Indexed LL(k) Grammars

R. PARCHMANN, J. DUSKE, J. SPECHT

1. Introduction

In the literature a number of extensions of context-free languages have been proposed. The motivation for this is to describe certain constructs of programming languages which are not context-free. An important class of such an extension are the indexed languages introduced by Aho [1].

In the area of context-free languages, the LL(k) languages are of special interest. In [10] and [11] proposals have been made to generalize this notion to the indexed case.

Indexed languages coincide with the IO-macro languages introduced in [2]. In [6], Mehlhorn defined the notion of a strong LL(k) macro grammar and investigated this class of grammars. Furthermore he defined the notion of a general LL(k) macro grammar and stated the following problems for these classes of grammars:

(a) Is the class of languages defined by the strong LL(k) condition equal to the class of languages defined by the general LL(k) condition?

(b) Is the general LL(k) condition decidable for a given k?

In [10], Sebesta and Jones gave a positive answer to question (b) for indexed grammars without e-productions in the case $k=1$.

In this paper we will answer completely these two questions for indexed LL(k) grammars.

In Section 2, basic notions and definitions will be given and compared with those introduced in [10] and [11].

In Section 3 it will be proved that the strong indexed LL(k) property is decidable for a given k.

In Section 4 we will show that the strong indexed LL(k) languages are deterministic indexed languages which were introduced in [7]. In Section 5 deterministic context-free languages will be characterized as right linear strong indexed LL(1) languages.

In Section 7 the main result of this paper, namely the decidability of the general indexed LL(k) property will be proved. This will be shown by using a general transformation on indexed grammars given in Section 6. This transformation converts an arbitrary indexed grammar into an equivalent grammar which is a strong indexed LL(k) grammar iff the original one is a general indexed LL(k) grammar. This answers the question (a) posed above. The decidability of the strong indexed LL(k) property then implies the decidability of the general indexed LL(k) property, which answers question (b).

2. Definitions of indexed LL(k) grammars

In this section we will consider subclasses of indexed grammars. Aho [1] introduced indexed grammars and languages. We will state these notions in the following form:

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, α), $A \in N$, $f \in I \cup \{e\}$, $\alpha \in (NI^* \cup T)^*$, the set of *productions.* (Af, α) is denoted by $Af \to \alpha$.

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

Let $\alpha = u_1 B_1 \beta_1 u_2 B_2 \beta_2 \dots B_k \beta_k u_{k+1}$ with $u_i \in T^*$ for $i \in [1: k+1]$, and $B_j \in N$, $\beta_j \in I^*$ for $j \in [1: k]$ with $k \geqq 0$ be an element of $(NI^* \cup T)^*$ and let $\gamma \in I^*$. Then we set

$$\alpha:\gamma = u_1 B_1 \beta_1 \gamma u_2 B_2 \beta_2 \gamma \dots B_k \beta_k \gamma u_{k+1}.$$

For $u, v \in (NI^* \cup T)^*$ we set $u \Rightarrow v$ iff $u = \varphi_1 Af\gamma\varphi_2$, $v = \varphi_1(\alpha:\gamma)\varphi_2$ with $\varphi_1, \varphi_2 \in (NI^* \cup T)^*$ and $Af \to \alpha \in P$. $\overset{n}{\Rightarrow}$ is the n-fold product and $\overset{*}{\Rightarrow}$ is the reflexive, transitive closure of $\Rightarrow$.

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

The subclasses of indexed grammars considered in this paper are the indexed LL(k) — and strong indexed LL(k) grammars, whose definitions are generalizations of the corresponding context-free notions. Furthermore we will compare these definitions with the corresponding definitions introduced in [10] and [11]. First we will introduce some basic notions.

Let Σ be an alphabet and let $k \geqq 1$ be an integer. ${}^{(k)}\Sigma^*$ denotes the set of all words w over Σ with $|w| \leqq k$, where $|w|$ denotes the length of w. The function ${}^{(k)}: \Sigma^* \to {}^{(k)}\Sigma^*$ is the identity on ${}^{(k)}\Sigma^*$ and assigns to each $w \in \Sigma^*$ with $|w| > k$ the prefix of w of length k.

Now let $G=(N, T, I, P, S)$ be an indexed grammar. Let π: $Af \to \beta$ be a production, let $k \geqq 1$, and let $\gamma \in I^*$. Then we set

$$\text{First}_k(\pi) = \{{}^{(k)}u \mid S \overset{*}{\Rightarrow} wA\alpha \overset{\pi}{\Rightarrow} w\theta \overset{*}{\Rightarrow} wu\},$$

$$\text{First}_k(\pi, \gamma) = \{{}^{(k)}u \mid A\gamma \overset{\pi}{\Rightarrow} \theta \overset{*}{\Rightarrow} u\},$$

where $w, u \in T^*$, and where $\overset{*}{\Rightarrow}$ and $\overset{\pi}{\Rightarrow}$ are leftmost derivations. Furthermore we set for $\theta \in (NI^* \cup T)^*$

$$\text{First}_k(\theta) = \{{}^{(k)}u \mid \theta \overset{*}{\Rightarrow} u\} \text{ with } u \in T^*.$$

From now on, all derivations are assumed to be leftmost derivations.

Now we define the notion of an indexed LL(k) grammar (ILL(k) grammar). This notion corresponds to the context-free LL(k) grammars.

Definition 2.2. Let $G=(N, T, I, P, S)$ be an indexed grammar and let $k \geqq 1$ be an integer. G is called an *ILL(k) grammar* if the following holds:
Let

$$S \stackrel{*}{\Rightarrow} wA\gamma\alpha \stackrel{\pi}{\Rightarrow} w\theta_1 \stackrel{*}{\Rightarrow} wx \quad \text{and}$$

$$S \stackrel{*}{\Rightarrow} wA\gamma\alpha \stackrel{\pi'}{\Rightarrow} w\theta_2 \stackrel{*}{\Rightarrow} wy$$

be two leftmost derivations with $A \in N, \gamma \in I^*$, $\alpha \in (NI^* \cup T)^*$, and $w, x, y \in T^*$. Then ${}^{(k)}x = {}^{(k)}y$ implies $\pi = \pi'$.

(*Note:* $I = \emptyset$ yields the context-free LL(k) grammars.)

Remark. Let G be an ILL(k) grammar. Then for each word $w \in L(G)$ there is exactly one leftmost derivation according to G.

Example 2.1. Consider the indexed grammar $G=(N, T, I, P, S)$ with $N = \{S, A, B, C\}$, $T = \{a, b, c\}$ and $I = \{f, g\}$. The productions in P are π_1: $S \to aAf$, π_2: $S \to bAg$, π_3: $A \to B$, π_4: $A \to C$, π_5: $Bf \to a$, π_6: $Cf \to b$, π_7: $Cg \to a$, and π_8: $Bg \to c$. Only the following derivations are possible:

$$S \Rightarrow aAf \Rightarrow aBf \Rightarrow aa,$$

$$S \Rightarrow aAf \Rightarrow aCf \Rightarrow ab,$$

$$S \Rightarrow bAg \Rightarrow bBg \Rightarrow bc,$$

$$S \Rightarrow bAg \Rightarrow bCg \Rightarrow ba.$$

Obviously G is an ILL(l) grammar.

As for the context-free case it is possible to define the notion of a strong ILL(k) grammar.

Definition 2.3. Let $G=(N, T, I, P, S)$ be an indexed grammar and let $k \geqq 1$ be an integer. G is called a *strong ILL(k) grammar* if $\text{First}_k(\pi) \cap \text{First}_k(\pi') = \emptyset$ holds for all productions $\pi \neq \pi'$ which possess the same lefthand side or are of the form π: $A \to \alpha$, π': $Af \to \alpha'$ with $f \in I$.

(*Note:* $I = \emptyset$ yields the context-free strong LL(k) grammars.)

Remark. It is easy to see that strong ILL(k) grammars are ILL(k) grammars. The ILL(1) grammar G of Example 2.1 is not a strong ILL(1) grammar because $\text{First}_1(\pi_3) = \{a, c\}$ and $\text{First}_1(\pi_4) = \{a, b\}$. This shows that an ILL(k) grammar is not necessarily a strong ILL(k) grammar even for $k=1$. This differs from the context free case.

We can state:

Theorem 2.1. 1) A strong ILL(k) grammar is an ILL(k) grammar. 2) An ILL(1) grammar is not necessarily a strong ILL(1) grammar.

We will call a language a *(strong) ILL(k) language* if there exists a (strong) ILL(k) grammar generating this language.

We will now compare our definition of ILL(k) — and strong ILL(k) grammars with the definitions of indexed LL2(k) — and indexed LL1(k) grammars introduced in [10].

Obviously, an ILL(k) grammar is an indexed LL2(k) grammar. On the other hand consider the indexed grammar given by the productions π_1: $S\to Af$, π_2: $A\to a$, and π_3: $Af\to a$. The two possible leftmost derivations

$$S \overset{\pi_1}{\Longrightarrow} Af \overset{\pi_2}{\Longrightarrow} a,$$

$$S \overset{\pi_1}{\Longrightarrow} Af \overset{\pi_3}{\Longrightarrow} a$$

of the same word a according to this grammar show that it is not an ILL(1) grammar, but the LL2(1) condition is still satisfied.

The notion of a strong ILL(k) grammar and that of an LL1(k) grammar are incomparable.

The indexed grammar G_1 given by the productions π_1: $S\to aAf$, π_2: $S\to bAg$, π_3: $Af\to ab$, and π_4: $Ag\to ac$ is obviously a strong ILL(1) grammar.

G_1 is not an indexed LL1(1) grammar, since BASE(G_1) (see [10]) is not a context-free strong LL(1) grammar. This stems from the fact that in leftmost derivations according to G_1, applicability of π_3 excludes applicability of π_4 and vice versa. On the other hand consider the indexed grammar G_2 given by the productions π_1: $S\to Af$, π_2: $S\to Ag$, and π_3: $A\to a$ in P. G_2 is not a strong ILL(1) grammar, since $\mathrm{First}_1(\pi_1)\cap\mathrm{First}_1(\pi_2)=\{a\}$. G_2 obviously is an indexed LL1(1) grammar.

Furthermore, G_2 is not an indexed LL2(1) grammar, which shows that Theorem 4 in [10] is false.

In [11], three types of indexed LL(k) grammars, the α, β, γ-ILL(k) grammars were introduced. It is easy to see that the definitions of α-ILL(k)- and ILL(k) grammars and those of β-ILL(k)- and strong ILL(k) grammars coincide. These notions are defined but are not investigated further in [11], where only γ-ILL(k) grammars are investigated. These grammars do not even include all context-free LL(k) grammars.

3. Properties of strong ILL(k) grammars and languages

In this section we will first show that, given an indexed grammar G and a production $\pi\in P$, the language $\mathrm{First}_k(\pi)$ can be given effectively. This result implies that it is decidable whether a given indexed grammar is a strong ILL(k) grammar for a given k. Furthermore, we will call an indexed grammar "reduced", if each production occurs in at least one derivation of a terminal word. With the aid of $\mathrm{First}_k(\pi)$, we can construct, given an indexed grammar G, an equivalent reduced indexed grammar.

Theorem 3.1. Let an indexed grammar $G=(N, T, I, P, S)$, a production $\pi\in P$, and an integer $k\geqq 1$ be given. Then an indexed grammar G'' with $L(G'')=\mathrm{First}_k(\pi)$ can be constructed effectively.

Proof. Construct the indexed grammar $G'=(N, T', I, P', S)$ with $T'=T\cup\{\#\}$, and $P'=P\cup\{\pi'\}$ where π': $Af\to \#\alpha$ if π: $Af\to\alpha$. It is easy to construct a finite transducer M with $M(L(G'))=\mathrm{First}_k(\pi)$. Here we use the notion "finite transducer" as given in [1].

From Theorem 3.1 and Lemma 3.2 in [1] it follows that we can construct effectively an indexed grammar G'' with $L(G'')=M(L(G'))=\text{First}_k(\pi)$.
Now we can state

Corollary 3.1. Let $G=(N, T, I, P, S)$ be an indexed grammar, π a production of G, $k\geqq 1$ an integer, $\gamma\in I^*$, and $\theta\in(NI^*\cup T)^*$. Given a $v\in{}^{(k)}T^*$ it is decidable whether

(1) $v\in\text{First}_k(\pi)$, (2) $v\in\text{First}_k(\theta)$, (3) $v\in\text{First}_k(\pi, \gamma)$ holds.

Proof. (1) Construct an indexed grammar G'' with $L(G'')=\text{First}_k(\pi)$. In [5] it is shown that the membership problem for indexed languages is decidable.

(2) With the aid of G construct the indexed grammar $G'=(N\cup\{S'\}, T, I, P', S')$ where S' is a new start variable and $P'=P\cup\{\pi'\}$ where π': $S'\to\theta$ is a new production. Then we have $\text{First}_k(\theta)=\text{First}_k(\pi')$.

(3) Let Af be the lefthand side of the production π. If π cannot be applied to $A\gamma$, then we have $\text{First}_k(\pi, \gamma)=\emptyset$. If $A\gamma \overset{\pi}{\Rightarrow} \hat{\theta}$, then $\text{First}_k(\pi, \gamma)=\text{First}_k(\hat{\theta})$ holds.

Corollary 3.2. Let $G=(N, T, I, P, S)$ be an indexed grammar and $k\geqq 1$ be an integer. It is decidable whether G is a strong ILL(k) grammar.

Since the language $\text{First}_k(\pi)$ can be given effectively for an indexed grammar G, it is possible to single out all productions of G which never appear in derivations of terminal words. We will call an indexed grammar without such productions "reduced".

Definition 3.1. An indexed grammar $G=(N, T, I, P, S)$ with $L(G)\neq\emptyset$ is called *reduced* if for each $\pi\in P$ there exists a derivation of a terminal word in which π is applied.

Theorem 3.2. Let $G=(N, T, I, P, S)$ be an indexed grammar with $L(G)\neq\emptyset$. Then it is possible to construct a reduced indexed grammar $G'=(N, T, I, P', S)$ which is equivalent to G.

Proof. Determine for each production π the language $\text{First}_1(\pi)$. If $\text{First}_1(\pi)=\emptyset$ then remove the production. The grammar G' obtained in this way is reduced and obviously $L(G)=L(G')$ holds.

4. Strong ILL(k) languages are deterministic indexed languages

In [7] an indexed pushdown automaton (IPDA) has been defined, and it has been shown that these automata accept exactly the indexed languages.

Furthermore, a deterministic IPDA (d-IPDA) has been introduced in [7]. The class of languages accepted by these automata is called the class of deterministic indexed languages (DIL's). This class has properties similar to those of the class of deterministic context-free languages [7, 8]. In this section we will show that the strong ILL(k) languages form a subclass of the DIL's.

Theorem 4.1. If $G=(N, T, I, P, S)$ is a strong ILL(k) grammar then $L(G)$ is a deterministic indexed language.

Proof. We will construct a d-IPDA K which accepts the language $L(G)\$^k$ where \$ is an endmarker not in T. Since the deterministic indexed languages are closed under right quotient with regular sets [8], the language $L(G)$ is a DIL.

The states of K are the contents of a buffer of length k. This buffer contains a lookahead of k input symbols. Furthermore, the automaton simulates leftmost derivations according to G.

Set $K=(Z, X, \Gamma_1, \Gamma_2, \delta, z_0, A_0, g_0, F)$ with $Z={}^{(k)}X^* \cup \{z_f\}$, $X=T\cup\{\$\}$, $\Gamma_1=N\cup T\cup\{A_0\}$, where A_0 is a new element, $\Gamma_2=I$, $z_0=g_0=e$, $F=\{z_f\}$, and δ will be defined as follows:

(1) For all $u\in X^*$ with $|u|\leqq k-2$ and $a\in X$ set

$$(ua, (A_0, e))\in\delta(u, a, (A_0, e)).$$

For all $u\in X^{(k-1)}$, $a\in X$ set

$$(ua, (S, e)(A_0, e))\in\delta(u, a, (A_0, e)).$$

(2) Let π: $Af\to B_1\gamma_1\ldots B_r\gamma_r$ be a production of G, $r\geqq 0$, $B_i\in N\cup T$, and $\gamma_i\in I^*$ for $i\in[1: r]$. Then set $(v, (B_1, \gamma_1)\ldots(B_r, \gamma_r))\in\delta(v, e, (A, f))$ if $|v|=k$ and $\bar{v}\in\mathrm{First}_k(\pi)$, where $\bar{v}$ is the maximal prefix of v with $\bar{v}\in T^*$.

(3) For all $b\in X$, $a\in T$, and $u\in X^{k-1}$ set $(ub, e)\in\delta(au, b, (a, e))$.

(4) $(z_f, e)\in\delta(\$^k, e, (A_0, e))$.

Obviously, we have $L(K)=L(G)\$^k$. Since G is a strong ILL(k) grammar, we have $|\delta(z, x, (B, g))|\leqq 1$ for all $z\in Z, x\in X\cup\{e\}$, $B\in\Gamma_1$, and $g\in I\cup\{e\}$. (For example: $\delta(v, e, (A, e))$ with $A\in N$ can only be defined in (2). If $|\delta(v, e, (A, e))|>1$ we have a contradiction to the strong ILL(k) condition.)

It is easy to see that in each configuration (z, w, θ) of K at most one move is possible. (For example: $\delta(v, e, (A, e))\neq\emptyset$ and $\delta(v, e, (A, f))\neq\emptyset$ in (2) for $A\in N$ and $f\in I$ leads to a contradiction to the strong ILL(k) condition.)

Therefore K is a deterministic IPDA.

5. Strong ILL(k) languages and deterministic context-free languages

Theorem 4.1. shows that the class of strong ILL(k) languages is contained in the class of DIL's. The DIL's include all deterministic context-free languages, which we will now characterize as a special class of strong ILL(1) languages.

Theorem 5.1. For each deterministic context-free language L there exists a strong ILL(1) grammar G with $L=L(G)$.

Proof. Choose a deterministic pda $K=(Z, T, \Gamma, \delta, z_0, A_0, F)$ with $L=L(K)$. We may assume that in a final state, K may make no e-move (see [4], p. 239). Now construct the following indexed grammar $G=(N, T, I, P, S)$ with $N=Z\cup\{S\}$ and $I=\Gamma$. The productions of G will be defined as follows:

1) $S\to z_0A_0$ is in P

2) If $\delta(z, a, A)=(z', B_1\ldots B_r)$, then the production $zA\to az'B_1\ldots B_r$ is in P.

3) For each $z\in F$ the production $z\to e$ is in P.

Obviously, G is a strong ILL(1) grammar generating L.

The productions of the indexed grammar G in the foregoing proof are of a special "right linear" form. Let us define:

Definition 5.1. An indexed grammar $G=(N, T, I, P, S)$ is called a *right linear indexed grammar*, if each production in P has one of the forms $Af \to aB\gamma$ or $Af \to a$ with $A, B \in N$, $f \in I \cup \{e\}$, $a \in T \cup \{e\}$, and $\gamma \in I^*$.

Recall that an indexed grammar $G=(N, T, I, P, S)$ is called an *RIR grammar* (right linear indexed right linear) if all productions in P are of one of the forms $Af \to aB$, $Af \to a$, or $A \to aBf$, where $A, B \in N$, $a \in T \cup \{e\}$, and $f \in I \cup \{e\}$.

RIR grammars generate exactly the context-free languages. (see [1]).

Obviously, each RIR grammar is a right linear indexed grammar. On the other hand, it is easy to show that for each right linear indexed grammar there is an equivalent RIR grammar.

Therefore we can state:

Theorem 5.2. Right linear indexed grammars generate exactly the context-free languages.

Now we can state:

Corollary 5.1. Each deterministic context-free language is generated by a right linear strong ILL(1) grammar.

To prove the converse of this statement, we first need the following lemma.

Lemma 5.1. For each right linear indexed grammar $G=(N, T, I, P, S)$ there exists an equivalent right linear indexed grammar G' with the following properties:

1) There is exactly one start production.
2) All the other productions are of the form $Af \to \alpha$ with $f \neq e$.
3) If G is a strong ILL(k) grammar, then G' is a strong ILL(k) grammar.

Proof. Set $G'=(N', T, I', P', S')$ with $N'=N \cup \{S'\}$, $I'=I \cup \{\#\}$ and $P'=\{S' \to S\#\} \cup P''$, where P'' is defined as follows:

a) If $Af \to \alpha \in P$, $f \neq e$, then $Af \to \alpha \in P''$.
b) If $A \to \alpha \in P$, then $Ag \to \alpha \colon g \in P''$ for all $g \in I'$.

Obviously, G' is a right linear indexed grammar which satisfies 1) and 2), and is equivalent to G. Furthermore, it is easy to see that if G is a strong ILL(k) grammar, then G' is a strong ILL(k) grammar too.

Now we can prove:

Theorem 5.3. Each right linear indexed grammar $G=(N, T, I, P, S)$ which is a strong ILL(1) grammar, generates a deterministic context-free language.

Proof. If $L(G)=\emptyset$ then $L(G)$ is a deterministic context-free language. If $L(G)\neq\emptyset$, construct $G'=(N', T, I', P', S')$ according to the proof of Lemma 5.1. Furthermore we can assume w.l.o.g. that G' is reduced. We will define a pda K which accepts the language $L(G')\$$, where $\$$ is a new symbol. K buffers a lookahead of length one in its states and simulates leftmost derivations according to G'. The strong ILL(1) property of G' then implies that K is a deterministic pda.

Set $K=(Z, X, \Gamma, \delta, z_0, \#, F)$ with $Z=N' \times (T \cup \{e, \$\}) \cup \{z_f\}$, $X=T \cup \{\$\}$, $\Gamma=I'$, $z_0=(S', e)$, $F=\{z_f\}$, and define δ as follows:

(1) For all $c \in X$ let $((S, c), \#) \in \delta((S', e), c, \#)$.
(2) Let π: $Af \to aB\gamma$ be a production in P with $A \neq S'$.
 (a) If $a \neq e$ then $((B, c), \gamma) \in \delta((A, a), c, f)$ for all $c \in X$.
 (b) If $a = e$ then $((B, b), \gamma) \in \delta((A, b), e, f)$ for all $b \in {}^{(1)}\mathrm{First}_1(\pi)\$$.
(3) Let π: $Af \to a$ be a production in P.
 (a) If $a \neq e$ then $(z_f, e) \in \delta((A, a), \$, f)$.
 (b) If $a = e$ then $(z_f, e) \in \delta((A, \$), e, f)$.

First we make the following observations concerning K:

Claim 1. If $\delta((A, a), c, f) \neq \emptyset$ for $a, c \in X$ then there is a production π: $Af \to a\alpha$ and $\{a\} = \mathrm{First}_1(\pi) = {}^{(1)}\mathrm{First}_1(\pi)\$$.

Claim 2. If $\delta((A, a), e, f) \neq \emptyset$ for $a \in X$ then there is a production π: $Af \to \alpha$ with $\alpha = e$ or the first symbol of α is in N and furthermore $a \in {}^{(1)}\mathrm{First}_1(\pi)\$$.

Claim 1 and Claim 2 correspond to the subcases (a) and (b) respectively in the above definition of δ.

Claim 3. For all $z \in Z$, $c \in X \cup \{e\}$, and $f \in \Gamma$ we have $|\delta(z, c, f)| \leq 1$.

If $z = (S', e)$ then $|\delta(z, c, f)| \leq 1$ obviously holds. Now assume $z = (A, a)$ with $a \in X$ and $|\delta(z, c, f)| > 1$. Then there are productions π and π' with $\pi \neq \pi'$ and $a \in {}^{(1)}(\mathrm{First}_1(\pi)\$) \cap {}^{(1)}(\mathrm{First}_1(\pi')\$)$. This is a contradiction to the strong ILL(1) property of G.

Now consider $z = (A, a)$ with $a \in X$ and $f \in \Gamma$. If $\delta(z, e, f) \neq \emptyset$ and $\delta(z, c, f) \neq \emptyset$ for a $c \in X$ then Claim 1 and Claim 2 state the existence of two productions π and π' with $\pi \neq \pi'$ and furthermore $a \in {}^{(1)}(\mathrm{First}_1(\pi)\$) \cap {}^{(1)}(\mathrm{First}_1(\pi')\$)$. But this is a contradiction to the strong ILL(1) property of G.

If $z = (S', e)$ then $\delta(z, e, f) = \emptyset$ holds. Together with Claim 3 this shows that K is a deterministic pda.

To prove $L(K) = L(G')\$$ we need

Claim 4. If $S\# \overset{n}{\Rightarrow} wA\gamma\# \overset{*}{\Rightarrow} wv$, $w, v \in T^*$, $A \in N$, according to G' then $(z_0, wv\$, \#) \overset{n+1}{\vdash} ((A, c), v', \gamma\#)$ according to K, and $cv' = v\$$ with $c \neq e$.

The claim will be proved by induction on n.

If $n = 0$ then $w = \gamma = e$ and $A = s$. According to K we have

$$((S', e), v\$, \#) \vdash ((S, c), v', \#) \quad \text{with} \quad cv' = v\$ \quad \text{and } c \neq e.$$

Assume the claim holds for all $k \leq n$.

Let $S\# \overset{n}{\Rightarrow} wA\gamma\# \overset{\pi}{\Rightarrow} waB\gamma'\# \overset{*}{\Rightarrow} wav$ be given where $a \in T \cup \{e\}$. From the induction hypothesis $(z_0, wav\$, \#) \overset{n+1}{\vdash} ((A, c'), v'', \gamma\#)$ with $c'v'' = av\$$ and $c' \neq e$ follows. If now $a \in T$ then $c' = a$ and $v'' = v\$$ holds. According to K the move $((A, a), v\$, \gamma\#) \vdash ((B, c), v', \gamma'\#)$ with $cv' = v\$$ and $c \neq e$ is possible. If $a = e$ then $c'v'' = v\$$ and $c' \in {}^{(1)}(\mathrm{First}_1(\pi)\$)$ holds. According to K the move $((A, c'), v'', \gamma\#) \vdash ((B, c'), v'', \gamma'\#)$ is possible. This completes the induction.

Now, by induction on n we will show

Claim 5. If $(z_0, wv\$, \#) \overset{n+1}{\vdash} ((A, c), v', \gamma\#)$ with $w, v \in T^*$, $cv' = v\$$ and $c \neq e$ holds according to K then $S\# \overset{n}{\Rightarrow} wA\gamma\#$ holds according to G.

If $n=0$ then we have $((S', e), wv\$, \#) \vdash ((S, c), v', \gamma\#)$ with $cv'=v\$$ according to K. This implies $w=\gamma=e$ and $S\# \overset{0}{\Rightarrow} wS\gamma\#$ holds according to G. Assume the claim holds for all $k \leq n$ and let $(z_0, wv\$, \#) \overset{n+1}{\vdash} ((A', c'), v'', \gamma'\#) \vdash$ $\vdash ((A, c), v', \gamma\#)$ with $cv'=v\$$ and $c\neq e$ be given.

If $v''=v'$ then $c=c'$ and $c'v''=v\$$ holds. From the induction hypothesis $S\# \overset{n}{\Rightarrow} wA'\gamma'\#$ follows. Since $((A', c), v', \gamma'\#) \vdash ((A, c), v', \gamma\#)$ holds, there is a production π with $wA'\gamma'\# \overset{\pi}{\Rightarrow} wA\gamma\#$.

If $v''\neq v'$ then $v''=cv'$ and $w=w'c'$. From the induction hypothesis $S\# \overset{n}{\Rightarrow}$ $\overset{n}{\Rightarrow} w'A'\gamma'\#$ follows. Since $((A', c'), v'', \gamma'\#) \vdash ((A, c), v', \gamma\#)$ holds there is a production π with $w'A'\gamma'\# \overset{\pi}{\Rightarrow} w'c'A\gamma\# = wA\gamma\#$. This completes the induction.

Now let $w \in L(K)$, i.e. $(z_0, w, \#) \overset{*}{\vdash} ((A, c), d, \gamma\#) \vdash (z_f, e, \gamma')$. Here we have either $d=\$$ or $d=e$ and $c=\$$. If $d=\$$ then $w=w'c\$$. Obviously, $w'c\in T^*$ holds. Therefore the derivation $S\# \overset{*}{\Rightarrow} w'A\gamma\#$ exists according to Claim 5. Since $((A, c), \$, \gamma\#) \vdash (z_f, e, \gamma')$ holds, there is a production π: $Af \to c$ with $w'A\gamma\# \overset{\pi}{\Rightarrow}$ $\overset{\pi}{\Rightarrow} w'c$. Therefore $w'c \in L(G)$ and $w=w'c\$ \in L(G')\$$. If $d=e$ and $c=\$$ then $w=w'\$$ with $w' \in T^*$, and the derivation $S\# \overset{*}{\Rightarrow} w'A\gamma\#$ exists according to Claim 5. Since $((A, \$), e, \gamma\#) \vdash (z_f, e, \gamma')$ holds there is a production $\pi: Af \to e$ with $w'A\gamma\# \overset{\pi}{\Rightarrow} w'$. Therefore $w' \in L(G')$ and $w=w'\$ \in L(G')\$$, and hence $L(K) \subseteq L(G)\$$.

Conversely, let $w \in L(G')$, i.e. $S\# \overset{*}{\Rightarrow} w'A\gamma\# \Rightarrow w=w'c'$, $c' \in T \cup \{e\}$. Then $(z_0, w'c'\$, \#) \overset{*}{\vdash} ((A, c), d, \gamma\#)$ with $cd=c'\$$ and $c\neq e$ holds according to Claim 4. If $c'=e$ then $c=\$$ and $d=e$ hold. Hence the move $((A, \$), e, \gamma\#) \vdash (z_f, e, \gamma')$ exists. If $c' \in T$ then $c=c'$, $d=\$$, and the move $((A, c), \$, \gamma\#) \vdash (z_f, e, \gamma')$ exists. Therefore $w\$ \in L(K)$ and hence $L(G)\$ \subseteq L(K)$.

Together with the inclusion $L(K) \subseteq L(G)\$$ this shows $L(K)=L(G)\$$.

Since K is a deterministic pda, $L(G)\$$ is a deterministic context-free language which implies that $L(G')$ is a deterministic context-free language as well (see [3], Theorem 11.2.2).

Combining Corollary 5.1 and Theorem 5.3 we can state

Theorem 5.4. The class of deterministic context-free languages is exactly the class of right linear indexed strong ILL(1) languages.

The indexed grammar for the language $\{a^nb^nc^n | n \geq 1\}$ given in [1] is a strong ILL(1) grammar. This proves

Theorem 5.5. The class of strong ILL(1) languages properly contains the class of the deterministic context-free languages.

6. A general transformation of indexed grammars

In Section 3 it was shown that the strong indexed LL(k) property is decidable. This will be used in Section 7 to prove the main result of this paper, namely the decidability of the (general) indexed LL(k) property.

To this end we first investigate in this section properties of two functions which are defined with respect to a given indexed grammar. Let $G=(N, T, I, P, S)$ be an indexed grammar and let λ: $I^* \to L$ and μ: $(NI^* \cup T)^* \to M$ be two functions in two finite nonempty sets L and M. λ and μ can be interpreted as assigning information

$\lambda(\gamma)$ and $\mu(\theta)$ from finite domains L and M to words $\gamma \in I^*$ and $\theta \in (NI^* \cup T)^*$, respectively. If λ and μ satisfy certain compatibility conditions C_λ and C_μ then it is possible to construct a grammar $G_{\lambda\mu}$ equivalent to G which simulates the derivations of G and attaches in left sentential forms $wA\gamma\theta$ of G the information $\lambda(\gamma)$ and $\mu(\theta)$ to the variable A. This means that the values of the functions λ and μ can be "computed" during leftmost derivations.

Let us now state the compatibility conditions C_λ and C_μ.

(C_λ) If $\lambda(\gamma_1)=\lambda(\gamma_2)$ then $\lambda(f\gamma_1)=\lambda(f\gamma_2)$ for all $\gamma_1, \gamma_2 \in I^*$, and $f \in I$.

(C_μ) If $\lambda(\gamma_1)=\lambda(\gamma_2)$ and $\mu(\theta_1)=\mu(\theta_2)$ then $\mu(\theta:\gamma_1\theta_1)=\mu(\theta:\gamma_2\theta_2)$ for all $\theta_1, \theta_2 \in (NI^* \cup T)^*$, $\theta \in NI^* \cup T$, and $\gamma_1, \gamma_2 \in I^*$.

Note: If C_λ is satisfied then it is easy to see that $\lambda(\gamma_1)=\lambda(\gamma_2)$ implies $\lambda(\alpha\gamma_1)=\lambda(\alpha\gamma_2)$ for all $\alpha \in I^*$. Furthermore it easy to prove that if C_μ holds then $\lambda(\gamma_1)=\lambda(\gamma_2)$ and $\mu(\theta_1)=\mu(\theta_2)$ imply $\mu(\theta:\gamma_1\theta_1)=\mu(\theta:\gamma_2\theta_2)$ for all $\theta \in (NI^* \cup T)^*$.

We will now give examples for functions λ and μ satisfying the conditions C_λ and C_μ.

Example 6.1. Let $G=(N, T, I, P, S)$ be an indexed grammar. Set $L=\mathcal{P}(N)$, the power set of N, and set $\lambda(\gamma)=\{A \mid A\gamma \overset{*}{\Rightarrow} e\}$. We will show that λ satisfies condition C_λ. For this purpose let $\gamma_1, \gamma_2 \in I^*$ with $\lambda(\gamma_1)=\lambda(\gamma_2)$, $f \in I$ and $A \in \lambda(f\gamma_1)$ be given. Since $A \in \lambda(f\gamma_1)$, there exists a derivation $Af\gamma_1 \overset{*}{\Rightarrow} e$. If $Af \overset{*}{\Rightarrow} e$, then $A \in \lambda(f\gamma_2)$ obviously holds. Otherwise there exists a derivation $Af \overset{*}{\Rightarrow} B_1 \dots B_n$ with $B_i \in N$, $i \in [1:n]$ and $B_i\gamma_1 \overset{*}{\Rightarrow} e$ for $i \in [1:n]$ holds. Therefore $B_i \in \lambda(\gamma_1)=\lambda(\gamma_2)$ and hence $B_i\gamma_2 \overset{*}{\Rightarrow} e$ for $i \in [1:n]$ holds too. Consequently we have $Af\gamma_2 \overset{*}{\Rightarrow} B_1\gamma_2 \dots B_n\gamma_2 \overset{*}{\Rightarrow} e$ and $A \in \lambda(f\gamma_2)$ follows. This shows $\lambda(f\gamma_1) \subseteq \lambda(f\gamma_2)$. The converse inclusion follows by symmetry.

Example 6.2. Let $G_c=(N, T, P, S)$ be a context-free grammar which can be interpreted as an indexed grammar $G=(N, T, I, P, S)$ with $I=\emptyset$. Let L be a finite nonempty set and let $\lambda\colon I^* \to L$ be defined by $\lambda(e)=q$ for a $q \in L$. Obviously λ satisfies condition C_λ. For a $k \geqq 1$ set $M={}^{(k)}T^*$ and define $\mu(\theta)=\{u \mid \theta \overset{*}{\Rightarrow} v,\ v \in T^*,\ u={}^{(k)}v\}=\mathrm{First}_k(\theta)$ for all $\theta \in (N \cup T)^*$. We will show that μ satisfies condition C_μ. Let $\theta_1, \theta_2 \in (N \cup T)^*$ with $\mu(\theta_1)=\mu(\theta_2)$ be given. Let $\theta \in (N \cup T)^*$ and let $v \in \mu(\theta\theta_1)$. Hence there exists a derivation $\theta\theta_1 \overset{*}{\Rightarrow} w$ with $v={}^{(k)}w$. This derivation can be written as $\theta\theta_1 \overset{*}{\Rightarrow} u_1\theta_1 \overset{*}{\Rightarrow} u_1u_2=w$. If $|u_1| \geqq k$ then $v={}^{(k)}u_1 \in \mu(\theta\theta_2)$ holds. Now assume $|u_1|<k$. We can state the existence of a derivation $\theta_2 \overset{*}{\Rightarrow} \hat{u}_2$ with ${}^{(k)}\hat{u}_2={}^{(k)}u_2$ since $\mu(\theta_1)=\mu(\theta_2)$. Hence $\theta\theta_2 \overset{*}{\Rightarrow} u_1\theta_2 \overset{*}{\Rightarrow} u_1\hat{u}_2$ and ${}^{(k)}u_1\hat{u}_2=v \in \mu(\theta\theta_2)$ holds. Therefore we have $\mu(\theta\theta_1) \subseteq \mu(\theta\theta_2)$. The converse inclusion simply holds by symmetry.

This completes the examples and we return to the general discussion.

Let $\bar{L}=L \cup \{\bar{q}\}$ where $\bar{q}$ is a new element. With the aid of λ the function $\bar{\lambda}\colon I^* \times \bar{L} \to \bar{L}$ is defined for all $\gamma \in I^*$ and $q \in \bar{L}$ by

$$\bar{\lambda}(\gamma, q) = \lambda(\gamma\gamma') \quad \text{if} \quad q=\lambda(\gamma') \quad \text{for a} \quad \gamma' \in I^*$$
$$= \bar{q} \quad \text{otherwise.}$$

If λ satisfies the condition C_λ then $\bar{\lambda}$ is well defined.

Now set $\bar{M}=M\cup\{\bar{m}\}$ where $\bar{m}$ is a new element and define the function $\bar{\mu}$: $(NI^*\cup T)^*\times\bar{M}\times\bar{L}\rightarrow\bar{M}$ for all $\theta\in(NI^*\cup T)^*$, $m\in\bar{M}$, and $q\in\bar{L}$ by

$$\begin{aligned}\bar{\mu}(\theta, m, q) &= \mu(\theta\colon\gamma\theta_1) \quad \text{if} \quad q=\lambda(\gamma) \quad \text{for a} \quad \gamma\in I^* \\ &\qquad\qquad \text{and} \quad m=\mu(\theta_1) \quad \text{for a} \quad \theta_1\in(NI^*\cup T)^* \\ &= \bar{m} \quad \text{otherwise.}\end{aligned}$$

If μ satisfies the condition C_μ then $\bar{\mu}$ is well defined. Now we can state

Lemma 6.1.

(1) If λ is effectively computable and satisfies the condition C_λ then $\lambda(I^*)$ and $\bar{\lambda}$ are effectively computable, too.

(2) If λ and μ are effectively computable and satisfy the conditions C_λ and C_μ then $\mu((NI^*\cup T)^*)$ and $\bar{\mu}$ are effectively computable, too.

Proof. (1) We will first show that the value of λ for an index word with length greater than or equal to $|L|$ can be obtained by applying λ to a word of length less than $|L|$.

Let λ be effectively computable, i.e. there is an algorithm A_λ which determines for each given $\gamma\in I^*$ the value $\lambda(\gamma)$. With the aid of A_λ determine the set $\lambda(^{(|L|-1)}I^*)$. (Recall that $^{(|L|-1)}I^*$ denotes the set of all words over I with length less than or equal to $|L|-1$.)

Let $\gamma=f_r\dots f_1\in I^*$ with $r\geqq|L|$ be given and let $q_k=\lambda(f_k\dots f_1)$ for $k\in[0\colon r]$. There exist $i,j\in[0\colon r]$ with $i<j$ and $q_i=q_j$ because $|L|=r$. Since λ satisfies the condition C_λ we have $\lambda(\gamma)=\lambda(f_r\dots f_{j+1}f_i\dots f_1)$. Hence $\lambda(I^*)=\lambda(^{(|L|-1)}I^*)$ and this implies that $\lambda(I^*)$ is effectively computable.

To show that $\bar{\lambda}$ is effectively computable let $\gamma\in I^*$ and $q\in\bar{L}$ be given. If $q\notin\lambda(I^*)$ then $\bar{\lambda}(\gamma, q)=\bar{q}$ holds. Otherwise, if $q\in\lambda(I^*)$, determine a $\gamma'\in{}^{(|L|-1)}I^*$ with $\lambda(\gamma')=q$ and compute the value $\lambda(\gamma\gamma')$ with the aid of A_λ. This completes the proof of (1).

(2) We will first show that for each $\theta\in(NI^*\cup T)^*$ there exists a $\theta'\in(NI^*\cup T)^*$ with $\mu(\theta)=\mu(\theta')$, and with the length of the index words in θ' restricted by $|L|$. In the next step we show that it suffices to compute the values of μ for words over $(NI^*\cup T)^*$ with length less than $|M|$.

Let λ be effectively computable and let λ and μ satisfy the conditions C_λ and C_μ. First we will show by induction on the length that for each $\theta\in(NI^*\cup T)^*$ there exists a $\theta'\in(N(^{(|L|-1)}I^*)\cup T)^*$ with $\mu(\theta)=\mu(\theta')$. In case $\theta=e$ the assertion is trivial. Let $\theta=\theta_1\theta_2$ with $\theta_1\in NI^*\cup T$ and $\theta_2\in(NI^*\cup T)^*$. The induction hypothesis guarantees the existence of a $\theta_2'\in(N(^{(|L|-1)}I^*)\cup T)^*$ with $\mu(\theta_2)=\mu(\theta_2')$.

If $\theta_1=a\in T$ then $\mu(\theta)=\mu(a\theta_2)=\mu(a\theta_2')=\mu(\theta')$ with $\theta'\in(N(^{(|L|-1)}I^*)\cup T)^*$ since μ satisfies condition C_μ.

Now let $\theta_1=A\gamma\in NI^*$. Part (1) of this lemma guarantees the existence of a $\gamma'\in{}^{(|L|-1)}I^*$ with $\lambda(\gamma)=\lambda(\gamma')$. Since μ satisfies the condition C_μ, we have $\mu(\theta)=$ $=\mu(A\colon\gamma\theta_2)=\mu(A\colon\gamma'\theta_2')=\mu(\theta')$ with $\theta'\in(N(^{(|L|-1)}I^*)\cup T)^*$.

Now let μ be effectively computable, i.e. there is an algorithm A_μ which determines for each $\theta\in(NI^*\cup T)^*$ the value $\mu(\theta)$. With the aid of A_μ determine the set $\mu(^{(|M|-1)}(N(^{(|L|-1)}I^*)\cup T)^*)$. We will show that this set equals $\mu((NI^*\cup T)^*)$.

Let a $\theta\in(NI^*\cup T)^*$ be given. We have proved above the existence of a $\theta'\in(N(^{(|L|-1)}I^*)\cup T)^*$ with $\mu(\theta)=\mu(\theta')$. Let $\theta'=\theta_r\ldots\theta_1$ with $\theta_k\in NI^*\cup T$ for $k\in[1: r]$ and $r\geqq|M|$. If we set $m_k=\mu(\theta_k\ldots\theta_1)$ for $k\in[0: r]$ then there exist $i,j\in[0: r]$ with $i<j$ and $m_i=m_j$. Since μ satisfies the condition C_μ, we have $\mu(\theta')=\mu(\theta_r\ldots\theta_{j+1}\theta_i\ldots\theta_1)$.

This implies $\mu((NI^*\cup T)^*)=\mu((^{|M|-1)}N(^{(|L|-1)}I^*)\cup T)^*)$ and therefore $\mu((NI^*\cup T)^*)$ is effectively computable.

It remains to be shown that $\bar{\mu}$ is effectively computable. Let $\theta\in(NI^*\cup T)^*$, $m\in\overline{M}$, and $q\in\overline{L}$ be given. If $m\in\mu((NI^*\cup T)^*)$ and $q\in\lambda(I^*)$ then determine a θ_1 with $\mu(\theta_1)=m$ and a γ with $\lambda(\gamma)=q$ and compute the value $\mu(\theta:\gamma\theta_1)=\bar{\mu}(\theta, m, q)$ with the aid of A_μ. Otherwise, if $m=\bar{m}$ or $q=\bar{q}$, we have $\bar{\mu}(\theta, m, q)=\bar{m}$. This completes the proof of the lemma.

Starting with an indexed grammar $G=(N, T, I, P, S)$ and functions λ and μ satisfying C_λ and C_μ, we will construct an indexed grammar $G_{\lambda\mu}=(N', T, I', P', S')$. This grammar is structurally equivalent to G, i.e. generates the same set of terminal strings and the same set of derivation trees (ignoring the labels of the intermediate nodes).

Define $N'=N\times\overline{M}\times\overline{L}$, $I'=I\times\overline{L}$ and $S'=(S, m_0, q_0)$ with $m_0=\mu(e)$ and $q_0=\lambda(e)$.

For the definition of P' we need two functions φ and ψ. The function φ: $I^*\times\overline{L}\to(I\times\overline{L})^*=I'^*$ attaches a second component to each index f_i in an index word $f_1\ldots f_n$. For a given $q\in\overline{L}$ the second component of f_i will be value $\bar{\lambda}(f_{i+1}\ldots\ldots f_n, q)$, i.e. φ is defined by

$$\varphi(e, q)=e, \quad\text{and}$$

$$\varphi(f\gamma, q)=(f, \bar{\lambda}(\gamma, q))\varphi(\gamma, q) \quad\text{for all } \gamma\in I^*,\ f\in I, \text{ and } q\in\overline{L}.$$

The function ψ: $(NI^*\cup T)^*\times\overline{M}\times\overline{L}\to((N\times\overline{M}\times\overline{L})(I\times\overline{L})^*\cup T)^*=(N'I'^*\cup T)^*$ attaches two components to the variables A_i in a word $A_1\gamma_1A_2\gamma_2\ldots A_n\gamma_n$ of $(NI^*\cup\cup T)^*$ with $A_i\in N\cup T, \gamma_i\in I^*$ for $i\in[1: n]$. For a given $m\in\overline{M}$ and $q\in\overline{L}$ the values of these components will be $\bar{\mu}(A_{i+1}\gamma_{i+1}\ldots A_n\gamma_n, m, q)$ and $\bar{\lambda}(\gamma_i, q)$ respectively. Furthermore the γ_j will be replaced by $\varphi(\gamma_j, q)$ for $j\in[1: n]$, i.e. ψ is defined by

$$\psi(e, m, q)=e,$$

$$\psi(a\theta, m, q)=a\psi(\theta, m, q) \quad\text{and}$$

$$\psi(A\gamma\theta, m, q)=(A, \bar{\mu}(\theta, m, q), \bar{\lambda}(\gamma, q))\varphi(\gamma, q)\psi(\theta, m, q),$$

for all $A\in N$, $\gamma\in I^*$, $\theta\in(NI^*\cup T)^*$, $m\in\overline{M}$, $q\in\overline{L}$, and $a\in T$.

Now we are able to define the productions of $G_{\lambda\mu}$. Let π: $Af\to\beta$ be a production of P. Then for all $m\in\mu((NI^*\cup T)^*)$ and $q\in\lambda(I^*)$ the production

$$\pi^{m,q}\colon\ \psi(Af, m, q)\to\psi(\beta, m, q) \quad\text{is in } P'.$$

Note that $\psi(Af, m, q)=(A, m, \bar{\lambda}(f, q))\varphi(f, q)\in N'\times(I'\cup\{e\})$, since $\varphi(f, q)=e$ if $f=e$, or $\varphi(f, q)=(f, q)$ if $f\neq e$.

$G_{\lambda\mu}$ is called the *$\lambda\mu$-grammar of G*. If the functions λ and μ are effectively computable, then the function ψ is also effectively computable and the grammar $G_{\lambda\mu}$ can be constructed effectively.

The homomorphism $\delta\colon (N'\cup I'\cup T)^* \to (N\cup I\cup T)^*$ defined by $\delta(A, m, q)=A$, $\delta(f, q)=f$, and $\delta(a)=a$ for all $A\in N, m\in\bar{M}$, $q\in\bar{L}, f\in I, a\in T$, deletes the components attached to variables $A\in N$ and indices $f\in I$. Obviously, if $\theta_1' \,{}^{\pi^{m,q}}\!\!\Rightarrow \theta_2'$ according to $G_{\lambda\mu}$, where $\theta_1', \theta_2' \in (N'I'^*\cup T)^*$, then $\delta(\theta_1')\,{}^{\pi}\!\!\Rightarrow \delta(\theta_2')$ holds according to G.

Furthermore for all $\theta\in(NI^*\cup T)^*$, $m\in\bar{M}$, $q\in\bar{L}$, we have $\delta(\psi(\theta, m, q))=\theta$.

If $S'=(S, m_0, q_0)\,{}^{\pi_1^{m_1,q_1}}\!\!\Rightarrow \theta_1' \ldots {}^{\pi_n^{m_n,q_n}}\!\!\Rightarrow \theta_n'$ is a leftmost derivation according to $G_{\lambda\mu}$, then

$$S\,{}^{\pi_1}\!\!\Rightarrow \delta(\theta_1') \Rightarrow \ldots {}^{\pi_n}\!\!\Rightarrow \delta(\theta_n')$$

is called the *corresponding leftmost derivation according to G.*

In the following two lemmas we will make precise the structural equivalence of the grammars G and $G_{\lambda\mu}$. To prove these lemmas we need a number of facts concerning the functions $\bar{\lambda}$, $\bar{\mu}$, φ, and ψ.

Claim 1. For all $\theta, \theta_1, \theta_2 \in (NI^*\cup T)^*$, and $\gamma, \gamma_1\in I^*$ we have

(i) $$\bar{\lambda}(e, \lambda(\gamma)) = \lambda(\gamma),$$

(ii) $$\bar{\lambda}(\gamma, q_0) = \lambda(\gamma) \quad \text{where} \quad q_0 = \lambda(e),$$

(iii) $$\bar{\lambda}(\gamma_1, \bar{\lambda}(\gamma, q)) = \bar{\lambda}(\gamma_1\gamma, q) \quad \text{for all} \quad q\in\bar{L},$$

(iv) $$\bar{\mu}(e, \mu(\theta), \lambda(\gamma)) = \mu(\theta),$$

(v) $$\bar{\mu}(\theta, m_0, q_0) = \mu(\theta) \quad \text{where} \quad m_0=\mu(e) \quad \text{and} \quad q_0=\lambda(e),$$

(vi) $$\bar{\mu}(\theta_1\theta, m, q) = \bar{\mu}(\theta_1, \bar{\mu}(\theta, m, q), q) \quad \text{for all} \quad m\in\bar{M} \quad \text{and} \quad q\in\bar{L},$$

(vii) $$\bar{\mu}(\theta{:}\gamma, m, q) = \bar{\mu}(\theta, m, \bar{\lambda}(\gamma, q)) \quad \text{for all} \quad m\in\bar{M} \quad \text{and} \quad q\in\bar{L}.$$

These identities are easily obtained from the definitions of the functions $\bar{\lambda}$ and $\bar{\mu}$.

The following three claims state properties of the functions φ and ψ. All claims are proved by induction on the length of words over I and $NI^*\cup T$ respectively.

Claim 2. For all $\gamma, \gamma_1\in I^*$ and $q\in\bar{L}$ we have

$$\varphi(\gamma\gamma_1, q) = \varphi(\gamma, \bar{\lambda}(\gamma_1, q))\varphi(\gamma_1, q).$$

Proof. The assertion holds for $\gamma=e$. If $\gamma=f\gamma'$ with $f\in I$ then

$$\varphi(f\gamma'\gamma_1, q) = (f, \bar{\lambda}(\gamma'\gamma_1, q))\varphi(\gamma'\gamma_1, q)$$
$$= (f, \bar{\lambda}(\gamma', \bar{\lambda}(\gamma_1, q)))\varphi(\gamma', \bar{\lambda}(\gamma_1, q))\varphi(\gamma_1, q)$$

(see Claim 1 (iii) and induction hypothesis)

$$= \varphi(f\gamma', \bar{\lambda}(\gamma_1, q))\varphi(\gamma_1, q)$$

(see definition of φ).

Claim 3. For all $\theta, \theta_1\in(NI^*\cup T)^*$, $m\in\bar{M}$, $q\in\bar{L}$ we have

$$\psi(\theta_1\theta, m, q) = \psi(\theta_1, \bar{\mu}(\theta, m, q), q)\psi(\theta, m, q).$$

Proof. The assertion holds for $\theta_1=e$. Assume that $\theta_1=a\theta_1'$ with $a\in T$ and that the assertion holds for $\theta_1'\theta$. Then, by the definition of ψ we can conclude

$$\begin{aligned}\psi(a\theta_1'\theta, m, q) &= a\psi(\theta_1'\theta, m, q)\\ &= a\psi(\theta_1', \bar{\mu}(\theta, m, q), q)\psi(\theta, m, q)\\ &= \psi(a\theta_1', \bar{\mu}(\theta, m, q), q)\psi(\theta, m, q)\end{aligned}$$

Now assume $\theta_1=A\gamma\theta_1'$ with $A\gamma\in NI^*$ and $\theta_1'\in(NI^*\cup T)^*$, and assume the assertion holds for $\theta_1'\theta$. Then, by the definition of ψ and Claim 1 (vi) we can conclude

$$\begin{aligned}\psi(A\gamma\theta_1'\theta, m, q) &= (A, \bar{\mu}(\theta_1'\theta, m, q), \bar{\lambda}(\gamma, q))\varphi(\gamma, q)\psi(\theta_1'\theta, m, q)\\ &= (A, \bar{\mu}(\theta_1'\theta, m, q), \bar{\lambda}(\gamma, q))\varphi(\gamma, q)\psi(\theta_1', \bar{\mu}(\theta, m, q), q)\psi(\theta, m, q)\\ &= (A, \bar{\mu}(\theta_1', \bar{\mu}(\theta, m, q), q), \bar{\lambda}(\gamma, q))\varphi(\gamma, q)\psi(\theta_1', \bar{\mu}(\theta, m, q), q)\psi(\theta, m, q)\\ &= \psi(A\gamma\theta_1', \bar{\mu}(\theta, m, q), q)\psi(\theta, m, q).\end{aligned}$$

Claim 4. For all $\theta\in(NI^*\cup T)^*$, $\gamma\in I^*$, $m\in\bar{M}$, and $q\in\bar{L}$ we have

$$\psi(\theta\colon \gamma, m, q) = \psi(\theta, m, \bar{\lambda}(\gamma, q))\colon\varphi(\gamma, q).$$

Proof. The assertion holds for $\theta=e$. Assume $\theta=a\theta_1$ with $a\in T$ and the assertion holds for θ_1. From the definition of ψ it follows:

$$\begin{aligned}\psi((a\theta_1)\colon\gamma, m, q) &= \psi(a(\theta_1\colon \gamma), m, q) = a\psi(\theta_1\colon \gamma, m, q)\\ &= a(\psi(\theta_1, m, \bar{\lambda}(\gamma, q))\colon \varphi(\gamma, q))=\\ &= (a\psi(\theta_1, m, \bar{\lambda}(\gamma, q)))\colon \varphi(\gamma, q) =\\ &= \psi(a\theta_1, m, \bar{\lambda}(\gamma, q))\colon \varphi(\gamma, q)\end{aligned}$$

Now assume $\theta=A\gamma_1\theta_1$ with $A\gamma_1\in NI^*$, $\theta_1\in(NI^*\cup T)^*$ and assume that the assertion holds for θ_1. Then, by the definition of ψ, Claim 1 (iii), Claim 1 (vii), and Claim 2 we have

$$\begin{aligned}&\psi((A\gamma_1\theta_1)\colon\gamma, m, q) = \psi(A\gamma_1\gamma(\theta_1\colon \gamma), m, q)\\ &= (A, \bar{\mu}(\theta_1\colon \gamma, m, q), \bar{\lambda}(\gamma_1\gamma, q))\varphi(\gamma_1\gamma, q)\psi(\theta_1\colon \gamma, m, q)\\ &= (A, \bar{\mu}(\theta_1, m, \bar{\lambda}(\gamma, q)), \bar{\lambda}(\gamma_1\gamma, q))\varphi(\gamma_1, \bar{\lambda}(\gamma, q))\varphi(\gamma, q)\,\psi(\theta_1, m, \bar{\lambda}(\gamma, q))\colon \varphi(\gamma, q)\\ &= ((A, \bar{\mu}(\theta_1, m, \bar{\lambda}(\gamma, q)), \bar{\lambda}(\gamma_1, \bar{\lambda}(\gamma, q)))\varphi(\gamma_1, \bar{\lambda}(\gamma, q))\psi(\theta_1, m, \bar{\lambda}(\gamma, q)))\colon \varphi(\gamma, q)\\ &= \psi(A\gamma_1\theta_1, m, \bar{\lambda}(\gamma, q))\colon \varphi(\gamma, q).\end{aligned}$$

For the remainder of this section we will use the following general assumptions:

(*) Let $G=(N, T, I, P, S)$ be an indexed grammar and let $\lambda\colon I^*\to L$ and $\mu\colon (NI^*\cup T)^*\to M$ be two functions in two finite sets L and M satisfying the conditions C_λ and C_μ. Let $\bar{L}=L\cup\{\bar{q}\}$ and $\bar{M}=M\cup\{\bar{m}\}$ where $\bar{q}$ and $\bar{m}$ are new elements and let $G_{\lambda\mu}=(N', T, I', P', S')$ be the $\lambda\mu$-grammar of G.

The next lemma establishes a correspondence between a leftmost derivation step in G and an analogous leftmost derivation step in $G_{\lambda\mu}$.

Lemma 6.2. Under the assumptions (*) for all $w\in T^*$, $\gamma\in I^*$, $\theta\in(NI^*\cup T)^*$ the following holds:

If π: $Af\to\beta\in P$ and $wAf\gamma\theta \overset{\pi}{\Rightarrow} w\beta:\gamma\theta$ according to G, then for all $m_1\in\mu((NI^*\cup T)^*)$ and $q_1\in\lambda(I^*)$ we have $\psi(wAf\gamma\theta, m_1, q_1) \overset{\pi^{m,q}}{\Rightarrow} \psi(w\beta:\gamma\theta, m_1, q_1)$, where $m=\bar{\mu}(\theta, m_1, q_1)$ and $q=\bar{\lambda}(\gamma, q_1)$.

Proof. Since $m_1\in\mu((NI^*\cup T)^*)$ and $q_1\in\lambda(I^*)$, the same holds for m and q, hence $\pi^{m,q}\in P'$. Consider $\pi^{m,q}$: $(A, m, \bar{\lambda}(f, q))\varphi(f, q)\to\psi(\beta, m, q)\in P'$. By the definitions of φ and ψ and Claim 1 (iii) we have

$$\psi(wAf\gamma\theta, m_1, q_1) = w(A, m, \bar{\lambda}(f\gamma, q_1))\varphi(f\gamma, q_1)\psi(\theta, m_1, q_1) =$$
$$= w(A, m, \bar{\lambda}(f, q))\varphi(f, q)\varphi(\gamma, q_1)\psi(\theta, m_1, q_1)$$

This shows that $\pi^{m,q}$ is applicable to $\psi(wAf\gamma\theta, m_1, q_1)$, i.e. the following leftmost derivation step is possible:

$$w(A, m, \bar{\lambda}(f, q))\varphi(f, q)\varphi(\gamma, q_1)\psi(\theta, m_1, q_1)$$
$$\overset{\pi^{m,q}}{\Rightarrow} w\psi(\beta, m, q):\ \varphi(\gamma, q_1)\psi(\theta, m_1, q_1).$$

Using the Claims 3 and 4 we have

$$w\psi(\beta, m, q):\varphi(\gamma, q_1)\psi(\theta, m_1, q_1) = w\psi(\beta:\gamma, m, q_1)\psi(\theta, m_1, q_1)$$
$$= \psi(w\theta_1\theta, m_1, q_1).$$

This completes the proof.

The following lemma is in some sense the converse of Lemma 6.2 and describes the simulation of a leftmost derivation step according to $G_{\lambda\mu}$ in the grammar G.

Lemma 6.3. Under the asumptions (*) the following holds for all $w\in T^*$, $\gamma\in I^*$, and $\theta\in(NI^*\cup T)^*$: If $\pi^{m,q}$: $\psi(Af, m, q)\to\psi(\beta, m, q)\in P'$ and $\psi(wAf\gamma\theta, m_1, q_1) \overset{\pi^{m,q}}{\Rightarrow} \bar{\theta}$, where $m_1\in\mu((NI^*\cup T)^*)$ and $q_1\in\lambda(I^*)$, then $m=\bar{\mu}(\theta, m_1, q_1)$, $q=\bar{\lambda}(\gamma, q_1)$, and $\bar{\theta}=\psi(w\beta:\gamma\theta, m_1, q_1)$.

Proof. Since $\psi(wAf\gamma\theta, m_1, q_1)=$

$$w(A, \bar{\mu}(\theta, m_1, q_1), \bar{\lambda}(f\gamma, q_1))\varphi(f\gamma, q_1)\psi(\theta, m_1, q_1) =$$
$$w(A, \bar{\mu}(\theta, m_1, q_1), \bar{\lambda}(f, \bar{\lambda}(\gamma, q_1)))\varphi(f, \bar{\lambda}(\gamma, q_1))\varphi(\gamma, q_1)\psi(\theta, m_1, q_1)$$

(see Claim 1 (iii)) we have $m=\bar{\mu}(\theta, m_1, q_1)$ and $q=\bar{\lambda}(\gamma, q_1)$.

Furthermore, with the aid of Claims 3 and 4 we have:

$$\bar{\theta} = w\psi(\beta, m, q):\ \varphi(\gamma, q_1)\psi(\theta, m_1, q_1) = w\psi(\beta:\gamma, m, q_1)\psi(\theta, m_1, q_1)$$
$$= \psi(w\theta_1\theta, m_1, q_1)$$

This completes the proof.

The repeated application of Lemma 6.2 yields

Corollary 6.1. Under the assumptions (*) we have: If $S \overset{*}{\Rightarrow} wA\gamma\theta$ acccording to G with $w\in T^*$, $A\in N$, $\gamma\in I^*$, and $\theta\in(NI^*\cup T)^*$ then $S'=(S, m_0, q_0) \overset{*}{\Rightarrow} \psi(wA\gamma\theta, m_0, q_0)$ according to $G_{\lambda\mu}$.

Remark. Since

$$\psi(wA\gamma\theta, m_0, q_0) = w(A, \bar{\mu}(\theta, m_0, q_0), \bar{\lambda}(\gamma, q_0))\varphi(\gamma, q_0)\psi(\theta, m_0, q_0) = \\ = w(A, \mu(\theta), \lambda(\gamma))\varphi(\gamma, q_0)\psi(\theta, m_0, q_0)$$

holds, we conclude: $G_{\lambda\mu}$ simulates the leftmost derivations of left sentential forms $wA\gamma\theta$ of G and attaches the information $\lambda(\gamma)$ and $\mu(\theta)$ to the variable A.

Repeated application of Lemma 6.3 yields

Corollary 6.2. Let $S'=(S, m_0, q_0) \overset{*}{\Rightarrow} \theta'$ be a leftmost derivation according to $G_{\lambda\mu}$ and let $S \overset{*}{\Rightarrow} \theta$ be the corresponding leftmost derivation according to G. Then $\theta'=\psi(\theta, m_0, q_0)$ holds.

Furthermore, with Lemma 6.2 we obtain

Corollary 6.3. Under the assumptions (*) we have: If $\theta\in(NI^*\cup T)^*$, $m\in\mu((NI^*\cup \cup T)^*)$, $q\in\lambda(I^*)$, and $v\in T^*$ then $\psi(\theta, m, q) \overset{*}{\Rightarrow} v$ according to $G_{\lambda\mu}$ iff $\theta \overset{*}{\Rightarrow} v$ according to G. In particular $L(G)=L(G_{\lambda\mu})$ holds.

Remark. The underlying principle of this construction is applied for example in [1, 9]. In [1] the function λ of Example 6.1 is used for constructing a normal form of an indexed grammar. In [9] the function μ of Example 6.2 is used in investigations of context-free LL(k) grammars.

A construction similar to the construction of $G_{\lambda\mu}$ can be applied to indexed pushdown automata and pushdown automata. The principles of this construction are used in [7, 8] in the indexed case and, for example, to prove closure properties of deterministic languages in the context-free case (cf. [3], Section 11.2).

7. Decidability of indexed LL(k)

In this section we will prove our main theorem concerning the connection between ILL(k) and strong ILL(k) grammars. For this purpose we will introduce two special functions λ and μ in the following manner.

Let $G=(N, T, I, P, S)$ be an indexed grammar with $P=\{\pi_1, \pi_2, \ldots, \pi_p\}$. Set $L=(\mathscr{P}({}^{(k)}T^*))^p$, the set of all p-vectors, whose components are subsets of ${}^{(k)}T^*$, and $M=\mathscr{P}({}^{(k)}T^*)$ with $k\geq 1$.

Define $\lambda\colon I^*\to L$ by $\lambda(\gamma)=(q_1, \ldots, q_p)$ where $q_i=\text{First}_k(\pi_i, \gamma)$ holds for $i\in[1:p]$. Furthermore define $\mu\colon (NI^*\cup T)^*\to M$ by $\mu(\theta)=\text{First}_k(\theta)$. From Corollary 3.1 we know that λ and μ are effectively computable.

First we want to show that λ satisfies condition C_λ. For this purpose let $\gamma_1, \gamma_2\in I^*$ with $\lambda(\gamma_1)=\lambda(\gamma_2)$ be given, i.e. $\text{First}_k(\pi_i, \gamma_1)=\text{First}_k(\pi_i, \gamma_2)$ holds for all $i\in[1:p]$. Under this assumption we will show by induction on n that for each $\gamma\in I^*$ and each production π_j the following holds:

If

$$A\gamma\gamma_1 \overset{\pi_j}{\Rightarrow} \theta_1 \overset{n}{\Rightarrow} u \quad \text{with} \quad u\in T^* \quad \text{then}$$

$$A\gamma\gamma_2 \overset{\pi_j}{\Rightarrow} \theta_2 \overset{*}{\Rightarrow} v \quad \text{with} \quad v\in T^* \quad \text{and} \quad {}^{(k)}u = {}^{(k)}v.$$

(Recall, $\overset{n}{\Rightarrow}$ denotes a leftmost derivation in n steps.)

Obviously this assertion holds for $n=0$. Now let $A\gamma\gamma_1 \overset{\pi_j}{\Rightarrow} \theta_1 \overset{n+1}{\Rightarrow} u$ with $u\in T^*$ be given.

If $\gamma=e$, the assertion follows from $\lambda(\gamma_1)=\lambda(\gamma_2)$.

If $\gamma\neq e$, no index of γ_1 can be consumed in the first step of the derivation. Therefore we have

$$\theta_1 = B_1\beta_1 \dots B_r\beta_r \quad \text{with} \quad B_i\in N\cup T, r>0, \quad \text{and}$$

$$\beta_i = \gamma_i'\gamma_1 \quad \text{with} \quad \gamma_i'\in I^* \quad \text{if} \quad B_i\in N$$

$$= e \quad \text{otherwise.}$$

In addition $B_i\beta_i \overset{n_i}{\Rightarrow} u_i$ with $n_i\leqq n+1$ holds for $i\in[1:r]$ where $u=u_1\dots u_r$. Since $\gamma\neq e$ it is possible to apply π_j to $A\gamma\gamma_2$ which yields

$$A\gamma\gamma_2 \overset{\pi_j}{\Rightarrow} B_1\beta_1' \dots B_r\beta_r' = \theta_2 \quad \text{with}$$

$$\beta_i' = \gamma_i'\gamma_2 \quad \text{with} \quad \gamma_i'\in I^* \quad \text{if} \quad B_i\in N$$

$$= e \text{ otherwise.}$$

The induction hypothesis quarantees the existence of the derivations

$$B_i\beta_i' \overset{*}{\Rightarrow} v_i \quad \text{with} \quad v_i\in T^* \quad \text{and} \quad {}^{(k)}v_i = {}^{(k)}u_i \quad \text{for} \quad i\in[1:r].$$

Consequently we have

$$A\gamma\gamma_2 \overset{\pi_j}{\Rightarrow} \theta_2 \overset{*}{\Rightarrow} v_1 \dots v_r = v \quad \text{with} \quad {}^{(k)}v = {}^{(k)}u.$$

This completes the induction, and hence

$$\text{First}_k(\pi_j, \gamma\gamma_1) \subseteqq \text{First}_k(\pi_j, \gamma\gamma_2).$$

The converse inclusion holds by symmetry. The special case $\gamma\in I$ shows that λ satisfies condition C_λ.

We now prove that μ satisfies condition C_μ.

Let $\gamma_1, \gamma_2\in I^*$ with $\lambda(\gamma_1)=\lambda(\gamma_2)$ and $\theta_1, \theta_2\in(NI^*\cup T)^*$ with $\mu(\theta_1)=\mu(\theta_2)$ be given, i.e. for all productions π_i we have $\text{First}_k(\pi_i, \gamma_1)=\text{First}_k(\pi_i, \gamma_2)$, and furthermore $\text{First}_k(\theta_1)=\text{First}_k(\theta_2)$ holds.

Now for each $\theta\in NI^*\cup T$ the equality $\text{First}_k(\theta:\gamma_1\theta_1)=\text{First}_k(\theta:\gamma_2\theta_2)$ has to be shown.

If $\theta\in T$ this assertion holds obviously.

If $\theta=A\gamma\in NI^*$ then observe that $\text{First}_k(A\gamma\gamma_1)=\bigcup_{i\in J}\text{First}_k(\pi_i, \gamma\gamma_1)$ holds where J is the set of all numbers of productions with lefthand side Af, $f\in I\cup\{e\}$. Since $\lambda(\gamma_1)=\lambda(\gamma_2)$ and λ satisfies condition C_λ, we have $\lambda(\gamma\gamma_1)=\lambda(\gamma\gamma_2)$. Hence we have

$$\text{First}_k(A\gamma\gamma_1) = \bigcup_{i\in J}\text{First}_k(\pi_i, \gamma\gamma_2) = \text{First}_k(A\gamma\gamma_2).$$

This equality enables us to conclude

$$\begin{aligned}\mathrm{First}_k(A\gamma\gamma_1\theta_1) &= {}^{(k)}(\mathrm{First}_k(A\gamma\gamma_1)\,\mathrm{First}_k(\theta_1))\\ &= {}^{(k)}(\mathrm{First}_k(A\gamma\gamma_2)\,\mathrm{First}_k(\theta_2))\\ &= \mathrm{First}_k(A\gamma\gamma_2\theta_2).\end{aligned}$$

Consequently μ satisfies the condition C_μ.

According to Lemma 6.1 we observe that $\lambda(I^*)$, $\bar{\lambda}$, $\mu((NI^*\cup T)^*)$, and $\bar{\mu}$ are effectively computable.

The $\lambda\mu$-grammar of G defined by these special functions λ and μ will be called the S_k*-grammar* of G and denoted by $S_k(G)$.

Now we can state our main theorem.

Theorem 7.1. Let $G=(N, T, I, P, S)$ be an indexed grammar and let $k\geqq 1$. $S_k(G)$, the S_k-grammar of G, is a strong ILL(k) grammar iff G is an ILL(k) grammar.

Proof. (a) Let G be an ILL(k) grammar. If $S_k(G)$ does not satisfy the strong ILL(k) condition, then the following two cases are possible.

(1) There are two productions $\pi_i^{m,q}$: $\psi(Af, m, q)\to\psi(\beta_1, m, q)$ and $\pi_j^{m,q}$: $\psi(Af, m, q)\to\psi(\beta_2, m, q)$ in P' with the same lefthand side, where $i\neq j$, $f\in I\cup\{e\}$, and $\mathrm{First}_k(\pi_i^{m,q})\cap\mathrm{First}_k(\pi_j^{m,q})\neq\emptyset$, i.e. there are two leftmost derivations

$$S' \overset{*}{\Rightarrow} w\psi(Af, m, q)\gamma'\theta' \overset{\pi_i^{m,q}}{\Rightarrow} w\theta_1'\theta' \overset{*}{\Rightarrow} wu\theta' \overset{*}{\Rightarrow} wuv \tag{7.1}$$

and

$$S' \overset{*}{\Rightarrow} \bar{w}\psi(Af, m, q)\bar{\gamma}'\bar{\theta}' \overset{\pi_j^{m,q}}{\Rightarrow} \bar{w}\bar{\theta}_1'\bar{\theta}' \overset{*}{\Rightarrow} \bar{w}\bar{u}\bar{\theta}' \Rightarrow \bar{w}\bar{u}\bar{v} \tag{7.2}$$

according to $S_k(G)$ with $w, \bar{w}, u, \bar{u}, v, \bar{v}\in T^*$, $\gamma', \bar{\gamma}'\in I'^*$, and

$$\theta', \theta_1', \bar{\theta}', \bar{\theta}_1'\in(N'I'^*\cup T)^*, \quad \text{where}$$

$$\theta_1' \overset{*}{\Rightarrow} u, \quad \theta' \overset{*}{\Rightarrow} v,$$

$$\bar{\theta}_1' \overset{*}{\Rightarrow} \bar{u}, \quad \bar{\theta}' \overset{*}{\Rightarrow} \bar{v},$$

and ${}^{(k)}uv={}^{(k)}\bar{u}\bar{v}$ holds.

Consider the derivation (7.1). We have the corresponding leftmost derivation

$$S \overset{*}{\Rightarrow} wAf\gamma\theta \overset{\pi_i}{\Rightarrow} w\theta_1\theta \overset{*}{\Rightarrow} wu\theta \overset{*}{\Rightarrow} wuv,$$

where $\gamma=\delta(\gamma')$, $\theta_1=\delta(\theta_1')$, and $\theta=\delta(\theta')$.

Then from Corollary 6.2 we conclude $w\psi(Af, m, q)\gamma'\theta'=\psi(wAf\gamma\theta, m_0, q_0)$. Since $\psi(wAf\gamma\theta, m_0, q_0)\overset{\pi_i^{m,q}}{\Rightarrow}w\theta_1'\theta'$ we have from Lemma 6.3 $m=\bar{\mu}(\theta, m_0, q_0)= =\mu(\theta)$, and $q=\bar{\lambda}(\gamma, q_0)=\lambda(\gamma)$.

Analogously, considering the derivation (7.2), we have a corresponding leftmost derivation

$$S \overset{*}{\Rightarrow} \bar{w}Af\bar{\gamma}\bar{\theta} \overset{\pi_j}{\Rightarrow} \bar{w}\bar{\theta}_1\bar{\theta} \overset{*}{\Rightarrow} \bar{w}\bar{u}\bar{\theta} \overset{*}{\Rightarrow} \bar{w}\bar{u}\bar{v},$$

where $\bar{\gamma}=\delta(\bar{\gamma}')$, $\bar{\theta}_1=\delta(\bar{\theta}_1')$, and $\bar{\theta}=\delta(\bar{\theta}')$. Furthermore we have $m=\mu(\bar{\theta})$ and $q=\lambda(\bar{\gamma})$.

Since $\lambda(\gamma)=\lambda(\bar{\gamma})$ we have in particular $\mathrm{First}_k(\pi_j, f\gamma)=\mathrm{First}_k(\pi_j, f\bar{\gamma})$ and since $\mu(\theta)=\mu(\bar{\theta})$ we have $\mathrm{First}_k(\theta)=\mathrm{First}_k(\bar{\theta})$.

If $|\bar{u}| \geqq k$ then ${}^{(k)}\bar{u} \in \mathrm{First}_k(\pi_j, f\bar{\gamma})$ and therefore ${}^{(k)}\bar{u} \in \mathrm{First}_k(\pi_j, f\gamma)$, i.e. $Af\gamma \overset{\pi_j}{\Rightarrow} \beta_2 : \gamma \overset{*}{\Rightarrow} \hat{u}$ according to G with ${}^{(k)}\hat{u} = {}^{(k)}\bar{u}$.

Then the following derivation according to G is possible:

$$S \overset{*}{\Rightarrow} wAf\gamma\theta \overset{\pi_j}{\Rightarrow} w\beta_2 : \gamma\theta \overset{*}{\Rightarrow} w\hat{u}v$$

with ${}^{(k)}\hat{u}v = {}^{(k)}\bar{u} = {}^{(k)}uv$. But this is a contradiction to the ILL(k) property of G.

If $|\bar{u}| < k$, then $\bar{u} \in \mathrm{First}_k(\pi_j, f\bar{\gamma})$ and therefore $\bar{u} \in \mathrm{First}_k(\pi_j, f\gamma)$, i.e. $Af\gamma \overset{\pi_j}{\Rightarrow} \overset{\pi_j}{\Rightarrow} \beta_2 : \gamma \overset{*}{\Rightarrow} \bar{u}$ according to G. Since $\mathrm{First}_k(\theta) = \mathrm{First}_k(\bar{\theta})$ there exists $\theta \overset{*}{\Rightarrow} \hat{v}$ according to G with ${}^{(k)}\hat{v} = {}^{(k)}\bar{v}$. Then the following derivation according to G is possible:

$$S \overset{*}{\Rightarrow} wAf\gamma\theta \overset{\pi_j}{\Rightarrow} w\beta_2 : \gamma\theta \overset{*}{\Rightarrow} w\bar{u}\theta \overset{*}{\Rightarrow} w\bar{u}\hat{v}$$

with ${}^{(k)}\bar{u}\hat{v} = {}^{(k)}\bar{u}\bar{v} = {}^{(k)}uv$. This is a contradiction to the ILL(k) property of G.

(2) There are productions $\pi_i^{m,q}$: $\psi(Af, m, q) \to \psi(\beta_1, m, q)$, $f \in I$, and $\pi_j^{m,q'}$: $\psi(A, m, q') \to \psi(\beta_2, m, q')$ in P' with $q' = \bar{\lambda}(f, q)$ and

$$S' \overset{*}{\Rightarrow} w\psi(Af, m, q)\gamma'\theta' \overset{\pi_i^{m,q}}{\Rightarrow} w\theta_1'\theta' \overset{*}{\Rightarrow} wu\theta' \overset{*}{\Rightarrow} wuv \tag{7.3}$$

and

$$S' \overset{*}{\Rightarrow} \bar{w}\psi(A, m, q')\bar{\gamma}'\bar{\theta}' \overset{\pi_j^{m,q'}}{\Rightarrow} \bar{w}\bar{\theta}_1'\bar{\theta}' \overset{*}{\Rightarrow} \bar{w}\bar{u}\bar{\theta}' \overset{*}{\Rightarrow} \bar{w}\bar{u}\bar{v} \tag{7.4}$$

according to $S_k(G)$ with $w, \bar{w}, u, \bar{u}, v, \bar{v} \in T^*$, $\gamma', \bar{\gamma}' \in I'^*$,

$$\theta_1', \theta', \bar{\theta}_1', \bar{\theta}' \in (N'I'^* \cup T)^* \quad \text{where}$$

$$\theta_1' \overset{*}{\Rightarrow} u, \quad \theta' \overset{*}{\Rightarrow} v,$$

$$\bar{\theta}_1' \overset{*}{\Rightarrow} \bar{u}, \quad \bar{\theta}' \overset{*}{\Rightarrow} \bar{v}$$

and ${}^{(k)}uv = {}^{(k)}\bar{u}\bar{v}$ holds.

Consider the derivation (7.3). We have the corresponding leftmost derivation

$$S \overset{*}{\Rightarrow} wAf\gamma\theta \overset{\pi_i}{\Rightarrow} w\theta_1\theta \overset{*}{\Rightarrow} wu\theta \overset{*}{\Rightarrow} wuv,$$

where $\gamma = \delta(\gamma')$, $\theta_1 = \delta(\theta_1')$, and $\theta = \delta(\theta')$. As above we conclude $m = \mu(\theta)$ and $q = \lambda(\gamma)$.

Considering the derivation (7.4), we have the corresponding leftmost derivation

$$S \overset{*}{\Rightarrow} \bar{w}A\bar{\gamma}\bar{\theta} \overset{\pi_j}{\Rightarrow} \bar{w}\bar{\theta}_1\bar{\theta} \overset{*}{\Rightarrow} \bar{w}\bar{u}\bar{\theta}' \overset{*}{\Rightarrow} \bar{w}\bar{u}\bar{v},$$

where $\bar{\gamma} = \delta(\bar{\gamma}')$, $\bar{\theta}_1 = \delta(\bar{\theta}_1')$, and $\bar{\theta} = \delta(\bar{\theta}')$. Furthermore $m = \mu(\bar{\theta})$ and $q' = \lambda(\bar{\gamma})$.

Since $\lambda(f\gamma) = \bar{\lambda}(f, q) = q' = \lambda(\bar{\gamma})$ we have in particular $\mathrm{First}_k(\pi_j, \bar{\gamma}) = \mathrm{First}_k(\pi_j, f\gamma)$ and since $\mu(\theta) = \mu(\bar{\theta})$ we have $\mathrm{First}_k(\bar{\theta}) = \mathrm{First}_k(\theta)$.

If $|\bar{u}| \geqq k$ then ${}^{(k)}\bar{u} \in \mathrm{First}_k(\pi_j, \bar{\gamma})$ and therefore ${}^{(k)}u \in \mathrm{First}_k(\pi_j, f\gamma)$, i.e. $Af\gamma \overset{\pi_j}{\Rightarrow} \overset{\pi_j}{\Rightarrow} \beta_2 : f\gamma \overset{*}{\Rightarrow} \hat{u}$ with ${}^{(k)}\hat{u} = {}^{(k)}\bar{u}$.

Hence the derivation

$$S \overset{*}{\Rightarrow} wAf\gamma\theta \overset{\pi_j}{\Rightarrow} w\beta_2 : f\gamma\theta \overset{*}{\Rightarrow} w\hat{u}\theta \overset{*}{\Rightarrow} w\hat{u}v$$

according to G is possible with ${}^{(k)}\hat{u}v = {}^{(k)}\hat{u} = {}^{(k)}\bar{u} = {}^{(k)}uv$. This is a contradiction to the ILL(k) property of G.

If $|\bar{u}|<k$ then $\bar{u}\in\mathrm{First}_k(\pi_j,\bar{\gamma})$ and therefore $\bar{u}\in\mathrm{First}_k(\pi_j,f\gamma)$, i.e. $Af\gamma \overset{\pi_j}{\Rightarrow} \overset{\pi_j}{\Rightarrow}\beta_2{:}\,f\gamma \overset{*}{\Rightarrow}\bar{u}$ holds according to G. Since $\mathrm{First}_k(\theta)=\mathrm{First}_k(\bar{\theta})$ we have $\theta \overset{*}{\Rightarrow}\hat{v}$ with ${}^{(k)}\hat{v}={}^{(k)}v$. Hence

$$S \overset{*}{\Rightarrow} wAf\gamma\theta \overset{\pi_j}{\Rightarrow} w\beta_2{:}f\gamma\theta \overset{*}{\Rightarrow} w\bar{u}\theta \overset{*}{\Rightarrow} w\bar{u}\hat{v}$$

is possible according to G with ${}^{(k)}\bar{u}\hat{v}={}^{(k)}\bar{u}\bar{v}={}^{(k)}uv$ which contradicts the ILL(k) poperty of G.

(b) Let $S_k(G)$ be a strong ILL(k) grammar. Assume that G is not an ILL(k) grammar, i.e. there are two leftmost derivations

$$S \overset{*}{\Rightarrow} wA\gamma\theta \overset{\pi_i}{\Rightarrow} w\theta_1 \overset{*}{\Rightarrow} wx \quad \text{and}$$

$$S \overset{*}{\Rightarrow} wA\gamma\theta \overset{\pi_j}{\Rightarrow} w\theta_2 \overset{*}{\Rightarrow} wy$$

with $A\in N$, $\gamma\in I^*$, $\theta_1,\theta_2\in(NI^*\cup T)^*$, $w,x,y\in T^*$, ${}^{(k)}x={}^{(k)}y$ and $i\neq j$.

Let us consider the case that the lefthand side of the two productions are different, i.e. π_i: $A\to\beta_1$ and π_j: $Af\to\beta_2$ with $f\in I$. Hence $\gamma=f\gamma'$ holds. Set $m=\mu(\theta)=\bar{\mu}(\theta,m_0,q_0)$, $q'=\lambda(\gamma')=\bar{\lambda}(\gamma',q_0)$ and $q=\lambda(\gamma)=\bar{\lambda}(\gamma,q_0)=\bar{\lambda}(f,q')$.

In P' there are the productions

$$\pi_i^{m,q}\colon\ \psi(A,m,q)\to\psi(\beta_1,m,q)$$

and

$$\pi_j^{m,q'}\colon\ \psi(Af,m,q')\to\psi(\beta_2,m,q').$$

With Corollary 6.1, Lemma 6.2 and Corollary 6.3 the existence of the leftmost derivations

$$S' \overset{*}{\Rightarrow} \psi(wA\gamma\theta,m_0,q_0) \overset{\pi_i^{m,q}}{\Rightarrow} \psi(w\theta_1,m_0,q_0) \overset{*}{\Rightarrow} wx \quad \text{and}$$

$$S' \overset{*}{\Rightarrow} \psi(wA\gamma\theta,m_0,q_0) \overset{\pi_j^{m,q'}}{\Rightarrow} \psi(w\theta_2,m_0,q_0) \overset{*}{\Rightarrow} wy$$

according to $S_k(G)$ follows.

Since ${}^{(k)}x={}^{(k)}y$ we have $\mathrm{First}_k(\pi_i^{m,q})\cap\mathrm{First}_k(\pi_j^{m,q'})\neq\emptyset$ which is a contradiction to the strong ILL(k) property of $S_k(G)$.

In a similar manner the case that π_i and π_j posess the same lefthand side yield a contradiction.

This completes the proof of the theorem.

Now we can easily derive the following decidability result.

Theorem 7.2. Given an indexed grammar G and an integer $k\geqq 1$. It is then decidable whether G is an ILL(k) grammar.

Proof. S_k-grammar of G is effectively constructable since the functions λ and μ defining this grammar are effectively computable. Furthermore it is decidable whether $S_k(G)$ is a strong ILL(k) grammar (cf. Corollary 3.2).

Clearly, given an indexed grammar G, it is not decidable whether there exists a k such that G is an indexed LL(k) grammar, for otherwise this question would be decidable in the contextfree case.

Furthermore, the construction used in the proof of Theorem 7.1 shows

Theorem 7.3. The classes of ILL(k) and strong ILL(k) languages coincide.

Abstract

The classes of indexed LL(k) grammars and strong indexed LL(k) grammars are defined. First the class of strong indexed LL(k) grammars is investigated. In particular it is shown that the strong indexed LL(k) property is decidable and that the class of strong indexed LL(k) languages is contained in the class of deterministic indexed languages. Furthermore it is proved that the deterministic contect-free languages coincide with the right linear strong indexed LL(1) languages and are a proper subclass of the strong indexed LL(1) languages. The remainder of the paper is devoted to proving the decidability of the (general) indexed LL(k) property. To prove this result, a general transformation of indexed grammars is introduced. This transformation unifies proof techniques used in the context-free and indexed areas.

PROF. DR. RAINER PARCHMANN
INSTITUT FÜR INFORMATIK, UNIVERSITÄT HANNOVER
WELFENGARTEN 1
D-3000 HANNOVER 1, WEST GERMANY

References

[1] AHO, A. V., "Indexed grammars", J. Assoc. Comput. Mach. 15 (1968), 647—671.
[2] FISCHER, M. J., "Grammars with macro-like productions", in IEEE Conference Record of the Ninth Annual Symposium on Switching and Automata Theory, (1969), 131—142.
[3] HARRISON, M. A., Introduction to Formal Language Theory, Addison Wesley, Reading, Mass., 1978.
[4] HOPCROFT, J. E. and ULLMAN, J. D., Introduction to Automata Theory, Languages and Computation, Addison Wesley, Reading, Mass., 1979.
[5] MAIBAUM, T. S. E., "Pumping lemmas for term languages", J. Comput. System Sci. 17 (1978), 319—330.
[6] MEHLHORN, K., "Parsing macro grammars top down", Inf. Contr. 40 (1979), 123—143.
[7] PARCHMANN, R., DUSKE, J. and SPECHT, J., "On deterministic indexed languages", Inform. Contr. 45 (1980), 48—67.
[8] PARCHMANN, R., DUSKE, J. and SPECHT, J., "Closure properties of deterministic indexed languages", Inform. Contr. 46 (1980), 200—218.
[9] ROSENKRANTZ, D. J. and STEARNS, R. E., "Properties of deterministic top-down grammars", Inform. Contr. 17 (1970), 226—256.
[10] SEBESTA, R. W. and JONES, N. D., "Parsers for indexed grammars", Internat. J. Comput. and Inform. Sci. 7 (1978), 345—359.
[11] WEISS, K., "Deterministische indizierte Grammatiken", Doctoral Dissertation, University of Karlsruhe, 1976.

(Received July 11, 1983)