Archivum Mathematicum

Milan Drášil; Miroslav Novotný
A characterization of harmonic languages

Archivum Mathematicum, Vol. 27 (1991), No. 1-2, 7--14

Persistent URL: http://dml.cz/dmlcz/107398

A CHARACTERIZATION OF HARMONIC LANGUAGES

MILAN DRÁŠIL AND MIROSLAV NOVOTNÝ

(Received July 15, 1987)

Abstract. The family of languages weakly grammatizable by means of derivatives is introduced and is proved to coincide with the family of linear harmonic languages. This family appears in connection with grammatical inference problem.

Key words. Generalized grammar with linear productions, permitting triple defined by means of derivatives, language weakly grammatizable by means of derivatives, harmonic grammar, harmonic language.

MS Classification. 68 Q 50.

Grammatical inference problem was solved for various families of languages in different ways. In [2] and [8], for any nontrivial language L, any finite nonempty set of nontrivial contexts C, and any nonnegative integer i a grammar $G(L, C, i)$ is effectively constructed in such a way that the following holds. If there exists a number i_0 such that $G(L, C, i_0) = G(L, C, i)$ for any $i \geqq i_0$, then $G(L, C, i_0)$ generates L; the existence of such i_0 is equivalent with the condition that L is a so called C-grammatizable language in the sense of [7]. In [9], the grammatical inference problem is solved for the so called (linear) harmonic languages in a different way.

The C-grammatizable languages are grammatizable by means of derivatives in the sense of [6]. In this note we define languages weakly grammatizable by means of derivatives and we prove that they coincide with harmonic languages. The family of languages grammatizable by means of derivatives was investigated in [1].

Constructions of grammars contained in [2] and [8] are based on constructions of generalized grammars introduced in [4] and [5]. Further refe rences to grammatical inference problem can be found in [3].

For any set V, we denote by V^* the set of all strings over V. Let V be a finite set and $(u, v) \in V^* \times V^*$; then the ordered pair (u, v) is said to be a *context* over V. For any context (u, v) over V and any set $Q \subseteq V^*$, we put

$$Q_{(u,v)} = \{x \in V^*; uxv \in Q\};$$

the set $Q_{(u,v)}$ is said to be the *derivative* of Q by the context (u, v).

If (u_1, v_1), (u_2, v_2) are contexts over V, we put $(u_1, v_1) \circ (u_2, v_2) = (u_1u_2, v_2v_1)$. Clearly, $\circ$ is a binary operation on the set $V^* \times V^*$ of all contexts over V. It is easy to see that $(V^* \times V^*, (\Lambda, \Lambda), \circ)$ is a monoid where Λ denotes the empty string; the context (Λ, Λ) is said to be *trivial*. For any set C of contexts, we denote by $[C]$ the carrier of the submonoid generated by C in the above mentioned monoid. We often write $(u_1, v_1)(u_2, v_2)$ for $(u_1, v_1) \circ (u_2, v_2)$.

By a *language*, we mean an ordered pair (V, L) where V is a finite set and $L \subseteq V^*$. This definition differs a little from that appearing in the literature and used in the introduction to the present article; it is advantageous for our pourposes because the set V appears in our constructions. A language (V, L) is said to be *nontrivial* if $V \neq \emptyset \neq L$; it is said to be *infinite* if so is the set L.

Let V be a finite set, S a set such that $S \cap V = 0$, $R \subseteq S \times V^* \cup S \times V^*SV^*$, $s_0 \in S$. Then the ordered quadruple (S, V, R, s_0) is said to be a *generalized grammar with linear productions*; it is said to be a *linear grammar* if the sets S and R are finite.

Let $G = (S, V, R, s_0)$ be a generalized grammar with linear productions. For s, t in $(S \cup V)^*$ and $(y, x) \in R$ we put $s \Rightarrow t$ $(\{(y, x)\})$ if there exist u, v in $(S \cup V)^*$ such that $s = uyv$, $uxv = t$. Furthermore, we set $s \Rightarrow t\,(R)$ if there exists $(y, x) \in R$ such that $s \Rightarrow t$ $(\{(y, x)\})$. If $s, t \in (S \cup V)^*$, if p is a nonnegative integer, and $t_0, t_1, \ldots, t_p$ are strings in $(S \cup V)^*$ such that $s = t_0$, $t_p = t$, and $t_{i-1} \Rightarrow t_i\,(R)$ for any i with $0 < i \leqq p$, then the sequence $(t_i)_{i=0}^p$ is said to be an *s-derivation* of t of length p in R. For s, t in $(S \cup V)^*$ we set $s \overset{p}{\Rightarrow} t\,(R)$ if there exists at least one s-derivation of t of length p in R. Furthermore, we put $s \overset{*}{\Rightarrow} t\,(R)$ if there exists at least one nonnegative integer p such that $s \overset{p}{\Rightarrow} t\,(R)$. We set $L(G) = \{t \in V^*;\ s_0 \overset{*}{\Rightarrow} t\,(R)\}$ and the language $(V, L(G))$ is said to be the language *generated* by G.

Let (V, L) be a nontrivial language. Furthermore, we suppose that C is a nonempty set of nontrivial contexts over V and $\boldsymbol{P}$ a set of nonempty derivatives of the set L by contexts of the set $[C]$ such that $L \in \boldsymbol{P}$. Finally, let N be a mapping of $\boldsymbol{P}$ into the set 2^C such that for any $Q \in \boldsymbol{P}$ and any $(a, b) \in N(Q)$ the condition $Q_{(a,b)} \in \boldsymbol{P}$ holds. Then $(C, \boldsymbol{P}, N)$ is said to be a *permitting triple* for (V, L) *defined by means of derivatives.*

Let (V, L) be a nontrivial language and $(C, \boldsymbol{P}, N)$ its permitting triple defined by means of derivatives. Let S be a set equipotent with $\boldsymbol{P}$ and disjoint to V and i a bijection of $\boldsymbol{P}$ onto S. We put

$$R_1 = \{(i(Q), ai(Q_{(a,b)})\,b);\ Q \in \boldsymbol{P}, (a, b) \in N(Q)\},$$

$$R_2 = \{(i(Q), t);\ Q \in \boldsymbol{P}, t \in Q - \bigcup_{(a,b) \in N(Q)} \{a\}\, Q_{(a,b)}\{b\}\},$$

$$G(C, \boldsymbol{P}, N) = (S, V, R_1 \cup R_2, i(L)).$$

Clearly, $G(C, P, N)$ is a generalized grammar with linear productions. It has the following properties.

1. Theorem. *Let (V, L) be a nontrivial language, (C, P, N) a permitting triple for (V, L) defined by means of derivatives. Then the generalized grammar with linear productions $G(C, P, N) = (S, V, R_1 \cup R_2, i(L))$ has the following properties.*

(i) For any $Q \in P$ and any $t \in V^*$ the conditions $t \in Q$ and $i(Q) \overset{*}{\Rightarrow} t$ $(R_1 \cup R_2)$ are equivalent.

(ii) $G(C, P, N)$ generates (V, L).

The proof can be found in [4]. □

Furthermore, we obtain

2. Lemma. *If (V, L) is a nontrivial language, (C, P, N) its permitting triple defined by means of derivatives, and $G(C, P, N) = (S, V, R_1 \cup R_2, i(L))$, then the following assertion holds. For any $Q \in P$ and any context $(u, v) \in [C]$ with the property $i(L) \overset{*}{\Rightarrow} ui(Q)\, v$ $(R_1 \cup R_2)$ the condition $L_{(u,v)} = Q$ is satisfied.*

The proof can be performed by induction on derivation length.

Indeed, if $i(L) \overset{0}{\Rightarrow} ui(Q)v$ $(R_1 \cup R_2)$, then $i(L) = ui(Q)v$ which implies that $(u, v) = (\Lambda, \Lambda)$ and $i(L) = i(Q)$ and, therefore, $L = Q$. It follows that $L_{(u,v)} = L_{(\Lambda,\Lambda)} = L = Q$.

Suppose that $n \geq 0$ and that $i(L) \overset{n}{\Rightarrow} ui(Q)v$ $(R_1 \cup R_2)$ implies $L_{(u,v)} = Q$. Let us have $i(L) \overset{n+1}{\Rightarrow} ui(Q)v$ $(R_1 \cup R_2)$. Thus, there exists an $i(L)$-derivation $(t_i)_{i=0}^{n+1}$ of $ui(Q)v$ in $R_1 \cup R_2$. Regarding the form of productions in $R_1 \cup R_2$ we have $t_n = u'i(Q')v'$ for some $(u', v') \in [C]$ and some $Q' \in P$. Since $i(L) \overset{n}{\Rightarrow} u'i(Q')v'$ $(R_1 \cup R_2)$, we have $Q' = L_{(u',v')}$ by induction hypothesis. Furthermore $t_n \Rightarrow t_{n+1}$ $(R_1 \cup R_2)$ implies the existence of $(a, b) \in N(Q')$ such that $(i(Q'), ai(Q'_{(a,b)})b) \in R_1$ and $ui(Q)v = t_{n+1} = u'ai(Q'_{(a,b)})bv'$ which implies that $u = u'a$, $v = bv'$, $i(Q) = i(Q'_{(a,b)})$. Thus, $Q = Q'_{(a,b)} = (L_{(u',v')})_{(a,b)} = L_{(u',v')(a,b)} = L_{(u,v)}$. □

Let (V, L) be a nontrivial language, (C, P, N) its permitting triple defined by means of derivatives. Then $G(C, P, N) = (S, V, R_1 \cup R_2, i(L))$ is a generalized grammar with linear productions generating (V, L). Hence, for any $t \in L$ we have $i(L) \overset{*}{\Rightarrow} t$ $(R_1 \cup R_2)$. Thus, there exists an $i(L)$-derivation $(t_i)_{i=0}^{p}$ of t in $R_1 \cup R_2$. Clearly, $t_{i-1} \Rightarrow t_i$ (R_1) for $i = 1, 2, \dots, p - 1$, and $t_{p-1} \Rightarrow t_p$ (R_2). Thus, there are $u, v \in V^*$ and $(i(Q), x) \in R_2$ such that $t_{p-1} = ui(Q)v$, $t = t_p = uxv$. We set $\| (t_i)_{i=0}^{p} \| = | x |$ and define $\| t \| = \min \{\| (t_i)_{i=0}^{p} \|; (t_i)_{i=0}^{p}$ is an $i(L)$-derivation of t in $R_1 \cup R_2\}$. Finally, we put $R_3 = \{(i(Q), x); (i(Q), x) \in R_2$ and there exists $t \in L$ such that $| x | \leq \| t \|\}$. Clearly, $(S, V, R_1 \cup R_3, i(L))$ is a generalized grammar with linear productions. We set

$$g(C, P, N) = (S, V, R_1 \cup R_3, i(L)).$$

3. Theorem. *Let (V, L) be a nontrivial language and (C, P, N) its permitting triple defined by means of derivatives. Then $g(C, P, N)$ generates (V, L).*

Proof. We set $g(C, P, N) = (S, V, R_1 \cup R_3, i(L))$, $G(C, P, N) = (S, V, R_1 \cup R_2, i(L))$. Since $R_3 \subseteq R_2$, we have $L(g(C, P, N)) \subseteq L(G(C, P, N)) = L$ by 1.

Let $t \in L$ be arbitrary. Then there exists an $i(L)$-derivation $(t_i)_{i=0}^p$ of t in $R_1 \cup R_2$ whose norm is minimal, i.e., $\| (t_i)_{i=0}^p \| = \| t \|$. Thus, $t_{p-1} \Rightarrow t_p$ $(\{(i(Q), x)\})$ where $(i(Q), x) \in R_2$ and $| x | = \| t \|$. It follows that $(i(Q), x) \in R_3$ and $(t_i)_{i=0}^p$ is an $i(L)$-derivation of t in $R_1 \cup R_3$ which implies that $t \in L(g(C, P, N))$. Thus, $L \subseteq L(g(C, P, N))$ which completes the proof. □

A nontrivial language (V, L) is said to be *weakly gramatizable by means of derivatives* if there exists its permitting triple (C, P, N) defined by means of derivatives such that $g(C, P, N) = (S, V, R_1 \cup R_3, i(L))$ is a grammar, i.e., that P, R_1, R_3 are finite.

Let $G = (S, V, R, s_0)$ be a linear grammar. An element $s \in S$ is said to be *harmonic* if $(u, v) \in V^* \times V^*$, $(u', v') \in V^* \times V^*$, $s_0 \overset{*}{\Rightarrow} usv\ (R)$, $s_0 \overset{*}{\Rightarrow} u'sv'\ (R)$ imply $(L(G))_{(u,v)} = (L(G))_{(u',v')}$. A linear grammar $G = (S, V, R, s_0)$ is said to be *harmonic* if all elements in S are harmonic. A language (V, L) is said to be *harmonic* if it is generated by a harmonic grammar.

4. Lemma. *To any harmonic grammar there exists a harmonic grammar (S, V, R, s_0) generating the same language and having the following properties.*

(i) For any $s \in S$ there are u, v, t in V^* such that $s_0 \overset{*}{\Rightarrow} usv\ (R)$, $s \overset{*}{\Rightarrow} t\ (R)$.

(ii) For any $(s, atb) \in R$ with $s, t \in S$ and $a, b \in V^*$, the condition $(a, b) \neq (\Lambda, \Lambda)$ is satisfied.

Proof. If the given harmonic grammar does not satisfy (*i*), we cancel all nonterminals s for which the conditions formulated in (*i*) are not satisfied and all productions where the cancelled nonterminals appear either in the left side member or in the right side member. We obtain a linear grammar (S, V, R', s_0) generating the same language; clearly, (S, V, R', s_0) is harmonic. We put $R^1 = \{(s, atb) \in R'; s, t \in S, a, b \in V^*, (a, b) \neq (\Lambda, \Lambda)\}$, $R^2 = \{(s, z) \in R'; s \in S, z \in V^*\}$, $R^3 = \{(s, t) \in R'; s \in S, t \in S\}$. Furthermore, we set $P_1 = \{(s, atb)$; there exists $r \in S$ such that $s \overset{*}{\Rightarrow} r\ (R^3)$, $(r, atb) \in R^1\}$, $P_2 = \{(s, z)$; there exists $r \in S$ such that $s \overset{*}{\Rightarrow} r\ (R^3)$, $(r, z) \in R^2\}$, $R = P_1 \cup P_2$. Clearly, (S, V, R, s_0) has the properties (*i*), (*ii*). Similarly as in Lemma 1 of § 2 in [6] we see that (S, V, R, s_0) and (S, V, R', s_0) generate the same language (V, L).

Since $(s, z) \in R$ implies $s \overset{*}{\Rightarrow} z\ (R')$, the conditions u, v, u', v' in V^*, $s_0 \overset{*}{\Rightarrow} usv\ (R)$, $s_0 \Rightarrow u'sv'\ (R)$ imply $s_0 \overset{*}{\Rightarrow} usv\ (R')$, $s_0 \overset{*}{\Rightarrow} u'sv'\ (R')$. Since (S, V, R', s_0) is harmonic, we obtain $L_{(u,v)} = L_{(u',v')}$ and, therefore, (S, V, R, s_0) is harmonic, too. □

A linear grammar (S, V, R, s_0) is said to have the property (*iii*) if there are $a, b \in V^*$ and $s, t \in S$ such that $(s, atb) \in R$.

5. Lemma. *Let* $G = (S, V, R, s_0)$ *be a harmonic grammar with the properties* (*i*), (*ii*), (*iii*) *and* (V, L) *be the language generated by* G. *We put*

$C = \{(a, b) \in V^* \times V^*;$ *there exist* $s, t \in$ *with* $(s, atb) \in R\}$,

$c(s) = L_{(u,v)}$ *for any* $s \in S$ *where* $s_0 \overset{*}{\Rightarrow} usv\ (R)$,

$P = \{c(s); s \in S\}$,

$N(Q) = \{(a, b) \in V^* \times V^*;$ *there are* $s, t \in S$ *with* $c(s) = Q, (s, atb) \in R\}$,

$p(G) = (C, P, N)$.

Then $p(G)$ *is a permitting triple defined by means of derivatives for* (V, L) *and* $g(p(G)) = (Z, V, R_1 \cup R_3, i(L))$ *is a grammar such that* $(y, x) \in R_3$ *implies* $|x| \leqq \max\{|w|; (s, w) \in R, s \in S, w \in V^*\}$.

Proof. By (*ii*) and (*iii*), C is a finite nonempty set of nontrivial contexts. Since G is harmonic, the definition of c is correct. For any $s \in S$, there exists $t \in V^*$ such that $s \overset{*}{\Rightarrow} t\ (R)$ by (*i*). Thus, by (*i*), there exists $(u, v) \in [C]$ such that $s_0 \overset{*}{\Rightarrow} utv\ (R)$ which implies that $utv \in L$ and, hence, $t \in L_{(u,v)} = c(s)$. Therefore, $c(s) \neq \emptyset$ for any $s \in S$ and, consequently, P is a finite set of nonempty derivatives of L by contexts in $[C]$. Clearly, $L = c(s_0) \in P$.

If $(a, b) \in N(Q)$, there are $s, t \in S$ with $c(s) = Q$, $(s, atb) \in R$. By definition of $c(s)$, there exists $(u, v) \in [C]$ such that $s_0 \overset{*}{\Rightarrow} usv\ (R)$, $Q = c(s) = L_{(u,v)}$. This implies that $s_0 \overset{*}{\Rightarrow} uatbv\ (R)$ and, hence, $c(t) = L_{(ua,bv)} = L_{(u,v)(a,b)} = Q_{(a,b)}$. Thus, $Q \in P$, $(a, b) \in N(Q)$ imply $Q_{(a,b)} \in P$.

We have proved that $p(G) = (C, P, N)$ is a permitting triple defined by means of derivatives for (V, L).

We have $g(C, P, N) = (Z, V, R_1 \cup R_3, i(L))$. By definition of R_1, we have $(a, b) \in N(Q)$ for any production $(i(Q), ai(Q_{(a,b)})b) \in R_1$ which means the existence of $s, t \in S$ with the properties $c(s) = Q$, $(s, atb) \in R$, $c(t) = Q_{(a,b)}$. It follows that the set R_1 is finite.

We set

$$n = \max\{|w|; (s, w) \in R, s \in S, w \in V^*\}.$$

Let $t \in L$ be arbitrary. Then there exist $s \in S$, $u, v, w \in V^*$ such that $s_0 \overset{*}{\Rightarrow} usv\ (R)$, $(s, w) \in R$, $uwv = t$. Since for any $(s', at'b) \in R$ with $s', t' \in S$ and $a, b \in V^*$ the condition $(i(c(s')), ai(c(t'))b) \in R_1$ holds, we have $i(c(s_0)) \overset{*}{\Rightarrow} ui(c(s))v\ (R_1)$. Furthermore; $uwv = t \in L$ implies that $w \in L_{(u,v)} = c(s)$ which entails $i(c(s)) \overset{*}{\Rightarrow} w\ (R_1 \cup R_2)$ by 1.

Two cases may occur.

(*a*) If $(i(c(s)), w) \in R_2$, then $\|t\| = \|uwv\| \leqq |w| \leqq n$.

(*b*) If $(i(c(s)), w) \notin R_2$, there exist $Q \in P$, $(u', v') \in [C]$, $w' \in V^*$ such that $i(c(s)) \overset{*}{\Rightarrow} u'i(Q)v'\ (R_1)$, $(i(Q), w') \in R_2$, $u'w'v' = w$. This implies that $t = uu'w'v'v$ and, therefore, $\|t\| \leqq |w'| \leqq |u'w'v'| = |w| \leqq n$.

Thus, we have $\| t \| \leqq n$ for any $t \in L$. As a consequence we obtain that for any $Q \in P$, there exists only a finite number of strings $w \in V^*$ such that $(i(Q), w) \in R_2$ and $| w | \leqq \| t \|$ for some $t \in L$. Thus, the set R_3 is finite and $g(C, P, N)$ is a grammar. Furthermore, $(y, x) \in R_3$ implies that $| x | \leqq \| t \|$ for some $t \in L$, which implies that $| x | \leqq n$. □

6. Lemma. *Let $G = (S, V, R, s_0)$ be a harmonic grammar that has property (i) and has not property (iii) and let the language (V, L) generated by G be nontrivial. Put $C = \{(a, A); a \in V\}$, $P = \{L\}$, $N(L) = \emptyset$. Then (C, P, N) is a permitting triple defined by means of derivatives for (V, L) and $g(C, P, N)$ is a grammar.*

Indeed, we have $S = \{s_0\}$ and R is a finite subset of $\{s_0\} \times V^*$. Thus L is finite, $R = \{s_0\} \times L$. Clearly, $g(C, P, N) = (S, V, R, s_0)$. □

7. Corollary. *Any nontrivial harmonic language is weakly grammatizable by means of derivatives.* □

8. Lemma. *Any nontrivial language that is weakly gramatizable by means o derivatives is harmonic.*

Proof. Let (V, L) be a nontrivial language that is weakly grammatizable by means of derivatives. Thus, there exists a permitting triple (C, P, N) defined by means of derivatives for (V, L) such that $g(C, P, N) = (S, V, R_1 \cup R_3, i(L))$ is a grammar. Let $s \in S$ be arbitrary. Then there exists $Q \in P$ such that $s = i(Q)$. Let u, u', v, v' be strings in V^* such that $i(L) \overset{*}{\Rightarrow} ui(Q)v$ $(R_1 \cup R_3)$, $i(L) \overset{*}{\Rightarrow} u'i(Q)v'$ $(R_1 \cup R_3)$. Clearly, $i(L) \overset{*}{\Rightarrow} ui(Q)v$ $(R_1 \cup R_2)$, $i(L) \overset{*}{\Rightarrow} u'i(Q)v'$ $(R_1 \cup R_2)$. By 2, we have $L_{(u,v)} = Q = L_{(u',v')}$ which implies that $g(C, P, N)$ is harmonic. By 3, (V, L) is a harmonic language. □

9. Main Theorem. *A nontrivial language is harmonic if and only if it is weakly grammatizable by means of derivatives.*

This is a consequence of 7 and 8. □

10. Example. Let us have $V = \{a, b\}$, $L = \{ab^m a; m \geqq 0\}$, $C = \{(a, A), (a, a), (A, b)\}$. Then we have $L_{(a,A)} = \{b^m a; m \geqq 0\}$, $L_{(a,a)} = \{b^m; m \geqq 0\} = L_{(a,a)(A,b)^n}$ for any $n \geqq 0$ and all remaining derivatives are empty. We put $P = \{L, L_{(a,A)}, L_{(a,a)}\}$, $N(L) = \{(a, A), (a, a)\}$, $N(L_{(a,a)}) = \{(A, b)\}$, $N(L_{(a,A)}) = \emptyset$. It follows that $G(C, P, N)$ has the following sets of productions.

$$R_1 = \{(L, aL_{(a,A)}), (L, aL_{(a,a)}a), (L_{(a,a)}, L_{(a,a)}b)\},$$

$$R_2 = \{(L_{(a,a)}, A)\} \cup \{(L_{(a,A)}, b^m a); m \geqq 0\},$$

where the set $\{\bar{L}, \bar{L}_{(a,\Lambda)}, \bar{L}_{(a,a)}\}$ is disjoint to V and the elements $\bar{L}$, $\bar{L}_{(a,\Lambda)}$, $\bar{L}_{(a,a)}$ are mutually different. Since R_2 is an infinite set, $G(C, P, N)$ is not a grammar.

On the other hand, we have $\bar{L} \Rightarrow a\bar{L}_{(a,a)}a$ (R_1); furthermore $a\bar{L}_{(a,a)}a \overset{*}{\Rightarrow} a\bar{L}_{(a,a)}b^m a$ (R_1), $a\bar{L}_{(a,a)}b^m a \Rightarrow ab^m a$ (R_2) for any $m \geqq 0$. It follows that $\| ab^m a \| = 0$ for any $m \geqq 0$ and, therefore $R_3 = \{(\bar{L}_{(a,a)}, \Lambda)\}$. Thus, R_3 is a finite set and $g(C, P, N)$ is a grammar. □

By [6], a nontrivial language (V, L) is said to be grammatizable by means of derivatives if there exists its permitting triple (C, P, N) defined by means of derivatives such that $G(C, P, N)$ is a grammar. Clearly, any nontrivial language that is grammatizable by means of derivatives is also weakly grammatizable by means of derivatives and, therefore, harmonic. We do not know whether any harmonic language is grammatizable by means of derivatives or not. It follows by 10 that there is a language (V, L) and its permitting triple (C, P, N) defined by means of derivatives such that $g(C, P, N)$ is a grammar while $G(C, P, N)$ is not. Thus, the language $(V, \bar{L})$ is harmonic. But if putting $C' = \{(a, a), (\Lambda, b)\}$, $P' = \{L, L_{(a,a)}\}$, $N'(L) = \{(a, a)\}$, $N'(L_{(a,a)}) = \{(\Lambda, b)\}$, we see that (C', P', N') is also a permitting triple of (V, L) defined by means of derivatives and that $G(C', P', N')$ is a grammar and, therefore, (V, L) is grammatizable by means of derivatives.

In [9] Tanatsugu proposed an effective procedure inferring a harmonic grammar from a finite sample of a language. In Example 3 (page 418), the language generated by the grammar $G = (\{s\}, \{a, b\}, \{(s, asb), (s, a^2sb), (s, \Lambda)\}, s)$ was stated to be nonharmonic; however the proof of this statement was missing. Following the ideas of [1] we show one possibility of proving this statement.

11. Lemma. *Let (V, L) be an infinite language, (C, P, N) its permitting triple defined by means of derivatives such that $g(C, P, N)$ is a grammar. Then there exist contexts (u, v), $(r, s) \in [C]$ with the following properties.*

(*i*) $L_{(u,v)} = L_{(u,v)(r,s)} \in P$ *and* $(r, s) \neq (\Lambda, \Lambda)$.

(*ii*) $L_{(u,v)}$ *is an infinite set of strings.*

Proof. There exist a nonterminal t of the grammar $g(C, P, N) = (S, V, R_1 \cup R_3, i(L))$, some contexts (u, v), $(r, s) \in [C]$, and a string $w \in V^*$ such that $i(L) \overset{*}{\Rightarrow} utv$ $(R_1 \cup R_3)$, $t \overset{*}{\Rightarrow} rts$ $(R_1 \cup R_3)$, $t \overset{*}{\Rightarrow} w$ $(R_1 \cup R_3)$, $(r, s) \neq (\Lambda, \Lambda)$, since otherwise $g(C, P, N)$ would generate only a finite set of strings. There exists $Q \in P$ such that $t = i(Q)$. Regarding that $i(L) \overset{*}{\Rightarrow} ui(Q)v$ $(R_1 \cup R_2)$, $i(L) \overset{*}{\Rightarrow} uri(Q)sv$ $(R_1 \cup R_2)$, we obtain $L_{(u,v)} = Q = L_{(ur,sv)} = L_{(u,v)(r,s)}$ by 2. Clearly, $i(Q) \overset{*}{\Rightarrow} r^k ws^k$ $(R_1 \cup R_3)$ for any nonnegative integer k which implies that $i(Q) \overset{*}{\Rightarrow} r^k ws^k$ $(R_1 \cup R_2)$ and, by 1, $r^k ws^k \in Q = L_{(u,v)}$ for any nonnegative integer k. Thus $L_{(u,v)}$ is infinite. □

By a slight modification of the proof of 4.1 in [1], we obtain

12. Lemma. *Let* (V, L) *be the language generated by the grammar* $G = (\{s\}, \{a, b\}, \{(s, a^i sb^j), (s, a^k sb^l), (s, \Lambda)\}, s)$ *where* $il \neq jk$, $(C, \boldsymbol{P}, N)$ *its permitting triple defined by means of derivatives. Then for no pair* $(u, v), (r, s) \in [C]$ *of contexts the conditions (i) and (ii) of 11 are satisfied.* □

13. Corollary. *The language generated by a grammar* $(\{s\}, \{a, b\}, \{(s, a^i sb^j), (s, a^k sb^l), (s, \Lambda)\}, s)$ *where* $il \neq jk$ *is not harmonic.*

Indeed, by 11 and 12, a language of the above described form is not weakly grammatizable by means of derivatives; by 10, it is not harmonic. □

REFERENCES

[1] M. Drášil, *On languages linearly grammatizable by means of derivatives*, Archivum Math. Brno 22 (1986), 139—144.

[2] M. Drášil, *A grammatical inference for C-finite languages*, Archivum Math. Brno 25 (1989), 163—173.

[3] R. C. Gonzalez, M. G. Thomason, *Syntactic pattern recognition; an introduction*, Addison—Wesley, Reading 1978.

[4] B. Kříž, *Zobecněné gramatické kategorie (Generalized grammatical categories)*, Thesis, University J. E. Purkyně, Brno, 1980.

[5] B. Kříž, *Generalized grammatical categories in the sense of Kunze*, Archivum Math. Brno 17 (1981), 151—158.

[6] M. Novotný, *On some constructions of grammars for linear languages*, Intern. J. Computer Math. 17 (1985), 65—77.

[7] M. Novotný, *On a construction of linear grammars*, Prague Studies in Math. Linguistics 10 (1990), 135—143.

[8] M. Novotný, *Grammatical inference problem for a certain family of languages*, Bulletin Polish Acad. Sci. Math. 36 (1988), 721—735.

[9] K. Tanatsugu, *A grammatical inference for harmonic linear languages*, Internat. J. Comput. Inform. Sci. 13, No. 5, 1984, 413—423.

Milan Drášil
Geographical Institute
of the Czechoslovak Academy
of Sciences
Mendlovo n. 1, 603 00 Brno
Czechoslovakia

Miroslav Novotný
Masaryk University
Department of Computer Science
Burešova 20, 602 00 Brno
Czechoslovakia