A NEW FORM OF HOMOMORPHIC CHARACTERIZATION OF CONTEXT-FREE LANGUAGES

To the honor of Professor Masami Ito on his 60^{th} birthday *

PÁL DÖMÖSI† AND SATOSHI OKAWA‡

Received January 27, 2002

ABSTRACT.
In this paper we give a special form of the Chomsky-Schützenberger-Stanley type characterization of context-free languages.

1 Introduction There have been many papers on the homomorphic characterization of language classes. Chomsky-Schützenberger-Stanley type characterization of the context-free languages [2, 4, 11] states as follows: for an alphabet Σ, an alphabet Δ, a homomorphism $h : \Delta^* \to \Sigma^*$ and a Dyck language D over Δ can be determined such that for every context-free language L over Σ, there can be found a regular language R over Δ satisfying $L = h(D \cap R)$.

Hirose and Yoneda proved in [6] that every context-free language L can be expressed in the above form by a minimal linear and regular language R. By results of Shamir [10] we can give a little bit stronger version of this statement such that we can write R in the form XK^*, or equivalently, in the form K^*X, where K is a finite set and X is a singleton. This statement can also be immediately derived from the new proof of the Chomsky-Schützenberger theorem (i.e., Chomsky-Schützenberger-Stanley theorem) given by Autebert, Berstel and Boasson in [1]. (Actually, we can also retrieve this result from [6].)

In this paper we give a special form of the Chomsky-Schützenberger-Stanley type characterization of context-free languages. Some consequences of this statement are also discussed.

This paper organizes as follows. In the next section, some fundamental concepts and notations related to formal language theory are given. In Section 3, a new homomorphic characterization of context-free languages will be shown. Some concluding remarks will also be stated in the last section.

2 Preliminaries In this section we provide some notions and notations on formal languages. (For notions and notations not defined here see, for example, [5, 9, 12, 7].) The elements of an *alphabet* Σ are called *letters* (Σ is supposed to be finite and nonempty). A *word* over an alphabet Σ is a finite string consisting of letters of Σ. The string consisting of zero letters is called the *empty word,* written by λ. The *length* of a word w, in symbols $|w|$, means the number of letters in w when each letter is counted as many times it occurs. By definition, $|\lambda| = 0$. At the same time, for any set H, $|H|$ denotes the cardinality of H. If u and v are words over an alphabet Σ, then their *catenation* uv is also a word over Σ. Catenation is an associative operation and the empty word λ is the identity with respect to catenation: $w\lambda = \lambda w = w$ for any word w. For a word w and positive integer n, the

2000 *Mathematics Subject Classification.* 68Q42, 68Q45, 68Q70 .
Key words and phrases. Context-Free Languages, Homomorphic Characterization .
*This research has been supported in part by the Project "Words and Languages" of the University of Aizu (No R-3-4) and the Hungarian National Foundation for Scientific Research (OTKA T030140).

notation w^n means the word obtained by catenating n copies of the word w. w^0 equals the empty word λ. w^m is called the *m-th power* of w for any non-negative integer m. We also use $u^* = \{u^m \mid m \geq 0\}$ and $u^+ = u^* \setminus \{\lambda\}$. Moreover, put w^R for the reverse of a word w. (Then $w^R = w$ if $w \in \Sigma \cup \{\lambda\}$.)

Let Σ^* be the set of all words over Σ, moreover, let $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$. Σ^* and Σ^+ are the *free monoid* and the *free semigroup,* respectively, generated by Σ under catenation. Every subset L of Σ^* is called a (formal) language over Σ. L is called λ*-free* if $\lambda \notin L$.

Let K and L be arbitrary languages. We put $KL = \{uv \mid u \in K, v \in L\}$, $K^+ = \{u_1 \dots u_m \mid u_1, \dots, u_m \in K\}$, $K^* = K^+ \cup \{\lambda\}$.

A *generative* (Chomsky-type) *grammar* [3] is an ordered quadruple $G = (V, \Sigma, S, P)$ where V and Σ are disjoint alphabets, $S \in V$, and P is a finite set of ordered pairs (u, v) such that u is a word over $V \cup \Sigma$ containing at least one letter of V and v is an arbitrary word over $V \cup \Sigma$. The elements of V are called *nonterminals* and those of Σ *terminals.* S is called the *start symbol.* Elements (u, v) of P are called *productions* and are written as $u \to v$. A word w over V *derives directly* a word z, in symbols, $w \Rightarrow z$, if there are words u_1, u_2, u_3, v_1 such that $w = u_2 u_1 u_3$, $z = u_2 v_1 u_3$, and $u_1 \to v_1$ belongs to P. w *derives* z, or in symbols, $w \overset{*}{\Rightarrow} z$ if there is a finite sequence of words $w_0, w_1, \dots, w_k, k \geq 0$ over $V \cup \Sigma$ where $w_0 = w, w_k = z$ and $w_i \Rightarrow w_{i+1}$ for $0 \leq i \leq k-1$. If $k > 0$ then we also say that w *really derives* z and we write $w \overset{+}{\Rightarrow} z$. In other words, $\overset{*}{\Rightarrow}$ is the reflexive and transitive closure and $\overset{+}{\Rightarrow}$ is the transitive closure of the binary relation $\Rightarrow$. We also say that $w_0 \Rightarrow w_1 \Rightarrow \dots \Rightarrow w_{k-1} \Rightarrow w_k$ is a *derivation* (of w_k from w_0). The (formal) *language* $L(G)$ *generated by* G is defined by $L(G) = \{w \mid w \in \Sigma^*, S \overset{+}{\Rightarrow} w\}$. G is *minimal* if V is a singleton (i.e., if $V = \{S\}$). G is *right-linear* or *regular* if each production is one of the two forms $X \to xY$ or $X \to x$ where $X, Y \in V, x \in \Sigma^*$. G is *left-linear* if each production is one of the two forms $X \to Yx$ or $X \to x$ where $X, Y \in V, x \in \Sigma^*$. G is *context-free* if each production is of the form $X \to u$ where $X \in V$ and $u \in (V \cup \Sigma)^*$. We say that G is a context-free grammar given in *Chomsky normal form* if $P \subseteq \{X \to YZ \mid X, Y, Z \in V\} \cup \{X \to x \mid X \in V, x \in \Sigma\}$. It is said that G is a context-free grammar given in *Greibach normal form* if $P \subseteq \{X \to xW \mid X \in V, x \in \Sigma, W \in V^*\}$.

Let $G = (V, \Sigma, S, P)$ be a context-free grammar and $w \Rightarrow z$ a direct derivation such that $w = pXu, z = pqu, p \in \Sigma^*, u \in (V \cup \Sigma)^*$, and of course, $X \to q \in P$. Then we say that $w \Rightarrow z$ is a *leftmost direct derivation* and sometimes we emphasize this fact by using the notation $w \Rightarrow_\ell z$. Moreover, if there is a finite sequence of words $w_0, w_1, \dots, w_k,\ k \geq 0$ over $V \cup \Sigma$ where $w_0 = w, w_k = z$ and $w_i \Rightarrow_\ell w_{i+1}$ for $0 \leq i \leq k-1$ then we say that w *derives* z *by leftmost derivation* and put $w \overset{*}{\Rightarrow}_\ell z$. If $k > 0$ then we also say that w *really derives* z *by leftmost derivation* and put $w \overset{+}{\Rightarrow}_\ell z$. Then it is said that $w_0 \Rightarrow w_1 \Rightarrow \dots \Rightarrow w_k$ is a *leftmost derivation* (of w_k from w_0), in symbols, $w_0 \Rightarrow_\ell w_1 \Rightarrow_\ell \dots \Rightarrow_\ell w_k$. It is well-known (see, for example, [7]) that for every $w \in (V \cup \Sigma)^*$ and $u \in \Sigma^*$, $w \overset{*}{\Rightarrow} u$ if and only if $w \overset{*}{\Rightarrow}_\ell u$. Thus $L(G) = \{u \mid u \in \Sigma^*, S \overset{+}{\Rightarrow}_\ell u\}$.

Let Δ be an alphabet. Then $\bar{\Delta}$ is defined by $\bar{\Delta} = \{\bar{d} \mid d \in \Delta\}$. Moreover, let S be a symbol with $S \notin \Delta \cup \bar{\Delta}$. A *Dyck language* D over Δ is a language generated by a *Dyck grammar* $G_\Delta = (\{S\}, \Delta, P_\Delta, S)$, where $P_\Delta = \{S \to SS\} \cup \{S \to xS\bar{x} \mid x \in \Delta\} \cup \{S \to \lambda\}$. Then it is clear that a Dyck language is minimal context-free. We will use the notation $\bar{u}$ for $\bar{u}_1 \dots \bar{u}_m$ whenever $u = u_1 \dots u_m, u_1, \dots, u_m \in \Delta$. Moreover, we put $\bar{\lambda} = \lambda$.

If G is regular or context-free then it is said that the language $L(G)$ is regular or context-free, respectively. In this sense we also speak about minimal regular and minimal context-free languages. It is well-known (see, for example, [7]) that every regular language can also be generated by left linear grammars. Moreover, it is also proved, if L is a λ-free context-free language then it can be generated by a grammar G given in either Chomsky normal form or Greibach normal form. (See [7] for a proof.) In addition, it is clear that every minimal right linear language has the form $L_1^*L_2$, where L_1, L_2 are finite languages. Similarly, every minimal left linear languages have the form $L_1L_2^*$, where L_1, L_2 are finite languages. Thus a minimal right linear language is not necessarily minimal left linear and vice versa. We say that a language is *minimal regular* if it is minimal right-linear and minimal left-linear.

Let Δ, Σ be two alphabets. A mapping $h : \Delta^* \to \Sigma^*$ is called a *homomorphism* if $h(\lambda) = \lambda$ and $h(pq) = h(p)h(q), p, q \in \Delta$. If $\Sigma \subseteq \Delta$ with $h(x) = x, x \in \Sigma$ and $h(y) = \lambda, y \in \Delta \setminus \Sigma$ then h is said to be a *trivial homomorphism.*

If $\Sigma, \bar{\Sigma} = \{\bar{x} \mid x \in \Sigma\}, \Delta'$ are pairwise disjoint sets and $\Delta = \Sigma \cup \bar{\Sigma} \cup \Delta'$, $h(\bar{x}) = x, \bar{x} \in \bar{\Sigma}, h(y) = \lambda, y \in \Sigma \cup \Delta'$, then we say that h is *quasi-trivial.*

Given an $L \in \Delta^*$, we shall also use $h(L) = \{h(w) \mid w \in L\}$ and we say that $h(L)$ is the homomorphic image of L (with respect to the homomorphism h).

Let $G = (V, \Sigma, S, P)$ be a context-free grammar in Greibach normal form. To each terminal letter $a \in \Sigma$, associate the finite set $\Phi(a) = \{\bar{X}w^R \in (V \cup \bar{V})^* \mid X \to aw \in P\}$. This defines *Shamir's homomorphism* $\Phi : \Sigma^* \to \mathcal{P}((V \cup \bar{V})^*)$, where $\mathcal{P}((V \cup \bar{V})^*)$ denotes the power set of $(V \cup \bar{V})^*$. An induction allows to prove Shamir's theorem [10]:

Theorem 1 (Shamir) *For every context-free language $L \subseteq \Sigma^*$ there exists an alpabet V, a letter $S \in V$ and a homomorphism $\Phi : \Sigma^* \to \mathcal{P}((V \cup \bar{V})^*)$ such that $u \in L(G) \iff \{S\}\Phi(u) \cap D_V \neq \emptyset$.*

3 Homomorphic Characterization Results

First we consider a context-free grammar $G = (V, \Sigma, S, P)$ in Greibach normal form. Then $P \subset \{X \to u \mid u \in \Sigma V^*\}$. We apply a similar construction given in [1].

For every leftmost derivation $S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k$ define a word $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k)$ in the following way. If $k = 1$ and $w_1 = aw,\ a \in \Sigma$ then let $\rho(S \Rightarrow_\ell w_1) = \bar{S}w^Ra\bar{a}$. Now we suppose $k > 1$ and that $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})$ is defined. Suppose that $w_{k-1} = pXu, w_k = pqu$, with $p \in \Sigma^*,\ u \in (V \cup \Sigma)^*, X \to aw \in P, a \in \Sigma$. Then let $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\bar{X}w^Ra\bar{a}$. By induction we get that $\psi(\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k)) = w_k \in \Sigma^*$ if and only if $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) \in D_{V \cup \Sigma} \cap \{\bar{X}w^Ra\bar{a} \mid X \to aw \in P, a \in \Sigma\}^*$, where ψ denotes either the trivial or the quasi-trivial homomorphism over $(V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma})^*$.

Thus we get the following statement.

Theorem 2 *Every context-free grammar $G = (V, \Sigma, S, P)$ given in Greibach normal form determines a finite subset of words $K = \{\bar{X}w^Ra\bar{a} \mid X \to aw \in P, a \in \Sigma\}$ over the alphabet $V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma}$ such that $L(G) = \psi(D_{\Sigma \cup V} \cap \{S\}K^*)$, where ψ denotes either the trivial or the quasi-trivial homomorphism over $V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma}$.*

Now we give a grammar $G = (V, \Sigma, S, P)$ in Chomsky normal form. Then $P \subseteq \{X \to u \mid u \in V^2 \cup \Sigma\}$. Similarly as before, for every leftmost derivation $S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k$ define a word $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k)$ in the following manner.

If $k = 1$ and $w_1 = a, a \in \Sigma$ then let $\rho(S \Rightarrow_\ell w_1) = \bar{S}a\bar{a}$. If $k = 1$ and $w_1 \in V^2$ then let $\rho(S \Rightarrow_\ell w_1) = {w_1}^R$. Now we suppose $k > 1$ and that $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})$ has already been defined. Suppose that $w_{k-1} = pXu, w_k = pau$ with $p \in \Sigma^*,\ u \in (V \cup \Sigma)^*, X \to a \in P, a \in \Sigma$. Then let $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\bar{X}a\bar{a}$. Now we assume $k > 1$ and that $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})$ has already been defined. Suppose $w_{k-1} = pXu, w_k = pvu$ with

$p \in \Sigma^*,\ u \in (V \cup \Sigma)^*, X \to v \in P, v \in V^2$. Then let $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\bar{X}v^R$.

Similarly as before, we obtain by induction that $\psi(\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k)) = w_k \in \Sigma^*$ if and only if $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) \in D_{V\cup\Sigma} \cap (\{\bar{X}a\bar{a} \mid X \to a \in P, a \in \Sigma\} \cup \{\bar{X}v^R \mid X \to v \in P, v \in V^2\})^*$, where ψ denotes either the trivial or the quasi-trivial homomorphism over $(V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma})^*$. Therefore, we get as follows.

Theorem 3 *Every context-free grammar $G = (V, \Sigma, S, P)$ given in Chomsky normal form determines a finite subset of words $K = \{\bar{X}a\bar{a} \mid X \to a \in P, a \in \Sigma\} \cup \{\bar{X}v^R \mid X \to v \in P, v \in V^2\}$ over the alphabet $V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma}$ such that $L(G) = \psi(D_{\Sigma\cup V} \cap \{S\}K^*)$, where ψ denotes either the trivial or the quasi-trivial homomorphism over $V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma}$.*

Now we take a context-free grammar $G = (V, \Sigma, S, P)$ given in either Greibach normal form or Chomsky normal form. We define $\rho(X \to u)$ for every rule $X \to u$ as before:

For every rule $X \to aw, a \in \Sigma, w \in V^*$, let $\rho(X \to aw) = \bar{X}w^R a\bar{a}$. Furthermore, let $\rho(X \to a) = \bar{X}a\bar{a}$. Finally, put $\rho(X \to v) = \bar{X}v^R$ if $v \in V^2$. Then we have $\rho(S \Rightarrow_\ell w_1) = \rho(S \to w_1)$ whenever $S \to w_1 \in P$. Let us consider a leftmost derivation $S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k$ such that $k > 1, w_{k-1} = pXu, w_k = pqu$ with $p \in \Sigma^*,\ u \in (V \cup \Sigma)^*, X \to q \in P$. Then we obtain $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\rho(X \to q)$.

We may obtain again by induction that $\psi(\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k)) = w_k \in \Sigma^*$ if and only if $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) \in D_{V\cup\Sigma} \cap (\{\bar{X}w^R a\bar{a} \mid X \to aw \in P, a \in \Sigma\} \cup \{\bar{X}a\bar{a} \mid X \to a \in P, a \in \Sigma\} \cup \{\bar{X}v^R \mid X \to v \in P, v \in V^2\})^*$, where ψ denotes either the trivial or the quasi-trivial homomorphism over $(V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma})^*$.

Thus we get the following statement by induction which also can be derived from the above two theorems.

Corollary 1 *Let G be a context-free grammar given in either Greibach normal form or Chomsky normal form. Then G determines a finite subset $K = (\{\bar{X}w^R a\bar{a} \mid X \to aw \in P, a \in \Sigma\} \cup \{\bar{X}a\bar{a} \mid X \to a \in P, a \in \Sigma\} \cup \{\bar{X}v^R \mid X \to v \in P, v \in V^2\})$ of words over the alphabet $V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma}$ such that $L(G) = \psi(D_{\Sigma\cup V} \cap \{S\}K^*)$, where ψ is either the trivial or the quasi-trivial homomorphism over $V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma}$.*

Now we turn to arbitrary context-free grammars. Let $G = (V, \Sigma, S, P)$ be a context-free grammar.

For every rule $X \to u \in P$ in G construct the string $\rho(X \to u)$ such that

$$\rho(X \to u) = \begin{cases} \bar{X}u^R & \text{if } u = vw, v \in V, \\ \bar{X}u^R\bar{u} & \text{if } u \in \Sigma^*, \\ \bar{X}u^R\bar{v} & \text{if } u = vwz, v \in \Sigma^*, wz \in V(V \cup \Sigma)^*. \end{cases}$$

We extend this definition to arbitrary leftmost derivation $S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k, w_1, \ldots, w_k \in (V \cup \Sigma)^*, k \geq 1$ in the following way.

- If $k = 1$ then let $\rho(S \Rightarrow_\ell w_1) = \rho(S \to w_1)$.
- Suppose that $k > 1$ and the word $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})$ has already defined. Moreover, suppose that $w_{k-1} = pXu, w_k = pqu$, with $p \in \Sigma^*,\ u \in (V \cup \Sigma)^*, X \to q \in P$.
  - •• If $u, q \in \Sigma^*$ then put $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\rho(X \to q)\bar{u}$.

- •• If $q \in \Sigma^*$ and $u = vw, v \in \Sigma^*, w \in V(V \cup \Sigma)^*$ then put $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\rho(X \to q)\bar{v}$.
- •• If $q \notin \Sigma^*$ then let $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\rho(X \to q)$.
- •• If $u \in V\Sigma^* \cup \{\lambda\}$ then let $\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k) = \rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-2} \Rightarrow_\ell w_{k-1})\rho(X \to q)$.

Consider the quasi-trivial homomorphism $\psi : (V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma})^* \to \Sigma^*$ with $\psi(\bar{x}) = x, \bar{x} \in \bar{\Sigma}, \psi(y) = \lambda, y \in \Sigma \cup \bar{\Sigma} \cup V \cup \bar{V}$. By induction we get the following statements.

Lemma 1 *For every context-free grammar $G = (V, \Sigma, S, P)$ and leftmost derivation $S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_{k-1} \Rightarrow_\ell w_k, k \geq 1$, it follows:*

- $w_k \in \Sigma^*$ *if and only if* $S\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow w_{k-1} \Rightarrow_\ell w_k) \in D_{V \cup \Sigma} \cap \{S\}(K \cup \bar{\Sigma})^*$, *where* $K = \{\rho(X \to u) \mid X \to u \in P\}$.
- $w_k \in \Sigma^*$ *implies* $\psi(\rho(S \Rightarrow_\ell w_1 \Rightarrow_\ell \ldots \Rightarrow_\ell w_k)) = w_k$, *where ψ is the quasi-trivial homomorphism of* $(V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma})^*$ *onto* Σ^*.

Theorem 4 *For every context-free grammar $G = (V, \Sigma, S, P)$ we can determine a finite subset K of words over the alphabet $V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma}$ such that $L(G) = \psi(D_{\Sigma \cup V} \cap \{S\}(K \cup \bar{\Sigma})^*)$, where ψ is the quasi-trivial homomorphism of $(V \cup \bar{V} \cup \Sigma \cup \bar{\Sigma})^*$ onto Σ^*, $K = \{\rho(X \to u) \mid X \to u \in P\}$, and*

$$\rho(X \to u) = \begin{cases} \bar{X}u^R & \text{if } u = vw, v \in V, \\ \bar{X}u^R\bar{u} & \text{if } u \in \Sigma^*, \\ \bar{X}u^R\bar{v} & \text{if } u = vwz, v \in \Sigma^*, wz \in V(V \cup \Sigma)^*. \end{cases}$$

We note that the above two statements does not hold in general if we consider the trivial homomorphism instead of the quasi-trivial one. We finish with the following example.

Example 1 *Let $G = (\{\mathcal{S}, \mathcal{A}, \mathcal{B}, \mathcal{D}, \mathcal{E}, \mathcal{F}\}, \{C, I, P, T, a, f, g, i, l, n, o, r, s, t, u, !\}, \mathcal{S}, \{\mathcal{S} \to Congr\mathcal{A}io\mathcal{B}\mathcal{D}o!, \mathcal{A} \to atulat, \mathcal{B} \to n\mathcal{E}of\mathcal{F}, \mathcal{D} \to t, \mathcal{E} \to sToPr, \mathcal{F} \to I\})$. Consider the leftmost derivation $\mathcal{S} \Rightarrow_\ell w_1, \ldots, w_5 \Rightarrow_\ell w_6$ using the consecutive rules $\mathcal{S} \to Congr\mathcal{A}io\mathcal{B}\mathcal{D}o!$, $\mathcal{A} \to atulat$, $\mathcal{B} \to n\mathcal{E}of\mathcal{F}$, $\mathcal{E} \to sToPr$, $\mathcal{F} \to I, \mathcal{D} \to t$. Then we get $\rho(S \Rightarrow_\ell w_1 \Rightarrow \ldots \Rightarrow w_5 \Rightarrow_\ell w_6) = \mathcal{S}\bar{\mathcal{S}}!o\mathcal{D}\mathcal{B}oi\mathcal{A}rgnoC\bar{C}\bar{o}\bar{n}\bar{g}\bar{r}\bar{\mathcal{A}}talutaa\bar{a}\bar{t}\bar{u}\bar{l}\bar{a}\bar{t}\bar{i}\bar{o}\bar{\mathcal{B}}\mathcal{F}fo\mathcal{E}nn\bar{n}\bar{\mathcal{E}}rPoTss\bar{s}\bar{T}\bar{o}\bar{P}\bar{r}\bar{o}\bar{f}\bar{\mathcal{F}}I\bar{I}\bar{\mathcal{D}}t\bar{t}\bar{o}\bar{!}$.*

Therefore, $\psi(\rho(S \Rightarrow_\ell w_1, \ldots, w_5 \Rightarrow_\ell w_6))$ will have the following value:

C o n g r a t u l a t i o n s T o P r o f I t o !

4 Concluding Remarks In this paper we studied a special form of homomorphic characterization of the Chomsky-Schützenberger-Stanley type for context-free languages. First we obtained special forms for context-free grammars given in Greibach normal form or Chomsky normal form. Later we generalized our constructions to arbitrary context-free grammars. Our results also show the following statement: For an alphabet Σ, an alphabet Δ, a trivial homomorphism $h : \Delta^* \to \Sigma^*$ and a Dyck language D over Δ can be determined such that for every context-free language L over Σ, there can be found a singleton X and a finite language K satisfying $L = h(D \cap XK^*)$. We note that we can also get this result in the form $L = h(D \cap K^*X)$ considering rightmost derivation in our treatments. It is clear that XK^* is a minimal left linear language and that K^*X is a minimal right linear language. Moreover, by Theorem 3 we may assume that for every $u \in D \cap XK^*$, $|u| = 3|\psi(u)| - 2$. This fact may be interesting in special applications.

It would be also interesting to extend our researches to larger classes of languages. For example, we would like to get similar results for the class of indexed and linear indexed languages. This is a challenging problem for the further researches.

References

[1] J. Autebert, J. Berstel, L Boasson, Context-free languages and pushdown automata, in: G. Rosenberg and A. Salomaa, eds., Handbook of Formal Languages, Vol.1, 111-174.

[2] N. Chomsky, Context-free grammars and pushdown storage, *MIT Res. Lab. Electron. Quart. Prog. Report.* **65** (1962) 187–194.

[3] N. Chomsky, Formal properties of grammars, *Handbook of Math. Psychology 2* (1963), 328–418.

[4] N. Chomsky, M. P. Schützenberger, The algebraic theory of context-free languages, in: P. Bradford, D. Hirschberg, eds., *Computer programming and formal systems,* North-Holland, Amsterdam, 1963, 118–161.

[5] S. Ginsburg, *The Mathematical Theory of Context-Free Languages* (McGraw-Hill, New York, 1966).

[6] S. Hirose, M. Yoneda, On the Chomsky and Stanley's homomorphic characterization of context-free languages, *Theoret. Comput. Sci.* **36** (1985) 109–112.

[7] Gy. E. Révész, *Introduction to Formal Languages* (McGraw-Hill, New York, 1983).

[8] G. Rosenberg, A. Salomaa, eds., *Handbook of Formal languages,* Vol.1, Vol.2, Vol. 3 (Springer-Verlag, 1997).

[9] A. Salomaa, *Formal Languages* (Academic Press, New York, 1973).

[10] E. Shamir, A representation theorem for algebraic and context-free power series in communicating variables, *Inform. Comput.,* 11 (1967), 239–254.

[11] R. J. Stanley, Finite state representations of context-free languages, *MIT Res. Lab. Electron. Quart. Prog. Report.* **76** (1964) 276–279.

[12] T. A. Sudkamp, *Languages and Machines: An Introduction to the Theory of Computer Science* (Addison-Wesley, 1997).

† Debrecen University, Debrecen, H-4010, POB 12, Hungary
e-mail: domosi@math.klte.hu
‡ The University of Aizu, Aizu-Wakamatsu, 965-8580, Japan
e-mail: okawa@u-aizu.ac.jp