*Electronic Notes in Theoretical Computer Science 53 (2002)*

*URL:* [*http://www.elsevier.nl/locate/entcs/volume53.html*](http://www.elsevier.nl/locate/entcs/volume53.html) *9 pages*

Chomsky-Schu¨tzenberger Type Characterizations of Poly-Slender and Parikh Slender Context-Free Languages 1

## Masami Ito 2

*Department of Mathematics Faculty of Science,*

*Kyoto Sangyo University Kyoto 603, Japan*

## Carlos Mart´in-Vide 3

*Research Group in Mathematical Linguistics Rovira i Virgili University*

*Pc¸a. Imperial Ta`rraco 1, 43005 Tarragona, Spain*

## Victor Mitrana 4

*Faculty of Mathematics University of Bucharest*

*Str. Academiei 14, 70109, Bucharest, Romania*

**Abstract**

In this paper we propose a Chomsky-Schu¨tzenberger type characterization of *k*- poly-slender context-free languages, as the homomorphical image of an intersec- tion of a Dyck language and a *(2k + 1)*-poly-slender regular language. A stronger result is provided, namely the homomorphism and the Dyck language are deter- mined irrespective of the given poly-slender context-free language, when consid- ering the family of all poly-slender context-free languages. A similar characteri- zation is obtained for Parikh slender context-free languages.

*1* Work supported by the Grants-in Aid for Scientific Research No. 10440034, Japan So- ciety for the Promotion of Sciences and the Direccio´n General de Ensen˜anza Superior e Investigacio´n Cientifica, SB 97–00110508

*2* Email: [*ito@ksuvx0.kyoto-su.ac.jp*](mailto:ito@ksuvx0.kyoto-su.ac.jp)

*3* Email: [*cmv@astor.urv.es*](mailto:cmv@astor.urv.es)

*4* Email: [*mitrana@funinf.cs.unibuc.ro*](mailto:mitrana@funinf.cs.unibuc.ro)

c 2002 Published by Elsevier Science B. V. *Open access under* [*CC BY-NC-ND license.*](http://creativecommons.org/licenses/by-nc-nd/3.0/)

# Introduction

Considerations concerning the length of words in languages have been in the focus of interest of many researchers for a long time. Thus, in [5] (see also [3,14]) for a given language *L*, one associates with each nonnegative integer *n*, the number of words of length at most *n* in *L*, denoted by *#L(n)*.

Then one defines the “generating function” of *L* by

*fL(z) = X #L(n)zn*

*n>0*

and proves that this function is an algebraic function over the field of ratio- nals, provided that *L* is an unambiguous context-free language. Moreover, the transcendentality of generating function is the main tool in the study of inherent ambiguity of context-free languages in [7].

Also deep results involving words length considerations in Lindenmayer systems have been reported in a series of papers [6,9,15,20].

Motivated by some problems arising in cryptography, [1] introduces *slender languages* as being those languages which contain at most a con- stant number of words of any length. Roughly speaking, the following for- mal language approach of the Richelieu cryptosystem is considered in [1]: a plaintext is encripted using an encription key of length *n* yielding a cryp- totext of the same length of the encription key. If the legal receiver knows the key, then it can easily recover the original text. The problem appears when transmitting all possible keys to the legal receiver. In order to de- crease the recovering complexity, the keys should come from a slender language such that the receiver may test a limited number of keys for each cryptotext. The reader interestead in further details may consult [1].

Characterizations of slender regular and context-free languages have been reported in [18,21] and [11], respectively. Many aspects of different types of slender languages have been investigated in a series of papers [16,17]

Raz extends in [19] the concept of slenderness to poly-slenderness, a poly-slender language being a language whose number of words of length *n* is bounded by the value of a polynomial in *n* for all *n 0*. A character- ization of poly-slender regular languages has been provided in [21] while a characterization of poly-slender context-free languages has been pro- posed in [12].

In this paper we propose a characterization of *k*-poly-slender context- free languages in the Chomsky-Schu¨tzenberger type [5], as the homomor- phical image of an intersection of a Dyck language with a *(2k + 1)*-poly- slender regular language. When dropping *k*, a stronger result is provided, namely the homomorphism and the Dyck language are determined irre- spective of the given poly-slender context-free language. Moreover, the converse assertion is also true. A similar characterization is obtained for

Parikh slender context-free languages, namely each Parikh slender context- free language is the homomorphical image of an intersection of a Dyck language with a Parikh slender regular language, but the converse does not hold.

# Basic Definitions and Previous Results

The set of all words over the alphabet *V* is denoted by *V* , while *V + =*

*V n f"g*, where *"* is the empty word. For each word *x 2 V +*, whose length is denoted by *jxj*, *jxja* delivers the number of occurrences of the letter *a* in

*x*. Any subset of *V* is called a language over *V* . For a nonnegative integer

*k* a language *L V* is called *k-bounded* if there are the words *x ;x ;::: ;x*

*1 2 k*

over *V* such that *L fx g fx g ::: fx g* . *L* is bounded if it is *k*-bounded for some *k 0*. For basic notions and results of combinatorics on words the reader is refered to [4,13].

*1 2 k*

Given a language *L* the mapping *#L(n)* gives the number of all words of length *n* in *L* for every *n*. Formally, *#L(n) =* card*(fx 2 L j jxj = ng):* For a nonnegative integer *k*, the language *L* is said to be *k-poly-slender* if *#L(n) 2 O(nk):* Furthermore, *L* is *poly-slender* if it is *k*-poly-slender for some *k 0*. Any *0*-poly-slender language is called also *slender*.

The following result appears in [19].

**Theorem 2.1** *The class of poly-slender context-free languages is exactly the class of bounded context-free languages.*

Moreover, a direct consequence of the proof of Proposition 1 in [19] useful in the sequel is

**Proposition 2.2** *A k-bounded language is (k-1)-poly-slender for any k 1.*

For an alphabet *V* we denote by *V = fa j a 2 V g*. The language over

*(V [ V )* generated by the context-free grammar having the set of rules *fS ! SS; S ! "g[ fS ! aSa j a 2 V g* is called the Dyck language over *V* denoted *DV* . We write *w = a 1a 2 ::: a n* for *w = a1a2 ::: an*, *ai 2 V* , *1 i n*.

Now we recall the definition of Dyck loops following [12]. Let *Vk = fa1; a2;::: ; akg* and *z = z1z2 ::: z2k 2 DVk* be a word with *zj 2 Vk [ V k* , *1 j 2k*, *jzjai = jzja i = 1* for all *1 i k*, and *U* be an alphabet disjoint from

*Vk [ V k*. For a homomorphism *h : (U [ Vk [ V ) ! U* and some integers

*k*

*ni 0*, *1 i k*, we define the homomorphism

*k*

*hn1;n2;:::;nk*

by

*: (U [ Vk*

*[ V ) ! U*

*hn1 ;n2 ;:::;nk (a)= a* for all *a 2 U;*

*h (a )= uni ; 1 i k*

*n1;n2;:::;nk i* *i*

*h (a )= vni ; 1 i k;*

*n1;n2;:::;nk i* *i*

for some *u ;v*

*2 U* , *1 i k*. Then for some words *w ;w ;::: ;w*

*2 U*

*i* *i*

the language

*0 1 2k*

*D = fhn1;n2;:::;nk (w0z1w1z2w2 ::: z2kwk) j ni 0; 1 i kg*

is called a *k-Dyck loop*. Clearly, the same set *D* can be obtained by chang- ing the homomorphism *h1;1;:::;1* or the Dyck word *z*. We say that *z* is a *un- derlying word* of *D* and *h = h1;1;:::;1* is a *underlying homomorphism* of *D*.

According to the above notations, the main results of [11] and [12] can be written as

**Theorem 2.3** *A context-free language is k-poly-slender if and only if it is a finite union of (k+1)-Dyck loops.*

# Characterizations ofPoly-Slender Context-Free Languages

We say that the class of languages *L1* is weakly characterized in the Chomsky- Schu¨tzenberger type with the class of languages *L2* if for any language

*L1 2 L1* over an alphabet *V* , there exist an alphabet *U* ,a language *L2 2 L2*

over *U* and a homomorphism *h* from *U* into *V* with *L1 = h(DU \ L2)*. Fur- ther, if *h(DU \ L2) 2 L1* for any homomorphism *h* from *U* into an alphabet

*V* , and any language *L2 2 L2* over *U* , then we say that the class of lan-

guages *L1* is strongly characterized in the Chomsky-Schu¨tzenberger type with the class of languages *L2*.

Such a strong characterization is well known for one of the most inves- tigated families of languages in the formal language theory [5].

**Theorem 3.1** *The family of context-free languages is strongly characterized in the Chomsky-Schu¨ tzenberger type with the family of regular languages.*

Now we are ready to prove the main result of this section.

**Theorem 3.2** *For each k 0, the family of k-poly-slender context-free lan- guages is weakly characterized in the Chomsky-Schu¨tzenberger type with the family of (2k+1)-poly-slender regular languages.*

**Proof.** Let *V* be a given alphabet and *L* be a *k*-poly-slender context-free language over *V* . By Theorem 2.3, *L* can be written as a finite union of

*(k+1)*-Dyck loops. Let us denote these *(k+1)*-Dyck loops by *D1; D2;::: ; Dm*

for some *m 1*. Further, assume that

*D = fh(Di )*

*(w(Di)z(Di)w(Di)z(Di)w(Di) ::: z(Di) w(Di) )*

*i n1;n2;:::;nk+1 0* *1*

*1 2 2*

*2k+2*

*2k+2*

*j nj 0; 1 j k + 1g;*

with *w(Di ); w(Di );::: ; w(Di ) 2 V* for any *1 i m*, and

*0 1 2k+2*

*V (Di) [ V (Di) = fz(Di); z(Di);::: z(Di) g;*

*k+1*

*k+1 1* *2*

*2k+2*

for all *1 i m*. Moreover, all alphabets *V (Di )* may be supposed mutually

*k+1*

disjoint.

We consider the alphabet

*m m*

*U = [ V (Di ) [ [fc(i); c(i);::: ; c(i) g;*

*i=1*

*k+1*

*0 1*

*i=1*

*2k+2*

*(i) m (Di)*

where *cj* , *1 i m; 0 j 2k + 2*, are new letters not in *Si=1(Vk+1 [*

*(Di) k+1*

*V )*.

Now we define the regular language

*m*

*R = [fc(i)c (i)gfz(Di )g fc(i)c (i)gfz(Di )g fc(i)c (i)g ::: fz(Di ) g fc(i)*

*c (i) g;*

*0 0 1*

*i=1*

*1 1 2 2 2*

*2k+2*

*2k+2*

*2k+2*

and the homomorphism *g : (U [ U ) ! V* defined by

*g(c(i))= w(Di );*

*j j*

*g(c (i))= ";*

*j*

*g(z(Di))= h(Di)*

*(z(Di ));*

*j 1;1;:::;1 j*

for all *1 i m* and *1 j 2k + 2*. Clearly, for each *1 i m* the following relation holds:

*D = g(D \ fc(i)c (i)gfz(Di )g fc(i)c (i)gfz(Di )g fc(i)c (i)g :::*

*i U 0 0 1 1 1 2 2* *2*

*fz(Di) g fc(i)*

*c (i)*

*g):*

*2k+2*

*2k+2*

*2k+2*

Consequently, *L = g(DU \ R)* holds as well. By Proposition 2.2 it follows that the regular language *R* from above is *(2k +1)*-poly-slender which con- cludes the proof. *2*

Since the homomorphical image of every *k*-Dyck loop is a *k*-Dyck loop, it follows that all families of *k*-poly-slender context-free languages, *k 0*, are closed under homomorphisms. We do not know whether the family of *(2k + 1)*-poly-slender regular languages in the above characterization can be replaced by the family of *k*-poly-slender regular languages. This would imply a strong characterization of the class of *k*-poly-slender context-free languages for each *k* *0*. Another related problem regards the minimal

value of *m k* such that each class of *k*-poly-slender languages can be weakly characterized in the Chomsky-Schu¨tzenberger type with the fam- ily of *m*-poly-slender regular languages. We conjecture that this minimal value is exactly *2k + 1*, so that the previous result is optimal in this respect. Several decidability problems, which remain *open* in this paper, natu-

rally arise if our conjecture is false:

* 1. Given a *k*-poly-slender language, is the minimal value of *m* from above computable?
  2. Are there *k*-poly-slender languages, which are not *(k 1)*-poly-slender, such that the aforementioned minimal value is *k*? In the affirmative, can we decide whether a given *k*-poly-slender language has this property, namely does it admit a Chomsky-Schu¨tzenberger characterization with a *k*-poly- slender regular language?

However, it is obvious that

**Theorem 3.3** *The class of poly-slender context-free languages is strongly characterized with the class of poly-slender regular languages.*

As one can easily see, in the above Chomsky-Schu¨tzenberger type char- acterizations, the alphabet of the regular and the Dyck language, together with the homomorphism depend on the given poly-slender context-free language. This result can be still improved such that all the aforemen- tioned objects can be determined irrespective of the given poly-slender context-free language. This improvment is based on the same characteri- zation of poly-slender context-free languages, by [8] via Theorem 2.1.

A family of languages *L* is closed under paired loop if for any language

*L 2 L* and any two words *x; y*, the language *S fxngLfyng* is in *L*.

*n 0*

**Theorem 3.4** *The family of poly-slender context-free languages is the small- est family which contains all singleton languages and is closed under the following operations:*

* + 1. *union,*
    2. *concatenation,*
    3. *paired loop.*

We call the minimal number of the above operations needed for ob- taining a poly-slender context-free language *L* by the *order* of *L*. Now we state

**Theorem 3.5** *Let V be an alphabet; a new alphabet U and a homomor- phism h from U into V can be determined such that for each poly-slender context-free language L over V there exists a poly-slender regular language R over U such that L = h(DU \ R). Moreover, h(DU \ R) is a poly-slender context-free for any homomorphism h from U into V and any poly-slender regular language R.*

**Proof.** For the given alphabet *V* we construct the alphabet *W = V [f#g[*

*V 0;* where *V 0 = fa0 j a 2 V g*, and *#* is a new symbol. For *w = a1a2 ::: an*, *ai 2 V* , *1 i n*, we denote by *w0* the word *a0 a0 ::: a0* . Put *U = W [ W* .

The homomorphism *h* is defined by

*1 2 n*

*h(a) = h(a 0) = a; a 2 V;* and *h(a) = "; a 2 U n (V [ V 0):*

Let *L* be a poly-slender context-free language over *V* . We prove the state- ment by induction on the order of *L*, say *k*. If *k = 0*, then *L* is a singleton

language, *L = fxg*. One takes the singleton regular language *R = fxx R g*, where *xR* delivers the mirror image of *x*. Clearly, *L = h(D \ R)*.

*U*

Let *L1; L2* be two poly-slender context-free languages over *V* of orders smaller than *k*. By the induction hypothesis,

*Li = h(DU \ Ri );i = 1; 2:*

If *L = L1 [ L2*, then *L = h(DU \ (R1 \ R2))* trivially holds. If *L = L1L2*, then *L = h(DU \ (R1f#gR2f# g))* holds.

If *L = S*

*1*

*1*

*U*

*i 0*

*fxi gL fyig*, then *L = h(D*

*\ (fxy0RgR fy 0x R g))* holds.

Consequently, the first part of the theorem is proved. The second part is

obvious. *2*

# Characterizations ofParikh Slender Languages

Instead of words of length *n*, in [10] one counts the number of words with the same Parikh vector. In particular, a very nice and simple proof of the result of Autebert, Flajolet, and Gabarro [2] concerning the inherent am- biguity of coprefix languages of infinite words is obtained.

By definition, if *V = fa1; a2;::: ; amg* is an alphabet and *w 2 V* is a word, the Parikh vector *V (w)* is defined by

*V (w) = (jwja1 ; jwja2 ;::: ; jwjam ):*

A language *L V* is termed *Parikh slender* if there is a positive integer *k* such that for each vector of positive integers *(i1; i2;::: ; im )* there are at most *k* words in *L* with the Parikh vector *(i1; i2;::: ; im )*. A natural ques- tion asks whether or not the family of Parikh slender context-free lan- guages admits a Chomsky-Schu¨tzenberger type characterization. As we have seen, this problem has been left open for slender languages in the previous section.

A language *L V* is said to be a *multiple loop language* if there exist *k 1* and the strings *u1; v1; u2; v2;::: ; uk; vk ; uk+1 2 V* such that the follow- ing two conditions are satisfied:

*(i) L = u1vu2v ::: uk vuk+1*

*1 2 k*

*(ii) V (v1); V (v2);::: ; V (vk )* are linearly independent over **Q***:*

The following characterization of Parikh slender regular languages ap- pears in [10].

**Theorem 4.1** *A regular language is Parikh slender if and only if it is a finite union of disjoint multiple loop languages.*

It is quite easy to notice that each Parikh slender context-free language is poly-slender and the regular language from the proof of Theorem 3.2 is Parikh slender in accordance with Theorem 4.1. Therefore, we have

just got a Chomsky-Schu¨tzenberger type characterization of Parikh slen- der context-free languages.

**Theorem 4.2** *The family of Parikh slender context-free languages is weakly characterized with the family of Parikh slender regular languages.*

However, the family of Parikh slender languages cannot be strongly characterized with the family of Parikh slender regular languages as shown by the following example. We take the Dyck language *DV* over the alphabet

*V = fa; b; cg*, the Parikh slender regular language *R = a+a + b+ b+ c+c +* and

the homomorphism *h* from *(V [V )* into *fa; bg* defined by *h(a) = h(c) = a*, *h(b) = b*, and *h(x ) = "*, *x 2 V* . We have

*h(D \ R) = a+ b+a+*

*V*

which is not a Parikh slender language.

# References

1. Andrasiu, M., J. Dassow, Gh. Pa˘un, and A. Salomaa, *Language-theoretic problems arising from Richelieu cryptosystems*, Theoret. Comput. Sci. **116** (1993), 339–357.
2. Autebert, J.-M., P. Flajolet, and J. Gabarro, *Prefixes of infinite words and ambiguous context-free languages*, Inform. Process. Letters **25** (1987), 211–

216.

1. Berstel, J. *Sur la densite´ asymptotique de langages formels*. In M. Nivat,

ed.,“Automata, Languages and Programming” North-Holland, 1972, 345– 358.

1. Choffrut, C., and J. Karhuma¨ki, *Combinatorics on words*. In G. Rozenberg and A. Salomaa, eds. “The Handbook of Formal Languages”, vol. I, Springer- Verlag, Berlin, 1997, 329–437.
2. Chomsky, N., and M. P. Schu¨tzenberger, *The algebraic theory of context-free languages*. In P. Bradford and D. Hirschberg, eds. “Computer Programming and Formal Systems”, North-Holland, Amsterdam, 1963, 118–161.
3. Dassow, J., Gh. Pa˘un, and A. Salomaa, *On thinness and slenderness of L languages*, EATCS Bulletin **49** (1993), 152–158.
4. Flajolet, P. *Analytic models and ambiguity of context-free languages*, Theoret. Comput. Sci. **49** (1987), 283–309.
5. Ginsburg, S. “The Mathematical Theory of Context-Free Languages”, McGraw-Hill, New York, 1966.
6. Honkala, J. *Decision problems concerning thiness and slenderness of formal languages*, Acta Inform. **35** (1998), 625–636.
7. Honkala, J. *On Parikh slender languages and power series*, J. Comput. System Sci. **52** (1996), 185–190.
8. Ilie, L. *On a conjectue about slender context-free languages*, Theoret. Comput. Sci. **132** (1994), 427–434.
9. Ilie, L., G. Rozenberg, and A. Salomaa, *A characterization of poly-slender context-free languages*, Theor. Inform. Appl. **34** (2000), 77–86.
10. Lothaire, M. “Combinatorics on Words”, Addison-Wesley, Reading, MA., 1983.
11. Maurer, H.A., and M. Nivat, *Rational bijection of rational sets*, Acta Inform.

**13** (1980), 365–378.

1. Nishida, T., and A. Salomaa, *Slender 0Llanguages*, Theoret. Comput. Sci. **158**

(1996), 161–176.

1. Pa˘un, Gh., and A. Salomaa, *Decision problems concerning the thinness of D0L languages*, EATCS Bulletin **46** (1992), 171–181.
2. Pa˘un, Gh., and A. Salomaa, *Closure properties of slender languages*, Theoret. Comput. Sci. **120** (1993), 293–301.
3. Pa˘un, Gh., and A. Salomaa, *Thin and slender languages*, Discrete Appl. Math.

**61** (1995), 257–270.

1. Raz, D. *Length considerations in context-free languages*, Theoret. Comput. Sci. **183** (1997), 21–32.
2. Rozenberg, G., and A. Salomaa, “The Mathematical Theory of L Systems”, Academic Press, New York, 1980.
3. Shallit, J. *Numeration systems, linear recurrences, and regular sets*, Inform. and Comput. **113** (1994), 331–347.