + rx, q) = gcd(sdx, de) \geq d > 1$ for infinitely many $r > 0$.

Case 2: $x = 0$, $0 < y < N + 1$.

Then $0 < y < N + 1$. From this follows that $1 < y + 1 \leq N + 1$, and therefore also $q + y = t(N+1)! + 1 + y = f(y+1)$, as well as $p = h(y+1)$. Hence $gcd(p, q + y) \geq y + 1 > 1$.

With $r = g(y+1) + 1$ and $g > 0$ follows that $q + ry = q + g(y+1) + 1 = (f + g)(y+1)$, hence $gcd(p, q + ry) \geq y + 1 > 1$ for infinitely many $r > 0$.

Case 3 : $0 < x$, $0 < y$, $x + y < N + 1$.

$0 < x < N+1$ implies $p = hx$, $h < (N+1)!$, and $y < N$. Therefore $y \leq N-1$ and $hy < (N-1)(N+1)!$. Now let $A = q - hy$. Since $t > N-1$, thus $t \geq N$, it follows that $A = t(N+1)! + 1 - hy > N(N+1)! - (N-1)(N+1)! = (N+1)!$.

Assume $A = de$ with $d > 1$ (e.g. $d = A$). Then, for any $r > 0$, we have the fact that $(p + rx, q + ry) = ((h+r)x, A + (h+r)y)$. Now, for infinitely many $r > 0$ we have $h + r = ds$, and therefore also $gcd(p + rx, q + ry) = gcd((h+r)x, A + (h+r)y) \geq d > 1$.

Thus, mQ_2 does not fulfill the iteration lemma for algebraic multiset languages. □

Immediate consequences are

Corollary 3.1 :
*$mQ_2 \notin$ **PsCF**, $mQ_k \notin$ **PsCF**.* □

and

Corollary 3.2 :
*$L_2 \notin$ **CF**, $L_k \notin$ **CF**, and $L'_2 \notin$ **CF**, $L'_k \notin$ **CF*** □

Theorem 3.2 :
*$mQ_k \in$ **PsDCS** $=$ **PsDLBA**.*

Proof : Consider $\nu(mQ_k)$. A **DLBA** A with input $w = s_1^{a_1} \cdots s_k^{a_k} = \nu(a_1, \cdots, a_k)$ can check whether $gcd(a_1, \cdots, a_k) = 1$. If so, A accepts w, otherwise A rejects w. Thus, A accepts exactly $\nu(mQ_k)$. Therefore, $mQ_k \in$ **PsDCS** $\subseteq$ **PsCS**. □

The next theorem is analogous to an earlier result of ours concerning the sets Q_k of primitive words [11].

Theorem 3.3 :
mQ_k is indecomposable, i.e. $mQ_k = A + B$ implies either $A = \{\mathbf{0}\}$ and $B = mQ_k$, or $B = \{\mathbf{0}\}$ and $A = mQ_k$.

Proof : We shall use induction on the length of multisets.

$A \cap B = \emptyset$. Since $\mathbf{0} + \mathbf{0} = \mathbf{0} \notin mQ_k$ it follows that $\mathbf{0} \notin A \cap B$. Assume $\alpha \in A \cap B$. Then $\alpha + \alpha \in mQ_k$, a contradiction.

Now $e_i \in mQ_k$. Then $e_i \in A$ or $e_i \in B$. Assume $e_i \in A$. This implies $\mathbf{0} \in B$, and furthermore $e_j \in A$ for all $1 \leq j \leq k$.

Thus for all $\alpha \in mQ_k$ with $|\alpha| = 1$ holds $\alpha \in A$ and B contains no multiset β with $|\beta| = 1$.

Now assume that for all $\alpha \in mQ_k$ with $|\alpha| \leq n$ holds $\alpha \in A$ and B contains no β with $0 < |\beta| \leq n$.

Let $\alpha \in mQ_k$ with $|\alpha| = n+1$. Then $\alpha = \gamma + \delta$ with $0 < |\delta| \leq n$. By the induction hypothesis follows that $\delta \notin B$. Therefore $\delta = \mathbf{0}$ and $\alpha = \gamma \in A$.

For $\beta \notin mQ_k$ with $|\beta| = n+1$ follows $\beta = s \cdot \gamma$ for some $s > 1$ and $\gamma \in mQ_k$. But then, by the induction hypothesis, $\gamma \in A$ and therefore $(s+1) \cdot \gamma \in mQ_k$, a contradiction. Thus B contains no β with $0 < |\beta| \leq n+1$.

The case $A = \{\mathbf{0}\}$ and $B = mQ_k$ is symmetric. □

In the following we introduce a large class of word languages with semilinear Parikh image, containing e.g. *COPY*, and being closed under word homomorphism.

Let $\hat{\mathbf{H}}_\mathbf{r}\mathbf{X}$ denote the following class of languages :

$\chi_r(L) = \{h_1(w) \cdots h_r(w) \mid w \in L\}$ with $L \in \mathbf{X}$ where

h_j ($j = 1, \cdots, r$) are word homomorphisms. The corresponding class with non-erasing homomorphisms is denoted by $\mathbf{H_r X}$. Define also $\mathbf{H'_r X} = \bigcup_{s=1}^{r} \mathbf{H_s X}$.

Especially, we will consider $\mathbf{X} = \mathbf{REG}$, $\mathbf{X} = \mathbf{LIN}$, and $\mathbf{X} = \mathbf{CF}$.

Let

$$\mathbf{\hat{H}_* X} = \bigcup_{r=1}^{\infty} \mathbf{\hat{H}_r X},$$

and

$$\mathbf{H_* X} = \bigcup_{r=1}^{\infty} \mathbf{H_r X}.$$

Note that $COPY \in \mathbf{H_2 REG}$, since $COPY = \{ww \mid w \in \Sigma^*\}$.

Clearly,

$$\mathbf{H_* X} = \bigcup_{r=1}^{\infty} \mathbf{H'_r X} = \lim_{r \to \infty} \mathbf{H'_r X}$$

and $\mathbf{REG} = \mathbf{\hat{H}REG} \subset \mathbf{H'_2 REG}$.

The following lemmata state some closure properties.

Lemma 3.2 :

$\mathbf{\hat{H}_r X}$ *is closed under word homomorphism.*

Proof : Consider $\chi_r(L) \in \mathbf{\hat{H}_r X}$, and let h be a word homomorphism. Then

$$\begin{aligned} h(h_1(w) \cdots h_r(w)) &= h(h_1(w)) \cdots h(h_r(w)) \\ &= h'_1(w) \cdots h'_r(w) \end{aligned}$$

with $h'_j = hh_j$.

Therefore $h(\chi_r(L)) = \{h'_1(w) \cdots h'_r(w) \mid w \in L\} \in \mathbf{\hat{H}_r X}$. □

Lemma 3.3 :

$\mathbf{H_r X}$ *and* $\mathbf{H'_r X}$ *are closed under nonerasing word homomorphism.*

Proof : As in the previous lemma. □

An immediate consequence is

Corollary 3.3 :

$\mathbf{\hat{H}_* X}$ *is closed under word homomorphism.*

$\mathbf{H_* X}$ *is closed under nonerasing word homomorphism.* □

The following iteration lemma holds :

Lemma 3.4 : *(Iteration lemma)*

For any $L' \in \mathbf{H_r REG}$ *there exists a* $N(L')$ *such that for any* $z \in L'$ *there exist words* u_j, v_j, w_j *(* $1 \le j \le r$ *) such that*

1. $z = u_1 v_1 w_j \cdots u_r v_r w_r$
2. $|u_j v_j| \le N(L')$ *for all* $1 \le j \le r$
3. $|v_j| > 0$ *for all* $1 \le j \le r$
4. $u_1 v_1^i w_j \cdots u_r v_r^i w_r \in L'$ *for all* $i \ge 0$.

Proof : Let $N(L)$ be the constant for the iteration lemma for **REG**. Define the constants $m = min\{h_j(s) \mid s \in \Sigma, 1 \leq j \leq r\}$, $M = max\{h_j(s) \mid s \in \Sigma, 1 \leq j \leq r\}$. Then we have $m|w| \leq |h_1(w) \cdots h_r(w)| \leq M|w|$. Choose $N(L') = MN(L)$.

Now, if $z \in L'$ with $|z| > N(L')$, and since $z = h_1(t) \cdots h_r(t)$ for some $t \in L$, it follows that $M|t| \geq |z| > MN(L)$, and therefore $|t| > N(L)$.

Thus, there exist u, v, w with $t = uvw$, $|uvw| \leq N(L)$, $|v| > 0$, and $uv^iw \in L$ for all $i \geq 0$.

Now define $u_j = h_j(u)$, $v_j = h_j(v)$, and $w_j = h_j(w)$ for $1 \leq j \leq r$.

Since $m > 0$ it follows that $|v_j| > 0$ for all $1 \leq j \leq r$.

Trivially, $u_1 v_1^i w_j \cdots u_r v_r^i w_r \in L'$ for all $i \geq 0$. □

From this follows the proper hierarchy

Lemma 3.5 :

$\mathbf{REG} = \mathbf{HREG} \subset \mathbf{H'_2REG} \subset \cdots \subset \mathbf{H'_rREG} \subset \mathbf{H'_{r+1}REG} \subset \cdots \subset \mathbf{H_*REG}$.

Proof : Consider the languages $\tilde{L}_r = \{(wc)^r \mid w \in \{a,b\}^*\}$ ($r > 0$). It is easy to see that $\tilde{L}_{r+1} \in \mathbf{H_{r+1}REG}$, but $\tilde{L}_{r+1} \notin \mathbf{H_sREG}$ for $s \leq r$.

To see this assume that $\tilde{L}_{r+1} \in \mathbf{H_sREG}$ with $s \leq r$. But the iteration lemma (Lemma 3.4) allows only s parts of a word $(wc)^{r+1}$ to be iterated, yielding a word not in $\tilde{L}_{r+1}$, a contradiction. □

Another iteration lemma is :

Lemma 3.6 : *(Iteration lemma)*

For any $L' \in \mathbf{H_rCF}$ there exists a $N(L')$ such that for any $z \in L'$ there exist words u_j, v_j, w_j, x_j, y_j ($1 \leq j \leq r$) such that

1. $z = u_1 v_1 w_j x_j y_j \cdots u_r v_r w_r x_r y_r$
2. $|v_j w_j x_j| \leq N(L')$ *for all* $1 \leq j \leq r$
3. $|v_j x_j| > 0$ *for all* $1 \leq j \leq r$
4. $u_1 v_1^i w_j x_j^i y_j \cdots u_r v_r^i w_r x_r^i y_r \in L'$ *for all* $i \geq 0$.

Proof : Let $N(L)$ be the constant for the iteration lemma for $L \in \mathbf{CF}$. Define constants $m = min\{h_j(s) \mid s \in \Sigma, 1 \leq j \leq r\}$, $M = max\{h_j(s) \mid s \in \Sigma, 1 \leq j \leq r\}$. Then we have $m|w| \leq |h_1(w) \cdots h_r(w)| \leq M|w|$. Choose $N(L') = MN(L)$.

Now, if $z \in L'$ with $|z| > N(L')$, and since $z = h_1(t) \cdots h_r(t)$ for some $t \in L$, it follows that $M|t| \geq |z| > MN(L)$, and therefore $|t| > N(L)$.

Thus, there exist u, v, w, x, y with $t = uvwxy$, $|uvw| \leq N(L)$, $|vx| > 0$, and $uv^iwx^iy \in L$ for all $i \geq 0$.

Now define $u_j = h_j(u)$, $v_j = h_j(v)$, $w_j = h_j(w)$, $x_j = h_j(x)$, and $y_j = h_j(y)$ for $1 \leq j \leq r$.

Since $m > 0$ it follows that $|v_j x_j| > 0$ for all $1 \leq j \leq r$.

Trivially, $u_1 v_1^i w_j x_j^i y_j \cdots u_r v_r^i w_r x_r^i y_r \in L'$ for all $i \geq 0$. □

In a similar way we get the hierachy

Lemma 3.7 :

$\mathbf{CF} = \mathbf{HCF} \subset \mathbf{H'_2CF} \subset \cdots \subset \mathbf{H'_rCF} \subset \mathbf{H'_{r+1}CF} \subset \cdots \subset \mathbf{H_*CF}$.

Proof : As in Lemma 3.5, showing that $L_{r+1} \in \mathbf{H_{r+1}CF}$ but $L_{r+1} \notin \mathbf{H_rCF}$. □

Lemma 3.8 :

$\mathbf{H_rCF} \subset \mathbf{CS}$.

Proof : Consider $\chi_r(L) = \{h_1(w) \cdots h_r(w) \mid w \in L\} \in \mathbf{H_rCF}$ with $L \subset \Sigma^*$.

Let Σ_i ($1 \leq i \leq r$) be disjoint alphabets. Define letter to letter homomorphisms f_i by $f_i(s) = s_i$ for $s \in \Sigma$. Then, trivially $L' = \{f_1(w) \cdots f_r(w) \mid w \in L\} \in \mathbf{CS}$.

Now define a homomorphism h on $\bigcup_{i=1}^{k} \Sigma_i$ by $h(f_i(s)) = h_i(s)$ for $s \in \Sigma$. This gives $\chi_r(L) = h(L')$. Since **CS** is closed under homomorphism, it follows that $\chi_r(L) \in CS$, and therefore $\mathbf{H_rCF} \subseteq \mathbf{CS}$.

$\mathbf{H_rCF} \subset \mathbf{CS}$ follows from the fact that **CS** contains word languages not being semilinear.

□

From this follows immediately

Corollary 3.4 :
$\mathbf{H_*CF} \subset \mathbf{CS}$. □

Theorem 3.4 :
$\mathbf{Ps\hat{H}_rREG} = \mathbf{Ps\hat{H}_rLIN} = \mathbf{Ps\hat{H}_rCF} = \mathbf{SLin}$.

Proof :
$$\begin{aligned}\psi(h_1(w) \cdots h_r(w)) &= \psi(h_1(w)) + \cdots + \psi(h_r(w)) \\ &= g_1(\psi(w)) + \cdots + g_r(\psi(w)) \\ &= g(\psi(w))\end{aligned}$$
where $g = g_1 + \cdots + g_r$ is a multiset homomorphism.

Therefore, $\psi(\chi_r(L)) = g(\psi(L))$ for any $L \in \mathbf{\hat{H}_rCF}$.

Since $\mathbf{PsREG} = \mathbf{PsLIN} = \mathbf{PsCF} = \mathbf{SLin}$, and **SLin** is closed under multiset homomorphism [20], it follows that $\mathbf{Ps\hat{H}_rCF} \subseteq \mathbf{SLin}$.

Trivially, $\mathbf{SLin} = \mathbf{PsREG} = \mathbf{PsHREG} \subseteq \mathbf{Ps\hat{H}_rCF}$.

Therefore, $\mathbf{Ps\hat{H}_rREG} = \mathbf{Ps\hat{H}_rLIN} = \mathbf{Ps\hat{H}_rCF} = \mathbf{SLin}$. □

An immediate consequence is that all the families of Parikh sets of such languages are identical, and that they don't contain L_k.

Corollary 3.5 :
$\mathbf{Ps\hat{H}_*REG} = \mathbf{Ps\hat{H}_*LIN} = \mathbf{Ps\hat{H}_*CF} = \mathbf{SLin}$. □

and

Corollary 3.6 :
$L_2 \notin \mathbf{\hat{H}_2CF}$, $L_k \notin \mathbf{\hat{H}_rCF}$, and $L_2 \notin \mathbf{\hat{H}_*CF}$, $L_k \notin \mathbf{\hat{H}_*CF}$ □

4 Multiset Languages Fulfilling the Iteration Lemma

Theorem 4.1 :
The cardinality of the set of multiset languages fulfilling the iteration lemma is the cardinality of the continuum, $2^{\aleph_0}$.

Proof : Let $P = \{p_1, \cdots, p_k\} \subseteq \mathbb{N}^k$ be a set of linearly independent vectors, i.e. we have $\sum_{j=1}^{k} m_j \cdot p_j = \sum_{j=1}^{k} n_j \cdot p_j$ implies $m_j = n_j$ for $1 \leq j \leq k$.

Let $H \subseteq \mathbb{N}$ be any infinite, also not recursively enumerable, subset of $\mathbb{N}$. Define the multiset language

$M_H = \{n_1 \cdot p_1 + \sum_{j=2}^{k} n_j \cdot p_j \mid n_1 \in H, n_j \in \mathbb{N}, (2 \leq j \leq k)\}$

$\cup \{n_1 \cdot p_1 \mid n_1 \in I\!N \setminus H\}$.

M_H fulfills the iteration lemma for algebraic multiset languages. To show this assume $|x| > 0$ for $x \in M_H$ where $N > max\{|p_j| \mid 1 \leq j \leq k\}$. Then there exists a j with $n_j > 0$. Trivially, $x + m \cdot p_j \in M_H$ for all $m \geq -1$.

$H \neq H'$ implies $M_H \neq M_{H'}$.

To show this assume $M_H = M_{H'}$. Consider $x = n_1 \cdot p_1 + n_2 \cdot p_2 \in M_H$ with $n_1 \in H$. It follows that $x = n_1 \cdot p_1 + n_2 \cdot p_2 = n_1' \cdot p_1 + \sum_{j=2}^{k} n_j' \cdot p_j \in M_{H'}$ implies, because of the linear indepence, $n_1' = n_1 \in H'$, $n_2' = n_2$, $n_j' = 0$ for $j > 2$. Hence $H \subseteq H'$. By symmetry also follows $H' \subseteq H$. Therefore $H = H'$, a contradiction.

Thus $|\{M_H \mid H \subseteq I\!N\}| = 2^{|I\!N|} = 2^{\aleph_0}$.

Note that also $2^{|I\!N^k|} = 2^{\aleph_0}$. □

Remark

By similar arguments we can prove also in the case of various iteration lemmata for word languages that each of the corresponding language classes fulfilling them has the cardinality of the continuum. E.g., in the case of the iteration lemma for regular languages, we get a completely different, alternative proof of the following result [1].

Theorem 4.2 :

The cardinality of the set of word languages fulfilling the iteration lemma for **REG** *is the cardinality of the continuum,* $2^{\aleph_0}$. □

All these results with detailed proofs will be presented in a forthcoming paper.

References

[1] L. Boasson, S. Horváth : *On Languages Satisfying Ogden's Lemma.* RAIRO Informatique Théorique, **12**, pp. 201-202, 1978.

[2] P. Dömösi, S. Horváth, M. Ito : *On the Connection between Formal Languages and Primitive Words.* In : Proc. *First Session on Scientific Communication* (Univ. of Oradea, Oradea, Romania, June 6-8, 1991), Ann. Univ. of Oradea, Fasc. Mat., pp. 19-67, 1991.

[3] P. Dömösi, S. Horváth, M. Ito : *Formal Languages and Primitive Words.* Publ. Math. (Debrecen), **42** pp. 315-321, 1993.

[4] P. Dömösi, S. Horváth, M. Ito, M. Katsura : *Some Remarks on Primitive Words and Palindromes.* J. Automata, Languages and Combinatorics, 2001.

[5] P. Dömösi, S. Horváth, M. Ito, L. Kászonyi, M. Katsura : *Formal Languages Consisting of Primitive Words.* In : Proc. Conf. *Fundamentals of Computation Theory* (Szeged, Hungary, 1993), Lec. Not. Comp. Sci., **710**, Springer-Verlag, Berlin, etc., pp. 194-203, 1993.

[6] P. Dömösi, S. Horváth, M. Ito, L. Kászonyi, M. Katsura : *Some Combinatorial Properties of Words, and the Chomsky-hierarchy.* In : Proc. Internat. Colloq. *Words, Languages and Combinatorics, II* (Kyoto, Japan, 1992), World Scientific, Singapore, New Jersey, etc., pp. 105-123, 1994.

[7] S. Horváth : *The Family of Languages Satisfying Bar-Hillel's Lemma.* RAIRO Informatique Théorique, **12**, pp. 193-199, 1978.

[8] S. Horváth : *A Comparison of Iteration Conditions on Formal Languages.* In : Proc. *Colloq. Algebra, Combinatorics and Logic in Comp. Sci.* (Győr, Hungary, 1983), Colloq. Math. Soc. J. Bolyai. **42**, J. Bolyai Math. Soc., Budapest, and North-Holland, Amsterdam, pp. 193-199, 1986.

[9] S. Horváth : *Strong Interchangeability and Nonlinearity of Primitive Words.* In : Proc. *Workshop AMiLP'95 (Algebraic Methods in Language Processing, 1995),* Univ. of Twente, Enschede, The Netherlands, 6-8 Dec., 1995, A. Nijholt et al., eds., Univ. Twente Service Centrum, pp. 173-178, 1995.

[10] S. Horváth, M. Ito : *Decidable and Undecidable Problems of Primitive Words, Regular and Context-free Languages.* J. Universal Comp. Sci. (an electronic journal, Techn. Univ. Graz, Austria) **5**, pp. 532-541, 1999.

[11] S. Horváth, M. Kudlek : *On Classification and Decidability Problems of Primitive Words.* In : Selected Papers of the First Joint Conf. Modern Appl. Math., 1995 (Ilieni/Illyefalva, County Covasna, Romania, 1995), PU.M.A. **6**, No. 2, pp. 171-189, 1995.

[12] L. Kászonyi : *On a Class of Stratified Linear Sets.* In : Selected Papers of the First Joint Conf. Modern Appl. Math., 1995 (Ilieni/Illyefalva, County Covasna, Romania, 1995), PU.M.A. **6**, No. 2, pp. 201-210, 1995.

[13] L. Kászonyi : *On Bounded Context-free Languages.* In : Proc. *First Symp. Algebra, Languages, and Computation*, Univ. Aizu, Japan, 1997, Imaoka, T., Nehaniv, C., eds.

[14] L. Kászonyi : *A Pumping Lemma for DLI-languages.* Accepted for Discrete Mathematics, 2002.

[15] L. Kászonyi, M. Holzer : *A Generalization of the Flip-flop Lemma.* In : Proc. AFL'96 (Salgótarján, Hungary, 1996), Publ. Math. (Debrecen) **54**, pp. 877-884, 1998.

[16] L. Kászonyi, M. Katsura : *On the Context-freeness of a Class of Primitive Words.* Publ. Math. (Debrecen) **51**, pp. 1-11, 1997.

[17] L. Kászonyi, M. Katsura : *Some New Results on the Context-freeness of Languages* $Q \cap (ab^*)^n$. In : Proc. AFL'96 (Salgótarján, Hungary, 1996), Publ. Math. (Debrecen) **54**, pp. 885-894, 1998.

[18] M. Kudlek : *Algebraic Characterization of Multiset Languages over Different Multiset Structures.* ROMJIST Vol. 4, Nrs. 1-2, pp. 125-134, 2001.

[19] M. Kudlek, C. Martín Vide, G. Păun : *Toward a Formal Macroset Theory.* In *Multiset Processing.*, eds. C. Calude, G. Păun, G. Rozenberg, A. Salomaa, LNCS 2235, pp. 123-133, Springer, 2001.

[20] M. Kudlek, V. Mitrana : *Normal Forms of Grammars, Finite Automata, Abstract Families, and Closure Properties of Macrosets.* LNCS 2235, pp. 135-146, 2001.

[21] H. Petersen : *The Ambiguity of Primitive Words.* In Proceedings of STACS 94, eds. P. Enjalbert, E. W. Mayr, K. W. Wagner, LNCS 775, pp. 679-690, Springer, 1994.