数理解析研究所講究録
1106 巻 1999 年 196-203



# On a homomorphic characterization of the class of slender context-free languages

Satoshi Okawa
The University of Aizu, Aizu-Wakamatsu, 965-8580, Japan
Pál Dömösi
Kossuth Lajos University, Debrecen, H-4032, Hungary

March 10, 1999

**Abstract**

We give the following homomorphic characterization of slender context-free languages. Let $\Sigma$ be an alphabet. Then, an alphabet $\Delta$, a homomorphism $h : \Delta^* \to \Sigma^*$ and a linear Dyck language $D_{\mathcal{L}}$ over $\Delta$ can be determined according to $\Sigma$ such that for every slender context-free language $L$ over $\Sigma$, there can be found a regular language $R \subseteq \Delta^*$ with $L = h(D_{\mathcal{L}} \cap R)$.

## 1 Introduction

Slender languages attract many researchers from the theoretical point of view as well as from the point of view of the application to the cipher systems. Recently, the loop chacacterizations of them are obtained, which will be stated in the next section.

On the other hand, many characterization results for language classes have been obtained from several viewpoints. Grammatical characterization, characterization by automata, homomorphic characterization, characterization with Dyck reduction, and characterization with equality sets are examples of them. The homomorphic characterization of language class is one of the most attractive one and many results are known. The clasical and famous homomorphic characterization is Chomsky and Stanley's one[1, 9], which states that any context-free language $L$ can be obtained with the form $L = h(D \cap R)$, where $D$ is a Dyck language, $R$ a regular language and $h$ a homomorphism.

In this paper, we investigate a Chomsky-Stanley type characterization for the class of slender context-free languages. However, Chomsky-Stanley type characterization of the class of slender context-free languages is almost meaningless, because a slender context-free language is linear but a Dyck language is not linear. If we use a Dyck language for characterization, then it becomes to the fact that we use structually complicated languages to characterize simple ones. We consider another characterization which is in parallel form to Chomsky-Stanley's one.

This paper is organized as follows. In Section 2, we introduce some fundamental notions, notations, definitons of slender languages, and the loop characterization results for slender languages. In Section 3, we give our main theorem, which characterize the class of slender context-free languages by Chomsky-Stanley type. Section 4 gives some concluding remarks.

# 2 Preliminaries

For all notions and notations not defined here, see [2, 3, 4, 8].

A language $L \subseteq \Sigma^*$ is said to be *k-slender* if $card\{w \in L \mid |w| = n\} \leq k$, for every $n \geq 0$. And a language is *slender* if it is $k$-slender for some positive integer $k$. Especially, a 1-slender language is called a *thin* language.

A language $L$ is said to be a *union of single loops* (or, in short, USL) if for some positive integer $k$ and words $u_i, v_i, w_i, 1 \leq i \leq k$,

$$(*) \quad L = \bigcup_{i=1}^{k} u_i v_i^* w_i.$$

A language $L$ is called a *union of paired loops* (or UPL, in short) if for some positive $k$ and words $u_i, v_i, w_i, x_i, y_i, 1 \leq i \leq k$,

$$(**) \quad L = \bigcup_{i=1}^{k} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\}.$$

A USL language $L$ is said to be a *disjoint union of single loops* (DUSL, in short) if the sets in the union $(*)$ are pairwise disjoint. The notion of a *disjoint union of paired loops* (DUPL) is defined analogously considering (**).

For a 2n-letter alphabet $\Sigma = \{a_i, a_i' \mid i = 1, 2, \cdots, n\}$, Dyck language $D$ over $\Sigma$ is a language generated with a Dyck grammar $G = (N, \Sigma, S, P)$ with $N = \{S\}$ and $P = \{S \to \lambda, S \to SS\} \cup \{S \to a_i S a_i' \mid i = 1, 2, \cdots, n\}$. Furthermore, if the set of productions of a grammar $G_{\mathcal{L}}$ is $P_{\mathcal{L}} = \{S \to \lambda\} \cup \{S \to a_i S a_i' \mid i = 1, 2, \cdots, n\}$, then $G_{\mathcal{L}}$ is called a linear Dyck grammar and its language $L(G_{\mathcal{L}})$ is called a linear Dyck language.

We shall use the following well-known results about slender languages.

**Proposition 1** *[6]*
*The next conditions, (i)-(iii), are equivalent for a language L.*
*(i) L is regular and slender.*
*(ii) L is USL.*
*(iii) L is DUSL.*

**Proposition 2** *[6]*
*Every UPL language is DUPL, slender, linear and unambiguous.*

**Proposition 3** *[5, 7]*
*Every slender context-free language is UPL.*

We have the following direct consequence of Propositions 2 and 3.

**Proposition 4** *The class of slender linear languages coincides with the class of slender context-fee languages. In addition, the class of slender context-free languages contains only unambiguous languages.*

# 3 Results

As stated in Introduction, when we consider the homomorphic characteization of the class of slender context-free languages, Chomsky-Stanley type characterization is almost meaningless, because a slender context-free language is linear but a Dyck language is in the bigger class than the linear one. Therefore, we adopt linear Dyck languages instead of Dyck languages in Chomsky-Stanley type characterization.

**Theorem 1** *Let $\Sigma$ be an alphabet. Then an alphabet $\Delta$, a homomorphism $h : \Delta^* \to \Sigma^*$ and a linear Dyck language $D_{\mathcal{L}}$ on $\Delta$ can be determined from $\Sigma$ such that for every slender context-free language $L \subseteq \Sigma^*$, there can be found a regular language $R \subseteq \Delta^*$ with $L = h(D_{\mathcal{L}} \cap R)$.*

**Proof.** Let $\Sigma$ be an alphabet. Then, we first define an alphabet $\Delta$, a homomorphism $h$, the linear Dyck language $D_{\mathcal{L}}$ on $\Delta$ as follows:

An alphabet $\Delta$ is defined by

$$\Delta = \Sigma \cup \Sigma' \cup \bar{\Sigma} \cup \bar{\Sigma}' \cup \{d, d', \bar{d}, \bar{d}'\},$$

where $d \notin \Sigma$, and for $\Sigma$,

$$\Sigma' = \{a' \mid a \in \Sigma\}, \bar{\Sigma} = \{\bar{a} \mid a \in \Sigma\}, \text{ and } \bar{\Sigma}' = \{\bar{a}' \mid a \in \Sigma\}.$$

The homomorphism $h : \Delta^* \to \Sigma^*$ is defined by

$$h(a) = h(\bar{a}') = a, \ a \in \Sigma \text{ and } h(x) = \lambda, \ x \in \Delta \setminus (\Sigma \cup \bar{\Sigma}').$$

The linear Dyck language $D_{\mathcal{L}}$ over $\Delta$ is the language generated by

$$G_{\mathcal{L}} = (\{S\}, \Delta, S, P_{\mathcal{L}}),$$

where

$$P_{\mathcal{L}} = \{S \to aSa', S \to \lambda \mid a \in \Sigma \cup \bar{\Sigma} \cup \{d, \bar{d}\}\}.$$

Let $L$ be any slender context-free language over $\Sigma$. By Proposition 3, we can find a finite index set $I$ and words $u_i, v_i, w_i, x_i, y_i$, for all $i \in I$ with $L = \bigcup_{i \in I} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\}$.

Furthermore, by a padding technique, we can assume $|u_i| = |y_i|$ and $|v_i| = |x_i|$. (Otherwise, for example, if $|u_i| < |y_i|$ then we set a new $u_i$ as $u_i d^{|y_i| - |u_i|}, d \notin \Sigma$.)

In order to simplify the notations, we use the following abbreviations. For a word $w = a_1 \ldots a_\ell \in \Sigma^*$, $w' = a_1' \ldots a_\ell', \bar{w} = \bar{a}_1 \ldots \bar{a}_\ell, \bar{w}' = \bar{a}_1' \ldots \bar{a}_\ell'$, and $w^R = a_\ell \ldots a_2 a_1$.

For $L = \bigcup_{i \in I} \{u_i v_i^n w_i x_i^n y_i \mid n \geq 0\}$, we define a regular grammar $G_R = (N, \Delta, A, P_R)$, where $N = \{A, B, C\}$, $P_R = P_1 \cup P_2 \cup P_3 \cup P_4 \cup P_5$ as

$P_1 = \{A \to u_i \bar{y}_i{}^R B \mid i \in I\}$,

$P_2 = \{B \to v_i \bar{x}_i{}^R B \mid i \in I\}$,

$P_3 = \{B \to w_i w_i'^R C \mid i \in I\}$,

$P_4 = \{C \to \bar{x}_i' v_i'^R C \mid i \in I\}$, and

$P_5 = \{C \to \bar{y}_i' u_i'^R \mid i \in I\}$.

Let $R$ be a language generated by $G_R$, i.e., $R = L(G_R)$. Then, $L = h(D_{\mathcal{L}} \cap R)$ can be proved by the following discussion.

a). $L \subset h(D_{\mathcal{L}} \cap R)$.

Suppose $w$ is in $L$, and $w$ is of the form $u_i v_i^n w_i x_i^n y_i$ for some $i$ and $n$.

By the definition of $G_R$, it is clear that a word

$$\xi = u_i \bar{y_i}^R (v_i \bar{x_i}^R)^n w_i w_i'^R (\bar{x}_i' v_i'^R)^n \bar{y}_i' u_i'^R$$

is generated by $G_R$ as follows.

$A \Rightarrow u_i \bar{y_i}^R B \Rightarrow u_i \bar{y_i}^R v_i \bar{x_i}^R B \Rightarrow^* u_i \bar{y_i}^R (v_i \bar{x_i}^R)^n B \Rightarrow u_i \bar{y_i}^R (v_i \bar{x_i}^R)^n w_i w_i'^R C$
$\Rightarrow u_i \bar{y_i}^R (v_i \bar{x_i}^R)^n w_i w_i'^R \bar{x}_i' v_i'^R C \Rightarrow u_i \bar{y_i}^R (v_i \bar{x_i}^R)^n w_i w_i'^R (\bar{x}_i' v_i'^R)^n C$
$\Rightarrow u_i \bar{y_i}^R (v_i \bar{x_i}^R)^n w_i w_i'^R (\bar{x}_i' v_i'^R)^n \bar{y}_i' u_i'^R$.

Moreover, it is clear that $\xi$ is in $D_{\mathcal{L}}$, and therefore, $\xi$ is in $D_{\mathcal{L}} \cap R$. By the definition of $h$, $h(\xi)$ is a word $u_i v_i^n w_i x_i^n y_i$, that is, $w$. So $w$ is a word in $h(D_{\mathcal{L}} \cap R)$.

b). $h(D_{\mathcal{L}} \cup R) \subset L$.

Let $w \in h(D_{\mathcal{L}} \cap R)$. Then, there exists a word $\xi$ in $D_{\mathcal{L}} \cap R$ such that $w = h(\xi)$. By the definition of $G_R$, $\xi$ should be of the form

$$\xi = u_{i_1} \bar{y}_{i_1}^R (v_{j_1} \bar{x}_{j_1}^R)(v_{j_2} \bar{x}_{j_2}^R) \ldots (v_{j_m} \bar{x}_{j_m}^R) w_{i_2} w_{i_2}'^R (\bar{x}_{k_1}' v_{k_1}'^R) \ldots (x\bar{_{k_m}}' v_{k_m}'^R) \bar{y}_{i_3} u_{i_3}'^R$$

for some $i_1, i_2, i_3, j_1, \ldots, j_m, k_1, \ldots k_n \in I$. By the definition of $D_{\mathcal{L}}$, these indices should be the same, say $i$, and $n = m$. Hence, $\xi = u_i \bar{y_i}^R (v_i \bar{x_i}^R)^n w_i w_i'^R$ $(\bar{x}_i' v_i'^R)^n \bar{y}_i' u_i'^R$ and $h(\xi) = u_i v_i^n w_i x_i^n y_i$. Therefore, $w = h(\xi)$ is in $L$.

This completes the proof. $\square$

**Remark.** There exists a regular language $R$ such that $h(D_{\mathcal{L}} \cap R)$ is not slender. For example, choose a regular language $\Delta^*$ as $R$. Then, by the fact that $D_{\mathcal{L}} \cap R$ is $D_{\mathcal{L}}$ and the fact $h(D_{\mathcal{L}})$ is $\Sigma^*$, the remark follows.

By the Remark, it is interesting to find a subclass $\mathcal{C}$ of regular languages satisfying the following condition.

**Condition.** For any slender context-free language $L$, we can find $R$ in $\mathcal{C}$ of regular languages such that $L = h(D_{\mathcal{L}} \cap R)$, and for any $R$ in $\mathcal{C}$, $h(D_{\mathcal{L}} \cap R)$ is slender context-free.

To determine such language class, we introduce a following notion.

A language L is called a *union of double loops* (or UDL, in short) if for words $u_i, v_i, w_i, x_i, y_i$ for $1 \leq i \leq k$,

$$L = \bigcup_{i=1}^{k} \{u_i v_i^* w_i x_i^* y_i\}.$$

Since $L$ can be generated with a regular grammar $G = (N, \Sigma, P, S)$, where $N = \{S\} \cup \{A_i, B_i \mid i = 1, 2, \cdots, k\}$, $P = \{S \to u_iA_i, A_i \to v_iA_i, A_i \to w_iB_i, B_i \to x_iB_i, B_i \to y_i \mid i = 1, 2, \cdots, k\}$, $L$ is regular. However, it is not slender by Proposition 1.

Then, we have the following result, a little stronger than Theorem 1.

**Theorem 2** *For an alphabet $\Sigma$, an alphabet $\Delta$, a homomorphism $h : \Delta^* \to \Sigma^*$ and a linear Dyck language $D_{\mathcal{L}}$ on $\Delta$ can be determined from $\Sigma$ such that for every slender context-free language $L$ over $\Sigma$, there can be found a UDL regular language $R$ over $\Delta$ such that $L = h(D_{\mathcal{L}} \cap R)$. Moreover, for any UDL regular language $R$, $h(D_{\mathcal{L}} \cap R)$ is slender context-free.*

**Proof.** We can find that a language $R$ employed in the proof of Theorem 1 is a UDL regular language. Therefore, the former part of the theorem holds.

We consider the latter part. Let $R$ be a UDL regular language. Then, since the class of linear context-free languages is closed under the operation of intersection with a regular set, $D_{\mathcal{L}} \cap R$ is linear. Furthermore, by counting the number of words of length $n$ in $D_{\mathcal{L}} \cap R$, we can find $D_{\mathcal{L}} \cap R$ slender. Since a homomorphism does not increase the number of words of length $n$ (and the class of context-free languages is closed under homomorphisms), $h(D_{\mathcal{L}} \cap R)$ is slender context-free.

This completes the proof. □

# 4 Concluding Remarks

In this paper, we investigated Chomsky-Stanley type homomorphic characterization for slender context-free languages and obtained the first characterization as Theorem 1 and the second characterization as Theorem 2, in which for any slender language can be represented by the homomorphic image of the intersection of a linear Dyck language and UDL regular language, and for any UDL regular language, the homomorphic image of the intersection of it with a linear Dyck language is slender. This means the second one is a stronger result than the fisrt one.

At last, we mention here the closure property under the homomorphisms, since the operation of homomorphisms is one of the key operations of this paper. In general, the class of slender languages is not closed under the operation of homomorphisms. In fact, let $\Sigma$ be a set $\{a, b\}$, and define a numbering function $f : \Sigma^* \rightarrow N$, a noninegative integers, with $\mapsto w_i$, the $i$-th word of $\Sigma^*$ in the lexicographical order. And let $n$ be an inverse of $f$. Then, a language L is defined as $L = \{wc^{n(w)} | w \in \Sigma^*\}$. Moreover, a homomorphism $h : (\Sigma \cup \{c\})^* \rightarrow \Sigma^*$ is defined by $h(a) = a, h(b) = b$, and $h(c) = \lambda$. Then, $h(L)$ is $\Sigma^*$, which is not slender[1] . However, this language $L$ is not context-free, but context-sensitive, if we restrict language class to context-free, the following proposition is holds.

**Proposition 5** *The class of slender context-free languages is closed under the operations of homomorphisms and intersection.*

# References


[1] Chomsky, N., Context-free grammars and pushdown storage, *M.I.T. Res. Lab., Electron Quart. Prog. Rept.,* **65** (1962) 187-194.

[2] Ginsburg, S., The Mathematical Theory of Context-Free Languages, *McGraw-Hill Book Company, New York, St Louis, San Francisco, Auckland, Bogota, Hamburg, Johannesburg, London, Madrid, Mexico, Montreal, New Delhi, Panama, Paris, Sao Paulo, Singapore, Sydney, Tokyo, Toronto,* 1966.

[3] Harrison, M. A., Introduction to Formal Language Theory, *Addison-Wesley Publishing Company, Reading , Massachusetts, Menlo Park, California, London, Amsterdam, Don Mils, Ontario, Sidney,* 1978.

[4] Hopcroft, J. E. & Ullmann, J. D., Introduction to Automata Theory, Languages, and Computation , *Addison-Wesley, Reading , Massachusetts, Menlo Park, California, London, Amsterdam, Don Mils, Ontario, Sidney,* 1979.

[5] Ilie, L., On a conjecture about slender context-free languages, Theoret. Comput.Sci. **132** (1994) 427-434.


[1] Prof. Nishida of Toyama Prefectural University pointed out that the class of slender languages is not closed under the $\lambda$-free homomorphisms.

[6] Păun, G., Salomaa, A., Thin and slender languages, *Discrete Appl. Math.*, **61** (1995) 257-270.

[7] Raz, D., Length considerations in context-free languages, *Theoret. Comput.Sci.* **183** (1997) 21-32.

[8] A. Salomaa, "Jewels of Formal Language Theory", Computer Science Press, Potomac,MD (1981).

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

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