INFORMATIQUE THÉORIQUE ET APPLICATIONS

LAURENT PIERRE
JEAN-MARC FARINONE

Context-free languages with rational index in $\Theta(n^\gamma)$ for algebraic numbers γ

Informatique théorique et applications, tome 24, nº 3 (1990), p. 275-322.
<http://www.numdam.org/item?id=ITA_1990__24_3_275_0>

© AFCET, 1990, tous droits réservés.

L'accès aux archives de la revue « Informatique théorique et applications » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/legal.php). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 24, n° 3, 1990, p. 275 à 322)

CONTEXT-FREE LANGUAGES WITH RATIONAL INDEX IN $\Theta(n^{\lambda})$ FOR ALGEBRAIC NUMBERS λ (*)

by Laurent PIERRE (1) and Jean-Marc FARINONE (2)

Communicated by A. ARNOLD

Abstract. – *The complexity of a non-empty language L may be estimated by the asymptotic behavior of its rational index, which is a function* $\rho_L : \mathbb{N}-\{0\} \to \mathbb{N}-\{0\}$. *For any positive integer* λ, *we knew a context-free language whose rational index is in* $\Theta(n^{\lambda})$. *In this paper we show context-free languages, whose rational indexes are in* $\Theta(n^{\lambda})$ *for other various values of* $\lambda>1$, *such as the rational numbers or the algebraic numbers or even some transcendental numbers.*

Résumé. – *La complexité d'un langage non vide L peut être estimée par le comportement asymptotique de son index rationnel, qui est une fonction* $\rho_L : \mathbb{N}-\{0\} \to \mathbb{N}-\{0\}$. *On connaissait déjà des langages algébriques d'index rationnel en* $\Theta(n^{\lambda})$ *pour tout entier positif* λ. *Dans cet article nous montrons qu'il existe des langages algébriques d'index rationnel en* $\Theta(n^{\lambda})$ *pour d'autres valeurs de* $\lambda>1$, *telles que les nombres rationnels, plus généralement les nombres algébriques, et même certains nombres transcendants.*

I. INTRODUCTION

There are many ways to measure the complexity of languages. The rational index introduced by L. Boasson, M. Nivat and B. Courcelle [3, 4] is one of them, that behaves well when combined with rational transductions: if $L \geqq L'$ (*i.e.* there exists a rational transduction τ, such that $\tau(L)=L'$), then the rational index ρ_L of L provides an upper bound on $\rho_{L'}$, since

$$\exists c \in \mathbb{N}-\{0\}, \quad \forall n \in \mathbb{N}-\{0\}, \qquad cn(\rho_L(cn)+1) \geqq \rho_{L'}(n).$$

This is why the rational index can prove helpful when studying sets of languages closed under rational transductions like the set of context-free

(*) Received February 1988, revised in January 1989.
(1) Université de Paris-X, Nanterre, U.F.R. de sciences économiques, 92001 Nanterre Cedex France.
(2) Université du Havre, L.A.C.O.S./I.T.E.P.E.A., Faculté des sciences, place Robert-Schumann, 76610 Le Havre, France.

0988-3754/90/03 275 48/$6.80/© AFCET-Gauthier-Villars

languages. We define the extented rational index $\bar{\rho}_L$ of a language L to be $\rho_{L \sqcup\!\sqcup s^*}$ for any letter s, which occurs in no word of L. The extended rational index $\bar{\rho}_L$ of a given language L is generally not harder to compute than its rational index ρ_L. Both indexes are related since

$$\forall n \in \mathbb{N} - \{0\}, \quad \rho_L(n) \leqq \bar{\rho}_L(n) < n(1 + \rho_L(n)),$$

but the extended one gives more information about the complexity of the language since

$$L' \leqq L \quad \Rightarrow \quad \exists c \in \mathbb{N}, \qquad \bar{\rho}_{L'}(n) \leqq \bar{\rho}_L(cn).$$

We denote by $\Theta(n^\lambda)$ the set of functions which are the products of $n \mapsto n^\lambda$ by positive bounded functions. Given two languages L_1 and L_2 and two numbers λ_1 and λ_2 such that $\bar{\rho}_{L_1} \in \Theta(n^{\lambda_1})$ and $\bar{\rho}_{L_2} \in \Theta(n^{\lambda_2})$ and $1 \leqq \lambda_1 < \lambda_2$, then you can conclude that L_2 does not belong to the rational cone generated by L_1. Note that this is true even if $\lambda_2 - \lambda_1 < 1$, but this case could not be handled with plain rational index. In reference [6] you can find a way to construct a context-free language with a rational index in $\Theta(n^k)$ for any positive even integer. For a long time the rational index of a context-free language was thought to necessarily behave asymptoticaly like a simple function, namely an exponential or a polynomial function. In this paper we give methods to construct context-free languages, whose rational indexes are in $\Theta(n^\lambda)$ for other various values of $\lambda > 1$, such as the rational numbers or the algebraic numbers or even some transcendental numbers. The technic used in this paper is strongly related to the one used in [10], where we proved that some context-free languages have rational indexes, which grow faster than any polynomial, but slower than any exponential function $\exp(\lambda n)$.

II. NOTATIONS AND DEFINITIONS

$\mathbb{N}$ will denote the set of non-negative integers, and $\mathbb{N}_+ = \mathbb{N} - \{0\}$ the set of positive integers.

$A \sqcup B$ will denote the union of the disjoint sets A and B.

An alphabet is a finite set of letters.

A language written over an alphabet T is a subset of T^*.

ε denotes the empty word.

$|u|$ is the length of the word u, *i.e.* the number of its letters. *E.g.* $|a^3 bac^2| = 7$. The function $u \mapsto |u|$ will be denoted $|\,.\,|$.

$|u|_x$ is the number of occurrences of the letter x in u. *E.g.* $|a^3bac^2|_a=4$. The function $u\mapsto|u|_x$ will be denoted $|\,.\,|_x$.

If X is an alphabet then $|u|_X$ is the number of occurrences of letters of X in u. *E.g.* $|a^3bac^2|_{\{b,\,c\}}=3$. The function $u\mapsto|u|_X$ will be denoted $|\,.\,|_X$.

$L(\mathcal{A})$ denotes the regular language recognized by the finite automaton $\mathcal{A}$.

A context-free language is a language generated by a context-free grammar. For instance

$$S_{\neq}=\{a^n b^m,\ n\neq m,\ n,\ m\in\mathbb{N}\}$$

is a context-free language, since it is generated by the grammar

$$\langle\{a, b\}, \{S, T, U\}, \{S\to aSb+T+U,\ T\to aT+a,\ U\to bU+b\}, S\rangle.$$

Similarly

$$S_{=}=\{a^n b^n,\ n\in\mathbb{N}\}$$

is a context-free language generated by the grammar

$$\langle\{a, b\}, \{S\}, \{S\to aSb+\varepsilon\}, S\rangle.$$

We shall use $S_{\neq}$ a lot in this paper.

Let r be a binary relation between the two free monoids X^* and Y^*. We say that r is a rational transduction, if its graph is a rational subset of the monoid $X^*\times Y^*$; *i.e.* it is the value of an expression containing only products, unions, stars (or$^+$ operation) and finite sets. The rational transductions may be characterised in another way:

THEOREM (Nivat) [9]: *For any rational transduction $r:X^*\to Y^*$ there exist an alphabet Z, a regular language $K\subset Z^*$ and two morphisms $\varphi:Z^*\to X^*$ and $\psi:Z^*\to Y^*$ such that:*

$$\forall L\subset X^*,\quad r(L)=\psi(K\cap\varphi^{-1}(L)).$$

Furthermore, we may assume the two morphisms to be alphabetic, i.e. $\varphi(Z)\subset X\cup\{\varepsilon\}$ *and* $\psi(Z)\subset Y\cup\{\varepsilon\}$. *We shall write*

$$\tau=\psi\circ\cap K\circ\varphi^{-1}.$$

Let L and L' be two languages. If L' is the image of L under a rational transduction, then we denote it $L\geqq L'$ and we say that L rationally dominates L'. For instance $S_{=}\geqq S_{\neq}$ since $S_{\neq}=a^+S_{=}\cup S_{=}b^+$.

The transformation $\tau: L \mapsto a^+ L \cup L b^+$ accords with the definition of a rational transduction, since its graph is

$$(\varepsilon, a)^+ \{(a, a), (b, b)\}^* \cup \{(a, a), (b, b)\}^* (\varepsilon, b)^+.$$

As an example of Nivat's theorem we can decompose it $\tau = \psi \circ \cap \mathrm{K} \circ \varphi^{-1}$, where $X = \{a, b\}$, $Z = \{a, b, a', b'\}$

$$\begin{array}{ll} \varphi: \quad Z^* \to X^*, & \psi: \quad Z^* \to X^* \\ a \mapsto a, & a \mapsto a \\ b \mapsto b, & b \mapsto b \\ a' \mapsto \varepsilon, & a' \mapsto a \\ b' \mapsto \varepsilon, & b' \mapsto b \end{array}$$

$$K = a'^+ X^* \cup X^* b'^+$$

If $L \geqq L'$ and $L' \ngeqq L$ then we say that L dominates strictly L' and we write $L > L'$. *E.g.* $S_= > S_{\neq}$.

Reference [1] holds the above definitions.

Every regular language is recognised by a finite automaton. $\mathscr{R}_n$ is the family of the regular languages recognized by a finite automaton. $\mathscr{R}_n$ is the family of the regular languages recognized by finite automata with at most n states.

A function $f: \mathbb{R} \to \mathbb{R}$ will be said increasing if

$$\forall x, \quad y \in \mathbb{R}, \qquad x < y \quad \Rightarrow \quad f(x) \leqq f(y).$$

You may notice that, according to this definition, a constant function is increasing.

Let f be a function $\mathbb{N} \to \mathbb{R}$. We shall use the Landau's notations o and O [8], § IV.7, and the Knuth's notations Ω and Θ [7]:

$$o(f) = \{g: \mathbb{N} \to \mathbb{R}, \forall c \in \mathbb{R}_+^*, \quad \exists n_0 \in \mathbb{N}, \forall n \geqq n_0, |g(n)| \leqq c |f(n)|\}$$
$$O(f) = \{g: \mathbb{N} \to \mathbb{R}, \exists c \in \mathbb{R}_+^*, \quad \exists n_0 \in \mathbb{N}, \forall n \geqq n_0, |g(n)| \leqq c |f(n)|\}$$
$$\Omega(f) = \{g: \mathbb{N} \to \mathbb{R}, \exists c \in \mathbb{R}_+^*, \quad \exists n_0 \in \mathbb{N}, \forall n \geqq n_0, |g(n)| \geqq c |f(n)|\}$$
$$\Theta(f) = O(f) \cap \Omega(f)$$

$g \sim f$ will stand for $g - f \in o(f)$.

Remark: If f does not take the value 0 then

$$g \sim f \quad \Leftrightarrow \quad \lim g/f = 1,$$

$$g \in o(f) \Leftrightarrow \lim g/f = 0,$$
$$g \in O(f) \Leftrightarrow \lim\sup |g/f| < \infty$$

and

$$g \in \Theta(f) \Leftrightarrow (\lim\inf |g/f| > 0 \text{ and } \lim\sup |g/f| < \infty).$$

$\lfloor x \rfloor$ is the floor of the real number x *i.e.* the greatest integer k such that $k \leqq x$.

$\lceil x \rceil$ is the ceiling of the real number x *i.e.* the lowest integer k such that $k \geqq x$.

If T is a sub-alphabet of an alphabet U, then π_T will denote the morphism $U^* \to (U-T)^*$, which erases the letters of T and keeps the letters of $U-T$. *E.g.*

$$\pi_{\{a,\, \bar{a}\}}(axayzx\bar{a}) = xyzx.$$

$|\pi_X|$ will stand for the morphism $|\,.\,| \circ \pi_X$, so that $|\pi_X| = |\,.\,| - |\,.\,|_X$.

$A \sqcup\!\sqcup B$ will denote the shuffle of the languages A and B, *i.e.* the set of the words produced when interspercing words of A in words of B. *E.g.*

$$a^* b^* \sqcup\!\sqcup c^* = c^* (ac^*)^* (bc^*)^* = \{a, c\}^* \{b, c\}^*.$$

III. DEFINITION AND BASIC PROPERTIES OF RATIONAL INDEX

1. Definition of ρ and $\bar{\rho}$

DEFINITION 1: *If L is a non-empty language then its rational index is the function $\rho_L : \mathbb{N}_+ \to \mathbb{N}$ defined by*

$$\rho_L(n) = \max_{\substack{K \in \mathcal{R}_n \\ K \cap L \neq \varnothing}} \ \min_{w \in K \cap L} |w|.$$

DEFINITION 2: *Let $L \subset X^*$ be a non-empty language. Let s be a letter which does not belong to X. We define the extended rational index of L to be the rational index of $L \sqcup\!\sqcup s^*$, and we denote it by $\bar{\rho}_L$.*

2. Basic properties

A morphism of free monoids $\varphi : X^* \to Y^*$ is said to be alphabetic if $\varphi(X) \subset Y \cup \{\varepsilon\}$, and strictly alphabetic if $\varphi(X) \subset Y$. In [2] Boasson *et al.* give the five following lemmas.

LEMMA 1: *If L and L' are two languages then* $\rho_{L \cup L'} \leqq \max(\rho_L, \rho_{L'})$.

LEMMA 2: *If L and L' are two languages then* $\rho_{LL'} \leqq \rho_L + \rho_{L'}$.

LEMMA 3: *Let $\varphi : X^* \to Y^*$ be an alphabetic morphism, and $L \subset X^*$. Then* $\rho_{\varphi(L)} \leqq \rho_L$.

LEMMA 4: *Let K be a regular language recognised by an m state automaton. Let L be a language. Then*

$$\forall n \in \mathbb{N}_+, \quad \rho_{L \cap K}(n) \leqq \rho_L(nm).$$

LEMMA 5: *Let φ be an alphabetic morphism from X^* to Y^*. Let L be a subset of Y^*. Then*

$$\forall n \in \mathbb{N}_+, \quad \rho_{\varphi^{-1}(L)}(n) < n(\rho_L(n) + 1).$$

Using the last three lemmas and Nivat's theorem they derive the theorem.

THEOREM 1: *If $L' \leqq L$, then there exists an integer c such that*

$$\forall n \in \mathbb{N}_+, \quad \rho_{L'}(n) < cn(\rho_L(cn) + 1).$$

Proof: According to Nivat's theorem there exist two alphabetic morphisms φ and ψ and a regular language K such that $L' = \varphi(K \cap \psi^{-1}(L))$. Let c be the number of states of an automaton recognising K. Then

$$\rho_{L'}(n) = \rho_{\varphi(K \cap \psi^{-1}(L))}(n) \leqq \rho_{K \cap \psi^{-1}(L)}(n) \leqq \rho_{\psi^{-1}(L)}(cn) < cn(1 + \rho_L(cn)). \quad \square$$

We can make a variation on lemma 5:

LEMMA 6: *Let φ be a strictly alphabetic morphism from X^* to Y^*. Let L be a subset of Y^*. Then* $\rho_{\varphi^{-1}(L)} \leqq \rho_L$.

The proof is left to the reader. This leads to the following theorem.

THEOREM 2: *If $L' \leqq L$, then there exists an integer c such that*

$$\forall n \in \mathbb{N}_+, \quad \rho_{L'}(n) \leqq \bar{\rho}_L(cn).$$

Proof: According to Nivat's theorem there exist two alphabetic morphisms φ and ψ and a regular language K such that $L' = \varphi(K \cap \psi^{-1}(L))$.

Let ψ' be the strictly alphabetic morphism defined by:

$$\psi'(a)=\psi(a) \qquad \text{if} \quad \psi(a)\neq\varepsilon$$

and

$$\psi'(a)=s \qquad \text{if} \quad \psi(a)=\varepsilon.$$

Then $\psi^{-1}(L)=\psi^{-1}(L \sqcup\!\sqcup s^*)$. Let c be the number of states of an automaton recognizing K. As in the proof of theorem 1 we have

$$\rho_{L'}(n)=\rho_{\varphi(K\cap\psi^{-1}(L))}(n)\leqq\rho_{K\cap\psi^{-1}(L)}(n)\leqq\rho_{\psi^{-1}(L)}(cn)$$

Hence

$$\rho_{L'}(n)\leqq\rho_{\psi'^{-1}(L \sqcup\!\sqcup s^*)}(cn)\leqq\rho_{L \sqcup\!\sqcup s^*}(cn)=\bar{\rho}_L(cn). \quad \square$$

This theorem has the corollary:

THEOREM 3: *If $L'\leqq L$ then there exists an integer c such that*

$$\forall n\in\mathbb{N}_+, \quad \bar{\rho}_{L'}(n)\leqq\bar{\rho}_L(cn).$$

Proof: We have $L' \sqcup\!\sqcup s^*\leqq L'\leqq L$. Hence theorem 2 yields that

$$\forall n\in\mathbb{N}_+, \quad \rho_{L' \sqcup\!\sqcup s^*}(n)\leqq\bar{\rho}_L(cn)$$

for some integer c. $\square$

$\pi_{\{s\}}$ is an alphabetic morphism verifying $\pi_{\{s\}}(L \sqcup\!\sqcup s^*)=L$ and $\pi_{\{s\}}^{-1}(L)=L \sqcup\!\sqcup s^*$. Hence lemmas 3 and 5 yield the theorem:

THEOREM 4: *If L is a language then*

$$\forall n\in\mathbb{N}_+ \quad \rho_L(n)\leqq\bar{\rho}_L(n)<n(\rho_L(n)+1).$$

Remark: In this paper, the rational index of a language and its extended rational index will be refered to as its rational indexes.

3. The rational come generated by $S_{\neq}$

In order to evaluate the rational indexes of $S_{\neq}$, we first give two lemmas.

LEMMA 7: $\forall n\in\mathbb{N}_+ \rho_{S_{\neq}}(n)\geqq 2n-1.$

Proof: Let n be a positive integer. The shortest word in $S_{\neq}$ recognised by the n state automaton drawn in figure 1 is $a^{n-1}b^n$. Its length is $2n-1$. Hence $\rho_{S_{\neq}}(n) \geqq 2n-1$. □

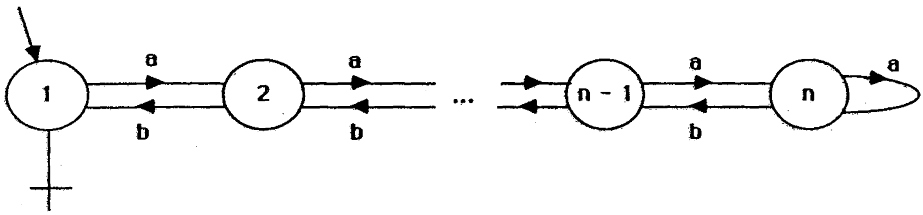

Figure 1.

LEMMA 8: $\forall n \in \mathbb{N}_+,\ \bar{\rho}_{S_{\neq}}(n) \leqq 2n-1$.

Proof: Let n be a positive integer. Let $\mathscr{A}$ be an n state automaton recognising at least one word in $S_{\neq} \sqcup\!\sqcup s^*$. Let w be a shortest word in $L(\mathscr{A}) \cap (S_{\neq} \sqcup\!\sqcup s^*)$. Let us assume that $|w| \geqq 2n$. Then a successful path in $\mathscr{A}$ labeled by w holds at least two disjoint loops. Hence $w = \alpha u \beta v \gamma$ for some words α, β, γ, u and v such that u and v are non-empty and $\mathscr{A}$ recognises $\alpha\beta v\gamma$, $\alpha u \beta\gamma$ and $\alpha\beta\gamma$. These three words belong obviously to $a^* b^* \sqcup\!\sqcup s^*$ but they do not belong to $S_{\neq} \sqcup\!\sqcup s^*$, since they are shorter than w. Hence they belong to $S_{=} \sqcup\!\sqcup s^*$. *I.e.* they hold as many a as b, and so do u, v and w. This is a contradiction to $w \in S_{\neq} \sqcup\!\sqcup s^*$. Hence we have proved that $|w| < 2n$. □

THEOREM 5: $\forall n \in \mathbb{N}_+,\ \bar{\rho}_{S_{\neq}}(n) = \rho_{S_{\neq}}(n) = 2n-1$.

Proof: Lemmas 7, 8 and theorem 4 yield

$$\forall n \in \mathbb{N}_+, \quad 2n-1 \leqq \rho_{S_{\neq}}(n) \leqq \bar{\rho}_{S_{\neq}}(n) \leqq 2n-1. \quad \square$$

Theorems 2 and 5 yield the proposition:

PROPOSITION 1: *If* $L \leqq S_{\neq}$, *then* $\exists c \in \mathbb{N},\ \forall n \in \mathbb{N}_+,\ \rho_L(n) < cn$.

We shall handle in this paper a lot of languages dominated by $S_{\neq}$. This is why we introduce a new notation:

DEFINITION 3: *Let* K_1, K_2, *and* K_3 *be three languages over the alphabet* X. *Let* φ_1, *and* φ_3 *be two morphisms* $X^* \to \mathbb{N}$. *Then we shall denote*

$$\nabla_{\neq}(K_1, \varphi_1, K_2, \varphi_3, K_3)$$

the language

$$\{ w_1 w_2 w_3 \mid w_1 \in K_1,\ w_2 \in K_2,\ w_3 \in K_3,\ \varphi_1(w_1) \neq \varphi_3(w_3) \}.$$

E.g. $S_{\neq} = \nabla_{\neq}(a^*, |\,.\,|, \varepsilon, |\,.\,|, b^*)$.

LEMMA 9: *Let* $K_1 K_2$ *and* K_3 *be three regular languages over the alphabet* X. *Let* φ_1 *and* φ_3 *be two morphisms* $X^* \to \mathbb{N}$. *Then* $\nabla_{\neq}(K_1, \varphi_1, K_2, \varphi_3, K_3) \leqq S_{\neq}$.

Proof: Let $\varphi'_1 : X^* \to a^*$ be the morphism such that $\varphi'_1(x) = a^{\varphi_1(x)}$ for every $x \in X$. Let $\varphi'_3 : X^* \to b^*$ be the morphism such that $\varphi'_3(x) = b^{\varphi_3(x)}$ for every $x \in X$. Let σ be the rational transduction, whose graph is the set of the couples $(w_1 w_2 w_3, \varphi'_1(w_1)\varphi'_3(\omega_3))$, when $w_1 w_2$ and w_3 range over $K_1 K_2$ and K_3. Then $\nabla_{\neq}(K_1, \varphi_1, K_2, \varphi_3, K_3) = \sigma^{-1}(S_{\neq})$. □

For instance this lemma proves that $S_{\neq}$ dominates the language

$$\{ a^\alpha c b^\beta c a^\gamma c b^\delta \mid \alpha + 2\beta \neq 2\gamma + 5\delta \}$$
$$= \nabla_{\neq}(a^* c b^*, |\,.\,|_a + 2|\,.\,|_b, c, 2|\,.\,|_a + 5|\,.\,|_b, a^* c b^*).$$

IV. STRUCTURE FUNCTIONS

1. Definitions of structure functions

We first define $S_{\neq}$-functions.

DEFINITION 4: A $S_{\neq}$*-function will be a partial function* $g : \mathbb{N}_+ \to X^*$, *where* X *is a finite alphabet, and*

$$X^* - g(\mathbb{N}_+) \leqq S_{\neq}.$$

Remarks:

– f is a partial function, *i.e.* $f(i)$ may not exist for some $i \in \mathbb{N}_+$.

– $X^* - g(\mathbb{N}_+)$ is a context-free language, since it is dominated by another context-free language.

– The choice of X does not matter. Indeed if Y is a superset of X, then g may be considered to be a partial function from $\mathbb{N}_+$ to Y^*. And, since

$$X^* - g(\mathbb{N}_+) = (Y^* - g(\mathbb{N}_+)) \cap X^*$$

and conversely

$$Y^* - g(\mathbb{N}_+) = (X^* - g(\mathbb{N}_+)) \cup (Y^* - X^*),$$

it is obvious that $X^* - g(\mathbb{N}_+) \leqq S_{\neq}$ if and only if $Y^* - g(\mathbb{N}_+) \leqq S_{\neq}$.

DEFINITION 5: *We define a structure function to be a* $S_{\neq}$*-function* $g : \mathbb{N}_+ \to X^*$ *verifying also the three following properties:*

- *for some unique letter* $x \in X$, *that we shall denote* x_g, *we have* $|g(i)|_x + 1 = i$ *for every* $i \in \mathbb{N}_+$, *for which* $g(i)$ exists.
- $g(\mathbb{N}_+)$ *does not contain any infinite regular language.*
- $g(i)$ *is defined for infinitely many* i.

Remark: In the first property uniqueness is supposed only for convenience: in order to specify a structure function g, we only have to give the value of $g(i)$ whenever it exists; we need not specify which letter is x_g.

The second property is easily checked by means of the following lemma:

LEMMA 10: *Let* $g : \mathbb{N}_+ \to X^*$ *be a partial function such that*

$$\lim_{i \to \infty} |g(i)|/i = \infty.$$

Then $g(\mathbb{N}_+)$ *does not contain any infinite regular language.*

Proof: Let assume $g(\mathbb{N}_+)$ to contain an infinite regular language. Then we can find three words α, u and β such that u is not empty and $\alpha u^+ \beta \subset g(\mathbb{N}_+)$. Hence for any positive integer i, there exists a positive integer j_i such that $\alpha u^i \beta = g(j_i)$. Let n be a positive integer. Then $j_1, \ldots, j_n$ are n pairwise distinct positive integers. So that

$$\prod_{i=1}^{n} j_i \geqq n!.$$

Thus

$$\prod_{i=1}^{n} \frac{|g(j_i)|}{j_i} \leqq \left(\prod_{i=1}^{n} |\alpha\beta| + i|u| \right) \Big/ n! = \prod_{i=1}^{n} \frac{|\alpha\beta| + i|u|}{i} \leqq |\alpha u \beta|^n$$

hence $\liminf |g(j_i)| / j_i \leqq |\alpha u \beta|$ and thus $\liminf |g(i)| / i \leqq |\alpha u \beta|$ which is not compatible with:

$$\lim_{i \to \infty} |g(i)| / i = \infty. \quad \square$$

For instance we shall prove later that

$$f_2 : \quad \mathbb{N}_+ \to \{x_1, x_2\}^*, \qquad i \mapsto x_1^{i-1} (x_2 x_1^{i-1})^{i-1}$$

is a structure function.

DEFINITION 6: *For any structure function g we define* $\tilde{g}$ *to be the partial function* $\mathbb{N}_+ \to \mathbb{N}_+$ *such that* $\tilde{g}(n)$ *is the largest integer p such that* $|g(p)| \leqq n-1$:

$$\tilde{g}(n) = \max\{p \mid |g(p)| \leqq n-1\}.$$

LEMMA 11: *If g is a structure function then:*

- *there exists an integer* n_0 *such that* $\tilde{g}(n)$ *is defined if and only if* $n \geqq n_0$;
- $\tilde{g}$ *is increasing;*
- *for any* $n \geqq n_0$ *we have* $\tilde{g}(n) \leqq n$;
- $\lim_{n \to \infty} \tilde{g}(n) = \infty$.

Proof: $g(\mathbb{N}_+)$ is not empty, since it is infinite. So we can consider the integer $n_0 = 1 + \min |g(\mathbb{N}_+)|$. Let us define $\tilde{G}(n)$ to be the set of numbers p such that $g(p)$ exists and $|g(p)| \leqq n-1$. Then obviously $\tilde{G}(n)$ is a increasing sequence of sets, which are non-empty if and only if $n \geqq n_0$. Furthermore, when $g(p)$ exists, we have $|g(p)|_{x_g} = p-1$, so that $|g(p)| \geqq p-1$. Hence, if $|g(p)| \leqq n-1$, then $p \leqq n$. This proves that $\tilde{G}(n) \subset [1, n]$. This completes the proof of the first three assertions of the lemma, since we may notice, that $\tilde{g}(n)$ is defined if and only if $\tilde{G}(n)$ is not empty, and then $\tilde{g}(n) = \max \tilde{G}(n)$.

Since $g(i)$ is defined for infinitely many i, for any integer j we can find a integer p such that $p \geqq j$ and $g(p)$ is defined. Then $p \in \tilde{G}(|g(p)|+1)$, so that

$$p \leqq \tilde{g}(|g(p)|+1).$$

Let n be an integer such that $n > |g(p)|$. Since $\tilde{g}$ is increasing, we have $\tilde{g}(n) \geqq \tilde{g}(|g(p)|+1)$ and thus

$$\tilde{g}(n) \geqq \tilde{g}(|g(p)|+1) \geqq p \geqq j.$$

We have proved that

$$\forall j, \quad \exists p, \quad \forall n, \quad n > |g(p)| \quad \Rightarrow \quad \tilde{g}(n) \geqq j.$$

Thus $\lim_{\infty} \tilde{g} = \infty$. □

DEFINITION 7: *Let f and g be two structure functions. We shall say that f dominates g and we shall write* $f \geqq g$, *if there exist two finite alphabets X and Y and a rational transduction* $\varphi_{f,g}: X^* \to Y^*$ *such that* $f(\mathbb{N}_+) \subset X^*$,

$g(\mathbb{N}_+)\subset Y^*$,

$$\varphi_{f,g}(X^*-f(\mathbb{N}_+))=Y^*-g(\mathbb{N}_+),$$
$$\varphi_{f,g}(X^*)=Y^*$$

and

$$\forall u\in X^*,\quad \forall v\in\varphi_{f,g}(u),\qquad |u|_{x_f}=|v|_{x_g}.$$

Obviously the domination between structure functions is a pre-order, *i.e.* it is reflexive and transitive.

DEFINITION 8: *Let f and g be two structure functions. If $f\geqq g$* and $\tilde{g}(n)\in o(\tilde{f}(n))$, *then we shall say that f dominates strictly g and we shall write $f>g$.*

Obviously the strict domination between structure functions is transitive.

2. Main example of structure function

DEFINITION 9: *We define $X_k=\{x_1,\ldots,x_k\}$, with $X_0=\varnothing$.*

DEFINITION 10: *We inductively define the sequence of functions $f_k:\mathbb{N}_+\to X_k^*$ by:*

$$f_0(i)=\varepsilon$$
$$f_k(i)=(f_{k-1}(i)\,x_k)^{i-1}f_{k-1}(i)\qquad \textit{if}\quad k>0.$$

In other words $f_k(i)$ is the word in $X_k^{i^k-1}$, whose l-th letter is x_j if i^{j-1} is the greatest power of i dividing l.

So we have

$$|f_k(i)|=i^k-1$$

and

$$|f_k(i)|_{x_j}=i^{k-j}(i-1).$$

E.g.

$$\begin{array}{lll} f_0(1)=\varepsilon, & f_0(2)=\varepsilon, & f_0(3)=\varepsilon\\ f_1(1)=\varepsilon, & f_1(2)=x_1, & f_1(3)=x_1x_1\\ f_2(1)=\varepsilon, & f_2(2)=x_1x_2x_1, & f_2(3)=x_1x_1x_2x_1x_1x_2x_1x_1\end{array}$$

$f_3(1)=\varepsilon, \qquad f_3(2)=x_1x_2x_1x_3x_1x_2x_1,$

$$f_3(3)=x_1^2x_2x_1^2x_2x_1^2x_3x_1^2x_2x_1^2x_2x_1^2x_3x_1^2x_2x_1^2x_2x_1^2.$$

DEFINITION 11: *Let i and k be two positive integers, such that $i \leq k$. Let w be a word of X_k^*. Then $\pi_{X_{i-1}}(w)$ can be written in a unique way*

$$\pi_{X_{i-1}}(w)=x_i^{\alpha_0}z_1x_i^{\alpha_1}z_2x_i^{\alpha_2}\dots z_jx_i^{\alpha_j}$$

where $\alpha_0, \alpha_1 \dots \alpha_j$ are non-negative integers and $z_1, z_2 \dots z_j$ are letters of X_k-X_i. Then $z_1z_2\dots z_j=\pi_{X_i}(w)$ and $j=|\pi_{X_i}(w)|$. *Let us define the sequence of the groups of x_i in w to be the finite sequence*

$$(x_i^{\alpha_0}, x_i^{\alpha_1}, \dots, x_i^{\alpha_j}).$$

There are exactly $|\pi_{X_i}(w)|+1$ groups of x_i's in w. Some of them may be empty. The length of the group of x_i's of rank p is the number of occurrences of x_i, which are preceded by exactly p occurrences of letters of X_k-X_i. *E.g.* Let $k=3$ and

$$w=x_1x_2x_1x_1x_3x_1x_1x_3x_1x_2x_1x_2x_2x_1x_1x_3x_1x_1x_1.$$

For $i=1$ we have

$$\pi_{X_0}(w)=w=x_1^1x_2x_1^2x_3x_1^2x_3x_1^1x_2x_1^1x_2x_1^0x_2x_1^2x_3x_1^3.$$

Note that there is an empty group of x_1 in the middle of the factor x_2^2. The lengths of the 8 groups of x_1 are 1 2 2 1 1 0 2 and 3. For $i=2$, we have

$$\pi_{X_1}(w)=x_2x_3^2x_2^3x_3=x_2^1x_3x_2^0x_3x_2^3x_3x_2^0,$$

hence there are 4 groups of x_2, whose lengths are 1 0 3 and 0. At last

$$\pi_{X_2}(w)=x_3^3,$$

hence w has 1 group of x_3, whose length is 3.

$f_k(n)$ is the only word of X_k^* such that for every $i \in [1, k]$ the lengths of all its groups of x_i are equal to $n-1$. And a word of X_k^* belongs to $f_k(\mathbb{N}_+)$ if and only if all its groups have the same length.

DEFINITION 12: *Let $A_k=X_k^*-f_k(\mathbb{N}_+)$.*

So a word belongs to A_k if and only if a group of x_i and the (only) group of x_k have different lengths for some i such that $1 \leq i < k$.

LEMMA 12: *For every $k \geq 2$,*

- f_k *is a structure function;*
- $\tilde{f}_k(n)=\lfloor \sqrt[k]{n} \rfloor$ *and*
- $f_k > f_{k+1}$.

The remaining of this section will be the proof of this lemma. For this we first prove two lemmas.

LEMMA 13: *Let* $k \geqq 2$. *There exists a rational transduction* $\sigma_{f_k, f_{k+1}} : X_k^* \to X_{k+1}^*$ *such that*

$$\text{If} \quad w' \in \sigma_{f_k, f_{k+1}}(w) \quad \text{then} \quad |w'|_{x_{k+1}} = |w|_{w_k} \tag{1}$$

$$\sigma_{f_k, f_{k+1}}(X_k^*) = X_{k+1}^*. \tag{2}$$

$$\sigma_{f_k, f_{k+1}}(A_k) = A_{k+1} \tag{3}$$

Proof: Let $\varphi : X_{k+1}^* \to X_k^*$ be the morphism defined by: $\varphi(x_1)=\varepsilon$ and $\varphi(x_{i+1})=x_i$ for $i \geqq 1$. Let $\varphi' : X_k^* \to X_{k+1}^*$ be the substitution defined by: $\varphi'(x_1)=x_1$ and $\varphi'(x_i)=(x_2 x_1^*)^* x_{i+1} (x_1^* x_2)^*$ for $i \geqq 2$. We define $\sigma_{f_k, f_{k+1}}$ by

$$\sigma_{f_k, f_{k+1}}(A) = \varphi^{-1}(A) \cup (x_1^* x_2)^* \varphi'(A) (x_2 x_1^*)^*.$$

(1) holds obviously, and (2) too, since $\varphi^{-1}(X_k^*) = X_{k+1}^*$.

DEFINITION 13: *If* $0 < i < k$, *we shall denote* $A_{k,i}$ *the set of the words* w *belonging to* X_k^* *holding a group of* x_i *whose length is not* $|w|_{x_k}$.

We have

$$A_k = A_{k,1} \cup \ldots \cup A_{k,k-1}.$$

If $w \in X_k^*$ then the groups of x_{i+1} in a word $w' \in \varphi^{-1}(w)$ have the lengths of the groups of x_i in w for every $i \in \{1, \ldots, k\}$. Its groups of x_1 have any lengths. Hence $\varphi^{-1}(A_k)$ is the set of the words of X_{k+1}^*, in which for some i such that $2 \leqq i < k+1$ a group of x_i and the group of x_{k+1} have different lengths. *I. e.* $\varphi^{-1}(A_{k,i}) = A_{k+1, i+1}$ and

$$\varphi^{-1}(A_k) = A_{k+1,2} \cup \ldots \cup A_{k+1,k}. \tag{4}$$

Similarly let w be a word in X_k^*. Let us consider the groups of x_1 in w:

$$w = x_1^{\alpha_1} x_{i_1} x_1^{\alpha_2} x_{i_2} \ldots x_1^{\alpha_k} x_{i_k} x_1^{\alpha_{k+1}}$$

where $k=|\pi_{x_1}(w)|$ and $\forall j,\ i_j>1$. Then

$$(x_1^* x_2)^* \varphi'(w)(x_2 x_1^*)^* = (x_1^* x_2)^* x_1^{\alpha_1}(x_2 x_1^*)^* x_{i_1}(x_1^* x_2)^* x_1^{\alpha_2}(x_2 x_1^*)^* x_{i_2} \dots$$
$$\dots (x_1^* x_2)^* x_1^{\alpha_k}(x_2 x_1^*)^* x_{i_k}(x_1^* x_2)^* x_1^{\alpha_{k+1}}(x_2 x_1^*)^* x_{i_{k+1}}.$$

Let w' be a word in $(x_1^* x_2)^* \varphi'(w)(x_2 x_1^*)^*$. The groups of x_{i+1} in w' have the lengths of the groups of x_i in w for every $i\in\{2, \dots, k\}$. The groups of x_2 in w' have any lengths. And the groups of x_1 of w appear among those of w'. More precisely every group x_1^j of x_1 in w becomes in w' a factor belonging to $(x_1^* x_2)^* x_1^j (x_2 x_1^*)^*$, *i.e.* a group of x_2 of any length λ, whose members alternate with $\lambda+1$ groups of x_1, among which one is x_1^j. Hence $(x_1^* x_2)^* \varphi'(A_k)(x_2 x_1^*)^*$ is the set of the words of X_{k+1}, in which for some $i\in\{1, 3, \dots, k\}$ a group of x_i and the group of x_{k+1} have different lengths. *I.e.*

$$(x_1^* x_2)^* \varphi'(A_k)(x_2 x_1^*)^* = A_{k+1,1} \cup A_{k+1,3} \cup \dots \cup A_{k+1,k}. \tag{5}$$

(4) and (5) add and yield

$$\varphi^{-1}(A_k) \cup (x_1^* x_2)^* \varphi'(A_k)(x_2 x_1^*)^* = A_{k+1,1} \cup \dots \cup A_{k+1,k},$$

i.e. $\sigma_{f_k, f_{k+1}}(A_k) = A_{k+1}$. □

Remark: This proof works only if $k\geqq 2$. For instance in a word of A_3 either a group of x_2 and the group of x_3 have different lengths and then it belongs to $\varphi^{-1}(A_2)$, or a group of x_1 and the group of x_3 have different lengths and then it belongs to $(x_1^* x_2)^* \varphi'(A_2)(x_2 x_1^*)^*$. On the other hand $A_1=\varnothing$. Hence $\sigma_{f_1, f_2}(A_1)=\varnothing\neq A_2$.

LEMMA 14: *$A_k \leqq S_{\neq}$ for any $k\geqq 2$.*

Proof: We shall prove it inductively.

• A_2 is the set of the words in $\{x_1, x_2\}^*$ in which two consecutive groups of x_1 have different lengths or the number of x_2 is not the length of the last group of x_1. *I.e.*

$$A_2 = (x_1^* x_2)^* \nabla_{\neq}(x_1^*, |\cdot|, x_2, |\cdot|, x_1^*)(x_2 x_1^*)^*$$
$$\cup \nabla_{\neq}((x_1^* x_2)^*, |\cdot|_{x_2}, \varepsilon, |\cdot|, x_1^*).$$

This proves that $A_2 \leqq S_{\neq}$.

• Let k be an integer greater than 2. Let us assume that $A_{k-1}\leqq S_{\neq}$. Lemma 13 yields that $A_k=\sigma_{f_{k-1}, f_k}(A_{k-1})$. Hence $A_k\leqq A_{k-1}$. This proves that $A_k\leqq S_{\neq}$. □

Proof of lemma 12 Let k be an integer such that $k \geqq 2$. According to lemma 14, f_k is a $S_{\neq}$-function. For any $j \in [1, k]$ and any $i \in \mathbb{N}_+$ we have

$$|f_k(i)|_{x_j} = i^{k-j}(i-1)$$

so that x_k is the only letter occuring $i-1$ times in $f_k(i)$ for every i. Hence $x_{f_k} = x_k$. Since

$$|f_k(i)| = i^k - 1, \tag{6}$$

we have

$$\lim_{i \to \infty} |f_k(i)|/i = \infty,$$

proving thereby that $f_k(\mathbb{N}_+)$ holds no infinite regular language. We have shown that f_k is a structure function. (6) results in the second assertion of lemma 12. So

$$\bar{f}_k(n) \sim n^{1/k}.$$

This proves that $\bar{f}_{k+1}(n) \in o(\bar{f}_k(n))$, while lemma 13 proves that $f_k \geqq f_{k+1}$. So the third assertion of lemma 12 holds. □

V. THE LANGUAGE RELATED TO A STRUCTURE FUNCTION

1. Definition of L_g

Let $g : \mathbb{N}_+ \to X^*$ be a structure function. Let b_1, a_∞ and b_∞ be three letters not belonging to X. We shall define a language $L_g \subset (X \cup \{b_1, a_\infty, b_\infty\})^*$. L_g is a subset of the regular language

$$F_g = (b_1^* \sqcup\!\sqcup\, X^*)(a_\infty b_\infty^*)^*,$$

that we shall call its frame. We define the structured part of L_g to be

$$S_g = \bigcup_{i \in \mathbb{N}_+} (b_1^* \sqcup\!\sqcup\, g(i))(a_\infty b_\infty^*)^i,$$

the unstructured part of L_g to be

$$U_g = (b_1^* \sqcup\!\sqcup\, (X^* - g(\mathbb{N}_+)))(a_\infty b_\infty^*)^*,$$

and the extended structured part of L_g to be

$$E_g = \{ w \in F_g, \ |w|_{x_g} + 1 = |w|_{a_\infty} \}.$$

These three languages are subsets of F_g. Since $|g(i)|_{x_g} + 1 = i$, we notice that $S_g = E_g - U_g$.

DEFINITION 14: *The above definitions of S_g, U_g and E_g allow us to define L_g as the union of E_g and U_g. It is also the disjoint union of S_g and U_g.*

$$L_g = E_g \cup U_g = S_g \sqcup U_g.$$

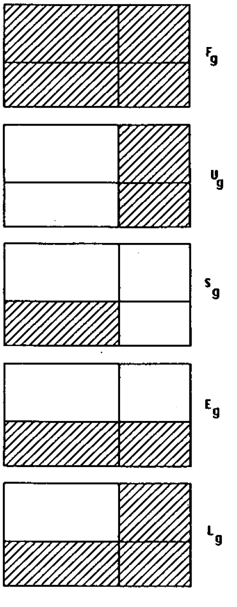

Figure 2.

Figure 2 represents the various languages, we just defined.

S_g is not a context-free language. (We shall not prove it.) But since g is a $S_{\neq}$-function, $U_g \leqq S_{\neq}$ and it is obvious that $E_g \leqq S_{=}$. Hence U_g and E_g are context-free languages, and so is L_g.

2. Lower bound on ρ_{L_g}.

Let $n \in \mathbb{N}_+$. Let us get a lower bound on $\rho_{L_g}(n)$. Let $p = \tilde{g}(n)$. Let $\mathscr{A}$ be the automaton depicted in figure 3.

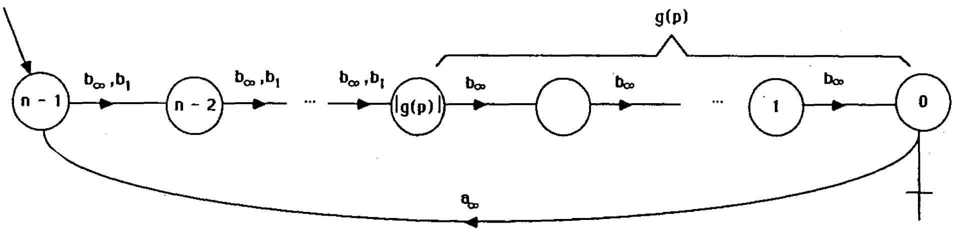

Figure 3.

In this figure

$$\overbrace{\bullet \xrightarrow{x_1} \bullet \ldots \bullet \xrightarrow{x_l} \bullet}^{w}$$

stands for

$$\bullet \xrightarrow{x_1,\, y_1} \bullet \ldots \bullet \xrightarrow{x_l,\, y_l} \bullet$$

where $w = y_1 \ldots y_l$.

This automaton has n states. It is made of a simple path of length $n-1$ leading from the only initial state to the only final state. Every arc of this path is labeled by two letters in such a way that the whole path is labeled by $b_1^{n-1-|g(p)|} g(p)$ and by b_∞^{n-1}. There is also an arc leading from the final state to the initial state labeled by a_∞. So $\mathscr{A}$ recognises a word of $(b_1^* \sqcup\!\sqcup X^*)(a_\infty b_\infty^*)^*$ if and only if it is

$$b_1^{n-1-|g(p)|} g(p)(a_\infty b_\infty^{n-1})^m$$

for some $m \in \mathbb{N}$. This word belongs to L_g only if $m = p$ and then it belongs to S_g. Thus the shortest (and only) word in $L(\mathscr{A}) \cap L_g$ is

$$w = b_1^{n-1-|g(p)|} g(p)(a_\infty b_\infty^{n-1})^p.$$

Hence

$$\rho_{L_g}(n) \geqq |w| = n - 1 + \tilde{g}(n)\, n. \qquad (7)$$

Remark: $|b_1^{n-1-|g(p)|} g(p)| = n-1$ and the letter b_1 is used to ensure that the path labeled by $b_1^{n-1-|g(p)|} g(p)$ is a simple path (*i.e.* a path holding no loops) of maximal length $(n-1)$ in an n state automaton. Similarly b_∞ is used to ensure that the loop labeled by $a_\infty b_\infty^{n-1}$ is a simple loop of maximal length.

3. Upper bound on $\bar{\rho}_{L_g}$.

Let $n \in \mathbb{N}_+$. Let $\mathscr{A}$ be any automaton with n states recognising at least one word in $L_g \sqcup\!\sqcup\ s^*$. Let w be a shortest word in $(L_g \sqcup\!\sqcup\ s^*) \cap L(\mathscr{A})$. We shall give an upper bound on $|w|$, that depends only on n and not on $\mathscr{A}$ so that it will be also an upper bound on $\bar{\rho}_{L_g}(n)$. Let us consider a successful path γ in $\mathscr{A}$ labeled by w.

- First let us assume that $(U_g \sqcup\!\sqcup\ s^*) \cap L(\mathscr{A}) \neq \varnothing$.

Let w' be a shortest word in $(U_g \sqcup\!\sqcup\ s^*) \cap L(\mathscr{A})$. Then $|w'| \leqq \bar{\rho}_{U_g}(n)$ because of the definition of rational index. w' belongs to $(L_g \sqcup\!\sqcup\ s^*) \cap L(\mathscr{A})$, whose shortest word is w. Hence $|w| \leqq |w'|$. Thus $|w| \leqq \bar{\rho}_{U_g}(n)$.

- Let us assume now that $U_g \sqcup\!\sqcup\ s^*$ and $L(\mathscr{A})$ are disjoint.

Then every word in $(L_g \sqcup\!\sqcup\ s^*) \cap L(\mathscr{A})$ belongs to $S_g \sqcup\!\sqcup\ s^*$. Thus w belongs to $S_g \sqcup\!\sqcup\ s^*$ and

$$w \in \underbrace{\overbrace{(b_1^* \sqcup\!\sqcup\ g(p) \sqcup\!\sqcup\ s^*)}^{|\cdot|<n} \left(a_\infty \overbrace{(b_\infty^* \sqcup\!\sqcup\ s^*)}^{|\cdot|<n}\right)^p}_{|\cdot| \leqq pn+n-1}$$

for some positive interger p. Braces show upper bounds on the lengths of parts of w, that we shall prove.

First let us prove that there are at most $n-1$ letters in w before the first a_∞. Let us assume that this part of w holds a loop. If the label of this loop belongs to $b_1^* \sqcup\!\sqcup\ s^*$ then it can be removed yielding a shorter word than w belonging to $S_g \sqcup\!\sqcup\ s^*$. This is a contradiction. Hence the label of this loop does not belong to $b_1^* \sqcup\!\sqcup\ s^*$. Since $g(\mathbb{N}_+)$ holds no infinite regular language, we can change $g(p)$ into a word of $X^* - g(\mathbb{N}_+)$ by iterating this loop. This transforms w into a word of $(U_g \sqcup\!\sqcup\ s^*) \cap L(\mathscr{A})$. This is a contradiction. Hence the prefix of w belonging to $b_1^* \sqcup\!\sqcup\ g(p) \sqcup\!\sqcup\ s^*$ holds no loop.

If we remove loops from the part of w belonging to $b_\infty^* \sqcup\!\sqcup\ s^*$, then w changes into a shorter word of $L(\mathscr{A}) \cap (S_g \sqcup\!\sqcup\ s^*)$. This is a contradiction. We have proved that the overbraced parts of w contain no loops. Hence their lengths are smaller than n. w is made of $p+1$ parts, whose lengths are at most $n-1$, and p times the letter a_{∞}. Hence its length is at most $pn+n-1$. We have $|g(p)|\leqq n-1$. Hence $p\leqq\tilde{g}(n)$. Thus in this case we have

$$|w|\leqq n-1+\tilde{g}(n)\,n.$$

The results in the two cases, we have looked at, can be summarized by

$$|w|\leqq\max(\bar{\rho}_{U_g}(n),\ n-1+\tilde{g}(n)\,n).$$

Hence

$$\bar{\rho}_{L_g}(n)\leqq\max(\bar{\rho}_{U_g}(n),\ n-1+\tilde{g}(n)\,n). \tag{8}$$

4. Value of ρ_{L_g}

Since $U_g\leqq S_{\neq}$ proposition 1 yields

$$\bar{\rho}_{U_g}(n)\in O(n),$$

while lemma 11 states $\lim_{n\to\infty}\tilde{g}(n)=\infty$. Hence

$$\bar{\rho}_{U_g}(n)\in o(n-1+\tilde{g}(n)\,n).$$

Hence for large enough n we have

$$\bar{\rho}_{U_g}(n)<n-1+\tilde{g}(n)\,n.$$

Hence (7) and (8) and theorem 4 yield

$$\rho_{L_g}(n)=\bar{\rho}_{L_g}(n)=n-1+\tilde{g}(n)\,n \quad \text{for large enough } n.$$

We have proved the theorem:

THEOREM 6: *If g is a structure function, then L_g is a context-free language, whose rational index is*

$$\rho_{L_g}(n)=\bar{\rho}_{L_g}(n)=n-1+\tilde{g}(n)\,n \quad \textit{for large enough } n.$$

DEFINITION 15: *If k is a integer greater than* 1, *then L_{f_k} will be denoted by L_k for simplicity.*

According to theorem 6, the language L_k is a context-free language, whose rational index is

$$\rho_{L_k}(n)=\bar{\rho}_{L_k}(n)=n-1+\lfloor\sqrt[k]{n}\rfloor\,n \quad \textit{for large enough } n.$$

$$\rho_{L_k}(n)\sim n^{1+1/k}.$$

The following section is concerned with relationship between domination of structure functions and domination of their related languages.

5. Comparison of the various L_g.

THEOREM 7: *Let f and g be two structure functions. If $f \geqq g$ then $L_f \geqq L_g$.*

Proof: Using the rational transduction $\varphi_{f,\,g}: X^* \to Y^*$, we shall build a rational transduction φ' such that

$$\varphi'(L_f) = L_g. \tag{9}$$

If $w \in F_f$ then it belongs to $(b_1^* \quad w_1) w_2$ for some unique $w_1 \in X^*$ and $w_2 \in (a_\infty b_\infty^*)^*$ and we define $\varphi'(w)$ to be $(b_1^* \quad \varphi_{f,\,g}(w_1)) w_2$.

If $w \notin F_f$ then we define $\varphi'(w)$ to be $\varnothing$. Since $\varphi_{f,\,g}$ is a rational transduction and F_f is a regular language, it follows that φ' is a rational transduction. The properties of $\varphi_{f,\,g}$ yield properties of φ':

- $\varphi_{f,\,g}(X^*) = Y^*$ hence $\varphi'(F_f) = F_g$.
- If $w_1 \in X^*$ and $w'_1 \in \varphi_{f,\,g}(w_1)$ then $|w_1|_{x_f} = |w'_1|_{x_g}$ hence $\varphi'(E_f) = E_g$.
- $\varphi_{f,\,g}(X^* - f(\mathbb{N}_+)) = Y^* - g(\mathbb{N}_+)$ hence $\varphi'(U_f) = U_g$.
- These last two points prove (9). □

We shall use the notation $\tilde{f}(n) \in O(\tilde{g}(O(n))$. It means that $\tilde{f}(n) \in O(\tilde{g}(h(n)))$ for some function $h \in O(n)$. In other words

$$\exists h: \mathbb{N}_+ \to \mathbb{N}_+, \quad \exists c > 0, \quad \exists n_0, \quad \forall n > n_0, \qquad h(n) \leqq cn \text{ and } \tilde{f}(n) \leqq c\tilde{g}(h(n)).$$

Eliminating h yields

$$\exists c > 0, \quad \exists n_0, \quad \forall n > n_0, \qquad \tilde{f}(n) \leqq c \max_{i \in [0,\, cn]} \tilde{g}(i).$$

Since $\tilde{g}$ is increasing, it becomes

$$\exists c > 0, \quad \exists n_0, \quad \forall n > n_0, \qquad \tilde{f}(n) \leqq c\tilde{g}(cn),$$

or in other words, for some positive c and large enough n we have $\tilde{f}(n) \leqq c\tilde{g}(cn)$. We can also write

$$\exists c > 0, \quad \limsup_{n \to \infty} \tilde{f}(n)/\tilde{g}(cn) < \infty.$$

Anyway, it is simpler to write $\tilde{f}(n) \in O(\tilde{g}(O(n))$ since it saves quantificators.

Similarly $\tilde{f}(n) \in o(\tilde{g}(O(n))$ means

$$\exists c>0, \quad \lim_{n \to \infty} \tilde{f}(n)/\tilde{g}(cn)=0,$$

or

$$\exists c>0, \quad \forall c'>0, \quad \exists n_0, \quad \forall n>n_0, \quad \tilde{f}(n) \leqq c' \tilde{g}(cn).$$

LEMMA 15: *Let f and g be two structure functions. If* $L_f \leqq L_g$, *then* $\tilde{f}(n) \in O(\tilde{g}(O(n)))$ [i. e. *for some c and for large enough n we have* $\tilde{f}(n) \leqq c\tilde{g}(cn)$].

Proof: According to theorem 6,

$$\bar{\rho}_{L_g}(n)=n-1+\tilde{g}(n)\,n \qquad \text{and} \qquad \bar{\rho}_{L_f}(n)=n-1+\tilde{f}(n)\,n$$

for large enough n. Since $L_f \leqq L_g$, theorem 3 proves that for some integer c we have

$$\forall n \in \mathbb{N}_+, \qquad \bar{\rho}_{L_f}(n) \leqq \bar{\rho}_{L_g}(cn).$$

So that for large enough n we have $n-1+\tilde{f}(n)\,n \leqq cn-1+\tilde{g}(cn)\,cn$ *i.e.* $\tilde{f}(n) \leqq c-1+\tilde{g}(cn)\,c$, which proves that $\tilde{f}(n)<2\,c\tilde{g}(cn)$, since $\tilde{g}(cn) \geqq 1$. □

Theorem 7 and lemma 15 combine immediatly into the lemma:

LEMMA 16: *Let f and g be two structure functions. If* $f \leqq g$ *then* $\tilde{f}(n) \in O(\tilde{g}(O(n))$.

LEMMA 17: *Let f and g be two partial increasing functions from* $\mathbb{N}_+$ *to* $\mathbb{N}_+$. *The three following properties cannot all be true.*

- *For some integer d,* $f(n) \in O(n^d)$.
- $g(n) \in o(f(O(n)))$.
- $f(n) \in O(g(O(n)))$.

Proof: Let assume all the three properties to be true. The last two properties result in $f(n) \in O(o(f(O(O(n)))))=o(f(O(n)))$. Since f is increasing, this means that for some positive integer c we have $\lim_{n \to \infty} f(cn)/f(n)=\infty$. So that we can find an integer n_0 such that for any $n \geqq n_0$, we have $f(cn)/f(n) \geqq 2\,c^d$. Then we can inductively prove that for any positive integer l we have $f(c^l n_0) \geqq 2^l c^{ld} f(n_0)$, so that

$$\lim_{l \to \infty} f(c^l n_0)/(c^l n_0)^d=\infty,$$

and thus $\limsup_{n \to \infty} f(n)/n^d = \infty$. This is contrary to the first property. □

Theorem 7 has the corollary:

THEOREM 8: *Let f and g be two structure functions. If $f > g$ then $L_f > L_g$.*

Proof: $f \geqq g$, hence $L_f \geqq L_g$. $\tilde{f}$ and $\tilde{g}$ are two increasing positive partial functions, verifying $\tilde{g} \in o(\tilde{f})$ and $\tilde{f}(n) \leqq n$. So that according to lemma 17, we cannot have $\tilde{f}(n) \in O(\tilde{g}(O(n)))$. Lemma 15 yields then that $L_g \ngeqq L_f$. □

For instance if $k \geqq 2$ then $L_{k+1} < L_k$.

VI. THE LANGUAGE RELATED TO A FINITE SEQUENCE OF STRUCTURE FUNCTIONS

The purpose of this section is to build for every finite sequence of structure functions $g_1, \ldots, g_e$ a context-free language whose rational index is $\Theta\left(n \prod_{i=1}^{e} \tilde{g}_i(n)\right)$. Hence it will follow that for every sequence $k_1, \ldots, k_e$ of integers greater than 1, the sequence of structure functions $f_{k_1}, \ldots, f_{k_e}$ yields a context-free language, whose rational index is $\Theta(n^{1+1/k_1+\ldots+1/k_e})$, so that for every rational number λ greater than 1, we can find a context-free language whose rational index is $\Theta(n^\lambda)$.

In order to avoid a lot of subscripts and ellipses (« . . . ») and to make the proofs clearer, we shall first handle a sequence f, g, h of three structure functions, and then we shall generalize the results to any sequence of structure functions.

1. Definition of $L_{f,g,h}$

Let $f: \mathbb{N}_+ \to X^*$, $g: \mathbb{N}_+ \to Y^*$ and $h: \mathbb{N}_+ \to Z^*$ be three structure functions. We assume that X, Y, Z and $\{b_1, a_2, b_2, a_3, b_3, a_\infty, b_\infty, \#\}$ are four disjoint alphabets. $L_{f,g,h}$ will be a language on the alphabet

$$X \cup Y \cup Z \cup \{b_1, a_2, b_2, a_3, b_3, a_\infty, b_\infty\},$$

but to define it we shall use the larger alphabet

$$\Omega = X \cup Y \cup Z \cup \{b_1, a_2, b_2, a_3, b_3, a_\infty, b_\infty, \#\}.$$

Let $A \subset \Omega^*$ and $B \subset \Omega^*$ be two languages and i be an integer greater than 1. We define $A \uparrow_i B$ to be the set of the words of A in which every factor $a_\infty b_\infty^*$ is replaced by a word of $a_i B$, in which every occurence of b_1 is replaced by an occurence of b_i. More precisely $A \uparrow_i B = \tau_{\uparrow iB}(A)$ where $\tau_{\uparrow iB}$ is the substitution defined by:

$$\tau_{\uparrow iB}(b_\infty) = \varepsilon$$

$$\tau_{\uparrow iB}(a_\infty) = a_i \varphi_{b_1, b_i}(B)$$

$$\tau_{\uparrow iB}(x) = x \quad \text{for any other letter}$$

where φ_{b_1, b_i} is the strictly alphabetic morphism, which replaces b_1 with b_i and keeps the other letters unchanged. $\uparrow$ has interesting obvious properties:

- $\uparrow$ is associative: For any languages A, B and C and any integers i and j greater than 1, the two languages $(A \uparrow_i B) \uparrow_j C$ and $A \uparrow_i (B \uparrow_j C)$ are equal, so that we can denote them $A \uparrow_i B \uparrow_j C$.
- If A and B are context-free languages, then so is $A \uparrow_i B$.
- If B is a regular language, then $A \uparrow_i B \leqq A$.
- If A and B are both regular languages, then so is $A \uparrow_i B$.

At last we define $\tau_\#$ to be the rational transduction, which keeps words containing at least one $\#$ and then erases all the $\#$ in the kept words. *I.e.* if $A \subset \Omega^*$ then $\tau_\#(A) = \tau_{\{\#\}}(A \cap \Omega^* \# \Omega^*)$. For instance

$$\tau_\#(\{dbc, dbb\#c, \#cb\#b\}) = \{dbbc, cbb\}.$$

We can now define $L_{f, g, h}$. As L_g is a subset of its frame $F_g = (b_1^* \sqcup\!\sqcup X^*)(a_\infty b_\infty^*)^*$, similarly $L_{f, g, h}$ will be a subset of its frame, which is to be the regular language

$$F_{f, g, h} = F_f \uparrow_2 F_g \uparrow_3 F_h = (b_1^* \sqcup\!\sqcup X^*)(a_2 (b_2^* \sqcup\!\sqcup Y^*)\, a_3 (b_3 \sqcup\!\sqcup Z^*)(a_\infty b_\infty^*)^*)^*)^*.$$

We define the structured part of $L_{f, g, h}$ to be

$$S_{f, g, h} = S_f \uparrow_2 S_g \uparrow_3 S_h$$

and the extended structured part of $L_{f, g, h}$ to be

$$E_{f, g, h} = E_f \uparrow_2 E_g \uparrow_3 E_h.$$

$S_{f, g, h}$ is not a context-free language, but $E_{f, g, h}$ is.

We define $U_{f, g, h}$, the unstructured part of $L_{f, g, h}$, to be the set of the words w in $F_f \uparrow_2 F_g \uparrow_3 F_h$ such that at least one of the words of F_f, F_g and

F_h involved in the construction of w is unstructured, *i. e.*

$$\begin{aligned} U_{f,g,h} &= \tau_{\#}((F_f \cup \# U_f) \uparrow_2 (F_g \cup \# U_g) \uparrow_3 (F_h \cup \# U_h)) \\ &= \tau_{\#}(((F_f \cup \# U_f) \uparrow_2 F_g \uparrow_3 F_h) \cup (F_f \uparrow_2 (F_g \cup \# U_g) \uparrow_3 F_h) \\ &\cup (F_f \uparrow_2 F_g \uparrow_3 (F_h \cup \# U_h))) \\ &= (U_f \uparrow_2 F_g \uparrow_3 F_h) \\ &\cup \tau_{\#}(F_f \uparrow_2 (F_g \cup \# U_g) \uparrow_3 F_h) \\ &\cup \tau_{\#}(F_f \uparrow_2 F_g \uparrow_3 (F_h \cup \# U_h)) \end{aligned} \tag{10}$$

Conversely $F_{f,g,h} - U_{f,g,h}$ is made of the words w belonging to $F_f \uparrow_2 F_g \uparrow_3 F_h$ such that none of the words of F_f, F_g and F_h involved in the construction of w is unstructured. *I. e.*

$$E_{f,g,h} - U_{f,g,h} = (F_f - U_f) \uparrow_2 (F_g - U_g) \uparrow_3 (F_h - U_h).$$

Hence

$$\begin{aligned} E_{f,g,h} - U_{f,g,h} &= E_{f,g,h} \cap (F_{f,g,h} - U_{f,g,h}) \\ &= (E_f \uparrow_2 E_g \uparrow_3 E_h) \cap ((F_f - U_f) \uparrow_2 (F_g - U_g) \uparrow_3 (F_h - U_h)) \\ &= (E_f \cap (F_f - U_f)) \uparrow_2 (E_g \cap (F_g - U_g)) \uparrow_3 (E_h \cap (F_h - U_h)) \\ &= S_f \uparrow_2 S_g \uparrow_3 S_h \\ &= S_{f,g,h}. \end{aligned}$$

DEFINITION 16: *The above definitions of $S_{f,g,h}$, $E_{f,g,h}$ and $U_{f,g,h}$ allow us to define $L_{f,g,h}$ as the union of its extended structured part and its unstructured part, and it is also the disjoint union of its structured part and its unstructured part.*

$$L_{f,g,h} = E_{f,g,h} \cup U_{f,g,h} = S_{f,g,h} \sqcup U_{f,g,h}.$$

Figure 2 still holds. U_f, U_g and U_h are dominated by $S_{\neq}$ and F_f, F_g and F_h are regular languages, hence (10) proves that $U_{f,g,h} \leqq S_{\neq}$. Hence $L_{f,g,h}$ is a context-free language.

We can express $L_{f,g,h}$ in an another way. $F_{f,g,h}$ is the union of the sets

$$(b_1^* \sqcup\!\sqcup \alpha) \prod_{i=1}^{p} \left(a_2 (b_2^* \sqcup\!\sqcup \beta_i) \prod_{j=1}^{q_i} (a_3 (b_3^* \sqcup\!\sqcup \gamma_{i,j}) (a_\infty b_\infty^*)^{r_{i,j}}) \right)$$

where

$$p \in \mathbb{N}, \qquad \alpha \in X^*,$$
$$q_i \in \mathbb{N}, \qquad \beta_i \in Y^* \qquad \text{for} \quad 1 \leqq i \leqq p,$$
$$r_{i,j} \in \mathbb{N}, \qquad \gamma_{i,j} \in Z^* \qquad \text{for} \quad 1 \leqq i \leqq p \quad \text{and} \quad 1 \leqq j \leqq q_i.$$

$U_{f,g,h}$ is made of those sets verifying the condition

$$\alpha \in X^* - f(\mathbb{N}_+)$$

or

$$\exists i, \quad \beta_i \in Y^* - g(\mathbb{N}_+) \tag{C_u}$$

or

$$\exists i, \quad \exists j, \qquad \gamma_{i,j} \in Z^* - h(\mathbb{N}_+)$$

$E_{f,g,h}$ is made of the sets verifying the condition

$$|\alpha|_{x_f} + 1 = p$$

and

$$\forall i, \quad |\beta_i|_{x_g} + 1 = q_i \tag{C_e}$$

and

$$\forall i, \quad \forall j, \qquad |\gamma_{i,j}|_{x_h} + 1 = r_{i,j}$$

$L_{f,g,h}$ is made of the sets verifying at least one of the two conditions (C_e) and (C_u). $S_{f,g,h}$ is made of the sets verifying (C_e) but not (C_u) *i.e.*

$$\alpha = f(r)$$

and

$$\forall i, \quad \beta_i = g(q_i) \tag{C_s}$$

and

$$\forall i, \quad \forall j, \qquad \gamma_{i,j} = h(r_{i,j})$$

Hence

$$S_{f,g,h} = \bigcup_{p \in \mathbb{N}_+} (b_1^* \sqcup\!\sqcup f(p)) \prod_{i=1}^{p} \left(a_2 \bigcup_{q_i \in \mathbb{N}_+} (b_2^* \sqcup\!\sqcup g(q_i)) \right.$$
$$\left. \prod_{j=1}^{q_i} \left(a_3 \bigcup_{r_{i,j} \in \mathbb{N}_+} (b_3^* \sqcup\!\sqcup h(r_{i,j}))(a_\infty b_\infty^*)^{r_{i,j}} \right) \right)$$

2. Lower bound on $\rho_{L_{f,\,g,\,h}}$

Let n be a large enough integer such that the three integers $p=\tilde{f}(n)$, $q=\tilde{g}(n)$ and $r=\tilde{h}(n)$ exist. We want to obtain a lower bound on $\rho_{L_{f,\,g,\,h}}(n)$. Let $\mathscr{A}$ be the automaton depicted in figure 4.

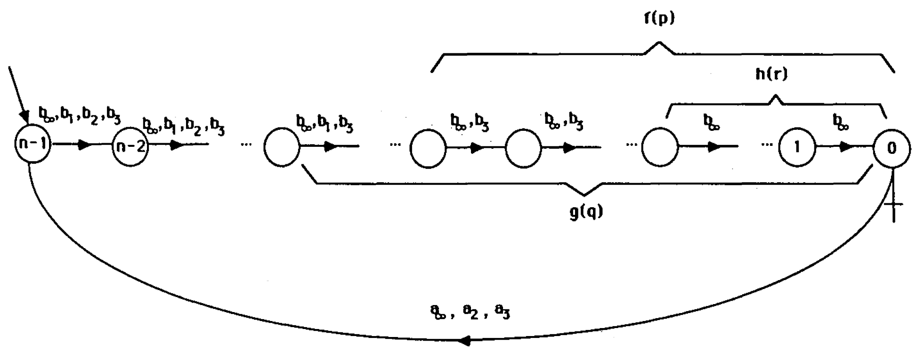

Figure 4.

This automaton has n states. It is made of a simple path of length $n-1$ leading from the only initial state to the only final state. Every arc of this path is labeled by four letters in such a way that the path is labeled by each of the four words $b_1^{n-1-|f(p)|}f(p)$, $b_2^{n-1-|g(q)|}g(q)$,$b_3^{n-1-|h(r)|}h(r)$ and b_∞^{n-1}. There is also an arc leading from the final state to the initial state labeled by the three letters a_2, a_3 and a_∞. So the set of the words of $F_{f,\,g,\,h}$ that $\mathscr{A}$ recognizes is

$$b_1^{n-1-|f(p)|}f(p)(a_2 b_2^{n-1-|g(q)|}g(q)(a_3 b_3^{n-1-|h(r)|}h(r)(a_\infty b_\infty^{n-1})^*)^*)^*.$$

It is disjoint with $U_{f,\,g,\,h}$, but it has exactly one element of $S_{f,\,g,\,h}$, which is

$$b_1^{n-1-|f(p)|}f(p)(a_2 b^{n-1-|g(q)|}g(q)(a_3 b_3^{n-1-|h(r)|}h(r)(a_\infty b_\infty^{n-1})^r)^q)^p,$$

whose length is $n-1+p(n+q(n+rn))$. Hence

$$\rho_{L_{f,\,g,\,h}}(n)\geqq n-1+\tilde{f}(n)(n+\tilde{g}(n)(n+\tilde{h}(n)n)). \tag{11}$$

3. Upper bound on $\rho_{L_{f,\,g,\,h}}$

Let $n\in\mathbb{N}_+$. Let $\mathscr{A}$ be any automaton with n states recognizing at least one word in $L_{f,\,g,\,h}\sqcup s^*$. Let w be a shortest word in $(L_{f,\,g,\,h}\sqcup s^*)\cap L(\mathscr{A})$. We shall give an upper bound on $|w|$, that depends only on n and not on $\mathscr{A}$ so that it will be also an upper bound on $\bar{\rho}_{L_{f,g,h}}(n)$. Let us consider a successful path γ in $\mathscr{A}$ labeled by w.

• First let us assume that $(U_{f,g,h} \sqcup\!\sqcup\, s^*) \cap L(\mathscr{A}) \neq \varnothing$.

As in the previous section, we can conclude that $|w| \leqq \bar{\rho}_{U_{f,g,h}}(n)$.

• Let us assume now that $U_{f,g,h} \sqcup\!\sqcup\, s^*$ and $L(\mathscr{A})$ are disjoint. Then every word in $(L_{f,g,h} \sqcup\!\sqcup\, s^*) \cap L(\mathscr{A})$ belongs to $S_{f,g,h} \sqcup\!\sqcup\, s^*$. Thus w belongs to $S_{f,g,h} \sqcup\!\sqcup\, s^*$ and

$$w \in \overbrace{(b_1^* \sqcup\!\sqcup\, f(p))}^{|\,.\,|<n}$$

$$\underbrace{\times \prod_{i=1}^{p} \left(a_2 \overbrace{(b_2^* \sqcup\!\sqcup\, g(q_i))}^{|\,.\,|<n} \underbrace{\prod_{j=1}^{q_i} (a_3 \overbrace{(b_3^* \sqcup\!\sqcup\, h(r_{i,j}))}^{|\,.\,|<n} \underbrace{(a_\infty \overbrace{(b_\infty^* \sqcup\!\sqcup\, s^*)}^{|\,.\,|<n})^{r_{i,j}}}_{|\,.\,|\leqq r_{i,j} n})}_{|\,.\,|\leqq q_i (n+\tilde{h}(n)\, n)} \right)}_{|\,.\,|\leqq p\,(n+\tilde{g}(n)\,(n+\tilde{h}(n)\, n))}$$

for some non negative integers $p, q_1, \ldots, q_p, r_{i,1}, \ldots, r_{i,q_i}$ for $1 \leqq i \leqq p$. As in the previous section overbraced parts of w hold no loops. Hence their lengths are smaller than n. As in the previous section we have $|f(p)| \leqq n-1$. Hence $p \leqq \tilde{f}(n)$. Similarly for every i in $\{1, \ldots, r\}$ we have $q_i \leqq \tilde{g}(n)$. And for every i and j we have $r_{i,j} \leqq \tilde{h}(n)$. All of this allows us to compute an upper bound on $|w|$. Indeed:

$$|w| \leqq n-1+\tilde{f}(n)(n+\tilde{g}(n)(n+\tilde{h}(n)\,n)).$$

The results in the two cases, we have looked at, can be summarized by

$$|w| \leqq \max(\bar{\rho}_{U_{f,g,h}}(n),\ n-1+\tilde{f}(n)(n+\tilde{g}(n)(n+\tilde{h}(n)\,n))).$$

This upper bound on $|w|$ is also an upper bound on $\rho_{L_{f,g,h}}(n)$.

4. Value of $\rho_{L_{f,g,h}}$

As in the previous section we can conclude that

$$\bar{\rho}_{L_{f,g,h}}(n) = \rho_{L_{f,g,h}}(n) = n-1 + \tilde{f}(n)(n+\tilde{g}(n)(n+\tilde{h}(n)\,n)) \quad \text{for large enough } n.$$

5. Generalization to more than three levels

In the same way we built $L_{f,g,h}$, we can define the language $L_{g_1,\ldots,g_e}$ for any sequence $g_1, \ldots, g_e$ of structure functions. In order to describe precisely this language we must change slightly the notations used so far. We assume that $g_i : \mathbb{N}_+ \to Y_i^*$ for any $i \in [1, e]$, and that $Y_1 \ldots Y_e$ and $\{b_1, a_2, b_2, \ldots, a_e, b_e, a_\infty, b_\infty, \#\}$ are disjoint. We define

$$\Omega = Y_1 \cup \ldots \cup Y_e \cup \{b_1, a_2, b_2, \ldots, a_e, b_e, a_\infty, b_\infty, \#\}.$$

Indeed these are the notations used so far except for Y_1, Y_2 and Y_3, which were called X, Y and Z.

We define

$$F_{g_1,\ldots,g_e} = F_{g_1} \uparrow_2 \ldots \uparrow_e F_{g_e}$$

$$S_{g_1,\ldots,g_2} = S_{g_1} \uparrow_2 \ldots \uparrow_e S_{g_e}$$

$$E_{g_1,\ldots,g_e} = E_{g_1} \uparrow_2 \ldots \uparrow_e E_{g_e}$$

$$U_{g_1,\ldots,g_e} = \tau_\# ((F_{g_1} \cup \# U_{g_1}) \uparrow_2 \ldots \uparrow_e (F_{g_e} \cup \# U_{g_e}))$$

$$L_{g_1,\ldots,g_e} = E_{g_1,\ldots,g_e} \cup U_{g_1,\ldots,g_e} = S_{g_1,\ldots,g_e} \sqcup U_{g_1,\ldots,g_e}.$$

Obviously the previous results generalize:

THEOREM 9: *If $g_1, \ldots, g_e$ are structure functions on disjoint alphabets, then $F_{g_1,\ldots,g_e}$ is a regular language, $E_{g_1,\ldots,g_e}$ and $L_{g_1,\ldots,g_e}$ are context-free languages, $U_{g_1,\ldots,g_e} \leq S_{\neq}$ and for large enough n we have*

$$\bar{\rho}_{L_{g_1,\ldots,g_e}}(n) = \rho_{L_{g_1,\ldots,g_e}}(n) = n - 1 + \tilde{g}_1(n)(n + \tilde{g}_2(n)(n + \ldots \tilde{g}_e(n)n) \ldots).$$

6. Main example

DEFINITION 17: *For any positive integers i and j we define the alphabet*

$$X_{i,j} = \{x_{1,j}, x_{2,j}, \ldots, x_{i,j}\}.$$

DEFINITION 18: *We define $\iota_{i,j} : X_i^* \to X_{i,j}^*$ to be the strictly alphabetic isomorphism, which adds the second subscript j to every letter. I. e. $\iota_{i,j}(x_l) = x_{l,j}$ for every $l \in [1, i]$.*

DEFINITION 19: *Let $k_1, \ldots, k_e$ be a finite sequence of integers greater than* 1. *Then $L_{k_1, \ldots, k_e}$ will be a short notation for*

$$L_{(\iota_{k_1, 1} \circ f_{k_1}), (\iota_{k_2, 2} \circ f_{k_2}), \ldots, (\iota_{k_e, e} \circ f_{k_e})}.$$

Remarks: This notation is compatible with the notation L_k defined in the previous section to mean L_{f_k} for an integer $k>1$, if we identify X_k and $X_{k,1}$.

– The functions ι's are needed only to ensure, that the structure functions $\iota_{k_1,1} \circ f_{k_1}, \iota_{k_2,2} \circ f_{k_2}, \ldots, \iota_{k_e,e} \circ f_{k_e}$ use disjoint alphabets $(X_{k_1,1}, \ldots, X_{k_e,e})$.

Theorem 9 yields that $L_{k_1, \ldots, k_e}$ is a context-free language, whose rational index is

$$\bar{\rho}_{L_{k_1, \ldots, k_e}}(n) = \rho_{L_{k_1, \ldots, k_2}}(n) = n - 1 + \lfloor \sqrt[k_1]{n} \rfloor (n + \lfloor \sqrt[k_2]{n} \rfloor (n + \ldots \lfloor \sqrt[k_e]{n} \rfloor n) \ldots)$$

for large enough n. So that

$$\bar{\rho}_{L_{k_1, \ldots, k_e}}(n) = \rho_{L_{k_1, \ldots, k_e}}(n) \sim n^{1 + 1/k_1 + \ldots + 1/k_e}.$$

THEOREM 10: *Let $r \in \mathbb{Q} \cap [1, +\infty[$. Then there exists a context-free language L such that $\rho_L(n) = \bar{\rho}_L(n) \in \Theta(n^r)$.*

Proof: If $r=1$ then $L = S_{\neq}$ works, since $\rho_{S_{\neq}}(n) = \bar{\rho}_{S_{\neq}}(n) = 2n - 1 \in \Theta(n)$.

• Let us assume $r>1$. Then $r = p/q$ for some integers p and q such that $0<q<p$. Hence $r = 1 + (p-q)\, 1/q$ and we can choose $L = L_{\underbrace{q, \ldots, q}_{p-q \text{ times}}}$. □

We study now the domination between the various $L_{g_1, \ldots, g_e}$. The three following theorems will provide an easy way to build infinite strictly increasing or strictly decreasing sequences of context-free languages.

THEOREM 11: *Let $g_1, \ldots, g_e$ and $h_1, \ldots, h_e$ be two sequences of structure functions. If $g_i \geqq h_i$ for all i, then $L_{g_1, \ldots, g_e} \geqq L_{h_1, \ldots, h_e}$, if these two languages exist.*

Proof: Let us assume that $g_i : \mathbb{N}_+ \to Y_i^*$ and $h_i : \mathbb{N}_+ \to Z_i^*$ for $i = 1, \ldots, e$. The existence of $L_{g_1, \ldots, g_e}$ means, that the $e+1$ alphabets $\{ b_1, a_2, b_2, \ldots, a_e, b_e, a_\infty, b_\infty, \# \}$ and $Y_1, \ldots, Y_e$ are disjoint. Similarly, the existence of $L_{h_1, \ldots, h_e}$ means, that the $e+1$ alphabets $Z_1, \ldots, Z_e$ and $\{ b_1, a_2, b_2, \ldots, a_e, b_e, a_\infty, b_\infty, \# \}$ are disjoint.

For every i in $\{ 1, \ldots, e \}$, we have $g_i \geqq h_i$. This means, by definition, the existence of a rational transduction $\sigma_{g_i, h_i} : Y_i^* \to Z_i^*$ with some properties. We define the rational transduction $\sigma_i : b_i^* \sqcup\!\sqcup Y_i^* \to b_i^* \sqcup\!\sqcup Z_i^*$ such that

$$\sigma_i = \pi_{\{b_i\}}^{-1} \circ \sigma_{g_i, h_i} \circ \pi_{\{b_i\}}.$$

It is the rational transduction which maps every word in $b_i^* \sqcup\!\sqcup w$ onto $b_i^* \sqcup\!\sqcup \sigma_{g_i, h_i}(w)$ for every word $w \in Y_i^*$.

We define

$$\Omega_g = Y_1 \sqcup \ldots \sqcup Y_e \sqcup \{ b_1, a_2, b_2, \ldots, a_e, b_e, a_\infty, b_\infty, \# \}$$

and

$$\Omega_h = Z_1 \sqcup \ldots \sqcup Z_e \sqcup \{ b_1, a_2, b_2, \ldots, a_e, b_e, a_\infty, b_\infty, \# \}.$$

We are now ready to define the rational transduction $\sigma'' : \Omega_g^* \to \Omega_h^*$ such that

$$\sigma''(L_{g_1, \ldots, g_e}) = L_{h_1, \ldots, h_e}.$$

- If $w \in \Omega_g^* - F_{g_1, \ldots, g_e}$ then $\sigma''(w) = \varnothing$.
- Let us assume now that $w \in F_{g_1, \ldots, g_e}$. Then we have

$$w \in \alpha \prod_{i_2=1}^{p} \left(a_2 \alpha_{i_2} \prod_{i_3=1}^{p_{i_2}} \left(a_3 \alpha_{i_2, i_3} \prod_{i_4=1}^{p_{i_2, i_3}} \times \left(\ldots \prod_{i_e=1}^{p_{i_2, \ldots, i_{e-1}}} (a_e \alpha_{i_2, \ldots, i_e} (a_\infty b_\infty^*)^{p_{i_2, \ldots, i_e}}) \ldots \right)\right)\right)$$

where

$$p \in \mathbb{N}, \qquad \alpha \in b_1^* \sqcup\!\sqcup Y_1^*$$

$$p_{i_2} \in \mathbb{N}, \qquad \alpha_{i_2} \in b_2^* \sqcup\!\sqcup Y_2^* \quad \text{for } 1 \leqq i_2 \leqq p,$$

$$p_{i_2, i_3} \in \mathbb{N}, \qquad \alpha_{i_2, i_3} \in b_3^* \sqcup\!\sqcup Y_3^* \quad \text{for } 1 \leqq i_2 \leqq p \quad \text{and} \quad 1 \leqq i_3 \leqq p_{i_2},$$

$$\vdots$$

$$p_{i_2, \ldots, i_e} \in \mathbb{N}, \qquad \alpha_{i_2, \ldots, i_e} \in b_e^* \sqcup\!\sqcup Y_e^* \qquad \text{for} \quad 1 \leqq i_2 \leqq p,$$

$$1 \leqq i_3 \leqq p_{i_2}, \ldots, 1 \leqq i_{e+1} \leqq p_{i_2, \ldots, i_e}.$$

Then we define

$$\sigma''(w)=\sigma_1(\alpha)\prod_{i_2=1}^{p}\left(a_2\,\sigma_2(\alpha_{i_2})\prod_{i_3=1}^{p_{i_2}}\left(a_3\,\sigma_3(\alpha_{i_2,i_3})\prod_{i_4=1}^{p_{i_2,i_3}}\right.\right.$$

$$\left.\left.\times\left(\ldots\prod_{i_e=1}^{p_{i_2,\ldots,i_{e-1}}}(a_e\,\sigma_e(\alpha_{i_2,\ldots,i_e})(a_\infty\,b_\infty^*)^{p_{i_2,\ldots,i_e}})\ldots\right)\right)\right).$$

The graph of the transduction σ'' is

$$\Sigma''=\Sigma_1((a_2,a_2)\Sigma_2((a_3,a_3)\Sigma_3(\ldots((a_e,a_e)\Sigma_e(a_\infty b_\infty^*\times a_\infty b_\infty^*)^*)^*\ldots)^*)^*)^*.$$

where Σ_i denotes the graph of the rational transduction σ_i. The product of the two regular sets $a_\infty b_\infty^*\times a_\infty b_\infty^*=(a_\infty,\varepsilon)(b_\infty,\varepsilon)^*(\varepsilon,a_\infty)(\varepsilon,b_\infty)^*$ and the graphs of rational transductions $\Sigma_1,\ldots,\Sigma_i$ are rational subsets of $\Omega_g^*\times\Omega_h^*$ and so Σ'' too. This proves that σ'' is a rational transduction.

As in the proof of theorem 7 the properties of the σ_i's result in $\sigma''(U_{g_1,\ldots,g_e})=U_{h_1,\ldots,h_e}$ and $\sigma''(E_{g_1,\ldots,g_2})=E_{h_1,\ldots,h_e}$ hence $\sigma''(L_{g_1,\ldots,g_e})=L_{h_1,\ldots,h_e}$ and $L_{g_1,\ldots,g_e}\geqq L_{h_1,\ldots,h_e}$. □

Theorem 11 has the corollary:

THEOREM 12: *Let $g_1,\ldots,g_e$ and $h_1,\ldots,h_e$ be two sequences of structure functions on disjoint alphabets such that $g_i\leqq h_i$ for all i, and $g_{i_0}<h_{i_0}$ for some i_0. Then $L_{g_1,\ldots,g_e}<L_{h_1,\ldots,h_e}$.*

Proof: This theorem can be proved in the same way as theorem 8:

$$\bar{\rho}_{L_{g_1,\ldots,g_e}}(n)\sim n\prod_{i=1}^{e}\tilde{g}_i(n)$$

$$\bar{\rho}_{L_{h_1,\ldots,h_e}}(n)\sim n\prod_{i=1}^{e}\tilde{h}_i(n).$$

For all i, since $g_i\leqq h_i$, lemma 16 yields

$$\tilde{g}_i(n)\in O(\tilde{h}_i(O(n)))$$

For i_0 we have

$$\tilde{g}_{i_0}(n)\in o(\tilde{h}_{i_0}(n)).$$

These facts result in

$$\bar{\rho}_{L_{g_1, \ldots, g_e}}(n) \in o(\bar{\rho}_{L_{h_1, \ldots, h_e}}(O(n))).$$

On the other hand we have $\bar{\rho}_{L_{h_1, \ldots, h_e}}(n) \in O(n^{e+1})$.

Lemma 17 yields then that $\bar{\rho}_{L_{h_1, \ldots, h_e}}(n) \notin O(\bar{\rho}_{L_{g_1, \ldots, g_e}}(O(n)))$ so that lemma 15 yields that $L_{g_1, \ldots, g_e} \not\geq L_{h_1, \ldots, h_e}$. □

Hence, if $k_1, \ldots, k_e$ and $l_1, \ldots, l_e$ are two different sequences of integers, such that for all i we have $2 \leq k_i \leq l_i$, then $L_{k_1, \ldots, k_e} > L_{l_1, \ldots, l_e}$.

NOTATION: Let $(g_1, \ldots, g_e)$ be a finite sequence of length e. We shall denote by $(g_1, \ldots, \hat{g}_{e'}, \ldots, g_e)$ the finite sequence of length $e-1$ obtained by the removal of $g_{e'}$.

THEOREM 13: *Let e be an integer greater than 1. Let $g_1, \ldots, g_e$ be a sequence of structure functions. Let $e' \in \{1, \ldots, e\}$. Then*

$$L_{g_1, \ldots, g_e} > L_{g_1, \ldots, \hat{g}_{e'}, \ldots, g_e}.$$

Proof: We shall only prove this theorem in the case $e=4$ and $e'=2$. The proof is similar in the general case.

Let $f: \mathbb{N}_+ \to X^*$, $g: \mathbb{N}_+ \to Y^*$, $h: \mathbb{N}_+ \to Z^*$ and $l: \mathbb{N}_+ \to T^*$ be four structure functions, such that X, Y, Z, T and $\{b_1, a_2, b_2, a_3, b_3, a_4, b_4, a_\infty, b_\infty, \#\}$ are five disjoint alphabets. We shall prove that

$$L_{f, g, h, l} > L_{f, h, l}.$$

For that we choose a word w_1 in $a_3 S_h \uparrow_4 S_l$ and a positive integer n_g such that $g(n_g)$ exists. Then we transform every word belonging to $L_{f, g, h, l} \cap F_f \uparrow_2 (g(n_g) w_1^{n_g - 1} a_3 F_h \uparrow_4 F_l)$ into a word of $F_f \uparrow_2 F_h \uparrow_3 F_l$ by removing all the factors of the form $g(n_g) w_1^{n_g - 1} a_3$ and then by decreasing by one the subscripts of the letters b_3, a_4 and b_4. The removed factors follow the occurrences of a_2.

Indeed this transformation is a bijection from

$$L_{f, g, h, l} \cap F_f \uparrow_2 (g(n_g) w_1^{n_g - 1} a_3 F_h \uparrow_4 F_l)$$

onto $L_{f, h, l}$, and it can be performed by the reciprocal of a morphisme φ.

Let us detail this. Let us define

$$\Omega = X \sqcup Y \sqcup Z \sqcup T \sqcup \{b_1, a_2, b_2, a_3, b_3, a_4, b_4, a_\infty, b_\infty, \#\}.$$

Let n_g (resp. n_h and n_l) be the least integer, for which g (resp. h and l) is defined. Let

$$w_1 = a_3\, h(n_h)\,(a_4\, l(n_l)\, a_\infty^{n_l})^{n_h}$$

be the word in $a_3\, S_h \uparrow_4 S_l$ having a minimal number of occurrences of a_∞.

Let

$$w_2 = g(n_g)\, w_1^{n_g - 1}\, a_3.$$

w_2 has been chosen such that

$$\begin{aligned}
\forall u \in \Omega^*, &\quad w_2\, u \in (S_g \uparrow_3 S_h \uparrow_4 S_l) &\Leftrightarrow&\quad u \in (S_h \uparrow_4 S_l),\\
\forall u \in \Omega^*, &\quad w_2\, u \in (U_g \uparrow_3 U_h \uparrow_4 U_l) &\Leftrightarrow&\quad u \in (U_h \uparrow_4 U_l),\\
\forall u \in \Omega^*, &\quad w_2\, u \in (E_g \uparrow_3 E_h \uparrow_4 E_l) &\Leftrightarrow&\quad u \in (E_h \uparrow_4 E_l),\\
\forall u \in \Omega^*, &\quad w_2\, u \in (F_g \uparrow_3 F_h \uparrow_4 F_l) &\Leftrightarrow&\quad u \in (F_h \uparrow_4 F_l).
\end{aligned}$$

We define the morphism

$$\varphi:\quad (X \sqcup Z \sqcup T \sqcup \{b_1, a_2, b_2, a_3, b_3, a_\infty, b_\infty, \#\})^* \to \Omega^*$$

by

$$\begin{aligned}
\varphi(x) = x \quad &\text{if}\quad x \in (X \sqcup Z \sqcup T)\\
\varphi(b_1) &= b_1\\
\varphi(a_2) &= a_2\, w_2\\
\varphi(b_2) &= b_3\\
\varphi(a_3) &= a_4\\
\varphi(b_3) &= b_4\\
\varphi(a_\infty) &= a_\infty\\
\varphi(b_\infty) &= b_\infty.
\end{aligned}$$

Then obviously

$$\begin{aligned}
\varphi^{-1}(F_{f,g,h,l}) &= F_{f,h,l},\\
\varphi^{-1}(S_{f,g,h,l}) &= S_{f,h,l},\\
\varphi^{-1}(U_{f,g,h,l}) &= U_{f,h,l},\\
\varphi^{-1}(E_{f,g,h,l}) &= E_{f,h,l},\\
\varphi^{-1}(L_{f,g,h,l}) &= L_{f,h,l}.
\end{aligned}$$

So that $L_{f,g,h,l} \geqq L_{f,h,l}$. On the other hand we have

$$\bar{\rho}_{L_{f,g,h,l}}(n) \sim \bar{\rho}_{L_{f,h,l}}(n)\,\tilde{g}(n)$$

so that

$$\bar{\rho}_{L_{f,h,l}}(n) \in o(\bar{\rho}_{L_{f,g,h,l}}(n))$$

and we can conclude as in proof of theorem 12, that $L_{f,h,l} \ngeqq L_{f,g,h,l}$. □

E. g. let $k_1, \ldots, k_e$ be a sequence of integers greater than 1. Let $e' \in \{1, \ldots, e\}$. Then $L_{k_1, \ldots, k_e} > L_{k_1, \ldots, \widehat{k_{e'}}, \ldots, k_e}$.

VII. OTHER EXAMPLES OF STRUCTURE FUNCTIONS

1. First example: a structure function leading to a context-free language whose rational index is $\Theta(n \ln n)$

DEFINITION 20: *Let* $X_{\exp} = \{a, b\}$ *and*

$$f_{\exp}: \quad \mathbb{N}_+ \to X^*_{\exp}$$

$$i \mapsto bab^1\, ab^3\, ab^7 \ldots ab^{2^{i-1}-1} = b \prod_{j=1}^{i-1} ab^{2^j-1}.$$

I. e.

$$f_{\exp}(1) = b$$

$$f_{\exp}(2) = bab$$

$$f_{\exp}(3) = babab^3$$

$$f_{\exp}(4) = babab^3\, ab^7$$

$$f_{\exp}(i+1) = f_{\exp}(i)\, ab^{|f_{\exp}(i)|}.$$

Let us show that $f_{\exp}$ is a structure function and $x_{f_{\exp}} = a$:

- $X^*_{\exp} - f_{\exp}(\mathbb{N}_+) = (X^*_{\exp} - b(ab^*)^*) \cup \nabla_{\neq}(X^*_{\exp}, |\cdot|, a, |\cdot|, b^*)(ab^*)^*$ so that according to lemma 9 $X^*_{\exp} - f_{\exp}(\mathbb{N}_+) \leqq S_{\neq}$.
- $\forall i \in \mathbb{N}_+, |f_{\exp}(i)|_a = i - 1$.
- $\forall i \in \mathbb{N}_+, |f_{\exp}(i)| = 2^i - 1$, so that

$$\lim_{i \to \infty} |f_{\exp}(i)|/i = \infty \qquad \text{and} \qquad \tilde{f}_{\exp}(n) = \lfloor \ln_2 n \rfloor.$$

Theorem 6 yields that $L_{f_{\exp}}$ is a context-free language and for large enough n we have

$$\rho_{L_{f_{\exp}}}(n)=\bar{\rho}_{L_{f_{\exp}}}(n)=n-1+n\tilde{f}_{\exp}(n)=n-1+n\lfloor \ln_2 n\rfloor \sim n\ln_2 n.$$

2. Second example: a structure function leading to a context-free language whose rational index is $\Theta(n\ln\ln n)$

Let us define a new notation in order to express the next examples.

DEFINITION 21: *If $i\in\mathbb{N}_+$ and w is a word, such that $|w|\leqq 2^{i-1}-2$, then we define*

$$F_{\exp}(i,w)=f_{\exp}(i)\,b^{-|w|-1}\,cw,$$

i. e. *a copy of $f_{\exp}(i)$ in which we have replaced the suffix $b^{|w|+1}$ with cw. If $|w|>2^{i-1}-2$ then $f_{\exp}(i)$ ends with too few b's and $F_{\exp}(i,w)$ is not defined.*

E.g. $F_{\exp}(4,d^2 f_{\exp}(2))=babab^3abcd^2bab$ and $F_{\exp}(3,d^2 f_{\exp}(2))$ is not defined.

Hence, in particular

$$|F_{\exp}(i,w)|=2^i-1$$

and

$$|F_{\exp}(i,w)|_a=i-1+|w|_a.$$

LEMMA 18: *Let $f:\mathbb{N}_+\to X$ be a $S_{\neq}$-function. Let X' be a subset of X. Then the function $g:i\mapsto F_{\exp}(|f(i)|_{X'}+1,f(i))$ is a $S_{\neq}$-function.*

Note that X and $\{a,b,c\}$ are not necessarily disjoint.

Proof: Let us define $Y=X\cup\{a,b,c\}$. Let us define the rational transduction $\tau:\{a,b\}^*\to Y^*$ whose graph is made of all the couples $(w_1 b^{1+|w_2|}, w_1 cw_2)$ for $w_1\in\{a,b\}^*$ and $w_2\in Y^*$. Then

$$\begin{aligned}Y^*-g(\mathbb{N}_+)=&(Y^*-X_{\exp}^*\,c\,Y^*)\\ &\cup\tau(X_{\exp}^*-f_{\exp}(\mathbb{N}_+))\\ &\cup X_{\exp}^*\,c\,(Y^*-f(\mathbb{N}_+))\\ &\cup\nabla_{\neq}(X_{\exp}^*,|\cdot|_a,c,|\cdot|_{X'},X^*).\end{aligned}$$

In this union the first term is regular. The two following terms are dominated by $S_{\neq}$, since $f_{\exp}$ and f are $S_{\neq}$-functions. And the last one is dominated by $S_{\neq}$. This proves that $Y^*-g(\mathbb{N}_+)\leqq S_{\neq}$. □

LEMMA 19: *Let $f:\mathbb{N}_+\to X$ be a $S_{\neq}$-function.* Let *X' be a subset of X. Let z be a letter, which does not belong to X. Then the function $g: i\mapsto f(i)\,z^{|f(i)|_{X'}}$ is a $S_{\neq}$-function.*

Proof: Let us define $Y=X\cup\{z\}$. Then

$$Y^*-g(\mathbb{N}_+)=(Y^*-X^*z^*)\cup(Y^*-f(\mathbb{N}_+))\,z^*\cup\nabla_{\neq}(X^*,|\,.\,|_{X'},\varepsilon,|\,.\,|,z^*).$$

In this union the first term is regular. The second term is dominated by $S_{\neq}$, since f is a $S_{\neq}$-function. And the last one is dominated by $S_{\neq}$. This proves that $Y^*-g(\mathbb{N}_+)\leqq S_{\neq}$. □

For $f=f_{\exp}$, $X=X_{\exp}$, $X'=\{a\}$ and $z=d$ this lemma yields, that

$$g_1:\quad i\mapsto f_{\exp}(i)\,d^{i-1}$$

is a $S_{\neq}$-function.

Lemma 18 yields for $f=g_1$, $X=\{a,b,d\}$ and $X'=\{a,b\}$, that

$$g_2:\quad i\mapsto F_{\exp}(2^i, f_{\exp}(i)\,d^{i-1})$$

is a $S_{\neq}$-function.

Indeed $g_2(i)$ is defined for every $i\in\mathbb{N}_+$ and $|g_2(i)|_d=i-1$ and $|g_2(i)|=2^{2^i}-1$. So that $\lim_{i\to\infty}|g_2(i)|/i=\infty$ and g_2 is a structure function. According to theorem 6, L_{g_2} is a context-free language, and for large enough n we have

$$\rho_{L_{g_2}}(n)=\bar\rho_{L_{g_2}}(n)=n-1+n\tilde g_2(n)=n-1+n\lfloor\ln_2\ln_2 n\rfloor\sim n\ln_2\ln n.$$

3. Third example: a structure function leading to a context-free language whose rational index is $\Theta(n\sqrt[k]{\ln n})$

Let k be an integer greater than 1. For $f=f_k$ and $X=X'=X_k$ lemma 18 yields, that the function $g_3: i\mapsto F_{\exp}(i^k, f_k(i))$ is a $S_{\neq}$-function. Indeed it is a structure function such that $x_{g_2}=x_k$ and $|g_3(i)|=2^{i^k}-1$. According to theorem 6, L_{g_2} is a context-free language, and for large enough n we have

$$\rho_{L_{g_3}}(n)=\bar\rho_{L_{g_3}}(n)=n-1+n\tilde g_3(n)=n-1+n\lfloor\sqrt[k]{\ln_2 n}\rfloor\sim n\sqrt[k]{\ln_2 n}.$$

4. Fourth example: a structure function leading to a context-free language whose rational index is $\Theta(n^{\sqrt{2}})$

We define g_4 to be the partial function such that $g_4(n)$ is defined only if n is a power of 2, and then

$$g_4(2^i)=F_{\exp}(i+j, d^{2^i-1} f_{\exp}(i) f_2(i) c f_2(j) a^{2i^2-j^2} b^{(j+1)^2-2i^2})$$

where $j=\lfloor\sqrt{2}\, i\rfloor$.

Remark: j is the only positive integer such that $j^2 \leqq 2i^2 < (j+1)^2$.

LEMMA 20: *g_4 is a structure function verifying* $|g_4(2^i)|=2^{|i(1+\sqrt{2})|}-1$ *and* $x_{g_4}=d$.

Proof: In order to prove that g_4 is a structure function, we define

$$g'_4:\quad i \mapsto d^{2^i-1} f_{\exp}(i) f_2(i) c f_2(j) a^{2i^2-j^2} b^{(j+1)^2-2i^2}.$$

Let $X=X_2 \sqcup \{a,b,c,d\}$. We have $g'_4(\mathbb{N}_+) \subset X^*$ and we are going to prove that $X^*-g'_4(\mathbb{N}_+)$ is equal to the union B of the following eight languages:

$$B_1=X^*-d^*\{a,b\}^* X_2^* c X_2^* a^* b^+$$
$$B_2=\nabla_{\neq}(d^*, |\,.\,|, \varepsilon, |\,.\,|, \{a,b\}^*) X_2^* c X_2^* a^* b^+$$
$$B_3=d^* \nabla_{\neq}(\{a,b\}^*, |\,.\,|_a, \varepsilon, |\,.\,|_{x_2}, X_2^*) c X_2^* a^* b^+$$
$$B_4=d^*(\{a,b\}^*-f_{\exp}(\mathbb{N}_+)) X_2^* c X_2^* a^* b^+$$
$$B_5=d^*\{a,b\}^*(X_2^*-f_2(\mathbb{N}_+)) c X_2^* a^* b^+$$
$$B_6=d^*\{a,b\}^* X_2^* c (X_2^*-f_2(\mathbb{N}_+)) a^* b^+$$
$$B_7=d^*\{a,b\}^* \nabla_{\neq}(X_2^* c, 2|\,.\,|_{X_2}+|\,.\,|_c, \varepsilon, |\,.\,|, X_2^* a^*) b^+$$
$$B_8=d^*\{a,b\}^* X_2^* \nabla_{\neq}(c X_2^*, 3|\,.\,|_c+2|\,.\,|_{x_2}, \varepsilon, |\,.\,|, a^* b^+).$$

- For any integer i, $g'_4(i)$ does not belong to this union because

$$g'_4(i) \in d^*\{a,b\}^* X_2^* c X_2^* a^* b^+$$
$$|d^{2^i-1}|=2^i-1=|f_{\exp}(i)|$$
$$|f_{\exp}(i)|_a=i-1=|f_2(i)|_{x_2}$$
$$f_{\exp}(i) \in f_{\exp}(\mathbb{N}_+)$$
$$f_2(i) \in f_2(\mathbb{N}_+)$$
$$f_2(j) \in f_2(\mathbb{N}_+)$$

$$2|f_2(i)c|_{X_2}+|f_2(i)c|_c=2|f_2(i)|+1=2(i^2-1)+1$$
$$=2i^2-1=(j^2-1)+(2i^2-j^2)=|f_2(j)a^{2i^2-j^2}|$$

$$3|cf_2(j)|_c+2|cf_2(j)|_{x_2}=3+2(j-1)$$
$$=2j+1=(2i^2-j^2)+((j+1)^2-2i^2)=|a^{2i^2-j^2}b^{(j+1)^2-2i^2}|.$$

This proves that $g'_4(\mathbb{N}_+)$ and B are disjoint, *i. e.*

$$g'_4(\mathbb{N}_+)\subset X^*-B.$$

• Conversely let w be a word in X^*-B. w belongs to X^*-B_1 *i. e.*

$$w\in d^*\{a,b\}^*X_2^*cX_2^*a^*b^+.$$

Since w belongs neither to B_4 nor to B_5 nor to B_6, we have

$$w\in d^*f_{\exp}(\mathbb{N}_+)f_2(\mathbb{N}_+)cf_2(\mathbb{N}_+)a^*b^+,$$

i. e.

$$w=d^pf_{\exp}(i')f_2(i)cf_2(j)a^qb^r,$$

for some i', i, j, $r\in\mathbb{N}_+$ and p, $q\in\mathbb{N}$.

Since w does not belong to B_2, we have $p=2^{i'-1}$.

Since w does not belong to B_3, we have $i'-1=i-1$ *i.e.* $i'=i$.

Since w does not belong to B_7, we have $2i^2-1=(j^2-1)+q$ *i.e.* $q=2i^2-j^2$.

Since w does not belong B_8, we have $2j+1=q+r$ *i.e.* $r=(2j+1)-(2i^2-j^2)=(j+1)^2-2i^2$.

$q\geqq 0$ and $r>0$ hence $j^2\leqq 2i^2<(j+1)^2$, *i.e.* $j=\lfloor\sqrt{2}\rfloor$. We have proved that $w=g'_4(i)$. Hence

$$g'_4(\mathbb{N}_+)\supset X^*-B.$$

We have proved that $g'_4(\mathbb{N}_+)=X^*-B$ *i.e.*

$$X^*-g'_4(\mathbb{N}_+)=B.$$

B_1 is a regular language, and $B_2\ldots B_8$ are languages dominated by $S_{\neq}$. This proves that g'_4 is a $S_{\neq}$-function.

Since $|g'_4(i)|_{\{x_2,c\}}=i+j-1$, lemma 18 yields that g_4 is a $S_{\neq}$-function too.

$$|g'_4(i)|=(2^i-1)+(2^i-1)+(i^2-1)+1+(j^2-1)$$
$$+(2i^2-j^2)+((j+1)^2-2i^2)\sim 2^{i+1}\in o(2^{i+j}).$$

Hence $g_4(2^i)=F_{\exp}(i+j,\ g'_4(i))$ is defined when i is large enough.

We have

$$|g_4(2^i)|_d=2^i-1$$

and

$$|g_4(2^i)|=2^{i+j}-1=2^{\lfloor i(1+\sqrt{2})\rfloor}-1$$

so that

$$\lim_{i\to\infty}|g_4(2^i)|/2^i=\infty.$$

Thus g_4 is a structure function and $x_{g_4}=d$. □

Let n be an integer large enough for $\tilde{g}_4(n)$ to exist. Then $\tilde{g}_4(n)$ is the largest integer p such that

$$|g_4(p)|\leqq n-1.$$

Hence p is the largest power of 2, say 2^i, such that

$$|g_4(2^i)|\leqq n-1.$$

This inequality is equivalent to the following ones:

$$2^{i+\lfloor\sqrt{2}\,i\rfloor}-1\leqq n-1,$$

$$\lfloor i+\sqrt{2}\,i\rfloor\leqq\log_2 n,$$

$$\lfloor i+\sqrt{2}\,i\rfloor\leqq\lfloor\log_2 n\rfloor,$$

$$i+\sqrt{2}i<1+\lfloor\log_2 n\rfloor,$$

$$i<(\sqrt{2}-1)\lfloor 1+\log_2 n\rfloor.$$

This upper bound on i cannot be an integer, so that the largest i is

$$\lfloor(\sqrt{2}-1)\lfloor 1+\log_2 n\rfloor\rfloor\in(\sqrt{2}-1)\log_2 n+O(1)$$

and the largest p is

$$\tilde{g}_4(n)=2^{\lfloor\lfloor 1+\log_2 n\rfloor(\sqrt{2}-1)\rfloor}\in n^{\sqrt{2}-1}\,2^{O(1)}=\Theta(n^{\sqrt{2}-1}).$$

Theorem 6 yields that L_{g_4} is a context-free language, such that for large enough n we have

$$\rho_{L_{g_4}}(n)=\bar{\rho}_{L_{g_4}}(n)=n-1+n\,\tilde{g}_4(n)=n-1+n\,2^{[[1+\log_2 n](\sqrt{2}-1)]}\in\Theta(n^{\sqrt{2}}).$$

This kind of construction may be generalized:

5. Fifth example: structure functions leading to a context-free language whose rational index is $\Theta(n^\lambda)$ for an algebraic number $\lambda>1$

The main example of structure functions was the family of f_k's. For any integer k greater than 1, we have $\tilde{f}_k(n)\in\Theta(n^{1/k})$. We extend this notation for other non integral numbers:

LEMMA 21: *Let λ be an irrational algebraic real number greater than* 1. *Then we can find a structure function f_λ such $\tilde{f}_\lambda(n)\in\Theta(n^{1/\lambda})$.*

Proof: Let P be a minimal polynomial of λ, *i.e.* a polynomial of minimal degree with integral coefficients such that $P(\lambda)=0$. Let m be the degree of P. Let us assume

$$P(t)=\alpha_0+\alpha_1 t+\ldots+\alpha_m t^m.$$

Since P is irreducible, λ is a simple root of P, *i.e.*

$$P(\lambda)=0$$

and $P'(\lambda)\neq 0$, where P' is the derivative of P. If $P'(\lambda)<0$, then we replace P by $-P$ in order to ensure that

$$P'(\lambda)>0.$$

P' is a continuous function. Hence we can find two rational numbers p_1/q_1 and p_2/q_2 such that

$$1\leqq\frac{p_1}{q_1}<\lambda<\frac{p_2}{q_2},$$

$$\forall t\in\left[\frac{p_1}{q_1},\frac{p_2}{q_2}\right],\quad P'(t)>0.$$

Hence

$$\forall t\in\left[\frac{p_1}{q_1},\lambda\right[,\quad P(t)<0,$$

$$\forall t\in\left]\lambda,\frac{p_2}{q_2}\right],\quad P(t)>0.$$

The integers p_1, q_1, p_2 and q_2 are now fixed, and we shall use them to define f_λ.

Let

$$n_1=\left\lceil 1/\min\left\{\frac{p_2}{q_2}-\lambda,\ \lambda-\frac{p_1}{q_1}\right\}\right\rceil.$$

Let i be a positive integer. An integer j verifies the conditions

$$\begin{aligned} q_1 j-p_1 i&\geqq 0\\ p_2 i-q_2(j+1)&\geqq 0\\ -i^m P(j/i)&>0\\ i^m P((j+1)/i)&>0 \end{aligned} \tag{12}$$

if and only if it verifies

$$\frac{p_1}{q_1}\leqq\frac{j}{i}<\frac{j+1}{i}\leqq\frac{p_2}{q_2},$$

$$P\left(\frac{j}{i}\right)<0<P\left(\frac{j+1}{i}\right),$$

i. e.

$$\frac{p_1}{q_1}\leqq\frac{j}{i}<\lambda<\frac{j+1}{i}\leqq\frac{p_2}{q_2},$$

and then

$$j=\lfloor i\lambda\rfloor. \tag{13}$$

Furthermore, if $i\geqq n_1$ then (13) and (12) are equivalent, *i.e.* $\lfloor i\lambda\rfloor$ is the only integer j verifying (12). If $i<n_1$ then (12) may have no solution or it may have the unique solution $\lfloor i\lambda\rfloor$.

We define the two alphabets

$$D=\{d_1,\ \ldots,\ d_9\}$$
$$X=\{x_{-1},\ \ldots,\ x_{m+1},\ a,\ b,\ c,\ c'\}.$$

The structure function f_λ will be defined on the alphabet

$$\Omega=D\sqcup X.$$

For every positive integer i for which (12) has a solution j we define

$$f'_\lambda(i)=d_1\, c'^{2^i-1}\, d_2\, f_{\exp}(i)\, d_3\, c^{j-1}\, d_4\, a^{p_2\, i-q_2\,(j+1)}\, d_5\, a^{q_1\, j-p_1\, i}\, d_6 f_m(i)$$

$$d_7 f_m(j)\, d_8\, b^{-i^m P(j/i)}\, d_9\left(x_{-1}\left(\prod_{k=0}^{m}(a\, x_k^{j^k})^{i^{m-k}}\right)a\, x_{m+1}\right)^2 d_8$$

$$d_7 f_m(j+1)\, d_8\, d_9\left(x_{-1}\left(\prod_{k=0}^{m}(a\, x_k^{(j+1)^k})^{i^{m-k}}\right)a\, x_{m+1}\right)^2 d_8\, b^{i^m P((j+1)/i)},$$

and

$$f_\lambda(2^i)=F_{\exp}(j, f'_\lambda(i)).$$

The letters of D are used as separators.

The factor $d_1\, c'^{2^i-1}$ ensures that a letter occurs 2^i-1 times in $f'_\lambda(i)$ and thus in $f_\lambda(2^i)$ too.

The factor $d_2 f_{\exp}(i)$ gives a relation between i and 2^i.

The factor $d_3\, c^{j-1}$ ensures that a letters occurs $j-1$ times in $f'_\lambda(i)$ so that we can define $F_{\exp}(j, f'_\lambda(i))$.

The factors $d_4\, a^{p_2 i-q_2\,(j+1)}$, $d_5\, a^{q_1 j-p_1 i}$, $d_8\, b^{-i^m P(j/i)}$ and $d_8\, b^{i^m P((j+1)/i)}$ correspond to (12).

The factor $d_6 f_m(i)$ gives a relation between i and i^k for every $k\in[0, m]$.

The factor $d_7 f_m(j)$ gives a relation between j and j^k for every $k\in[0, m]$.

The factor $x_{-1}\left(\prod_{k=0}^{m}(a\, x_k^{j^k})^{i^{m-k}}\right)a\, x_{m+1}$ is used to construct the number $(j/i)^k\, i^m$, which is the number of occurrences of x_k, from the numbers j^k and i^{m-k}, for every k in $[0, m]$. The factor $(a\, x_k^{j^k})^{i^{m-k}}$ is preceded by x_{k-1} and followed by $a\, x_{k+1}$ for every k in $[0, m]$. This explains what x_{-1} and $a\, x_{m+1}$ are for. $i^m P(j/i)$ is the linear combination of these numbers $i^{m-k} j^k$, whose coefficients are those of P. These coefficients may not have all the same sign, but in the equality

$$-i^m P\left(\frac{j}{i}\right)+\sum_{k=0}^{m}\max(0, \alpha_k)\, i^{m-k} j^k=\sum_{k=0}^{m}\max(0, -\alpha_k)\, i^{m-k} j^k+0$$

both sides are sums of non-negative numbers. This is why this factor appears twice.

In the same way the number $i^m P((j+1)/i)$ is built in the third line of the expression of $f'_\lambda(i)$.

Let $K=(DX^*)^*$. The language $\Omega^*-f'_\lambda(\mathbb{N}_+)$ is the union of the following languages $G_1, \ldots, G_{12}$.

$$G_1=\Omega^*-\Big(d_1\,c'^*\,d_2\,\{a, b\}^*\,d_3\,c^*\,d_4\,a^+\,d_5\,a^+\,d_6\,X_m^*$$

$$d_7\,X_m^*\,d_8\,b^+\,d_9\left(x_{-1}\left(\prod_{k=0}^{m} a\,(b\,x_k^+)^+\right)a\,x_{m+1}\right)^2 d_8$$

$$d_7\,X_m^*\,d_8\,d_9\left(x_{-1}\left(\prod_{k=0}^{m} a\,(b\,x_k^+)^+\right)a\,x_{m+1}\right)^2 d_8\,b^+\Big)$$

$$G_2=K\,d_2\,(\{a, b\}^*-f_{\exp}(\mathbb{N}_+))\,K$$

$$G_3=K\{d_6, d_7\}\,(X_m^*-f_m(\mathbb{N}_+))\,K$$

$$G_4=\nabla_{\neq}\,(d_1\,c'^*, |\,.\,|, \varepsilon, |\,.\,|, d_2\,\{a, b\}^*)\,K$$

$$G_5=K\nabla_{\neq}\,(d_3\,c^*\,d_4\,a^+, q_2\,|\,.\,|_{\{d_3, c, d_4\}^+}|\,.\,|_a, K, p_2\,|\,.\,|_{\{d_6, x_m\}}, d_6\,X_m)\,K$$

$$G_6=K\nabla_{\neq}\,(d_3\,c^*, q_1\,|\,.\,|, K, |\,.\,|_a+p_1\,|\,.\,|_{\{d_6, x_m\}}, d_5\,a^+\,d_6\,X_m)\,K$$

$$G_7=K\nabla_{\neq}\,(d_2\,\{a, b\}^*, |\,.\,|_a, K, |\,.\,|_{x_m}, d_6\,X_m^*)\,K$$

$$G_8=K\nabla_{\neq}\,(d_3\,c^*, |\,.\,|_c, K, |\,.\,|_{x_m}, d_7\,X_m^*)\,d_8\,b^+\,K$$

$$G_9=K\nabla_{\neq}\,(d_3\,c^*, |\,.\,|, K, |\,.\,|_{x_m}, d_7\,X_m^*)\,d_8\,K$$

$$G_{10}=K\bigcup_{k=0}^{m}\nabla_{\neq}\,(d_6\,X_m^*, |\pi_{X_k}|, K\,d_9\,X^*\,x_{k-1}, |\,.\,|_a, (a\,x_k^+)^+)\,a\,x_{k+1}\,\Omega^*\,K$$

$$G_{11}=K\bigcup_{k=0}^{m}\nabla_{\neq}\,(d_7\,X_m^*, |\pi_{X_{m-k}}|, d_8\,b^*\,d_9\,X^*\,a, |\,.\,|, x_k^+)\,a\,\Omega^*\,K$$

$$G_{12}=K\nabla_{\neq}\Big(d_8\,b^*\,d_9\,X^*\,x_{m+1}, |\,.\,|_b+\sum_{k=0}^{m}\max(0, \alpha_k)\,|\,.\,|_{x_k}, \varepsilon,$$

$$|\,.\,|_b+\sum_{k=0}^{m}\max(0, -\alpha_k)\,|\,.\,|_{x_k}, x_{-1}\,X^*\,d_8\,b^*\Big)K.$$

These twelve languages are dominated by $S_{\neq}$. Hence $\Omega^*-f'_\lambda(\mathbb{N}_+)=\bigcup_{i=1}^{12} G_i$ is dominated by $S_{\neq}$ too. So f'_λ is a $S_{\neq}$-function.

Since $|f'_\lambda(i)|_c=j-1$, lemma 18 yields that f_λ is a $S_{\neq}$-function.

$$|f'_\lambda(i)|\sim 2^{i+1}\in o\,(2^j)$$

and $f'_{\lambda}(i)$ is defined when $i \geqq n_1$. Hence $f_{\lambda}(2^i)$ is defined when i is large enough.

We have

$$|f_{\lambda}(2^i)|_{c'} = 2^i - 1$$

and

$$|f_{\lambda}(2^i)| = 2^j - 1 = 2^{\lfloor i\lambda \rfloor} - 1$$

so that

$$\lim_{i \to \infty} |f_{\lambda}(2^i)|/2^i = \infty.$$

Thus f_{λ} is a structure function and $x_{f_{\lambda}} = c'$.

Like in the fourth example we get

$$\tilde{f}_{\lambda}(n) = 2^{\lfloor \lfloor 1 + \log_2 n \rfloor / \lambda \rfloor} \in \Theta(n^{1/\lambda})$$

and

$$\rho_{L_{f_{\lambda}}}(n) \in \Theta(n^{1+1/\lambda}). \quad \square$$

THEOREM 14: *Let λ be an algebraic number greater than 1. Then there exists a context-free language L such that $\rho_L(n) = \bar{\rho}_L(n) \in \Theta(n^{\lambda})$.*

Proof: λ may be expressed as $\lambda = 1 + 1/\lambda_1 + \ldots + 1/\lambda_e$, where every λ_i is an irrational algebraic number greater than 1. Then lemma 21 and theorem 9 can be applied to copies of $f_{\lambda_1}, \ldots, f_{\lambda_e}$ on disjoint alphabets. This completes the proof. $\square$

THEOREM 15: *Let λ and μ be two algebraic numbers such that $1 < \lambda < \mu$. Then there exist two context-free languages L_{λ} and L_{μ} such that:*

$$\rho_{L_{\lambda}}(n) = \bar{\rho}_{L_{\lambda}}(n) \in \Theta(n^{\lambda}),$$

$$\rho_{L_{\mu}}(n) = \bar{\rho}_{L_{\mu}}(n) \in \Theta(n^{\mu}),$$

$$L_{\lambda} < L_{\mu}.$$

Proof: We may have $\mu = \lambda + 1/\lambda_{e+1} + \ldots + 1/\lambda_{e'}$ for some irrational algebraic numbers $\lambda_{e+1} \ldots \lambda_{e'}$ greater than 1. We define L_{λ} and L_{μ} like in the previous proof. Theorem 13 yields, that $L_{\lambda} < L_{\mu}$.

We can also build structure functions f_λ such that $\tilde{f}_\lambda(n) \in \Theta(n^{1/\lambda})$ for some transcendental numbers λ, e.g. $\pi/\sqrt{6}$:

6. Sixth example: a structure function leading to a context-free language whose rational index is $\Theta(n^{1+\sqrt{6}/\pi})$.

The construction of this structure function is based upon the equality

$$\frac{\pi^2}{6} = \sum_{j=1}^{\infty} \frac{1}{j^2}.$$

First we define the function

$$\alpha: \quad \mathbb{N}_+ \to \mathbb{N}_+$$
$$i \mapsto \sum_{j=1}^{i} \left\lfloor \frac{i^2}{j^2} \right\rfloor.$$

We define then g_6 to be the partial function such that $g_6(n)$ is defined only if n is a power of 2, and then

$$g_6(2^i) = F_{\exp}\Bigg(\left\lfloor \sqrt{\alpha(i)} \right\rfloor,\ x_3^{\lfloor\sqrt{\alpha(i)}\rfloor - 1} f_2(\lfloor \sqrt{\alpha(i)} \rfloor)\, c\, a^{\alpha(i) - \lfloor\sqrt{\alpha(i)}\rfloor^2}\, b^{(\lfloor\sqrt{\alpha(i)}\rfloor+1)^2 - 1 - \alpha(i)} x_4^{2^i - 1} f_{\exp}(i) f_2(i) \prod_{j=1}^{i} ((x_5 f_2(j))^{\lfloor i^2/j^2 \rfloor} a^{i^2 \bmod j^2} b^{j^2 - 1 - (i^2 \bmod j^2)}) \Bigg).$$

We can prove easily that g_6, like g_4, is a structure function, that $x_{g_6} = x_4$, and that $|g_6(2^i)| = 2^{\lfloor\sqrt{\alpha(i)}\rfloor} - 1$. We have

$$\alpha(i) \in i^2 \sum_{j=1}^{\infty} \frac{1}{j^2} + O(i) = \frac{\pi^2}{6} i^2 + O(i)$$

and thus

$$\left\lfloor \sqrt{\alpha(i)} \right\rfloor \in \frac{\pi}{\sqrt{6}} i + O(1)$$

so that

$$\tilde{g}_6(n) \in \Theta(n^{\sqrt{6}/\pi})$$

and

$$\bar{\rho}_{L_{g_6}}(n)\in\Theta(n^{1+\sqrt{6}/\pi}).$$

7. Other examples and generalization

• Let $\mathscr{C}_\lambda$ be the set of context-free languages, whose extended rational index is in $O(n^\lambda)$ for any real number greater than 1. It is a rational cone, *i.e.* it is closed for rational transductions. If $1<\lambda<\mu$ then you can find a rational number p/q between λ and μ. There exists a context-free language whose rational index is in $\Theta(n^{p/q})$. This language belongs to $\mathscr{C}_\mu-\mathscr{C}_\lambda$. This proves that $\mathscr{C}_\lambda$ is a proper sub-cone of $\mathscr{C}_\mu$. Hence the family $(\mathscr{C}_\lambda)_{\lambda\in]1,\infty[}$ is a strictly increasing family of cones with the same cardinality as $\mathbb{R}$.

• The structure functions g_2 and g_4 of second and fourth examples, and theorem 9 yield for instance that there exists a context-free language whose rational indexes for large enough n are:

$$n-1+\tilde{g}_2(n)(n+\tilde{g}_4(n)(n+n\tilde{f}_5(n)))$$
$$=n-1+\lfloor\ln_2\ln_2 n\rfloor(n+2^{\lfloor\lceil\ln_2 n\rceil(\sqrt{2}-1)\rfloor}(n+n\lfloor\sqrt[5]{n}\rfloor))$$
$$\in\Theta(n^{\sqrt{2}+1/5}\ln_2\ln_2 n).$$

• We could, with this technique, build a context-free language, whose rational indexes are in $\Theta(n^\pi)$.

• The technique used in this paper can be sophisticated: We can replace the language $S_{\neq}$, omnipresent in this paper, by a generator of the rational cone of linear languages, like the only language solution of the equation $L=aL\bar{a}\cup bL\bar{b}\cup\{\varepsilon\}$, whose rational index is in $\Theta(n^2)$. Then the structure functions could involve decimal numbers and arithmetical computations on these numbers. In this way we can obtain a context-free language L such that $\rho_L(n)=\bar{\rho}_L(n)$ and $|\rho_L(n)-n^\pi|<1$ for large enough n.

• Let Λ be the set of all the numbers $\lambda\in]1,\infty[$ such that there exists a context-free language whose rational index is $\Theta(n^\lambda)$. Since the non-isomorphic context-free languages form a denumerable set, Λ is denumerable too. However it holds all the algebraic numbers greater than 1, and seemingly any computable number greater than 1 like π, e, $e+\pi$, $2+\cos\sqrt[3]{e}+2+\ln 2$ or $2+\ln\int_0^\pi\sqrt{8+\cos x}\,dx$, for which there exists an efficient algorithm to compute as many of its digits as you wish. So here is an open problem: can we find an explicit number in $]1, \infty[-\Lambda$?

REFERENCES

1. J. Berstel, *Transductions and Context-free languages*, Teubner Verlag, 1979.
2. L. Boasson, *Non-générateurs algébriques et substitution*, RAIRO Informatique Théorique, Vol. 19, No. 2, 1985, pp. 125-136.
3. L. Boasson and M. Nivat, *Ordres et types de language, I, II, III, C. R. Acad. Sci. Paris*, série A, Tome 284, 1977, pp. 559-562, 625-628, 703-705.
4. L. Boasson, B. Courcelle and M. Nivat, *The Rational Index, a Complexity measure for Languages*, S.I.A.M. J. Comput, Vol. 10, (2), 1981, pp. 284-296.
5. J. M. Farinone, *Langages algébriques d'index rationnel singulier*, Thèse de 3e cycle, Rapport L.I.T.P., No. 86-64, Univ. de Paris-VII, 1986.
6. J. Gabarro, *Index rationnel, centre et langages algébriques*, Thèse de 3e cycle, Rapport L.I.T.P., No. 81-54, Univ. de Paris-VI, 1981.
7. D. E. Knuth, *Big Omicron and Big Omega and Big Theta*, Sigact News, April-June 1976, pp. 18-24.
8. J. Lelong-Ferrand and J. M. Arnaudiès, *Cours de mathématiques*, Tome 2, Analyse, Dunod, Paris.
9. M. Nivat, *Transductions des langages de Chomsky*, Ann. de l'Inst. Fourier, Vol. 18, 1968, pp. 339-456.
10. L. Pierre and J. M. Farinone, *Rational Index of Context-free Languages in* $\exp \Theta(\sqrt[p]{n})$ and $n^{\Theta(\sqrt[p]{\ln n})}$, TCS, Vol. 57, 1988, pp. 185-204.

IMPRIMERIE GAUTHIER-VILLARS, France
Dépôt légal 1990 : Imprimeur : 3017 — Éditeur : 029 — CPPP 63422
3161-90 août 1990 *Imprimé en France*