Milan Drášil

On languages linearly grammatizable by means of derivatives

Archivum Mathematicum, Vol. 22 (1986), No. 3, 139--144

Persistent URL: http://dml.cz/dmlcz/107257

ARCHIVUM MATHEMATICUM (BRNO)
Vol. 22, No. 3 (1986) 139–144

ON LANGUAGES LINEARLY GRAMMATIZABLE BY MEANS OF DERIVATIVES

MILAN DRÁŠIL

(Received April 18, 1985)

Abstract. It is proved in the paper that the family of languages linearly grammatizable by means of derivatives is properly included in the family of languages linearly grammatizable by means of categories.

Key words. Generalized grammar with linear productions, permitting triple, linearly grammatizable language, language linearly grammatizable by means of categories, language linearly grammatizable by means of derivatives.

1. INTRODUCTION

The notion of grammatical category was introduced in 50-tices as a way of natural languages investigation (Kunze, Trybulec, Dobrušin). Later the concepts of related problems have been based on algebra. From this point of view the studies have been continued by prof. M. Novotný and his collaborators. They generalized the notion of grammatical category and dealt with possibilities of constructing grammars where the role of nonterminals was played by grammatical categories. As that time the concepts of languages linearly grammatizable by means of categories and languages linearly grammatizable by means of derivatives were introduced. The family of languages linearly grammatizable by means of derivatives is included in the family of languages linearly grammatizable by means of categories which is included in the family of linear languages [4]. In the present paper we prove that there exists a family of languages that are linearly grammatizable by means of categories but not linearly grammatizable by means of derivatives.

2. LINEARLY GRAMMATIZABLE LANGUAGES

By V^* we denote the set of all strings over the set V and we put $XY = \{xy; x \in X, y \in Y\}$ for any $X \subseteq V^*$ and any $Y \subseteq S^*$. Let V, S be disjoint sets, V nonempty and finite, $s_0 \in S$ and $R \subseteq S \times V^* \cup S \times V^*SV^*$. Then the ordered quadruple $G = \langle V, S, R, s_0 \rangle$ is said to be the *generalized grammar with linear productions*. The relation $\Rightarrow (R)$ of direct derivation is defined as usual. Let $\overset{*}{\Rightarrow} (R)$ be the reflexive and transitive closure of $\Rightarrow (R)$. We put $L(G) = \{w \in V^*; s_0 \overset{*}{\Rightarrow} w(R)\}$ and $(V, L(G))$ is said to be the language generated by G. Generalized grammar with linear productions is clearly a grammar in the usual sense if the sets R and S are finite.

An element $(u, v) \in V^* \times V^*$ is said to be a *context* over the set V. For two contexts $w_1 = (u_1, v_1)$ and $w_2 = (u_2, v_2)$ we put $w_1 \circ w_2 = (u_1 u_2, v_2 v_1)$. It is easy to see that $(V^* \times V^*, (\lambda, \lambda), \circ)$ is a monoid. Let C be a set of contexts. By $[C]$ we denote the submonoid of $(V^* \times V^*, (\lambda, \lambda), \circ)$ generated by the set C.

Let (V, L) be a nontrivial language, C nonempty set of nontrivial contexts (i.e. $(\lambda, \lambda) \in C$), P a set of nonempty subsets of V^* such that $L \in P$. Let us have a mapping N of the set P into the set $2^{C \times P}$ such that for any $Q \in P$ and any $((u, v), T) \in N(Q)$ the condition $\{u\}\, T\{v\} \subseteq Q$ holds. Then the ordered triple (C, P, N) is said to be the *permitting triple of the language* (V, L). Let S be a set equivalent to P and disjoint with V, b a bijection of P onto S. We set

$$R_1 = \{(b(Q), ub(T)v);\ Q, T \in P,\ ((u, v), T) \in N(Q)\},$$
$$R_2 = \{(b(Q), z);\ Q \in P,\ z \in Q - \bigcup_{((u,v),\,T) \in N(Q)} \{u\}\, T\{v\}\},$$
$$G(C, P, N) = \langle V, S, R_1 \cup R_2, b(L) \rangle.$$

Then $G(C, P, N)$ is generalized grammar with linear productions. Kříž in [1] proved that $G(C, P, N)$ generates the language (V, L).

A nontrivial language (V, L) is said to be *linearly grammatizable* if there exists its permitting triple such that $G(C, P, N)$ is a grammar.

In [2] syntactic categories of an arbitrary language have been defined as follows: Let (V, L) be a language. We set

$$r = \{(x, (u, v)) \in V^* \times (V^* \times V^*);\ uxv \in L\}$$

and for any $X \subseteq V^*$ and any $Y \subseteq V^* \times V^*$ we put

$$m(X) = \{(u, v) \in V^* \times V^*;\ (x, (u, v)) \in r \text{ for any } x \in X\},$$
$$n(Y) = \{x \in V^*;\ (x, (u, v)) \in r \text{ for any } (u, v) \in Y\},$$
$$p(X) = n(m(X)).$$

The set $p(X)$ is called the *syntactic category* of the language (V, L) generated by the set X.

Let (C, P, N) be a permitting triple such that the elements of P are syntactic categories of the language (V, L). Then (C, P, N) is called the *permitting triple with categories*. A nontrivial language is said to be *linearly grammatizable by means of categories* if there exists its permitting triple with categories (C, P, N) such that $G(C, P, N)$ is a grammar.

Let $Q \subseteq V^*$, $w = (u, v) \in V^* \times V^*$. The set $Q_w = \{x \in V^*;\ uxv \in Q\}$ is said to be the *derivative* of the set Q by the context w. Let (C, P, N) be a permitting triple such that the elements of P are derivatives of L by contexts of $[C]$ and for any $Q \in P$ $N(Q) \subseteq \{((u, v), Q_{(u,v)});\ (u, v) \in C\}$. Then (C, P, N) is called the *permitting triple with derivatives.* A nontrivial language is said to be *linearly grammatizable by means of derivatives* if there exists its permitting triple with derivatives (C, P, N) such that $G(C, P, N)$ is a grammar.

3. SOME PROPERTIES OF LINEARLY GRAMMATIZABLE LANGUAGES

3.1. Lemma. *Let (V, L) be a linearly grammatizable language, (C, P, N) its permitting triple such that $G(C, P, N)$ is a grammar. Then there exists its permitting triple (C, P_1, N_1) with the following properties:*

(i) $P_1 \subseteq P$,
(ii) *Every set $Q \in P_1 - \{L\}$ is infinite,*
(iii) *$G(C, P_1, N_1)$ is a grammar.*

Proof. We set $P_1 = \{Q \in P;\ Q \text{ is infinite}\} \cup \{L\}$ and let N_1 be the restriction of the mapping N. Obviously (C, P_1, N_1) is a permitting triple with the properties (i) and (ii). Let $G(C, P, N) = \langle V, S, R_1 \cup R_2, b(L)\rangle$ and $G(C, P_1, N_1) = \langle V, \bar{S}, \bar{R}_1 \cup \bar{R}_2, \bar{b}(L)\rangle$. $\bar{R}_1 \subseteq R_1$ implies that $\bar{R}_1$ is finite. For every set $Q \in P_1$ we have $(Q - \bigcup_{((u,v),T) \in N_1(Q)} \{u\}\, T\{v\}) - (Q - \bigcup_{((u,v),T) \in N(Q)} \{u\}\, T\{v\}) \subseteq \bigcup_{\substack{((u,v),T) \in N(Q) \\ T \text{ is finite}}} \{u\}\, T\{v\} = Q_1$ and Q_1 is clearly finite. Hence $\bar{R}_2 - R_2$ is finite, $\bar{R}_2$ also finite and $G(C, P_1, N_1)$ is a grammar. □

3.2. Corollary. *Let (V, L) be a language linearly grammatizable by means of categories (derivatives). Then there exists its permitting triple with categories (derivatives) (C, P, N) such that every set $Q \in P - \{L\}$ is infinite and $G(C, P, N)$ is a grammar.* □

3.3. Lemma. *Let (V, L) be an infinite language, (C, P, N) its permitting triple with derivatives such that $G(C, P, N)$ is a grammar. Then there exist a context $(u_0, v_0) \in [C]$ and some contexts $(u_1, v_1), \ldots, (u_n, v_n) \in C (n \geqq 1)$ such that*

$$L_{(u_0, v_0)} = L_{(u_0, v_0) \circ \ldots \circ (u_n, v_n)} \in P.$$

Proof. There exist $(u_0, v_0) \in [C]$ and $(u_1, v_1), \ldots, (u_n, v_n) \in C$ $(n \geqq 1)$ such that the rules

$(b(L_{(u_0, v_0)}), u_1 b(L_{(u_0, v_0) \circ (u_1, v_1)})\, v_1)$,

$$(b(L_{(u_0, v_0)\circ(u_1, v_1)}), u_2 b(L_{(u_0, v_0)\circ(u_1, v_1)\circ(u_2, v_2)}) v_2), \ldots,$$
$$(b(L_{(u_0, v_0)\circ\ldots\circ(u_{n-1}, v_{n-1})}), u_n b(L_{(u_0, v_0)\circ\ldots\circ(u_n, v_n)}) v_n)$$

are elements of the set R_1 and $L_{(u_0, v_0)} = L_{(u_0, v_0)\circ\ldots\circ(u_n, v_n)}$. Otherwise $G(C, P, N)$ would generate only a finite set of strings. □

4. EXAMPLE OF LINEAR LANGUAGES NOT LINEARLY GRAMMATIZABLE BY MEANS OF DERIVATIVES

A language (V, L) is said to be *contextual* if there exist a finite set of contexts C over the set V and a finite set $L_1 \subseteq V^*$ such that for any $x \in L$ there exists $z \in L_1$ and a finite sequence of contexts $(u_1, v_1), \ldots, (u_n, v_n) \in C$ such that $x = u_n \ldots \ldots u_1 z v_1 \ldots v_n$. The ordered triple $\langle V, C, L_1 \rangle$ is said to be a *contextual grammar* [3]. A language is contextual if and only if it is generated by a linear grammar $\langle V, S, R, s_0 \rangle$ with exactly one nonterminal s_0. This implies that any contextual language is linearly grammatizable by means of categories [4]. In [4] Novotný put the question whether the language generated by the contextual grammar $\langle \{a, b\}, \{(a^5, b^5), (a^2, b)\}, \{\lambda\} \rangle$ is linearly grammatizable by means of derivatives or not. Next theorem shows that this language is not linearly grammatizable by means of derivatives.

4.1. Theorem. *Let $\langle V, C, L_1 \rangle$ be a contextual grammar with the following properties:*

(i) $V = \{a, b\}$,

(ii) $C = (a^{e_1}, b^{e_2}), (a^{\bar{e}_1}, b^{\bar{e}_2})$ *for some positive integers* $e_1, e_2, \bar{e}_1, \bar{e}_2$ *such that* $e_1 \bar{e}_1 \neq \bar{e}_1 e_2$,

(iii) $L_1 = \{\lambda\}$.

Then the language (V, L) generated by the contextual grammar $\langle V, C, L_1 \rangle$ is not linearly grammatizable by means of derivatives.

Proof. (*a*) Assume that the language (V, L) is linearly grammatizable by means of derivatives. Clearly

$L = \{a^{me_1 + n\bar{e}_1} b^{me_2 + n\bar{e}_2}; n, m \geqq 0\}$ is an infinite set. Hence by **3.2.** there exists its permitting triple with derivatives (C, P, N) such that every $Q \in P$ is an infinite set and $G(C, P, N)$ is a grammar.

(*b*) Let us investigate the set C. The elements of the set L are of the form $a^p b^q$, $p, q \geqq 0$. Consequently the set L has nonempty derivatives by the contexts of the form $(a^i b^k, b^j)$, $(a^i, a^k b^j)$, (a^i, b^j). Now we prove that the derivatives of L by the contexts of the form $(a^i b^k, b^j)$ and $(a^i, a^k b^j)$ are finite. If $t \in L_{(a^i b^k, b^j)}$ then $t = b^h$ where $h \geqq 0$ and $a^i b^{k+h+j} \in L$. This implies that there exists integers $m, n \geqq 0$ such that $i = me_1 + n\bar{e}_1$ and $k + h + j = me_2 + n\bar{e}_2$. There exists however only a finite number of integers $m, n \geqq 0$ such that $i = me_1 + n\bar{e}_2$ since the integers i, j, k are constant for a given context $(a^i b^k, b^j)$. Consequently there exists only a finite number of integers h such that $k + h + j = me_2 + n\bar{e}_2$. Hence the set $L_{(a^i b^k, b^j)}$ is finite. The fact that any set $L_{(a^i, a^k b^j)}$ is finite too can be proved similarly. Thus by (*a*) any set $Q \in P$ is of the form $Q = L_{(a^{i_Q}, b^{j_Q})}$.

(*c*) By 3.3 there exist some contexts $(u_0, v_0) \in [C]$ and $(u_1, v_1), \ldots, (u_n, v_n) \in C$ such that $L_{(u_0, v_0)} = L_{(u_0, v_0) \circ \cdots \circ (u_n, v_n)} \in P$. By (*b*) $(u_0, v_0) = (a^i, b^j)$ and $(u_r, v_r) = (a^{k_r}, b^{l_r})$ $(1 \leqq r \leqq n)$ since otherwise the derivative $L_{(u_0, v_0) \circ \ldots \circ (u_n, v_n)}$ would be finite. Let us set $(u_0, v_0) \circ \ldots \circ (u_n, v_n) = (a^k, b^l)$. Thus there exist two nontrivial contexts $(a^i, b^j), (a^k, b^l)$ such that $(a^i, b^j) \neq (a^k, b^l)$ and

$$L_{(a^i, b^j)} = L_{(a^k, b^l)} \in P \tag{1}$$

(*d*) The equation (1) means that every element $u \in L_{(a^i, b^j)} = L_{(a^k, b^l)}$ can be expressed both in the form

$$u = a^{me_1 + n\bar{e}_1 - i} b^{me_2 + n\bar{e}_2 - j} \quad \text{and} \quad u = a^{m'e_1 + n'\bar{e}_1 - k} b^{m'e_2 + n'\bar{e}_2 - 1}$$

Let us put

$$A = \begin{pmatrix} e_1 & \bar{e}_1 \\ e_2 & \bar{e}_2 \end{pmatrix}, \qquad I = \begin{pmatrix} i \\ j \end{pmatrix}, \qquad K = \begin{pmatrix} k \\ l \end{pmatrix}, \qquad B = I - K.$$

For any two vectors $x = \begin{pmatrix} m \\ n \end{pmatrix}$ and $x' = \begin{pmatrix} m' \\ n' \end{pmatrix}$ we define the relation $R = \{(x, x');\ Ax - I = Ax' - K\}$ and for any vector $x = \begin{pmatrix} m \\ n \end{pmatrix}$ we denote $x \geqq 0$ if $m, n \geqq 0$. Now the equation (1) can be expressed in the following way. For any vector $x = \begin{pmatrix} m \\ n \end{pmatrix} \geqq 0$ (resp. $x' = \begin{pmatrix} m' \\ n' \end{pmatrix} \geqq 0$) such that $Ax - I \geqq 0$ (resp. $Ax' - K \geqq 0$) there exist a vector $x' = \begin{pmatrix} m' \\ n' \end{pmatrix} \geqq 0$ (resp. $x = \begin{pmatrix} m \\ n \end{pmatrix} \geqq 0$) such that $Ax' - K \geqq 0$ (resp. $Ax - I \geqq 0$) and $(x, x') \in R$. The condition $(x, x') \in R$ is equivalent with the condition $x - x' = A^{-1}(B)$ where $A^{-1}(B) = C \neq 0$ since A is regular, A^{-1} is an isomorphism and $B \neq 0$. Thus we can express the condition $(x, x') \in R$ in the form $x = x' + C$ where $C \neq 0$. Assume that $C = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ and $c_1 < 0$. We set $m' = 0$ and $n' \geqq 0$ sufficiently large such that $Ax' - K \geqq 0$. Then there exists a vector $x = \begin{pmatrix} m \\ n \end{pmatrix} \geqq 0$ such that $(x, x') \in R$. However $m = 0 + c_1 < 0$ leads to a contradiction. Let $c_1 > 0$. We set $m = 0$ and $n \geqq 0$ sufficiently large such that $Ax - I \geqq 0$. Then there exists a vector $x' = \begin{pmatrix} m' \\ n' \end{pmatrix} \geqq 0$ such that $(x, x') \in R$. However $0 = m' + c_1$, i.e. $m' = -c_1 < 0$ leads to contradiction too. Hence $c_1 = 0$. Assume that $c_2 < 0$. We set $n' = 0$ and $m' \geqq 0$ sufficiently large such that $Ax' - K \geqq 0$. Then there exists a vector $x = \begin{pmatrix} m \\ n \end{pmatrix} \geqq 0$ such that $(x, x') \in R$. However $n = 0 + c_2 < 0$ leads to a contradiction. Let $c_2 > 0$. We set $n = 0$ and $m \geqq 0$ sufficiently large such that $Ax - I \geqq 0$. Then there exists a vector $x' = \begin{pmatrix} m' \\ n' \end{pmatrix} \geqq 0$ such that $(x, x') \in R$. However $0 = n' + c_2$, i.e. $n' = -c_2 < 0$ leads to a contradiction too. Hence $c_2 = 0$.

Thus we have $c_1 = 0 = c_2$, i.e. $C = 0$ which is a contradiction. □

REFERENCES

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

[2] B. Kříž: *Generalized grammatical categories in the sense of Kunze*, Archivum Mathematicum (Brno) 17 (1981), 151–158.

[3] S. Marcus: *Contextual grammars*, Rev. Roum. Pures et Appl. 14, 1969, 250–273.

[4] M. Novotný: *On some constructions of grammars for linear languages*, to appear in Intern. J. Comp. Math. 1985.

[5] M. Novotný: *Remarks on linearly grammatizable languages*, to appear in Prague Studies in Math. Linguistic 9.

Milan Drášil
Institute of Geography
of the Czechoslovak Academy of Sciences
Mendlovo nám. 1
603 00 Brno
Czechoslovakia