INFORMATIQUE THÉORIQUE ET APPLICATIONS

J. DUSKE
M. MIDDENDORF
R. PARCHMANN

Indexed counter languages

Informatique théorique et applications, tome 26, nº 1 (1992), p. 93-113
<http://www.numdam.org/item?id=ITA_1992__26_1_93_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 26, n° 1, 1992, p. 93 à 113)

INDEXED COUNTER LANGUAGES (*)

by J. DUSKE (1), M. MIDDENDORF (1) and R. PARCHMANN (1)

Communicated by J. BERSTEL

Abstract. – *Starting with the characterization of context-free counter languages by rightlinear indexed grammars with one index, indexed counter grammars are introduced and investigated. The family of indexed counter languages is a full AFL properly contained in the family of indexed languages and incomparable with the full trio of linear indexed languages. Furthermore by modifying the derivation mode, a characterization of type-*0 *languages by indexed counter grammars is given.*

Résumé. – *Après une caractérisation des langages algébriques à compteurs par des grammaires indexées linéaires droites d'index* 1, *on introduit et étudie les grammaires indexées à compteurs. La famille des langages indexés à compteurs est une AFL proprement contenue dans la famille des langages indexés, et incomparable au cône rationnel des langages indexés linéaires. De plus, en modifiant le mode de dérivation, on obtient une caractérisation des langages de type* 0 *par des grammaires indexées à compteurs.*

1. INTRODUCTION

Indexed grammers have been introduced by Aho [1] as an extension of context-free grammars. In the study of indexed grammars the question arises whether the generative power of these grammars depends on the number of indices.

It is obvious, that an indexed grammar with an empty set of indices can only generate context-free languages. On the other hand two indices suffice to generate all indexed languages, because the indices of a general indexed grammar can be coded by words of two indices.

The concept of indexed grammars permits two principle ways of representing context-free languages; first by using a context-free grammar, which is a special form of an indexed grammar, and second by using a rightlinear

(*) Received August 1990, revised November 1990.
(1) Institut für Informatik, Universität Hannover, D-3000 Hannover, Germany.

0988-3754/92/01 93 21/$4.10/© AFCET-Gauthier-Villars

indexed grammar, which can be interpreted as a grammatical description of a pushdown automaton.

Hence there is a natural way of representing context-free counter languages as rightlinear indexed grammars with only one index and a special "endmarker". Extending this concept to the case of a (general) indexed grammar, the use of only one index symbol can be interpreted as "counting while performing context-free derivations". The "endmarker" allows to detect a count of zero and to start counting again. In Section 2 we introduce such grammars formally and give examples. These grammars are called indexed counter grammars or ic-grammars and the corresponding languages are called indexed counter or ic-languages. Our investigations show that this subclass of indexed languages shares many properties with the context-free counter languages.

In Section 3 we first give normal forms of ic-grammars. In particular, using regularity properties of the index words appearing in derivations of indexed grammars, we arrive at the e-free standard form of ic-grammars. Then it is shown that grammars in this standard form allow derivations of terminal words such that the lengths of the appearing index words in these derivations are linear bounded by the lengths of the derived words. With the aid of this result we give in Section 4 an example of an indexed language which is not an ic-language, *i. e.* the family of ic-languages (which is a full AFL) is properly contained in the family of indexed languages.

We also investigate the linear and rightlinear forms of ic-grammars and completely characterize the subset and proper subset relation between the classes of (linear, rightlinear) indexed, (linear, rightlinear) indexed counter and (linear, rightlinear) context-free grammars.

In the final Section 5 we consider indexed grammars with another derivation mode introduced in [6]. Under this derivation mode ic-grammars generate the same languages as phrase structure grammars. This result shows that type-0 languages can be obtained by "counting on leftmost derivations of a context-free grammar".

2. FORMAL DEFINITIONS AND BASIC PROPERTIES

Let us first define the notion of an (linear, rightlinear) indexed grammar:

DEFINITION 2.1: An *indexed grammar* is a 5-tuple $G=(N, T, I, P, S)$ where

(1) N, T, I are finite, pairwise disjoint sets; the sets of *variables*, *terminals* and *indices* respectively;

(2) P is a finite set of pairs (Af, Θ), $A \in N$, $f \in I \cup \{e\}$, $\Theta \in (NI^* \cup T)^*$, the set of *productions*; (Af, Θ) is denoted by $Af \to \Theta$;

(3) $S \in N$, the *start variable*.

Let $\Theta = u_1 B_1 \beta_1 u_2 B_2 \beta_2 \ldots B_n \beta_n u_{n+1}$ with $u_i \in T^*$ for $i \in [1:n+1]$, $B_j \in N$, and $\beta_j \in I^*$ for $j \in [1:n]$ with $n \geqq 0$, be an element of $(NI^* \cup T)^*$, and let $\gamma \in I^*$, then we set

$$\Theta : \gamma = u_1 B_1 \beta_1 \gamma u_2 B_2 \beta_2 \gamma \ldots B_n \beta_n \gamma u_{n+1}.$$

For Θ', $\Theta'' \in (NI^* \cup T)^*$, we set $\Theta' \Rightarrow \Theta''$ iff $\Theta' = \Theta_1 Af\gamma\Theta_2$, $\Theta'' = \Theta_1 (\Theta : \gamma) \Theta_2$ with Θ_1, $\Theta_2 \in (NI^* \cup T)^*$ and $Af \to \Theta \in P$, $f \in I \cup \{e\}$.

$\overset{n}{\Rightarrow}$ is the n-fold product, $\overset{+}{\Rightarrow}$ is the transitive and $\overset{*}{\Rightarrow}$ is the reflexive, transitive closure of $\Rightarrow$.

An indexed grammar $G = (N, T, I, P, S)$ is called a *linear indexed grammar*, iff each production in P is of one of the forms $Af \to uB\gamma v$ or $Af \to u$ with A, $B \in N$, $f \in I \cup \{e\}$, u, $v \in T^*$ and $\gamma \in I^*$.

An indexed grammar $G = (N, T, I, P, S)$ is called a *rightlinear indexed grammar*, iff each production in P is of one of the forms $Af \to uB\gamma$ or $Af \to u$ with A, $B \in N$, $f \in I \cup \{e\}$, $u \in T^*$ and $\gamma \in I^*$.

The language $L(G)$ generated by an (linear, rightlinear) indexed grammar $G = (N, T, I, P, S)$ is the set $L(G) = \{w \mid w \in T^*, S \overset{*}{\Rightarrow} w\}$. A language L is called an (*linear, rightlinear*) *index language* iff $L = L(G)$ for an (linear, rightlinear) indexed grammar G.

The index words in a derivation of such a rightlinear grammar can be interpreted as the pushdown list; the nonterminals can be interpreted as states of a suitable pushdown automaton. Vice versa for a given pushdown automaton a rightlinear indexed grammar can be constructed which generates the language that is accepted by that automaton. It follows that the rightlinear indexed languages are exactly the languages accepted by pushdown automata, *i.e.* the context-free languages, as has been shown by Aho [1].

A pushdown automaton with only one pushdown symbol is a *counter*. Such a device must stop if the pushdown store is empty, *i.e.* it has counted to zero.

An *iterated counter* may count down to zero several times. For this purpose there is a bottom marker # in the pushdown store of an iterated counter.

The formal definition of such an automaton is as follows [4]:

DEFINITION 2.2: An *iterated counter* is a pushdown automaton $K=(Z, T, \Gamma, \delta, z_0, \#, F)$ with $\Gamma=\{f, \#\}$ and

$$\delta(z, a, f) \subseteqq Z \times f^*$$
$$\delta(z, a, \#) \subseteqq Z \times f^* \# \cup Z \times \{e\}$$

where $z \in Z$ and $a \in T \cup \{e\}$ (e denotes the empty word).

The classes of languages accepted by these automata with final state, empty store or both coincide [4].

It is easy to construct a rightlinear indexed grammar for the iterated counter $K=(Z, T, \Gamma, \delta, z_0, \#, F)$ of Definition 2.2 which generates the set of all words accepted by K with final states. Set $G=(N, T, I, P, S)$ with $N=Z \cup \{S\}$, $I=\Gamma=\{f, \#\}$ and the following set of productions:

(a) $S \to z_0 \#$

(b) if $(z', f^i) \in \delta(z, a, f)$ then $zf \to az' f^i \in P$

(c) if $(z', f^i \#) \in \delta(z, a, \#)$ then $z\# \to az' f^i \# \in P$

(d) for all $z \in F$ the production $z \to e$ is in P
and if $(z', e) \in \delta(z, a, \#)$ with $z' \in F$, then $z\# \to a \in P$.

The productions of the form $z\# \to az' f^i \#$ represent the capability of K to start counting again, *i.e.* the iteration capability.

Counting with the help of the pushdown store of K corresponds to counting on derivations of a rightlinear grammar. It is now interesting to investigate the problem of counting on derivations of (linear) context-free grammars. This leads to the definition of an (rightlinear, linear) indexed counter grammar.

DEFINITION 2.3: An indexed grammar $G=(N, T, I, P, S)$ is called *indexed counter grammar* (*ic-grammar*) iff $I=\{f, \#\}$ and the productions in P are of one of the forms:

(a)
$$S \to A\#$$
$$S \to e$$

where S does not appear in any other production in P

(b) $\quad Ag \to \Theta, \quad g \in \{f, e\}, \quad \Theta \in (Nf^* \cup T)^*$

(c) $\quad A\# \to \Theta, \quad \Theta \in (Nf^* \# \cup T)^*$.

G is called *linear indexed counter grammar* (*linear ic-grammar*) iff in the above definition $\Theta \in T^* Nf^* T^* \cup T^*$ in (b) and $\Theta \in T^* Nf^* \# T^* \cup T^*$ in (c). G is called *rightlinear indexed counter grammar* (*rightlinear ic-grammar*) iff in the above definition $\Theta \in T^* Nf^* \cup T^*$ in (b) and $\Theta \in T^* Nf^* \# \cup T^*$ in (c).

A language L is called (*rightlinear*, *linear*) *ic-language* if $L=L(G)$ for an (rightlinear, linear) ic-grammar G.

Remark: If G is an ic-grammar, the index words in sentential forms are of the form $f^i\#$, $i \geqq 0$.

As shown above the rightlinear ic-languages are exactly the languages accepted by iterated counters.

Example 2.4:

1. Let $G_1=(\{S, A, B\}, \{a\}, \{f, \#\}, P, S)$ with

$$P=\{S \to A\#, A \to Af, A \to B, Bf \to BB, B\# \to a\}.$$

G_1 is an ic-grammar with $L(G_1)=\{a^{2^n} \mid n \geqq 0\}$.

2. Let $G_2=(\{S, A, B\}, \{a, b, c\}, \{f, \#\}, P, S)$ with

$$P=\{S \to A\#, A \to aAfc, A \to B, Bf \to bB, B\# \to e\}.$$

G_2 is a linear ic-grammar with $L(G_2)=\{a^n b^n c^n \mid n \geqq 0\}$.

3. Let $G_3=(\{S, A, B, C, D\}, \{a, b, \$\}, \{f, \#\}, P, S)$ with

$$P=\{S \to A\#, A \to aAf, A \to B, Bf \to bB, B\# \to \$C\#,$$
$$C \to aCf, C \to D, Df \to bD, D\# \to e\}.$$

G_3 is a rightlinear ic-grammar with $L(G_3)=\{a^n b^n \$ a^k b^k \mid n, k \geqq 0\}$.

3. PROPERTIES OF INDEXED COUNTER GRAMMARS

Definition 3.1: An ic-grammar $G=(N, T, I, P, S)$ is in *standard form* if all productions are of one of the forms:

(*a*) $$S \to A\#$$
$$S \to e$$

where S does not appear in any other production in P

(*b*) $$A \to BC$$

(*c*) $$A \to a$$

$A \to Bf$ (*f-producing production*)

$Af \to B$ (*f-consuming production*)

(*d*) $$A\# \to B\#$$

where A, $B \in N$ and $a \in T \cup \{e\}$.

A linear ic-grammar $G=(N, T, I, P, S)$ is in *linear standard form* if all productions are of one of the forms (*a*), (*c*) or (*d*) as above or $A \to bC$ or $A \to Bc$, A, B, $C \in N$, b, $c \in T$.

A rightlinear ic-grammar $G=(N, T, I, P, S)$ is in *rightlinear standard form* if all productions are in one of the forms (*a*), (*c*) or (*d*) as above or $A \to bC$, A, $C \in N$, $b \in T$.

Remark: Applications of f-producing or f-consuming productions give rise to nodes with only one son in the corresponding derivation tree, *i.e.* all changes of the index words attached to a variable occur on paths with no branches.

THEOREM 3.2: *For every (linear, rightlinear) ic-grammar $G=(N, T, I, P, S)$ an (linear, rightlinear) ic-grammar G' in standard form with $L(G)=L(G')$ can effectively be constructed.*

Proof: First replace each production of the form

$$A\# \to u_1 B_1 \gamma_1 u_2 \ldots u_l B_l \gamma_l u_{l+1}, \qquad l \geqq 0$$

by the productions $A\# \to B\#$ and $B \to u_1 B_1 \gamma_1 u_2 \ldots u_l B_l \gamma_l u_{l+1}$ where B is a new nonterminal. Then eliminate in a standard way all singular productions (e. g. [5]). Now all productions are of one of the forms:

(*a*) $$S \to A\#$$
$$S \to e$$

(*b*) $$Ag \to \Theta$$

(*c*) $$A\# \to B\#,$$

where $\Theta \in (Nf^* \cup T)^*$, $g \in \{f, e\}$ and A, $B \in N$.

Finally all productions of the form (*b*) are transformed by standard constructions (*see* [1], [2]) into the desired form. □

DEFINITION 3.3: An (linear, rightlinear) ic-grammar $G=(N, T, I, P, S)$ is in *e-free standard form* if it is in standard form and there are no productions of the form $A \to e$, $A \in N$, $A \neq S$.

Given a context-free grammar, it is easy to construct an equivalent e-free grammar – determine all variables producing the empty word and delete variables with this property on the righthand side of productions.

In the case of indexed grammars, we have to consider the difficulty that derivations of the empty word from a variable depend on the attached index words.

In [1] a construction of an e-free indexed grammar is given. This construction increases the number of indices and therefore is not appropriate for transforming an ic-grammar into an equivalent e-free ic-grammar.

In the following we give a suitable construction for ic-grammars based on the fact, that for indexed grammars $G=(N, T, I, P, S)$ the sets

$$EMPTY_G(A)=\{\gamma \mid A\gamma \overset{*}{\Rightarrow} e,\ \gamma \in I^*\}$$

are regular for all $A \in N$ [7] and that regular sets over a one element alphabet are ultimately periodic (e. g. [4]). Now we can state:

LEMMA 3.4: *Let $G=(N, T, I, P, S)$ be an (linear, rightlinear) ic-grammar. There are two constants $n_0 \geqq 1$ and $p_0 \geqq 1$ such that for all $B \in N \setminus \{S\}$ the following holds:*

If $Bf^r \# \overset{*}{\Rightarrow} e$ then $r < n_0$ or $r \geqq n_0$ and $Bf^{n_0+k} \# \overset{*}{\Rightarrow} e$ with $k=(r-n_0)$ mod p_0.

Remark: The Lemma states that if it is possible to derive the empty word using a long index word $f^r \#$, then it is also possible to derive the empty word using an index word $f^s \#$, where s is in the small interval $[n_0 : n_0+p_0-1]$.

Proof: Let $B \in N \setminus \{S\}$. Consider

$$\hat{L}(B)=\{f^r \# \mid Bf^r \# \overset{*}{\Rightarrow} e\}=EMPTY_G(B) \cap f^* \#.$$

This set is regular and hence the set $L(B)=\{f^r \mid Bf^r \# \overset{*}{\Rightarrow} e\}$ is regular over $\{f\}^*$ and thus is ultimately periodic, *i. e.* there are constants $n_B \geqq 0$ and $p_B \geqq 1$ with:

$$\text{if } r \geqq n_B \text{ then we have } f^r \in L(B) \text{ iff } f^{r+p_B} \in L(B)$$

Now set

$$n_0=\max\{1, \max\{n_B \mid B \in N \setminus \{S\}\}\}$$

and

$$p_0 = \mathrm{lcm}\{p_B \mid B \in N \setminus \{S\}\}$$

where lcm denotes the lowest common multiple. □

We will now give a construction of an e-free (linear, rightlinear) ic-grammar G' from a general (linear, rightlinear) ic-grammar G. The variables of G' will be triples such that a term $Bf^r\#$ in a derivation of G will be transformed into the triplel

$$[B, f^r, 0]\# \qquad \text{if} \quad r < n_0$$

or

$$[B, f^{n_0}, k]f^{r-n_0}\# \qquad \text{with} \quad k = (r - n_0) \bmod p_0 \quad \text{if } r \geqq n_0.$$

Here n_0 and p_0 are the constants of Lemma 3.4. With Lemma 3.4 it is obvious that the triple alone determines, whether $Bf^r\# \overset{*}{\Rightarrow} e$ is possible.

THEOREM 3.5: *For every (linear, rightlinear) ic-grammar $G = (N, T, I, P, S)$ an (linear, rightlinear) ic-grammar $\bar{G}$ in e-free standard form can effectively be constructed such that $L(G) = L(\bar{G})$ holds.*

Proof: Let G be an ic-grammar. W.l.o.g. G is in standard form. Let n_0 and p_0 be as in Lemma 3.4. Set $G' = (N', T, I, P', S)$ with

$$N' = \{S\} \cup \{[A, f^j, 0] \mid A \in N, j \leqq n_0\} \cup \{[A, f^{n_0}, k] \mid A \in N, k \in [0 : p_0 - 1]\}$$

and P' is defined as follows:

(*a*) $S \to e$ is in P' if $S \overset{*}{\Rightarrow} e$

(*b*) $S \to [A, e, 0]\#$ is in P' if $S \to A\#$ is in P.

(*c*) $[A, e, 0]\# \to [B, e, 0]\#$ is in P' if $A\# \to B\#$ is in P.

(*d*) If $A \to Bf$ is in P, then for all $j \in [0 : n_0 - 1]$, $i \in [0 : p_0 - 1]$ the productions $[A, f^j, 0] \to [B, f^{j+1}, 0]$ and $[A, f^{n_0}, i] \to [B, f^{n_0}, (i+1) \bmod p_0]f$ are in P'.

(*e*) If $Af \to B$ is in P then for all $j \in [1 : n_0]$, $i \in [0 : p_0 - 1]$ the productions

$$[A, f^j, 0]\# \to [B, f^{j-1}, 0]\#$$

and

$$[A, f^{n_0}, i]f \to [B, f^{n_0}, (i-1) \bmod p_0] \text{ are in } P'.$$

(f) If $A \to BC$ is in P, then for all $j \in [0:n_0-1]$ the productions

$$[A, f^j, 0] \to [B, f^j, 0][C, f^j, 0]$$

and

$$[A, f^j, 0] \to [C, f^j, 0] \qquad \text{if} \quad Bf^j \# \overset{*}{\Rightarrow} e$$

and

$$[A, f^j, 0] \to [B, f^j, 0] \qquad \text{if} \quad Cf^j \# \overset{*}{\Rightarrow} e \text{ are in } P'.$$

For all $i \in [0:p_0-1]$ the productions

$$[A, f^{n_0}, i] \to [B, f^{n_0}, i][Cf^{n_0}, i]$$

and

$$[A, f^{n_0}, i] \to [C, f^{n_0}, i] \qquad \text{if} \quad Bf^{n_0+i} \# \overset{*}{\Rightarrow} e$$

and

$$[A, f^{n_0}, i] \to [B, f^{n_0}, i] \qquad \text{if} \quad Cf^{n_0+i} \# \overset{*}{\Rightarrow} e \text{ are in } P'.$$

(g) If $A \to a$, $a \neq e$, is in P, then for all $j \in [0:n_0-1]$, $i \in [0:p_0-1]$ the productions

$$[A, f^j, 0] \to a \text{ and } [A, f^{n_0}, i] \to a \text{ are in } P'.$$

G' is effectively constructable since the problem "$e \in L(G)$" is decidable for an arbitrary indexed grammar G. Delete all single productions of G' to obtain $\bar{G}$ with $L(G') = L(\bar{G})$ [*see* (6), p. 29].

To establish a correspondence between derivations according to G and derivations according to G' we need the following function

$$\Psi: \quad (N \setminus \{S\}) f^* \# \to (N' \setminus \{S\}) f^* \#$$

with

$$\Psi(Af^r \#) = \begin{cases} [A, f^r, 0] \# & \text{if} \quad r \leqq n_0 \\ [A, f^{n_0}, k] f^{r-n_0} & \text{if} \quad r \geqq n_0 \quad \text{where } k = (r - n_0) \bmod p_0 \end{cases}$$

Now we will prove:

For all $A \in N \setminus \{S\}$, $w \in T^*$

If $Af^r \# \overset{n}{\Rightarrow} w$, $w \neq e$ according to G then $\Psi(Af^r \#) \overset{*}{\Rightarrow} w$ according to G'.

The case $n=1$ obviously holds.

Now assume $A\# \Rightarrow B\# \overset{n}{\Rightarrow} w$ according to G. We have

$$[A, e, 0]\# \quad \Rightarrow \quad [B, e, 0]\#$$

and with the induction hypothesis $\Psi(B\#)=[B, e, 0]\# \overset{*}{\Rightarrow} w$ holds according to G'. Now let $Af^r\# \Rightarrow Bf^{r+1}\# \overset{n}{\Rightarrow} w$ according to G.

We have to consider two cases:

1. $r<n_0$:

P' contains the production $[A, f^r, 0] \to [B, f^{r+1}, 0]$.

Therefore

$$\Psi(Af^r\#)=[A, f^r, 0]\# \quad \Rightarrow \quad [B, f^{r+1}, 0]\# = \Psi(Bf^{r+1}\#) \quad \overset{*}{\Rightarrow} \quad w$$

holds according to G'.

2. $r \geqq n_0$:

P' contains the productions $[A, f^{n_0}, i] \to [B, f^{n_0}, (i+1) \bmod p_0]f$ for all $i \in [0 : p_0 - 1]$.

Therefore

$$\Psi(Af^r\#)=[A, f^{n_0}, k]f^{r-n_0}\# \quad \Rightarrow \quad [B, f^{n_0}, (k+1) \bmod p_0]\, f^{r-n_0+1}\#$$

$$= \Psi(Bf^{r+1}\#) \quad \overset{*}{\Rightarrow} \quad w \text{ holds according to } G'.$$

Now let $Af^r\# \Rightarrow Bf^{r-1}\# \overset{n}{\Rightarrow} w$ according to G.

Again we have to consider two cases:

1. $r \leqq n_0$:

P' contains the production $[A, f^r, 0]\# \to [B, f^{r-1}, 0]\#$.

Therefore

$$\Psi(Af^r\#)=[A, f^r, 0]\# \quad \Rightarrow \quad [B, f^{r-1}, 0]\# = \Psi(Bf^{r-1}\#) \quad \overset{*}{\Rightarrow} \quad w$$

holds according to G'.

2. $r > n_0$:

P' contains the productions $[A, f^{n_0}, i]f \to [B, f^{n_0}, (i-1) \bmod p_0]$ for all $i \in [0 : p_0 - 1]$.

Therefore

$$\Psi(Af^r\#)=[A,f^{n_0},k]\,f^{r-n_0}\# \quad\Rightarrow\quad [B,f^{n_0},(k-1)\bmod p_0]\,f^{r-n_0-1}\#$$

$$=\Psi(Bf^{r-1}\#) \quad\stackrel{*}{\Rightarrow}\quad w \text{ holds according to } G'.$$

Finally assume

$$Af^r\# \Rightarrow Bf^r\#Cf^r\# \stackrel{n}{\Rightarrow} w_1 w_2 = w \text{ with } Bf^r\# \stackrel{n_1}{\Rightarrow} w_1 \text{ and } Cf^r\# \stackrel{n_2}{\Rightarrow} w_2.$$

The case $w_1 \neq e$ and $w_2 \neq e$ obviously holds.

Now consider $w_1 = e$. We have $w_2 \neq e$ and $Bf^r\# \stackrel{n_1}{\Rightarrow} e$.

1. $r < n_0$:

P' contains the production $[A,f^r,0] \to [C,f^r,0]$.

Therefore

$$\Psi(Af^r\#)=[A,f^r,0]\# \quad\Rightarrow\quad [C,f^r,0]\#=\Psi(Cf^r\#) \quad\stackrel{*}{\Rightarrow}\quad w_2=w$$

holds according to G'.

2. $r \geqq n_0$:

Since $Bf^r\# \stackrel{*}{\Rightarrow} e$ holds according to G, with lemma 3.4 follows

$$Bf^{n_0+k}\# \quad\stackrel{*}{\Rightarrow}\quad e \qquad \text{with} \quad k=(r-n_0) \bmod p_0.$$

Hence the production $[A,f^{n_0},k] \to [C,f^{n_0},k]$ is in P'.

Therefore

$$\Psi(Af^r\#)=[A,f^{n_0},k]f^{r-n_0}\# \quad\Rightarrow\quad [C,f^{n_0},k]\,f^{r-n_0}\#$$

$$=\Psi(Cf^r\#) \quad\stackrel{*}{\Rightarrow}\quad w_2=w \text{ holds according to } G'$$

The case $w_2 = e$ is similar.

Altogether we have shown:

If $S \Rightarrow A\# \stackrel{*}{\Rightarrow} w$, $w \neq e$ holds according to G, $S \Rightarrow [A,e,0]\# = \Psi(A\#) \stackrel{*}{\Rightarrow} w$ holds according to G', *i.e.* $L(G) \subseteqq L(G')$.

In a similar way one can show:

For all $A \in N \setminus \{S\}$, $r \geqq 0$ and $w \in T^*$ we have: If $\Psi(Af^r\#) \overset{n}{\Rightarrow} w$ holds according to G' then $Af^r\# \overset{*}{\Rightarrow} w$ holds according to G.

Hence $L(G') \subseteqq L(G)$.

The proof for linear and rightlinear ic-grammars is similar. □

For technical reasons it is convenient to have an ic-grammar with the property that a final derivation step of the form $Af^k\# \Rightarrow a$ is possible only if $k=0$. For this we have:

LEMMA 3.6: *Let $G=(N, T, I, P, S)$ be an (linear, rightlinear) ic-grammar. An equivalent (linear, rightlinear) ic-grammar $G'=(N', T, I, P', S)$ can effectively be constructed such that in a derivation of a terminal word each final derivation step is of the form $A\# \Rightarrow a$.*

Proof: W.l.o.g. we assume G in standard form.

Set $N' = N \cup \{T_a, F_a \mid a \in T \cup \{e\}\}$. Furthermore introduce the new productions

$$F_a f \rightarrow F_a, \qquad F_a\# \rightarrow T_a\# \qquad \text{and} \qquad T_a \rightarrow a \quad \text{for each } a \in T \cup \{e\}.$$

Now replace each production $A \rightarrow a$, $a \in T \cup \{e\}$ by $Af \rightarrow F_a$ and $A\# \rightarrow T_a\#$. □

Remark: If the grammar G in the foregoing lemma is in (e-free) standard form then the grammar G' is in (e-free) standard form too.

In the next section we need the property that for words w of an ic-language there are derivations such that the index words ocurring in these derivations are linear bounded by $|w|$.

Let us first define the following notion:

DEFINITION 3.7: Let $G=(N, T, I, P, S)$ be an ic-grammar, $w \in L(G)$, and $\mathcal{T}_w$ be a derivation tree of w. Set

$$\operatorname{maxind}(\mathcal{T}_w) = \max\{k \mid Af^k\# \text{ is a label of a node in } \mathcal{T}_w\}$$

and

$$\operatorname{maxind}(w) = \min\{\operatorname{maxind}(\mathcal{T}_w) \mid \mathcal{T}_w \text{ is a derivation tree of } w\}.$$

THEOREM 3.8: *Let $G=(N, T, I, P, S)$ be an (linear, rightlinear) ic-grammar in e-free standard form which satisfies the condition of the foregoing lemma.*

Then there is a constant $c>0$ *such that* $\operatorname{maxind}(w) \leqq c \max\{|w|, 1\}$ *for all* $w \in L(G)$.

Proof: We will only give a proof of the general case because the proof for linear and rightlinear grammars is a slight modification thereof.

In the proof we use the fact that due to the special form of the grammar G a derivation tree $\mathscr{T}$ of a word $w \in L(G)$, $w \neq e$ has $|w|-1$ nodes with two sons whereas all other inner nodes have only one son and that the length of the index words of a node and it's parent differ by at most one. It follows that if $\operatorname{maxind}(\mathscr{T})>c|w|$, where c is a suitably choosen constant, then $\mathscr{T}$ must contain long paths of nodes with only one son. It is shown that some of those paths which correspond to derivations of the form $Bf^i \# \stackrel{*}{\Rightarrow} Bf^j \#$ are in fact useless since they lengthen index words that are already long enough. By deleting such "useless" paths we can get a new derivation tree $\mathscr{T}'$ for w such that $\operatorname{maxind}(\mathscr{T}') \leqq c|w|$.

We now start the proof.

For each $n>0$ set $\operatorname{LCM}(n)=\operatorname{lcm}\{1,2,\ldots,n\}$. Set $c'=|N|^2 \operatorname{LCM}(|N|)$ and $c=2c'$. Now let $w \in L(G)$ and $k=\operatorname{maxind}(w)$. Since the case $w=e$ is obvious, assume $w \neq e$. Let $\mathscr{T}$ be a corresponding derivation tree with $k=\operatorname{maxind}(\mathscr{T})$ and a minimal number of nodes labeled with $Af^k \#$.

Now assume $k>c|w|$.

In $\mathscr{T}$ there is a node n_k with label $A_k f^k \#$. Consider the path $\mathscr{P}$ from $S\#$ to this node and let $n_1, n_2, \ldots, n_{k-1}$ be the nodes on $\mathscr{P}$ determined by:

n_i is labeled with $A_i f^i \#$ and

each successor of n_i on $\mathscr{P}$ is labeled with $Bf^j \#$ with $i<j \leqq k$ for $i \in [1:k-1]$. The nodes n_i, $i \in [1:k-1]$, have exactly one son in $\mathscr{T}$. Let $h=\lceil k/2 \rceil$ and let $\mathscr{P}'$ be the subpath of $\mathscr{P}$ starting with n_h. Since G is e-free, $\mathscr{P}'$ contains at most $|w|-1$ nodes with two sons in $\mathscr{T}$.

Therefore, using the pigeon-hole principle, there is a subpath $\mathscr{P}''$ of $\mathscr{P}'$ starting with n_r and ending with n_{r+s}, which contains no node with two sons in $\mathscr{T}$ such that

$$s+1 \geqq \frac{k-h+1}{|w|}=\frac{\lceil (k+1)/2 \rceil}{|w|} \geqq \frac{\lceil c'|w|+(1/2) \rceil}{|w|}>c' \quad i.e.\ s \geqq c'.$$

Now consider the sequence of nonterminals $A_r, \ldots, A_{r+s}$. In each subsequence of length $|N|+1$ of this sequence a nonterminal must occur twice. Hence, there are at least $\lfloor s/|N| \rfloor \geqq c'/|N|$ disjoint subsequences $A_j, \ldots, A_{j+i-1}$ with $A_j=A_{j+i}$, $i \in [1:|N|]$ and $j \in [r:r+s-|N|]$. Since each of these disjoint subsequences has a length less than or equal $|N|$ there is a $k_1 \in [1 : |N|]$ such that at least $c'/|N|^2$ of them have the length k_1.

Consider the derivation determined by $\mathscr{P}'' : A_r f^r \# \overset{*}{\Rightarrow} A_{r+s} f^{r+s} \#$.
This derivation contains $c'/|N|^2$ subderivations of the form

$$A_j f^j \# \quad \overset{*}{\Rightarrow} \quad A_{j+k_1} f^{j+k_1} \# \qquad \text{with} \quad A_j = A_{j+k_1}.$$

Omitting

$$\frac{1}{k_1} \times \frac{c'}{|N|^2} = \frac{\mathrm{LCM}(|N|)}{k_1}$$

subderivations of this form one obtains a derivation

$$A_r f^r \# \quad \overset{*}{\Rightarrow} \quad A_{r+s} f^{r+s-\mathrm{LCM}(|N|)} \#.$$

Let $\tilde{\mathscr{P}}$ be the path determined by this derivation.

Now consider the subtree $\mathscr{T}_1$ of $\mathscr{T}$ with root $n_{r+s} = m_t$ and a path $\mathscr{P}_1^a$ of m_t to a leave a. Let $m_{t-1}^a, m_{t-2}^a, \ldots, m_0^a$ be the nodes and $B_{t-1}^a, B_{t-2}^a, \ldots, B_0^a \in N$ determined by:

m_i^a is labeled with $B_i^a f^i \#$ and

each predecessor of m_i^a on $\mathscr{P}_1^a$ is labeled with $C f^j \#$ for $i < j \leqq k$, $i \in [0 : t-1]$. The father of m_i^a has exactly one son in $\mathscr{T}$ for $i \in [0 : t-1]$. Since G is e-free, $\mathscr{P}_1^a$ contains at most $|w| - 1$ nodes with two sons in $\mathscr{T}$. Since

$$\frac{t}{|w|} = \frac{r+s}{|w|} > \frac{h}{|w|} \geqq c'$$

there is a uniquely determined subpath $\tilde{\mathscr{P}}_1^a$ of $\mathscr{P}_1^a$ starting with $m_{d+c'}^a$ and ending with m_d^a, where d is maximal, which contains no node with two sons in $\mathscr{T}$ (with the possible exception of m_d^a). Now mark all nodes on the path from m_t to $m_{d+c'}^a$, and denote the subtree of $\mathscr{T}_1$ with root m_d^a by $\mathscr{T}_1^a$.

If b is another leave of $\mathscr{T}_1$, then we have either $\tilde{\mathscr{P}}_1^a = \tilde{\mathscr{P}}_1^b$ and $\mathscr{T}_1^a = \mathscr{T}_1^b$ or $\tilde{\mathscr{P}}_1^a$, $\tilde{\mathscr{P}}_1^b$ and $\mathscr{T}_1^a$, $\mathscr{T}_1^b$ are disjoint.

Now assume that for each leave a the paths $\tilde{\mathscr{P}}_1^a$, $\mathscr{P}_1^a$ are determined and all nodes on the path from m_t to the first node of $\tilde{\mathscr{P}}_1^a$ are marked.

Consider the sequence of nonterminals $B_{d+c'}^a, B_{d+c'-1}^a, \ldots, B_d^a$ of a path $\tilde{\mathscr{P}}_1^a$. With the same argument as above there must be at least $c'/|N|$ disjoint subsequences $B^a_j, B^a_{j-1}, \ldots, B^a_{j-i+1}$ with $B^a_j = B^a_{j-i}$ and therefore there is a $k_2 \in [1 : |N|]$ such that at least $c'/|N|^2$ of them have the length k_2.

Consider the derivation $B^a_{d+c'} f^{d+c'} \# \overset{*}{\Rightarrow} B^a_d f^d \#$ determined by $\tilde{\mathscr{P}}^a_1$. This derivation contains $c'/|N|^2$ subderivations of the form

$$B^a_j f^j \# \quad \overset{*}{\Rightarrow} \quad B^a_{j-k_2} f^{j-k_2} \# \qquad \text{with} \quad B^a_j = B^a_{j-k_2}.$$

Omitting

$$\frac{1}{k_2} \times \frac{c'}{|N|^2} = \frac{\mathrm{LCM}(|N|)}{k_2}$$

subderivations of this form yields a derivation

$$B^a_{d+c'} f^{d+c'-\mathrm{LCM}(|N|)} \# \quad \overset{*}{\Rightarrow} \quad B^a_d f^d \#.$$

Let $\hat{\mathscr{P}}^a_1$ be the path determined by this derivation.

Now perform the following operation on $\mathscr{T}$:

1. Substitute $\mathscr{P}''$ by $\tilde{\mathscr{P}}$.
2. If a marked node is labeled by $Df^t \#$, change the label by $Df^{t-\mathrm{LCM}(|N|)} \#$.
3. Substitute each $\tilde{\mathscr{P}}^a_1$ by $\hat{\mathscr{P}}^a_1$.

This yields a derivation tree of w which contains less nodes labeled with $Af^k \#$ (since n_k is a marked node) and we have a contradiction to the definition of $\mathscr{T}$. $\square$

4. PROPERTIES OF INDEXED COUNTER LANGUAGES

Recall that a family of languages is called a full trio if it is closed under homomorphisms, inverse homomorphisms and intersection with regular sets. If it is futhermore closed under union, concatenation and Kleene closure it is called a full AFL. It is easy to give a grammar based proof of the following theorems.

THEOREM 4.1: *The family of ic-languages is a full AFL.*

THEOREM 4.2: *The family of linear ic-languages is a full trio.*

Remark: It is known that the family of rightlinear ic-languages, *i.e.* the family of iterated counter languages is a full AFL [3].

In the following we will show that the family of ic-languages is properly contained in the family of indexed languages.

To this, consider the language $L=\{u\$_1 u\$_2 u^R \mid u\in\{0,1\}^+\}$. L is generated by the indexed grammar $\bar{G}=(N,T,I,P,S)$ with $N=\{S,A,B\}$, $I=\{f,g,\#\}$ and the following set of productions:

$$\begin{array}{cc} S\to 0Af\#0 & Bf\to B0 \\ S\to 1Ag\#1 & Bg\to B1 \\ A\to 0Af0 & B\#\to e \\ \multicolumn{2}{c}{A\to 1Ag1} \\ \multicolumn{2}{c}{A\to \$_1 B \$_2} \end{array}$$

$\bar{G}$ is a linear indexed grammar, thus L is a linear indexed language.

Now we will show:

LEMMA 4.3: *There are linear indexed languages which are not ic-languages.*

Proof: We show that the linear indexed language

$$L=\{u\$_1 u\$_2 u^R \mid u\in\{0,1\}^+\}$$

is not an ic-language.

Assume, L is generated by the ic-grammar $G=(N,T,I,P,S)$. W.l.o.g. let G be in e-free standard form.

Let $w=u\$_1 u\$_2 u^R\in L$ and let $\mathscr{T}_w$ be a derivation tree of w according to G with maxind $(\mathscr{T}_w)=\mathrm{maxind}(w)$.

Consider the two paths from the root to the leaves labeled $\$_1$ and $\$_2$. Let the last common node of these two paths be labeled $Af^k\#$.

Hence there is a derivation $Af^k\#\Rightarrow Bf^k\#Cf^k\#\overset{*}{\Rightarrow} w_1 Cf^k\#\overset{*}{\Rightarrow} w_1 w_2$ with $w_i\in\{0,1\}^*\$_i\{0,1\}^*$, $i\in[1:2]$ and $w=v_1 w_1 w_2 v_2$ with $v_1, v_2\in\{0,1\}^*$.

The entire word w is determined by the subword $w_2 v_2$ (or $v_1 w_1$), hence each derivation $Bf^k\#\overset{*}{\Rightarrow} w_1'$ according to G implies $w_1'=w_1$.

The same holds for each derivation $Cf^k\#\overset{*}{\Rightarrow} w_2'$.

Therefore, if $Af^k\#\Rightarrow Bf^k\# Cf^k\#\overset{*}{\Rightarrow} w'$, we have $w'=w_1 w_2$. Since $w_1 w_2$ also determines w, we have:

$$\text{If } S\overset{*}{\Rightarrow}\Theta_1 Af^k\#\Theta_2\Rightarrow\Theta_1 Bf^k\# Cf^k\#\Theta_2\overset{*}{\Rightarrow}\tilde{w} \text{ then } \tilde{w}=w.$$

Now, for each $w \in L(G)$ we choose a derivation tree $\mathcal{T}_w$ as above which determines a pair $p(w) = (A \to BC, k)$ where $k \leqq \text{maxind}(w)$. We have $k \leqq \text{maxind}(w) \leqq c|w|$, where c is the constant of Theorem 3.8 for G and if $w \neq \tilde{w}$ then $p(w) \neq p(\tilde{w})$. There are 2^n words in L with length $3n+2$, $n \geqq 1$. Therefore there are at least 2^n words $w \in L$ with $p(w) = (A \to BC, k)$ and $k \leqq \text{maxind}(w) \leqq c(3n+2)$.

But there are at most $\rho[c(3n+2)+1]$ pairs of the form $(A \to BC, k)$ where ρ is the number of productions of the form $A \to BC$ and $k \in [0 : c(3n+2)]$. Hence we have $2^n \leqq \rho[c(3n+2)+1]$ for all $n \geqq 1$, which is a contradiction. □

LEMMA 4.4: *There are rightlinear indexed languages,* i.e. *context-free languages, which are not linear ic-languages.*

Proof: We show that the context-free language

$$L = \{ u \$_1 u^R \$ v \$_2 v^R \mid u, v \in \{0, 1\}^+ \}$$

is not a linear ic-language.

Assume, L is generated by the linear ic-grammar $G = (N, T, I, P, S)$. W.l.o.g. let G be in e-free standard form.

Let $w = u \$_1 u^R \$ v \$_2 v^R \in L$.

Since there is a derivation of the form $Af^k \# \Rightarrow \$_1 Bf^k \# \Rightarrow \$_1 u^R \$ v \$_2 v'$ where v' is a prefix of v^R or $Af^k \# \Rightarrow Bf^k \# \$_2 \Rightarrow u' \$_1 u^R \$ v \$_2$ where u' is a suffix of u, each derivation $Bf^k \# \Rightarrow w'$ implies u^R is a prefix of w', v is a suffix of w' respectively. Now proceed similar to the proof of lemma 4.3. □

Remark: It is known that there are linear context-free languages which are not rightlinear ic-languages, *i.e.* iterated counter languages.

E. g. consider the language $L = \{ u \$ u^R \mid u \in \{0, 1\}^+ \}$. That L is not a rightlinear ic-language can be seen with arguments similar to that used in the proof of lemma 4.3.

On the other hand, consider the ic-language $L = \{ a 2^n \mid n \geqq 1 \}$ of example 2.4.1. This language is not a linear indexed language since Parikh mappings of linear indexed languages are semilinear [2]. Therefore we have:

COROLLARY 4.5: *There are ic-languages, which are not linear indexed languages.*

From example 2.4.2 follows:

COROLLARY 4.6: *There are linear ic-languages which are not rightlinear indexed languages,* i.e. *context-free languages.*

Remark: It is known that there are rightlinear ic-languages, *i.e.* iterated counter languages, which are not linear context-free languages.

E.g. consider the language of example 2.4.3 which is not linear context-free (*see* [4], p. 224).

Now it is easy to show with the help of the above lemmata, corollarys, remarks and known facts the following theorem:

THEOREM 4.7: *Let* $(L, RL) I$ *be the class of* (*linear, rightlinear*) *indexed languages, let* $(L, RL)\ IC$ *be the class of* (*linear, rightlinear*) *ic-languages and let* $(L, RL)\ CF$ *be the class of* (*linear, rightlinear*) *context-free languages.*

Let $X \rightarrow Y$ *mean* Y *is a proper subset of* X *and let* $X \not\rightleftarrows Y$ *mean* $X \not\subset Y$ *and* $Y \not\subset X$.

The following diagram holds:

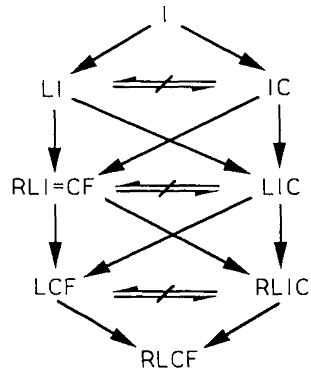

Figure 1.

5. R-MODE DERIVATIONS OF INDEXED COUNTER GRAMMARS

Let us now alter the derivation mode for indexed grammars $G=(N, T, I, P, S)$ in the following manner. For

$$\Theta=u_1 B_1 \beta_1 u_2 B_2 \beta_2 \ldots B_n \beta_n u_{n+1} \quad \text{with } u_i \in T^* \quad \text{for } i \in [1:n+1],$$

$B_j \in N$ and $\beta_j \in I^*$ for $j \in [1:n]$ with $n \geqq 0$ and $\gamma \in I^*$ we set

$$\Theta :_R \gamma = u_1 B_1 \beta_1 \gamma u_2 B_2 \beta_2 \ldots B_n \beta_n u_{n+1}.$$

In $\Theta :_R \gamma$ the indexword γ is appended to the indexword of the leftmost variable of Θ only, whereas in $\Theta : \gamma$, as defined in Section 2, γ is appended to the indexwords of all variables appearing in Θ.

Now let us define the R-mode derivation for G:

For Θ', $\Theta'' \in (NI^* \cup T)^*$ we set $\Theta' {}^R\!\Rightarrow \Theta''$ iff $\Theta' = wAf\gamma\Theta_1$, $\Theta'' = w(\Theta\Theta_1 :_R \gamma)$ with $w \in T^*$, $\Theta_1 \in (NI^* \cup T)^*$ and $Af \to \Theta \in P$, $f \in I \cup \{e\}$.

${}^R\!\overset{n}{\Rightarrow}$, ${}^R\!\overset{+}{\Rightarrow}$, ${}^R\!\overset{*}{\Rightarrow}$ are defined as usual.

Furthermore set $L_R(G) = \{w \mid w \in T^*,\ S\, {}^R\!\overset{*}{\Rightarrow} w\}$.

Remark: The R-mode derivation is a leftmost derivation. Our definition corresponds to R-mode derivations for extended grammars as defined in [6].

In [6] it was shown, that type-0 languages are exactly the languages $L_R(G)$, where G is an indexed grammar. Now the question arises, which class of languages is obtained if we restrict G to an ic-grammar. Here we have

THEOREM 5.1: *For each type-0 language L there is an ic-grammar G with $L = L_R(G)$.*

Proof: Let $L = L_R(G')$ for an indexed grammar $G' = (N', T, I', P', S')$ with $I' = \{g_1, \ldots, g_{m-1}\}$. W.l.o.g., *see* [6], we can assume that the productions in P' are of one of the forms

(a) $A \to BC$

(b) $A \to a$

(c) $A \to Bf$

(d) $Af \to B$

where A, B, $C \in N'$, $f \in I'$ and $a \in T \cup \{e\}$.

We will now construct an equivalent ic-grammar G. For this, we represent an indexword $g_{i_0} g_{i_1} \cdots g_{i_n} \in I'^*$ in G' by the indexword $f^k \#$ in G where $k = \sum_{j=0}^{n} i_j m^j$.

Let $G = (N, T, I, P, S)$ with

$$N = N' \cup \{T_i^j \mid i \in [0 : m-1], j \in [1 : m-1]\} \cup \{U_j \mid j \in [1 : m-1]\} \cup \{S, E, E'\},$$
$$I = \{f, \#\},$$

and P is defined as follows:

1. $S \to S'\#$ is in P,
2. $A \to BC$ is in P if $A \to BC$ is in P',
3. $A \to a$ is in P if $A \to a$ is in P',
4. $A \to T_0^j B$ is in P if $Ag_j \to B$ is in P',
5. $A \to U_j B$ is in P if $A \to Bg_j$ is in P',

6. for all $j\in[1:m-1]$, $i\in[0:m-2]$ the productions

$$T_i^j f\to T_{i+1}^j,$$
$$T_{m-1}^j f\to T_0^j E$$

and

$$T_j^j\# \to e$$

are in P.

7. $E\to E'f$ and $E'\to e$ are in P,
8. for all $j\in[1:m-1]$ the productions

$$U_j f\to U_j \underbrace{E\ldots E}_{m-\text{times}}$$

and

$$U_j\# \to E\# \underbrace{E\ldots E}_{(j-1)-\text{times}} \text{ are in } P$$

Now set $\Phi: I'^*\to\mathbb{N}$ with $\Phi(e)=0$ and $\Phi(g_i\gamma)=i+m\Phi(\gamma)$, *i.e.*, if $\gamma=g_{i_0}g_{i_1}\ldots g_{i_n}\in I'^*$, then $i_n i_{n-1}\ldots i_0$ is the m-adic representation of $\Phi(\gamma)$. With induction on the lenght of derivation it is possible to show

(*a*) if $S' \overset{R}{\Rightarrow}{}^* w A\gamma\Theta$ according to G' then $S \overset{R}{\Rightarrow}{}^* w A f^{\Phi(\gamma)}\#\Theta$ according to G and

(*b*) if $S \overset{R}{\Rightarrow}{}^* w A f^k\#\Theta$ according to G then there is a $\gamma\in I'^*$ with $\Phi(\gamma)=k$

and $S' \overset{R}{\Rightarrow}{}^* w A\gamma\Theta$ according to G' where $w\in T^*$, $A\in N'$, $\gamma\in I'^*$ and $\Theta\in(N' I'^*\cup T)^*$.

From this, $L_R(G)=L_R(G')=L$ can easily be shown. □

This theorem can be interpreted in the way that each type-0 language can be obtained by counting on leftmost derivations of context free grammars. This is similar to a grammatical version of the simulation of Turing-machines by two-counter machines, *see* [5], p. 172.

REFERENCES

1. A. V. Aho, Indexed Grammars, *J.A.C.M.*, 1968, *15*, p. 647-671.

2. J. DUSKE and R. PARCHMANN, Linear Indexed Languages, *Theoret. Comput. Sci.*, 1984, *32*, p. 47-60.
3. S. A. GREIBACH, An Infinite Hierarchy of Context-Free Languages, *J.A.C.M.*, 1969, *16*, p. 91-106.
4. M. A. HARRISON, Introduction to Formal Language Theory, *Addison-Wesley*, Reading, MA, 1978.
5. J. E. HOPCROFT and J. D. ULLMAN, Introduction to Automata Theory, Languages and Computation, *Addison-Wesley*, Reading, MA, 1979.
6. R. PARCHMANN and J. DUSKE, Grammars, Derivation Modes and Properties of Indexed and Type-0 Languages, *Theoret. Comput. Sci.*, 1987, *49*, p. 23-42.
7. R. PARCHMANN and J. DUSKE, The Structure of Index Sets and Reduced Indexed Grammars, *R.A.I.R.O. Inform. Theor. Appl.*, 1990, *24*, p. 89-104.