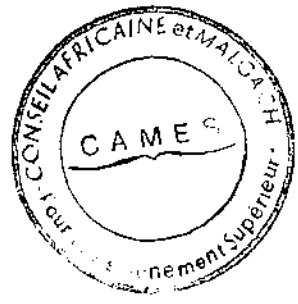
UNIVERSITÉ D' AIX - MARSEILLE II - LUMINY

THÈSE

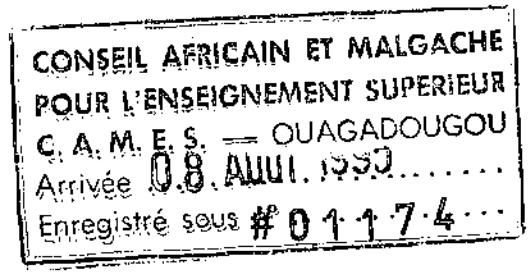
présentée

pour l'obtention du

DOCTORAT DE MATHÉMATIQUES
(3e CYCLE)



COMPLEXITÉ DE SUITES AUTOMATIQUES



soutenue par

Théodore TAPSOBA

le 1er octobre 1987, devant le Jury :

Président : **Gérard RAUZY**
Examinateurs : **Pierre LIARDET**
Christian SAMUEL

Monsieur RAUZY, professeur à l'Université d'Aix-Marseille II m'a fait l'honneur de diriger mes travaux de recherche et par les nombreux conseils qu'il n'a cessé de me prodiguer et sa totale disponibilité, a toujours montré à mon égard une bienveillante attention. Au delà des remerciements d'usage, je tiens à lui exprimer ici ma très profonde gratitude et ma respectueuse reconnaissance.

Monsieur LIARDET, professeur à l'Université d'Aix-Marseille I a bien voulu s'intéresser à mon travail et n'a ménagé ni ses suggestions et ses conseils ni son temps. Qu'il trouve ici l'expression de toute ma gratitude.

Je remercie particulièrement Monsieur SAMUEL, maître de conférence à l'Université d'Aix-Marseille II qui a bien voulu porter son attention sur cette thèse et accepter de participer au jury d'examen.

Je remercie également Monsieur SMADJA de l'Université d'Aix-Marseille II pour les indications en informatique qu'il m'a donné.

Je remercie enfin Madame LIARDET qui a mené à bien la réalisation matérielle de ce mémoire.

TABLE DES MATIÈRES

RESUME (i)

I.- INTRODUCTION

1.- Mots infinis 1

2.- Définitions et notations 3

3.- Résultats 6

II.- MINIMALITE

1.- Système symbolique d'un mot infini 8

2.- Mots minimaux 8

III.- COMPLEXITE DU LANGAGE

1.- Le Théorème d'unicité 12

2.- Commutation 15

3.- Majoration des facteurs puissances 17

4.- Démonstration du Théorème 1 pour deux lettres 20

5.- Démonstration du Théorème 1 dans le cas général 22

IV.- CALCUL AUTOMATIQUE

1.- Automates 26

2.- Etude des $F_n(n)$ 27

3.- Fin de la démonstration du Théorème 3 30

4.- Complexité polynômiale 31

V.- EXEMPLES

1.- Suite de Morse 32

2.- Suite de Rudin-Shapiro 34

VI.- UN EXEMPLE NON MINIMAL 36

BIBLIOGRAPHIE 40

THÈSE de 3e Cycle
Mathématique

COMPLEXITÉ DE SUITES AUTOMATIQUES

par Théodore TAPSOBA

RÉSUMÉ

Soit $\mathcal{A}$ un *alphabet* d'un nombre fini de *lettres* 1,2,...,q. Une suite finie $m := m_1m_2...m_k$ de k lettres dans $\mathcal{A}$ est appelée *mot* (fini) sur $\mathcal{A}$ de *longueur* $|m| := k$. L'ensemble $\mathcal{A}^*$ de ces mots est regardé comme le monoïde libre engendré sur $\mathcal{A}$. Un mot infini $u := u_0u_1u_2u_3...$ sur $\mathcal{A}$ est une suite $u : \mathbb{N} \to \mathcal{A}$; l'ensemble des mots infinis est noté $\mathcal{A}^\infty$. L'ensemble $\mathcal{M}(\mathcal{A}) := \mathcal{A}^\infty \cup \mathcal{A}^*$ est muni de la distance ultramétrique habituelle qui fait de $\mathcal{M}(\mathcal{A})$ un espace compact (avec $\mathcal{A}^*$ sous-ensemble dense). Soit S le shift sur $\mathcal{A}^\infty$ qui consiste à effacer la première lettre. Pour un mot infini u donné, la fermeture K_u de l'ensemble $\{S^k u ; k \in \mathbb{N}\}$ est invariante par S. Le couple (S, K_u) est appelé *système symbolique* associé à u. Un mot fini $m := m_1m_2...m_k$ est dit *facteur* d'un mot u (fini ou pas) s'il existe r tel que

$$m_1m_2...m_k = u_ru_{r+1}...u_{r+k-1}.$$

L'ensemble des facteurs de u de longueur k est noté $F_k(u)$. La suite $k \to \text{card}(F_k(u))$ donne une bonne indication sur la complexité du langage de u ainsi que, dans une large mesure, du système dynamique qui lui est associé.

Une *substitution* f sur $\mathcal{A}$ est une application $f : \mathcal{A} \to \mathcal{A}^*$ que l'on prolonge de manière naturelle en une application continue de $\mathcal{M}(\mathcal{A})$, encore notée f et définie par

$$f(u_0u_1u_2...) := f(u_0)f(u_1)f(u_2)...$$

La substitution est dite *uniforme* de *module* ρ si $\rho = |f(a)|$ pour tout $a \in \mathcal{A}$. Les mots infinis engendrés par itérations d'une substitution sur $\mathcal{A}$ est une méthode simple pour construire des suites intéressantes pour leurs régularités. Par exemple, sur les lettres 1 , 2 , la substitution s donnée par

$$s(1) := 121 \quad , \quad s(2) := 212 \, ,$$

conduit au mot périodique $x := 12121212...$ (et au mot périodique x' obtenu en échangeant les lettres) . Par contre la substitution σ donnée par

$$\sigma(1) := 12 \quad , \quad \sigma(2) := 21 \, ,$$

conduit au mot infini

$$\mu := 2112122112...$$

et au mot μ' obtenu par échange des lettres qui ne sont pas périodiques. Le mot μ est point fixe de σ, en ce sens que

$$\mu = \sigma(\mu_0)\, \sigma(\mu_1)\, \sigma(\mu_2)\, \sigma(\mu_3)\, \sigma(\mu_4)...,$$

les 2^k premiers termes de μ étant donnés par le k-ième itéré $\sigma^k(2)$ de 2 par σ . Le mot infini μ correspond à la célèbre suite de Morse (1921) qui a fait l'objet de nombreux travaux et suscité des investigations plus générales (suites de Morse généralisées de M. Keane (1968), systèmes symboliques associés (J. Martin (1970, 1973) et plus systématiquement M. Queffelec (Thèse, 1984)). Un mot infini u point fixe d'une substitution uniforme est encore dit *automatique* . On sait en effet d'après A. Cobham (1972) qu'un tel mot est reconnu par un automate fini .

Soit u un mot infini, point fixe d'une substitution uniforme f sur l'alphabet $\mathfrak{a}$. Dans ce travail, nous étudions la suite des entiers $p(u,n) := \text{card}(F_k(u))$ .On a évidemment $p(u,n) \leq (\text{Card}(\mathfrak{a}))^n$. A. Cobham a montré en 1972 qu' il existe une constante C (= C(u)) telle que $p(u,n) \leq C.n$ et N. Bleuzen-Guernalec (1986) a précisé la constante C en la majorant par $\rho(\text{Card}(\mathfrak{a}))^2$. Notre objectif principal est de démontrer l'existence d'un automate donnant la suite $n \rightarrow p(u,n+1) - p(u,n)$ à partir de l'écriture de n en base ρ , ceci dans le cas où u est une suite *minimale* (i.e. le système (S, K_u) est minimal) et f injective sur $\mathfrak{a}$.

Après l'introduction nous rappelons dans la partie II quelques propriétés des mots infinis u

minimaux obtenus par substitutions f . Nous donnons un critère effectif de minimalité. On se libère

ici de la condition restrictive $\lim_{k:\infty} | f^k(a) | = +\infty$ pour tout $a \in \mathfrak{a}$, habituellement faite. Lorsque u est minimal, la fonction p(v,.) est indépendante de $v \in K_u$ et cette propriété caractérise la minimalité de u .

La Partie III est consacrée à la démonstration du théorème suivant :

Théorème .- *Soit* u *point fixe de la substitution uniforme* f *de module* ρ. *On suppose* f *injective et* u *minimale mais pas périodique. Alors il existe une constante* L_0 *ne dépendant que de* ρ *telle que pour tout facteur* M *de* u *il existe* A , B *et* C, *facteurs de* u , *tels que :*

$$M = Bf(A)C, \ |B| < \rho , |C| < \rho \ ;$$

le triplet (A,B,C) *étant unique.*

La démonstration se fait tout d'abord pour un alphabet à deux lettres. C'est le cas essentiel que nous obtenons après une évaluation explicite du nombre maximum de facteurs successifs d'un même mot de deux lettres apparaissant dans u . On trouve un résultat analogue dans J. Martin (1970) et notamment pour les substitutions non nécessairement uniformes (1973), mais la démonstration dans ce cas général n'est pas correcte et nous donnons un contre-exemple. Notre méthode donne un meilleur résultat en fournissant explicitement la longueur du mot à lire pour avoir l'unicité.

La partie IV traite du calcul automatique des p(n) :=p(u,n) . On montre que la suite q(n) := p(n+1) - p(n) ne prend qu'un nombre fini de valeurs . Nous pécisons ce résultat sous la forme suivante :

Théorème.- *Il existe un* ρ-*automate tel que* q(n) *soit la sortie de l'automate lorsqu'il lit l'entier* n *écrit en base* ρ .

La suite de Morse et la suite de Rudin-Shapiro sont examinées en particulier dans la partie

suivante . Rappelons que la suite de Rudin-Shapiro , qui compte modulo 2 le nombre de "11" dans l'écriture des entiers en base 2, est aussi l'image par un morphisme littéral d'un point fixe de la substitution suivante τ sur un alphabet à 4 lettres :

$$\tau(1) = 12 \quad , \quad \tau(2) = 13 \quad , \quad \tau(3) = 42 \quad , \quad \tau(4) = 43 \quad .$$

Nous terminons en montrant que le caractère minimal de la suite est une condition suffisante mais non nécessaire. Un exemple simple est fourni par la substitution f sur trois lettres 1, 2 , 3 donnée par :

$$f(1) = 121 \quad , \quad f(2) = 232 \quad , \quad f(3) = 323.$$

COMPLEXITÉ DE SUITES AUTOMATIQUES

I. - INTRODUCTION

I.- 1. MOTS INFINIS.

Une suite finie $m := (m_0, m_1, ..., m_n)$ dans un ensemble fini $\mathcal{A}$ de symboles (ou lettres) sera vue le plus souvent comme un mot (fini) $m_0 m_1 ... m_n$ sur l'*alphabet* $\mathcal{A}$ et par extension, une suite $u : \mathbb{N} \to \mathcal{A}$ sera vue comme un mot infini $u := u_0 u_1 u_2 u_3 ...$ Une méthode simple pour construire de tels mots est de procéder par itérations d'une *substitution* . Chaque lettre d'un mot m est remplacée par un mot et ainsi de suite. On obtient ainsi des mots infinis dont les régularités ont retenues en premier lieu l'attention [1,8,16]. Par exemple, sur les symboles 1 , 2 , la substitution s donnée par

$$s(1) := 121 \quad , \quad s(2) := 212 \, ,$$

conduit au mot périodique $x := 12121212...$ (et au mot périodique x' obtenu en échangeant des symboles 1, 2 entre eux) par itération de s puisque $s^k(1)$ est formé de la répétition successive ℓ_k - fois du mot 12 suivi de 1. Un calcul simple montre d'ailleurs que $2\,\ell_k + 1 = 3^k$, d'où $\ell_k = 3^{k-1} + 3^{k-2} + ... + 3 + 1$. Par contre la substitution σ donnée par

$$\sigma(1) := 12 \quad , \quad \sigma(2) := 21 \, ,$$

conduit au mot infini

$$\mu := 2112122112...$$

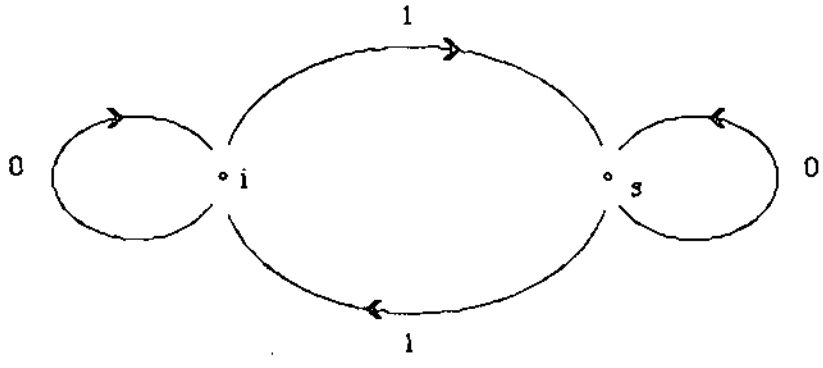
(et au mot μ' obtenu par échange les lettres). Le mot μ est point fixe de σ, en ce sens que

$$\mu = \sigma(\mu_0)\, \sigma(\mu_1)\, \sigma(\mu_2)\, \sigma(\mu_3)\, \sigma(\mu_2)...$$

les 2^k premiers termes de μ étant donnés par le k-ième itéré $\sigma^k(2)$ de 2 par σ . Le mot infini μ n'est pas périodique; il correspond à la célèbre suite de Morse [11] qui a fait l'objet de nombreux travaux et suscité des investigations plus générales, notamment dans [2,3,7,8,13,14,15]. Une des

propriétés les plus étonnantes du *langage* de μ, découverte par Morse et Hedlund en 1938 [6] est qu'aucun mot $\mu_n \mu_{n+1} \ldots \mu_{n+2k}$ lu dans μ n'est de la forme $(a_1 a_2 \ldots a_k)(a_1 a_2 \ldots a_k)a_1$. La réciproque de cette propriété , à savoir les mots qui n'ont pas de *sous-mots* de la forme $(a_1 a_2 \ldots a_k)(a_1 a_2 \ldots a_k)a_1$ sont lus dans μ , a été établie par Gottschalk et Hedlund [4].

Lorsqu'un mot infini u est point fixe d'une substitution uniforme sur un ensemble fini de symboles on le dit encore *automatique* . On sait en effet [3] qu'un tel mot est reconnu par un automate fini . Par exemple pour la suite de Morse, l'automate déterminé par les deux états i , s, le graphe orienté suivant



et la fonction de sortie $\tau : \{i , s\} \rightarrow \{1 , 2\}$ donnée par $\tau(i) = 2$, $\tau(s) = 1$ reconnait la suite de Morse. Précisons la signification de ceci. Les arcs indexés par les chiffres 0 et 1 correspondent à des instructions I_0 et I_1 : l'instruction I_0 est l'identité sur l'ensemble des états et I_1 représente l'échange entre les deux états. Si maintenant l'entiers n est représenté en base deux par la suite de ses chiffres $e_k e_{k-1} \ldots e_0$ ($n = 2^k e_k + \ldots + 2 e_1 + e_0$), alors

$$\mu_n = \tau(I_{e_k}(I_{e_{k-1}} \ldots (I_{e_0}(i))\ldots)) .$$

Sous cette forme on reconnaît une autre définition classique de μ , en introduisant la somme des chiffres en base deux. Classiquement, avec les notations précédentes, la suite de Morse est caractérisée par :

$$\mu_n = 2 \Leftrightarrow e_k + e_{k-1} + \ldots + e_0 \in 2\mathbb{N} .$$

I.-2. NOTATIONS ET DÉFINITIONS.

Espaces de mots .- Dans toute la suite $\mathfrak{a}$ désigne un ensemble fini de q symboles appelés *lettres*. L'ensemble $\mathfrak{a}$ est appelé *alphabet* et nous fixons $\mathfrak{a} := \{1,2,...,q\}$. Soit $\mathfrak{a}^*$ le monoide libre engendré par $\mathfrak{a}$. Un élément m de $\mathfrak{a}^*$ est donc un produit fini $m = m_1...m_k$ d'éléments de $\mathfrak{a}$ et que l'on regarde alors comme un mot sur l'alphabet $\mathfrak{a}$. Le nombre k de lettres pour écrire m est appelé *longueur* de m et se note $|m|$. L'élément neutre de $\mathfrak{a}^*$ est le mot vide, on le note $\wedge$. Par définition, $|\wedge| = 0$ et on pose

$$\mathfrak{a}^k := \{ m \in \mathfrak{a}^* ; |m| = k \}.$$

L'ensemble des mots infinis sur $\mathfrak{a}$ ($= \mathfrak{a}^{\mathbb{N}}$) est noté $\mathfrak{a}^\infty$. Soit a une lettre de $\mathfrak{a}$. On note simplement a* pour l'ensemble {a}* des mots finis formés de la seule lettre a. Le mot infini formé par la lettre a est noté $a^{(\infty)}$ ou simplement a^∞. Plus généralement, la lettre a peut être remplacée par un mot m et définir ainsi l'ensemble m* et le mot infini m^∞. Ce mot est dit périodique, de période m. D'autre part, nous définissons par récurrence les mots $m^{(k)}$ en posant $m^{(0)} = \wedge$ et $m^{(k)} = m^{(k-1)}m$ pour $k \in \mathbb{N}^*$.

Soit maintenant l'ensemble $\mathfrak{M}(\mathfrak{a}) = \mathfrak{a}^* \cup \mathfrak{a}^\infty$. Un mot fini $m := m_0...m_{k-1}$ dans $\mathfrak{a}^*$ est dit *préfixe* d'un mot $u := u_0 u_1...u_{k-1}...$ de $\mathfrak{M}(\mathfrak{a})$, si $m_i = u_i$ pour tout i = 0, 1,...,k-1. Si u est fini avec $u := u_0 u_1...u_{\ell-1}$ et $|u| \geq |m|$, alors le mot m est dit *suffixe* de u si $m_i = u_i$ pour tout $i = \ell-k, \ell-k-1,...,\ell-1$. D'une manière générale, pour tout mot m de $\mathfrak{a}^*$, on note s(m) (resp. p(m)) un suffixe (resp. un préfixe) de m. Plus précisément, $s_k(m)$ (resp. $p_k(m)$) désigne le suffixe (resp. le préfixe) de m de longueur k ($\leq |m|$). Un préfixe et un suffixe de m sont dits stricts s'ils sont distincts de m. On note $m\mathfrak{a}^*$ (resp. $\mathfrak{a}^*m$) l'ensemble des mots finis de préfixe m (resp. de suffixe m). Plus généralement, si A et B sont des parties de $\mathfrak{a}^*$ on note AB l'ensemble des mots $\alpha\beta$ tels que $\alpha \in A$ et $\beta \in B$.

Introduisons une topologie sur $\mathfrak{M}(\mathfrak{a})$. Soient u , v des mots distincts dans $\mathfrak{M}(\mathfrak{a})$ et

posons

$$\omega(u,v) := \text{Max}\ \{\ |\ m\ |\ ;\ m \in \mathfrak{a}^*\ \text{et}\ \ m\ \text{préfixe commun de}\ u\ \text{et}\ v\ \}.$$

L'application

$$d : \mathfrak{M}(\mathfrak{a}) \times \mathfrak{M}(\mathfrak{a}) \to \mathfrak{M}(\mathfrak{a})$$

définie par $d(u,v) := 2^{-\omega(u,v)}$ si $u \neq v$ et $d(u,v) := 0$ si $u = v$ est une distance sur $\mathfrak{M}(\mathfrak{a})$. L'espace métrique ainsi obtenu est compact ; $\mathfrak{a}^\infty$ est un sous-espace compact et $\mathfrak{a}^*$ est un sous-espace discret dense de $\mathfrak{M}(\mathfrak{a})$.

Substitution .- On apelle *substitution* une application $f : \mathfrak{a} \to \mathfrak{a}^*$. Cette application se prolonge de manière naturelle en morphisme $\mathfrak{a}^* \to \mathfrak{a}^*$ de monoïde puis en une application de l'espace $\mathfrak{M}(\mathfrak{a})$ sur lui-même, encore notée f, et définie sur $\mathfrak{a}^\infty$ par

$$f(u_0u_1...u_k...) = f(u_0)f(u_1)...f(u_k)...$$

On obtient ainsi une application continue puisque pour tout u, v de $\mathfrak{M}(\mathfrak{a})$ on a de manière évidente :

(1) $$d(f(u),f(v)) \leq d(u,v)$$

Notons que tout morphisme de monoïde $f : \mathfrak{a}^* \to \mathfrak{a}^*$ est déterminé par sa restriction sur $\mathfrak{a}$, restriction que l'on appelle aussi *substitution* f *sur* $\mathfrak{a}$. Selon la terminologie de Cobham [3], une substitution (ou le morphisme qu'elle définit) est dite *uniforme de module* ρ si $\rho = |\ f(a)\ |$ pour toutes les lettres de l'alphabet. Un morphisme uniforme de module 1 est appelé *morphisme littéral*. Une substitution f sur $\mathfrak{a}$ est dite :

- *croissante*, si pour toutes les lettres a de $\mathfrak{a}$ on a $|\ f(a)\ | \geq 2$;
- *non effaçante*, si pour toutes les lettres a de $\mathfrak{a}$ on a $f(a) \neq \wedge$.

Supposons qu'il existe une lettre a telle que $f(a) = aA$ où A est un mot non vide et soit P_a l'ensemble des mots finis ou infinis de préfixe a. Il est clair que $f(P_a) \subset P_a$ et que P_a est compact. Le morphisme f est en fait une contraction sur P_a, la formule (1) donnant ici :

(2) $$\forall\ u, v \in P_a\ :\ d(f(u),f(v)) \leq 2^{-|A|}\, d(u,v)\ .$$

En particulier, f possède un point fixe c dans P_a donné par la limite :

$$c = \lim_{k \to \infty} f^k(a) = \lim_{k \to \infty} aAf(A)f^2(A)...f^k(A).$$

Facteurs .- Soit u un mot dans $\mathcal{M}(\mathcal{A})$. Un mot $m := m_0 m_1 ... m_s$ dans $\mathcal{A}^*$ est dit *facteur* de $u := u_0 u_1 u_2...$ s'il existe un rang k tel que $m_i = u_{k+i}$ pour tout $i = 0,1,...,s.$ (en particulier, on a $|m| + k \leq |u|$). Si m est facteur de u , on note simplement $m \mid u$. La relation $(. \mid .)$ est une relation d'ordre sur $\mathcal{A}^*$. L'ensemble des facteurs de u est noté F(u) et l'ensemble des mots de longueur n qui sont facteurs de u est noté $F_n(u)$. On pose maintenant

$$p(u,n) := \mathrm{Card}(F_n(u)) .$$

Si aucune confusion n'est possible l'entier p(u,n) sera noté simplement p(n). D'autre part nous désignons par $F_n^{(k)}(u)$ l'ensemble des facteurs de u de longueur n qui sont prolongeables à droite de k manières différentes, c'est-à-dire :

$$F_n^{(k)}(u) := \{ m \in \mathcal{A}^* , |m| = n \text{ et } \mathrm{Card}\{a \in \mathcal{A} ; ma \mid u\} = k \} .$$

s'il n'y a pas d'ambiguïté, les ensembles $F_n(u)$ et $F_n^{(k)}(u)$ seront notés respectivement F_n et $F_n^{(k)}$. Nous avons par définition :

$$F_n(u) = \bigcup_{k \geq 1} F_n^{(k)}(u) ,$$

l'union étant disjointe.

Remarques :

1) Soit m un facteur de u . Si m est prolongeable à droite de k manières différentes, alors tout suffixe s de m est prolongeable à droite de k manières différentes au moins, ces k manières se faisant avec les mêmes lettres que pour m .

2) Il découle de la remarque 1) que si aucun facteur de u de longueur donnée n n'est pas prolongeable de k manières différentes, alors aucun facteur de u de longueur supérieure à n ne peut être prolongeable de k manière au plus.

3) Si u est point fixe d'une substitution f , alors toute image par f d'un facteur de u est aussi un

facteur de u. La réciproque est fausse en général si on ne fait pas d'hypothèses sur f .

I.- 3. RÉSULTATS.

Soit u un mot infini, point fixe d'une substitution uniforme f sur l'alphabet $\mathcal{A}$. Dans ce travail, nous étudions la suite des entiers $p(u,n) := \text{card}(F_k(u))$ .On a évidemment

$$p(u,n) \leq (\text{Card}(\mathcal{A}))^n.$$

A. Cobham a montré (1972 [3]) qu' il existe une constante C (= C(u)) telle que

$$p(u,n) \leq C.n$$

et N. Bleuzen-Guernalec (1986 [2]) a précisé la constante C en la majorant par $\rho(\text{Card}(\mathcal{A}))^2$.

Notre objectif principal est de démontrer l'existence d'un automate donnant la suite

$$n \to p(u,n+1) - p(u,n)$$

à partir de l'écriture de n en base ρ , ceci dans le cas où u est un mot infini *minimal* (i.e. le système dynamique (S, K_u) associé à u est minimal, cf. II.- 1).

Dans la partie II nous donnons quelques propriétés des mots infinis u minimaux obtenus par substitutions f . Nous donnons un critère effectif de minimalité. On peut se libérer de la condition restrictive $\lim_{k:\infty} | f^k(a) | = +\infty$ pour tout $a \in \mathcal{A}$, habituellement faite [10,13]. Lorsque u est minimal, la fonction p(v,.) est indépendante de $v \in K_u$ et cette propriété caractérise la minimalité de u .

La Partie III est consacrée à la démonstration du théorème suivant :

Théorème .- *Soit* u *point fixe de la substitution uniforme* f *de module* ρ. *On suppose* f *injective et* u *minimale mais pas périodique. Alors il existe une constante* L_0 *telle que pour tout facteur* M *de* u , *la factorisation* M = Bf(A)C avec $|B| < \rho$, $|C| < \rho$ *et* A *facteur de* u , *est unique* .

La démonstration se fait tout d'abord pour un alphabet à deux lettres. C'est le cas essentiel que

nous obtenons après une évaluation explicite du nombre maximum d'itérations successives des mots de deux lettres. On trouve un résultat analogue dans J. Martin (1970) [9] et notamment pour les substitutions non nécessairement uniformes (1973)[10], mais la démonstration dans ce cas général n'est pas correcte et nous donnerons un contre-exemple. Notre méthode, qui ne s'applique qu'aux subtitutions uniformes, conduit à un meilleur résultat en fournissant explicitement la longueur du mot à lire pour avoir l'unicité.

La partie IV traite du calcul automatique des $p(n) := p(u,n)$. la suite $q(n) := p(n+1) - p(n)$ ne prend qu'un nombre fini de valeurs et nous pécisons ce résultat sous la forme suivante :

Théorème.- *La suite* $n \to q(n)$ *est reconnaissable par une* ρ - *automate.*

La suite de Morse et la suite de Rudin-Shapiro sont examinées en détail dans la partie suivante. Rappelons que la suite de Rudin-Shapiro , qui compte modulo 2 le nombre de "11" dans l'écriture des entiers en base 2, est aussi l'image par un morphisme littéral d'un point fixe de la substitution suivante τ sur un alphabet à 4 lettres :

$$\tau(1) = 12 \quad , \quad \tau(2) = 13 \quad , \quad \tau(3) = 42 \quad , \quad \tau(4) = 43 \quad .$$

Nous terminerons en montrant que le caractère minimal de la suite est une condition suffisante mais non nécessaire.

II.- MINIMALITÉ

II.- 1. SYSTEME SYMBOLIQUE D'UN MOT INFINI.

Soit $S : \mathcal{A}^\infty \to \mathcal{A}^\infty$ l'application *shift* qui consiste à effacer la première lettre :

$$S(a_0a_1a_2...) = a_1a_2a_3... \ ;$$

S est une application continue sur $\mathcal{A}^\infty$, partie compacte de $\mathcal{M}(\mathcal{A})$. Le couple $(S, \mathcal{A}^\infty)$ est appelé *shift symbolique (plein)* sur $\mathcal{A}$. Un *système symbolique*, sur l'ensemble des symboles $\mathcal{A}$, est un sous-shift de $(S, \mathcal{A}^\infty)$, défini par une partie compacte K de $\mathcal{A}^\infty$ stable par S. On le note (S,K). Un tel système est dit *minimal* si les seules parties fermées F de K stables par S sont l'ensemble vide $\emptyset$ et K.

Soit u dans $\mathcal{A}^\infty$ et notons K_u l'orbite fermée de u sous l'action de S c'est-à-dire l'adhérence dans $\mathcal{A}^\infty$ de l'ensemble $\{S^k(u) ; k \in \mathbb{N}\}$. On a $S(K_u) \subset K_u$. Le mot infini u est ainsi associé au système symbolique $\mathcal{K}_u := (S, K_u)$.

II.-2. MOTS MINIMAUX.

Critères de minimalité.

Définition **.** - Un mot infini u sur l'alphabet $\mathcal{A}$ est dit *minimal* si le système symbolique associé $\mathcal{K}_u$ est minimal.

La caractérisation suivante est classique [4] :

Théorème A .- *Le mot* u *est minimal si et seulement si pour tout facteur* m *de* u, *il existe un entier* $\ell\ (= \ell(m))$ *tel que :*

(1) $$\forall\ k \in \mathbb{N}\ ,\ m \mid u_k u_{k+1} ... u_{k+\ell}\ .$$

La condition (1) du théorème A exprime que tout mot du langage de u apparaît dans u avec des lacunes bornées (par ℓ). Lorsque u est point fixe d'une substitution f il est intéressant de donner un critère de minimalité en fonction de f (cf. [13]).

Théorème B.- *Soit* u *point fixe de la substitution* f *sur l'alphabet* $\mathcal{A}$. *On suppose* $u \in l\mathcal{A}^{\infty}$, $|f(1)| \geq 2$ *et que toutes les lettres de* $\mathcal{A}$ *soient dans* u . *Alors les propriétés suivantes sont équivalentes :*

(i) u *est minimal et* $\lim_{k:\infty} |f^k(a)| = +\infty$ *pour toutes les lettres* a *de* $\mathcal{A}$.

(ii) $\exists\ L\ (\leq \mathrm{card}(\mathcal{A}))\ ,\ \forall\ a \in \mathcal{A}\ ,\ 1 \mid f^L(a)$

(iii) $\forall\ a \in \mathcal{A}\ ,\ \exists\ k(a) \in \mathbb{N}^*\ ,\ 1 \mid f^{k(a)}(a)$.

Démonstration.- L'implication (i) $\Rightarrow$ (iii) résulte directement du Théorème A. Supposons (iii) et soit $L := \mathrm{Max}\{\ k(a)\ ;\ a \in \mathcal{A}\}$. Puisque $f(1) \in l\mathcal{A}^*$, on a $1 \mid f^s(a)$ pour tout $s \geq k(a)$. Reste donc à montrer que $L \leq \mathrm{card}(\mathcal{A})$ pour obtenir (ii). Définissons les ensembles

$$\mathcal{A}_s = \{\ a \in \mathcal{A}\ ;\ k(a) \leq s\ \}.$$

Il est clair que si pour une lettre a , on a $k(a) = s$, alors il existe des lettres distinctes $b_1, b_2, \ldots, b_{s-1}$ telles que $k(b_i) = s - i$ et $b_i \mid f^i(a)$. On a donc dans ce cas $s \leq \mathrm{card}(\mathcal{A}_s) \leq \mathrm{card}(\mathcal{A})$ et en particulier $L \leq \mathrm{card}(\mathcal{A})$ et $\mathcal{A}_L = \mathcal{A}$.

Supposons (ii) et soit m un facteur de u . Alors il existe ℓ tel que $m \mid f^{\ell}(1)$ et comme $f^{\ell}(f^L(u)) = u$, on a $f^{\ell}(1) \mid f^{\ell + L}(u_n)$ pour tout indice n . Le mot m apparaît donc dans u avec des lacunes bornées par $(\mathrm{Max}\{|f(a)|\ ;\ a \in \mathcal{A}\})^{\ell + L} - |m|$. Par ailleurs $|f^{k+L}(a)| \geq |f^k(1)|$ pour toute lettre a et tout entier $k \geq 0$. Les hypothèses faites sur f et u impliquent donc $\lim_{k:\infty} |f^k(1)| = +\infty$.

♦

Lorsque $\mathcal{A} = \{1,2\}$, on a un critère de minimalité très simple :

Théorèmé C.- *Soit* u *un mot infini point fixe d'une substitution* f *sur* $\mathcal{A} = \{1,2\}$ *tel que* 1 *soit préfixe de* u *et* $u \neq 1^{\infty}$.

(i) *Supposons* f *croissante, alors :*

$$u \text{ minimal} \Leftrightarrow f(2) \notin 2^*.$$

(ii) *Supposons* $|f(1)| \geq 2$ *et* $f(2) = 2$, *alors :*

$$u \text{ minimal} \Leftrightarrow f(1) \in 1\mathcal{A}^*1.$$

Démonstration.

(i): Si le mot u est minimal, n'étant pas formé que de 1, il admet les facteurs $f^k(2)$. Donc si $f(2) \in 2^*$, le mot 2 apparaît dans u avec des lacunes arbitrairement grandes, contrairement au théorème A. Réciproquement si $f(2) \notin 2^*$, alors $1 \mid f(2)$ et la minimalité résulte du théorème B.

(ii): Si 2 est suffixe de f(1) alors on peut écrire $f(1) = 1A12^{(b)}$ avec $b \geq 1$ et plus généralement $f^k(1) = 1A_k12^{(kb)}$. Cette forme s'obtient par récurrence puisque $f^{k+1}(1) = (1A12^{(b)})f(A_k)(1A12^{(b)})2^{(kb)}$.

Ainsi les mots 2^s, $s \geq 0$, sont tous facteurs de u qui n'est donc pas minimale. Supposons maintenant que 1 soit suffixe de f(1). Posons $r := |f(1)|$. Le mot $2^{(r-1)}$ n'est pas facteur de f(1) et s'il n'est pas facteur d'un mot m, il n'est pas facteur du mot f(m). Il en résulte que 1 apparaît dans u avec des lacunes bornées par r-1. Par suite $f^k(1)$ apparaît aussi dans u avec des lacunes bornées et donc u est minimale.

♦

III.- COMPLEXITE DU LANGAGE

III. 1.- LE THEOREME D'UNICITE.

Dans la suite f désigne une substitution telle que $f(1) \in 1\mathcal{A}^*$ et $|f(1)| \geq 2$. Alors f admet un seul point fixe dans $1\mathcal{A}^\infty$ noté u . Sauf mention explicite du contraire, on supposera f *uniforme* de module ρ et le mot infini u *minimal* avec toutes les lettres de $\mathcal{A}$ apparaissant dans u .

Soit m un facteur de u . Il peut se factoriser sous la forme

(1) $$m = Bf(A)C$$

avec les conditions :

(2) B suffixe strict d'un mot f(b) , C préfixe strict d'un mot f(c) et bAc facteur de u.

Définition.- Un facteur m de u est dit *mot rythmé* sur l'alphabet $\mathcal{A}$ s'il n'existe qu'un seul triplet (A,B,C) vérifiant (1) et (2) .

On se propose de démontrer que tout facteur de u assez long est un mot rythmé. Ce résultat est énoncé dans [9] dans le cas de deux lettres. Nous donnerons ici une démonstration utilisant une méthode différente avec une information précise sur la longueur minimale du mot m pour que celui-ci soit rythmé. Le résultat est ensuite étendu au cas d'un alphabet à plus de deux lettres.

On trouve aussi chez Martin [10] l'énoncé du même théorème dans le cas des suites de longueurs non constantes sur un alphabet à deux lettres. Cette généralisation est fausse comme le montre la proposition suivante :

PROPOSITION 1 .- *Supposons* f *une substitution non uniforme sur* $\mathcal{A} = \{1,2\}$, f(2) *préfixe*

de f(1) *et* u *point fixe minimal non périodique de* f . *Alors il existe une infinité de facteurs* m *de* u *tels que :*

m2 | u *et* f(m2) *n'est pas rythmé.*

Démonstration.- En effet on a f(m2) = f(m)f(2) = f(m)p(f(1)). Il suffit donc de démontrer l'existence d'une infinité de mots m tels que m2 et m1 soient des facteurs de u . Soit w un préfixe de u . Par minimalité de u , il existe k tel que w soit aussi préfixe de $S^k u$ (= $u_k u_{k+1}$...). Comme u n'est pas périodique, on a $u \neq S^k u$ et il existe donc un mot w' tel que ww'1 et ww'2 soient préfixes pour u ou $S^k u$. Le mot m = ww'2 satisfait aux conditions de la proposition.

◆

Il suffit donc pour obtenir un contre-exemple de choisir f vérifiant les hypothèses de la proposition 1 avec des conditions assurant la minimalité de u (par exemple, f croissante, f(1) ∈ 1$\mathcal{A}$* et 1 | f(2)) . On peut cependant espérer que sous certaines conditions peu restrictives sur f (et u), les facteurs de longueurs assez grandes soient rythmés.

THEOREME 1.- *Soit* f *une substitution uniforme de module* $\rho \geq 2$ *injective sur l'alphabet* $\mathcal{A} = \{1,2, \ldots, q\}$. *On suppose* u ∈ 1$\mathcal{A}$* *et* u *minimale mais pas périodique. Alors il existe* $L_0 = L_0(\rho,q)$ ($\leq \rho(\rho^3 + \rho^2 + \rho + 1)$ *pour un alphabet à deux lettres) tel que tout facteur* m *de* u *de longueur plus gande que* L_0 *soit un mot rythmé.*

Avant de démontrer ce théorème, il est intéressant de remarquer la propriété générale suivante :

PROPOSITION 2 .- *Soit* u *un mot infini point fixe d'une substitution uniforme* f *de module* ρ *(* u *n'est pas ici supposé minimal) . Supposons f injective sur* $\mathcal{A}$. *S'il existe un facteur* R *de* u *rythmé et de longueur* $\geq \rho$ *alors tout facteur de* u *dont* R *est un facteur , est aussi un*

mot rythmé.

Démonstration. Soit m un facteur de u dont R est facteur. La factorisation de u par bloc de ρ lettres , i.e. $u = f(u_0)f(u_1)f(u_2)f(u_3)\ldots$ donne une factorisation de R lu dans m. Celle-ci étant unique, elle détermine alors un préfixe B d'un mot de $f(\mathcal{A})$, de longueur $|B| < \rho$ et tel que si $R = u_k u_{k+1} u_{k+2} \cdots u_{k+r}$, alors $k + |B|$ est un multiple de ρ . La position de $u_{k+|B|}$ dans R , et par suite dans m , fournit une factorisation de m induite par celle de u , mais celle-ci ne dépend pas de la lecture de m dans u . Soit

$$m = \pi(m_1)\,(m_2)\ldots\,(m_t)\sigma$$

cette factorisation unique où les mots m_i sont dans $f(\mathcal{A})$ et les mots π , σ ont des longueurs strictement inférieures à ρ . Le caractère injectif de f donne la factorisation *rythmée* de m.

♦

Notons que si f est injective et u mot infini laissé fixe par f , alors la relation $f(u) = u$ détermine une factorisation de u en facteur de longueur ρ. Appelons cette factorisation *canonique*. Dire qu'un facteur de u est rythmé sur $\mathcal{A}$ signifie que la factorisation canonique de u induit sur m une seule factorisation, quel que soit les diverses positions de m dans u .

Illustrons la proposition 2 et le théorème 1 par un exemple :

Le mot de Morse. Soit σ la substitution sur $\{1, 2\}$ définie par $\sigma(1) = 12$ et $\sigma(2) = 21$. σ est injective. Le mot de Morse envisagé dans l'introduction, est le point fixe μ de σ de préfixe 2 (cf. I.1). On observe facilement que les seuls mots rythmés dans μ de longueur 3 sont 221, 122, 112 , 211, les autres ne sont pas rythmés (121, 212) ou ne sont pas des facteurs (222 , 111). Soit m un mot de longueur 4 . Supposons qu'il n'admette pas de facteurs rythmés de longueur 3. Alors 121 ou 212 sont les seuls facteurs possibles de longueur 3 de sorte que m = 1212 ou m

= 2121. On vérifie directement que ces deux mots sont des facteurs de μ. Supposons m = 1212 = 1f(2)2. Etant facteur de μ le mot 21f(2)21 = f(222) est aussi facteur de μ et ne provient que de l'image par f de 222. Par suite le mot 222 est lui aussi un facteur de μ, ce qui est faux. Ainsi 1212 est rythmé de même que 2121. Les autres facteurs de μ de longueur 4 contiennent nécessairement l'un des quatre mots rythmés de longeur 3. Finalement on a montré :

PROPOSITION 3.- *Dans le mot de Morse tout facteur de longueur supérieure ou égale à 4 est rythmé et la valeur 4 est optimale.*

Remarque : La constante L_0 du théorème 1 vaut 32 dans le cas de la suite de Morse. Mais la démonstration donne en fait un procédé calculatoire simple qui dans la pratique fournit une constante plus faible (cf. Théorème 2).

III.2.- COMMUTATION.

Soient m et u des mots dans $\mathcal{A}^*$. On dit que u est un commutant de m s'il est à la fois préfixe et suffixe de m. Il existe donc v et v' dans $\mathcal{A}^*$ tel que m = uv = v'u. Les deux lemmes suivants décrivent la stucture des mots ayant des commutants identiques.

LEMME 1.- *Soient* m, u, v *des mots de l'alphabet* $\mathcal{A}$ *tels que* m = uv = vu, *et soit* d := (|u|, |v|) *le* pgcd *des longueurs de* u *et* v. *Alors* $m = p^{(r)}$ *où* p *est le préfixe de* m *de longueur* d *et* r = |m| / d.

Démonstration. - Commençons par démontrer que pour k = [|m| / |v|] on a $m = p(v)v^{(k)} = vp(v)v^{(k-1)}$ où p(v) est le préfixe de v de longueur |m| - k.|v|. Cette décomposition est

évidente si $|u| \leq |v|$ car alors $u = p(v)$ et $k = 1$. Supposons $|u| > |v|$. De $uv = vu$ on déduit $u = vu'$, d'où $vu'v = vvu'$ et par suite $u'v = vu'$. Par récurrence, on se ramène donc au cas où $u = v^{(k)}u''$ avec $|u''| \leq |v|$ et $u''v = vu''$ de sorte que $u'' = p(v)$ et m a la forme requise.

Le lemme est évident si $|u| = |v|$, ou si l'un des mots u, v est une lettre.Supposons le lemme vrai pour les mots m tels que $|m| < K$. Soit maintenant un mot $m = uv = vu$ de longueur K. On peut supposer $|u| > |v|$ et par le raisonnement précedent, on peut écrire $m = p(v)v^{(k)} = vp(v)v^{(k-1)}$. Alors le mot $m' = p(v)v$ $(= vp(v))$ est de longueur $|m'| < K$ et de plus $(|u|, |v|) = (|v|, |p(v)|) = d$ d'où par hypothèse de récurrence, $m' = \pi^{(r')}$ avec $r' = |m'|/d$ et π préfixe de v. Il existe donc s et t entiers tels que $p(v) = \pi^{(sd)}$ et $v = \pi^{(td)}$. Par suite $m = \pi^{(sd + ktd)}$.

◆

LEMME 2.- *Soient* m, m', u, v *et* v' *des mots dans* $\mathfrak{a}^*$ *tels que*

$$m = uv = v'u \quad , \quad m' = uv' = vu \quad et \quad |m| = |m'| .$$

Alors $m = m'$, $v = v'$ *et* $m \in p^*$ *où* p *est le préfixe de* m *de longueur* $(|u|, |v|)$.

Démonstration .- Le lemme est évident si $|u| \geq |v|$ $(= |v'|)$ car dans ce cas v et v' sont tous les deux préfixes de u et de même longueur . Ils sont donc égaux, et le lemme 1 s'applique. Supposons $|u| \leq |v|$. Posons $v = v_1u$ et $v' = v'_1u$. Alors $u\,v_1u = v'_1uu$ et $u\,v'_1u = v_1uu$ d'où $u\,v_1 = v'_1u$ et $u\,v'_1 = v_1u$. On se ramène donc au cas précédent avec $|v_1| = |v| - |u|$. Par récurrence on obtient pour $k = [\,|v|/|u|\,]$:

$$v = v_k u^{(k)} \quad \text{et} \quad v' = v'_k u^{(k)} ,$$

avec $uv_k = v'_k u$ et $uv'_k = v_k u$. Mais $|u| \geq |v_k|$, d'où $v_k = v'_k$ et le lemme 2 résulte maintenant directement du lemme 1.

◆

III.3.- MAJORATION DES FACTEURS PUISSANCES.

Mots périodiques.- Soit u point fixe d'une subtitution f de module ρ sur un alphabet $\mathcal{A}$ et supposons que toute les lettres de $\mathcal{A}$ apparaissent dans u . Il est clair que si u est purement périodique , de mot période de longueur ρ^ν alors tous les mots $f^\nu(a)$, $a \in \mathcal{A}$, sont identiques. Réciproquement, s'il existe ν tel que tous les mots $f^\nu(a)$, $a \in \mathcal{A}$, soient identiques, alors $f^\nu(1)$ est un mot période de u de longueur ρ^ν .

Si aucune puissance de ρ n'est période de u la propriété précédente est alors mise en défaut. Par exemple la substitution suivante :

$$f(1) = 121 \ , \ f(2) = 212$$

qui a pour point fixe $(12)^\infty$, vérifie $f^k(1) \neq f^k(2)$ pour tout entier k . Cette propriété est évidemment conservée dans le cas où u n'est pas périodique sur un alphabet à deux lettres; elle traduit le caractère injectif de f . En d'autres termes :

LEMME 3. - *Soit* f *une substitution uniforme sur* $\mathcal{A} = \{1, 2\}$ *de point fixe* u *dans* $1\mathcal{A}^*$, *non périodique. Alors* $f^k(1) \neq f^k(2)$ *pour tout entier* k *et pour tous mots* m , m' *dans* $\mathcal{A}^*$ *on a*

$$f(m) = f(m') \Rightarrow m = m' \ .$$

Définition.- Pour tout mot fini ou infini u et pour tout mot fini m non vide sur l'alphabet $\mathcal{A}$, on pose

$$L_u(m) = \text{Sup} \{k \ ; \ m^{(k)} \mid u \}.$$

Majoration de $L_u(m)$.- Nous étudions seulement $L_u(m)$ lorsque m est une lettre ou un mot de deux lettres distinctes.

LEMME 4 .- *Soit* u *un mot minimal non périodique, point fixe d'une substitution uniforme* f *de module* ρ *sur* $\mathcal{A}$ ($= \{1, 2\}$). *Alors* $L_u(1)$ *et* $L_u(2)$ *(resp.* $L_u(12)$ *et* $L_u(21)$) *sont majorés par* $\rho^2 + \rho$ *(resp. par* $\rho^3 + \rho^2 + \rho$ *)* .

Démonstration .- Nous supposons u dans $1\mathcal{A}^*$ et comme u est minimal on a $f(1) \notin 1^*$ et $f(2) \notin 2^*$ (Théorème C). Nous distinguons deux cas, suivant que $f(2)$ est ou n'est pas dans 1^* .

Premier cas : Supposons $f(2) \notin 1^*$, alors pour tout $i \in A$ on a $L_u(i) < 2\rho$ ($\leq \rho^2 + \rho$). En effet, dans le cas contraire, il existe une lettre j telle que $L_u(j) \geq 2\rho$ et $j^{(2\rho)}$ est un facteur de u. Le morphisme f étant de module ρ l'un des mots $f(1)$, $f(2)$ est facteur de $j^{(2\rho)}$. Comme $f(1) \notin 1^*$ mais $f(1) \in 1\mathcal{A}^*$ et $f(2) \notin 2^*$, on obtient $f(2) = 1^{(\rho)}$ contrairement à l'hypothèse.

Montrons maintenant que pour tout mot $ij \in \mathcal{A}^2$ tel que $i \neq j$ on a $L_u(ij) < 2\rho^2 + \rho$ ($\leq \rho^3 + \rho^2 + \rho$). Supposons que pour un choix de ij on ait $L_u(ij) \geq 2\rho^2 + \rho$. Le mot $(ij)^{2\rho^2 + \rho}$ est alors facteur de u et se décompose sous la forme $s(f(a))f(b_1)...f(b_{2\rho})p(f(c))$, avec $b_k \in \mathcal{A}$ pour $k \in \{1,...,2\rho\}$ et a , b dans $\mathcal{A} \cup \{\wedge\}$. Si ρ est pair, alors $f(b_k) = (ij)^{\rho/2}$ ou $f(b_k) = (ji)^{\rho/2}$ pour tout k et par le lemme 3 on a $b_1 = b_2 = ... = b_{2\rho}$ ce qui est contraire à $L_u(b) < 2\rho$ pour toute lettre b de $\mathcal{A}$. Si ρ est impair, nous avons maintenant soit $f(b_{2k-1}) = (ij)^{(\rho-1)/2}i$ et $f(b_{2k}) = (ji)^{(\rho-1)/2}j$, soit $f(b_{2k-1}) = (ji)^{(\rho-1)/2}j$ et $f(b_{2k}) = (ij)^{(\rho-1)/2}i$, pour $1 \leq k \leq \rho$. Dans tous les cas, on a $f(1) = (12)^{(\rho-1)/2}(1)$ et $f(2) = (21)^{(\rho-1)/2}(2)$. Mais alors u est périodique, ce qui est exclu.

Deuxième cas : Supposons $f(2) \in 1^*$. Comme $f(1) \in 1\mathcal{A}^*$, la lettre 2 n'apparaît que dans f(1) et par suite $L_u(2) < \rho$.

Montrons que $L_u(1) < \rho^2 + \rho$. Dans le cas contraire, $1^{(\rho^2 + \rho)}$ est un facteur de u et sa décomposition sous la forme $s(f(a))f(b_1)...f(b_\rho)p(f(c))$ donne $f(b_k) \in 1^*$ pour $k = 1, ..., \rho$. Par suite $b_1 = ... = b_\rho = 2$ ce qui contredit l'inégalité $L_u(2) < \rho$.

Examinons maintenant le cas de $ij \in \mathbf{a}^2$, $i \neq j$ et supposons $L_u(ij) \geq \rho^3 + \rho^2 + \rho$. Alors, comme précédemment, il existe un facteur m de u de longueur $\rho^2 + \rho$ tel que $f(m) \in (ij)^*$ ou $f(m) \in (ji)^*$. Par hypothèse, la lettre 1 est préfixe de $f(1)$ et de $f(2)$, d'où $f(m) \in (12)^*$. Alors ρ est pair, $f(1) = (12)^{(\rho/2)}$ et $m = 1^{(\rho^2 + \rho)}$ ce qui contredit la majoration déjà obtenue pour $L_u(1)$.

◆

III.- 4. DÉMONSTRATION DU THÉORÈME 1 POUR DEUX LETTRES.

Soit f une substitution uniforme de module $\rho \geq 2$ sur l'alphabet $\mathcal{A} = \{ 1 , 2 \}$ et soit u un point fixe minimal non périodique de f .

Soit m un facteur de u tel que $| m | \geq L_0$ et supposons m non rythmé. Il existe alors deux factorisations distinctes Bf(A)C et B'f(A')C' de m vérifiant les conditions données par (2). On peut supposer $| C | \neq | C' |$. En effet, dans le cas contraire, on a C = C' puis nécessairement $| B | = | B' |$, donc B = B'. D'où f(A) = f(A') et par le lemme 3, A = A' . On peut donc choisir $| C | < | C' |$ et poser

$$C' = X_0 C \ , \ | X_0 | = r \ , \ 0 < r < \rho - 1.$$

Posons maintenant

$$A := a_k ... a_1 \ , \ A' := a'_\ell ... a'_1 \ ;$$

on a

$$Bf(a_k)...f(a_1) = B'f(a'_\ell)...f(a'_1)X_0 \ , \ k - 1 \leq \ell \leq k \ ,$$

B et B' étant suffixes de f(b) et f(b') respectivement , X_0 préfixe de $f(a'_0)$, de telle sorte que les mots $ba_k...a_1$ et $b'a'_\ell...a'_1a'_0$ soient des facteurs de u . Posons $a_{k+1} = b$ et $a'_{\ell+1} = b'$.

Ecrivons

$$f(a_1) := X_1X_0 \ \text{ avec } \ | X_1 | = \rho - r \ \text{ et } \ f(a'_1) := X_2X_1 \ \text{ avec } \ | X_2 | = r \ .$$

On obtient ainsi

$$Bf(a_k)...f(a_2) = B'f(a'_\ell)...f(a'_2)X_2 \ ,$$

et de proche en proche, pour $i = 1, \ldots, k$ on peut écrire :

(3) $$f(a_i) = U_iV_i \ , \ f(a'_i) = V'_iU_i \ , f(a_{i+1}) = U_{i+1}V'_i \ ,$$

avec $| U_i | = | U'_i | = r$ et $| V_i | = | V'_i | = \rho - r$.

Par choix de L_0 , on a $k \geq \rho^3 + \rho^2 + \rho$. Le mot A n'est donc pas facteur d'un mot dans l'ensemble $1^* \cup 2^* \cup (12)^*$. D'autre part, écrivons A = WW' avec $| W | = [k/2]$. Comme on a $\rho \geq 2$, les mots W et W' sont de longueur plus grande que $\rho^2 + \rho$ et par le Lemme 4 il contiennent les lettre 1 et 2 . Si le mot 21 n'est pas facteur de W , alors W est facteur d'un

mot de l'ensemble 1^*2^* et par suite le mot 21 est facteur de W' . De même si le mot 12 n'est pas facteur de W, il est certainement facteur de W'. D'autre part si les mots 11 et 22 ne sont pas facteurs de A , alors celui-ci est facteur d' un mot dans $(12)^*$, ce qui est contraire au Lemme 4 . En définitive, on a deux cas possibles :

$$\{12, 21, 11\} \subset f_2(A) \text{ ou } \{12, 21, 22\} \subset f_2(A).$$

Posons $f(1) := UV$ et $f(2) := U'V'$ avec $|U| = |U'| = r$. D'après (3) , l'ensemble $E = \{VU', V'U, VU\}$ ou l'ensemble $E' = \{VU', V'U, V'U'\}$ est contenu dans $\{UV, U'V'\}$. L'hypothèse $f(1) \in 1a^*$ qui distingue la lettre 1 de la lettre 2 n'étant pas utilisé par la suite, on peut se contenter d'étudier le cas où $E \subset \{UV, U'V'\}$. Si E se réduit à un seul élément, alors $V = V'$, $U = U'$ et par suite $f(1) = f(2)$. Le mot u est alors périodique, contradiction. Ainsi E n'est formé que de deux élements. Si $U = U'$, alors $E = \{VU, V'U\}$ ce qui donne $UV = VU$ et $UV' = V'U$ ou bien $UV = V'U$ et $UV' = VU$. Le Lemme 1 dans le premier cas et le Lemme 2 dans le second donne la contradiction $f(1) = f(2)$. Si $V = V'$, le même raisonnement donne la même contradiction.

Nous pouvons donc maintenant supposer $U \neq U'$ et $V \neq V'$. Alors l'ensemble E est formé de trois éléments distints, ce qui est impossible.

La démonstration précédente montre que si A est un facteur de u contenant les mots 12, 21, et l'un des mots 11 ou 22 , alors le facteur f(A) est rythmé. Résumons:

THEOREME 2 .- *Soit* f *une substitution uniforme de module* ρ *sur* $\{1, 2\}$ *et admettant un point fixe minimal* u *non périodique et soit* L_1 *le plus petit entier* ℓ *tel que tout facteur de* u *de longueur* $\geq \ell$ *contienne les mots* 12 , 21 *et l'un des mots* 11 *ou* 22 . *Alors les facteurs de* u *de longueur* $\geq \rho(L_1 + 1)$ *sont rythmés.*

Remarques : 1) Si on ne fait pas l'hypothèse que u est minimale, l'existence de L_1 implique la

minimalité d'après le Théorème B .

2) Dans le cas de la suite de Morse, un calcul direct donne $L_1 = 5$. En effet, 1212 et 2121 sont des facteurs du mot de Morse et les mots 12121 et 21212 n'en sont pas. Dans ce cas la constante $\rho(L_1 + 1)$ (= 12) est bien meilleure que L_0 (= 32).

III.-5. DEMONSTRATION DU THEOREME 1 DANS LE CAS GENERAL.

Indication sur la démonstration .

Soit f substitution uniforme de module $\rho \geq 2$ sur l' alphabet $\mathcal{A} = \{1, 2, \ldots, q\}$, avec $q \geq 3$ et supposons f injective et qu'elle admette un point fixe u minimal non périodique. Nous allons montrer qu'il existe une constante $H = H(\rho,q)$ telle que s'il existe un facteur m de u non rythmé et de longueur plus grande que H , alors il existe un entier d diviseur de ρ et une substitution F uniforme de module ρ sur un alphabet $\mathcal{P}$ tel que $\mathcal{P} \subset \mathcal{A}^d$, $\text{card}(\mathcal{P}) \leq q - 1$ et $F(u) = u$ en regardant u comme un mot infini sur l'alphabet $\mathcal{P}$. De plus le mot m peut se lire comme un élément de $\mathcal{P}^*$ et on montre qu'il n'est pas rythmé sur l'alphabet $\mathcal{P}$. Sa longueur sur cet alphabet est $|m| / d$. Donc si $|m| \geq (\rho/2)L_0(\rho, q-1)$, on en tire une contradiction.

Le cas d'un alphabet à deux lettres revenait essentiellement à montrer que l'existence d'un mot non rythmé assez long impliquait l'égalité f(1) = f(2). En d'autres termes, la substitution f pouvait se voir comme étant définie sur un alphabet à une lettre. Nous allons suivre cette idée en deux étapes, en commençant par généraliser le Lemme 2.

étape **1.**

LEMME 4.- *Soient* $\mathcal{M}$ *un ensemble de mots de même longueur* ρ *sur un alphabet* $\mathcal{A}$. *Soit* $\mathcal{U} := p_r(\mathcal{M})$ *(resp.* $\mathcal{V} := s_{\rho-r}(\mathcal{M})$*) l'ensemble des préfixes (resp. suffixes) de longueur* r *(resp. de longueur* ρ - r*) des mots de* $\mathcal{M}$ *. On suppose de plus que*

(4) $$\mathcal{V} := p_{\rho-r}(\mathcal{M}) \text{ et } \mathcal{U} := s_r(\mathcal{M}) .$$

Soient $d := \text{pgcd}(\rho, r)$ *et* $\mathfrak{p}$ *l'ensemble des préfixes de longeur* d *des mots de* $\mathfrak{m}$. *Alors* $\mathfrak{m} \subset \mathfrak{p}^*$.

Démonstration. Quitte à échanger les rôles de $\mathfrak{u}$ et $\mathfrak{v}$ on peut supposer $r \leq \rho - r$. Soit v un mot dans $\mathfrak{v}$. Il existe $m \in \mathfrak{m}$ et $u \in \mathfrak{u}$ tel que $m = uv$ et comme v est préfixe d'un mot m_1 de $\mathfrak{m}$, il existe u' dans $\mathfrak{u}$ tel que $m_1 = vu'$, et d'après (4) il existe u_1 dans $\mathfrak{u}$ et v_1 dans $\mathfrak{v}$ tels que $m_1 = vu' = u_1v_1$. On peut alors écrire $v = u_1w_1$ et par suite $m = u\,u_1w_1$. En d'autres termes, $\mathfrak{m} \subset \mathfrak{u}\mathfrak{u}\mathfrak{w}_1 (= \mathfrak{u}^2\mathfrak{w}_1)$ où $\mathfrak{w}_1$ est l'ensemble des suffixes de longueur $\rho - 2r$ des mots de $\mathfrak{m}$. Soit $\mathfrak{w}_{k-1}$ l'ensemble des suffixes de longueur $\rho - kr$ (≥ 0) des mots de $\mathfrak{m}$. Supposons maintenant $\mathfrak{m} \subset \mathfrak{u}^k\mathfrak{w}_{k-1}$ pour un entier k tel que $\rho \geq (k + 1)r$. Alors, de $\mathfrak{v} = p_{\rho-r}(\mathfrak{m})$ on déduit comme précédemment $\mathfrak{v} \subset \mathfrak{u}^k\mathfrak{w}_k$. La relation $\mathfrak{m} \subset \mathfrak{v}'$ entraîne $\mathfrak{m} \subset \mathfrak{u}^{(k+1)}\mathfrak{w}_k$ et par suite $\mathfrak{v} \subset \mathfrak{u}^{(k+1)}\mathfrak{w}_{k+1}$. Itérons le procédé jusqu'à $t = [\rho/r]$. Tout mot m de $\mathfrak{m}$ s'écrivant sous les formes uv et $v'u'$ respectivement dans $\mathfrak{u}\mathfrak{v}$ et $\mathfrak{v}\mathfrak{u}$ avec $\mathfrak{v} = \mathfrak{u}^{(t-1)}\mathfrak{w}_{t-1}$, on a :

(5) $$m = u_1 \ldots u_t w = u'_1 \ldots u'_{t-1} w' u'_t,$$

les u_i, u'_i dans $\mathfrak{u}$ et w, w' dans $\mathfrak{w}_{t-1}$. En particulier $u_t w = w'u'_t$. Si $\rho = tr$, alors $\mathfrak{w}_{t-1}$ est vide et le lemme est démontré. Dans le cas contraire, soit $\mathfrak{m}' = s_\tau(\mathfrak{m})$ pour $\tau = \rho - (t - 1)r$, $\mathfrak{u}' = \mathfrak{u}$ et $\mathfrak{v}' = \mathfrak{w}_{t-1}$. Par hypothèses sur $\mathfrak{u}$ et $\mathfrak{v}$ on a $s_{\tau-r}(\mathfrak{m}') = s_{\tau-r}(\mathfrak{m}) = s_{\tau-r}(\mathfrak{v}) = \mathfrak{v}'$ et par suite (5) donne aussi $p_{\tau-r}(\mathfrak{m}') = \mathfrak{v}'$. D'autre part $\mathfrak{u} = s_r(\mathfrak{m}) = s_r(\mathfrak{m}') = s_r(\mathfrak{v})$ et par (5) $s_r(\mathfrak{v}) = p_r(\mathfrak{m}')$. Au total :

$$\mathfrak{u}' = p_r(\mathfrak{m}') = s_r(\mathfrak{m}') \quad \text{et} \quad \mathfrak{v}' = s_{\tau-r}(\mathfrak{m}') = p_{\tau-r}(\mathfrak{m}').$$

Nous sommes ainsi ramenés au cas précédent avec τ au lieu de ρ et $\tau - r$ au lieu de r, le pgcd restant inchangé. On voit donc que si lemme est démontré pour $\mathfrak{m}'$, alors $\mathfrak{u}'$ et $\mathfrak{v}'$ sont dans $\mathfrak{p}^*$ (notation du lemme) et par suite $\mathfrak{m}$ est aussi dans $\mathfrak{p}^*$. Le lemme est évident pour $\rho = 2d$ et s'il est vrai pour les multiples $2d, 3d, \ldots, kd$, alors le raisonnement précédent montre qu'il est encore vrai pour $\rho = (k + 1)d$. Le lemme est donc établi par récurrence sur les multiples de ρ.

♦

étape 2.

Les hypothèses et notations sont celles du Théorème 1. Soit $H_2 = H_2(\rho,q)$ une constante telle que pour tout facteur m de u on ait

$$|m| \geq H_2 \Rightarrow F_2(m) = F_2(u) .$$

Cette constante peut être choisie au mieux en fonction de f mais dans tous les cas la valeur $2\rho^{2q} - 1$ convient. En effet, si $m \mid u$ et $|m| \geq 2\rho^{2q} - 1$, alors m possède un facteur de la forme $f^{2q}(a)$. Désignons par $E_k(a)$ l'ensemble des facteurs de longueur deux apparaissant dans l'un des mots f(a), ..., $f^k(a)$. Si $E_{k+1}(a) = E_k(a)$ cela signifie que pour tout $uv \in E_k(a)$ on a $F_2(uv) \subset E_k(a)$. Il en résulte donc que $E_{k+n}(a) = E_k(a)$ pour tout $n \geq 0$ et cette égalité à lieu dès que $k \geq \text{card}(\mathcal{A}^2) = q^2$.

Choisissons maintenant $H = \rho(H_2 + 1)$ et supposons que u admette un facteur m non rythmé de longueur $\geq H$. Alors m admet deux factorisations distinctes

$$m = Bf(A)C = B'f(A')C' \tag{6}$$

vérifiant les conditions données par (2) . Comme dans le cas de la démonstration pour deux lettres (cf. III.-4), on peut écrire en adoptant les mêmes notations :

$$A := a_k...a_1 \quad , \quad A' := a'_\ell...a'_1 \ ;$$

avec $k - 1 \leq \ell \leq k$ et

$$Bf(a_k)...f(a_1) = B'f(a'_\ell)...f(a'_1)X_0 \ ,$$

où B et B' sont suffixes de f(b) et f(b') respectivement , X_0 préfixe de $f(a'_0)$ de longueur r et les mots $ba_k...a_1$ et $b'a'_\ell...a'_1a'_0$ facteurs de u .Posons $b = a_{k+1}$, $b' = a'_{\ell+1}$. On a encore les relations (3) , à savoir :

$$f(a_i) = U_iV_i \ , \ f(a'_i) = V'_iU_i \ , f(a_{i+1}) = U_{i+1}V'_i \ , \tag{3}$$

pour $1 \leq i \leq k$, avec $|U_i| = |U'_i| = r$ et $|V_i| = |V'_i| = \rho - r$.

Soit $\mathbf{m} = f(\mathbf{a})$, $\mathbf{u} = p_r(\mathbf{m})$ et $\mathbf{v} = s_{\rho-r}(\mathbf{m})$. Par choix de H , on a $k \geq H_2$. Les mots $a_k...a_1$ et $a'_k...a'_1$ contiennent toutes les lettres de $\mathbf{a}$, d'où $s_r(\mathbf{m}) = \mathbf{u}$ grâce aux relations (3). D'autre part le mot $a_{k+1}...a_2$ contient aussi toutes les lettres de l'alphabet et (3) donne donc

$p_{\rho-r}(\mathcal{m}) = \mathcal{v}$. D'après le Lemme 4 on a

$$\mathcal{m} \subset \mathcal{p}^* \quad \text{pour} \quad \mathcal{p} = p_d(\mathcal{m}) = s_d(\mathcal{m})$$

avec $d = \text{pgcd}(\rho,r)$.

Supposons $\text{card}(\mathcal{p}) = \text{card}(\mathcal{a})$. Alors les images $f(a)$ sont déterminées par leurs préfixes (ou suffixes) de longueur d et il existe des bijections $\beta : \mathcal{a} \to \mathcal{u}$, $\gamma : \mathcal{a} \to \mathcal{v}$ et $\beta' : \mathcal{a} \to \mathcal{u}$, $\gamma' : \mathcal{a} \to \mathcal{v}$ telles que

$$\forall\, a \in \mathcal{a},\ f(a) = \beta(a)\gamma(a) = \gamma'(a)\beta'(a).$$

Si le mot de deux lettres $a'a$ est facteur de A, alors les relations (3) donnent $\beta'(a') = \beta(a)$ d'où $a = \beta^{-1}\beta'(a')$. Posons $\delta = \beta^{-1}\beta'$, alors δ est une bijection dans $\mathcal{a}$ et

$$bA = b\delta(b)\ldots\delta^k(b).$$

Toutes les lettres étant dans A, la bijection δ est d'ordre q. D'autre part, tous les facteurs de longueur deux de u étant dans A, ils sont nécessairement de la forme $a\delta(a)$, d'où $u = 1\delta(1)\delta^2(1)\ldots$ c'est-à-dire $u_k = \delta^k(u_0)$. La suite est donc périodique (de période q), contradiction.

Supposons $\text{card}(\mathcal{p}) < \text{card}(\mathcal{a})$. Désignons par ψ le plongement canonique de $\mathcal{p}^*$ dans $\mathcal{a}^*$ que l'on prolonge à $\mathcal{M}(\mathcal{p})$. Comme $f(\mathcal{a}) = \mathcal{m} \subset \mathcal{p}^*$, on peut regarder les mots de $\mathcal{M}(\mathcal{m})$ comme des mots de $\mathcal{M}(\mathcal{p})$; notons $\varphi : \mathcal{M}(\mathcal{m}) \to \mathcal{M}(\mathcal{p})$ cette identification et finalement notons par $\mathbf{f}$ l'aplication de $\mathcal{M}(\mathcal{a})$ dans $\mathcal{M}(\mathcal{m})$ induite par la substitution f. Soit $F := \varphi\circ\mathbf{f}\circ\psi$ et $U = \varphi(\mathbf{f}(u))$. Par construction U est la suite u après regroupement par mots successifs de d lettres, c'est-à-dire

$$U_n = u_{nd}\ldots u_{(n+1)d-1},$$

et par construction F est une substitution uniforme de module ρ sur l'alphabet $\mathcal{p}$ et le mot infini U de $\mathcal{p}^\infty$ est point fixe de F. Le caractère minimal de U résulte du Théorème A.

Fin de la démonstration.

Montrons que le mot m vérifiant (6) peut se voir comme un facteur de U. En effet les mots B et B' sont des suffixes de mots de $\mathcal{m}$, de longueurs multiples de d et C, C' sont des préfixes de mots de $\mathcal{m}$, également de longueurs multiples de d.

Maintenant notons que les factorisations de m s'obtiennent à partir de la factorisation de u en blocs successifs de ρ lettres données par la relation $u = f(u)$. En considérant U et F, la ou les factorisations de m se déduisent de la factorisation de U en blocs de ρd lettres :

$$U = [(u_0 \cdots u_{d-1}) \cdots (u_{d(\rho-1)-1} \cdots u_{d\rho-1})][(u_{d\rho} \cdots u_{d(\rho+1)-1}) \cdots (u_{d(2\rho-1)} \cdots u_{2d\rho-1})] \cdots$$

chaque bloc pouvant se voir comme produit de ρ facteurs de mots dans $\mathbf{p}$ (de d lettres) ou comme produit de d facteurs de mots dans $\mathbf{m}$ (de ρ lettres). Il en résulte que la factorisation $Bf(A)C$ entraîne une factorisation de m sous la forme $BB_1F(A_1)C_1C$ avec $B_1 = f(\beta)$ (resp. $C_1 = f(\gamma)$) où β (resp. γ) est préfixe (resp. suffixe) de A. De même la factorisation $B'f(A')C'$ donne une factorisation de m de la forme $B'f(\beta')F(A'_1)f(\gamma')C'$. En particulier, si m est rythmé sur l'alphabet $\mathbf{p}$, alors $|Bf(\beta)| = |B'f(\beta')|$ d'où $|B| - |B'|$ est multiple de ρ ce qui est exclu. le mot m n'est donc pas rythmé sur l'alphabet $\mathbf{p}$ et sa longueur en tant que mot sur $\mathbf{p}$ est au moins $2|m|/\rho \geq 2H/\rho$.

Choisissons maintenant $L_0(\rho,q) = \rho^{2(q+1)}$ pour $q \geq 3$ (la valeur pour $q = 2$ est plus grande que celle déja obtenue). Comme $H(\rho,q) \leq 2\rho^{2q+1}$ si $|m| \geq L_0(\rho,q)$, on a $|m| \geq H$ et $2|m|/\rho \geq 2\rho^{2q+1} \geq L_0(\rho,q')$ pour tout $q' = 2,\ldots, q-1$. Si le théorème est démontré dans le cas d'alphabets d'au plus q-1 lettres alors tout mot m de longueur supérieure à $L_0(\rho,q)$ est rythmé dans u sur l'alphabet $\mathbf{a}$. Ceci termine la démonstration et donne une valeur pour la constante $L_0(\rho,q)$.

◆

IV.- CALCUL AUTOMATIQUE

IV.1.-AUTOMATES.

Définition.- Un g- automate est un quadruplet $(\mathbf{S}, \mathfrak{g}, \mathbf{I}, \mathbf{i})$ où

$$\mathbf{S} := \{s_1, \ldots, s_d\}$$

est un ensemble fini appelé espace des *états* , $\mathbf{i}$ est un état donné appelé état *initial* ,

$$\mathfrak{g} := \{0, 1, \ldots, g-1\},$$

avec $g \geq 2$, est un ensemble ordonné fini appelé *alphabet* ou ensemble des *entrées* , enfin $\mathbf{I}$ est un ensemble d'applications $I_0, \ldots, I_{g-1}$ de $\mathbf{S}$ dans $\mathbf{S}$ (appelées *instructions*).

A tout entier n correspond un mot $e_t \ldots e_0$ dans $\mathfrak{g}^*$ donné par le développement de n en base g :

$$n = e_t g^t + \ldots + e_1 g + e_0 .$$

On associe alors à n une *sortie* définie par :

$$I(n) := I_{e_t} \ldots I_{e_0}(s) .$$

Définition .- Soit E un ensemble fini . Une suite $q : \mathbb{N} \to E$ est dite g-reconnaissable s'il existe un g-automate $(\mathbf{S}, \mathfrak{g}, \mathbf{I}, s)$ et un application $\tau : \mathbf{S} \to E$ tels que

$$q(n) = \tau \circ I (n) .$$

Nous allons démontrer le théorème suivant :

THEOREME 3 .- *Soit* f *une substitution uniforme de module* ρ , *sur l'alphabet* $\mathfrak{a}$, *injective et admettant un point fixe* u *minimal, non périodique . Soit* $p(n)$ *le nombre de facteurs de* u *de longueur* n *et soit* $q(n) = p(n + 1) - p(n)$. *Alors la suite* $n \to q(n)$:

(i) *ne prend qu'un nombre fini de valeurs;*

(ii) *est* ρ*-reconnaissable par un automate.*

IV.2.- ETUDE DES $F_n(u)$.

L'évaluation de $q(.)$ se fera par récurrence sur la longueur n des facteurs de u. Avec les notations de l'introduction et en notant simplement le cardinal d'un ensemble fini par $| E |$ (au risque de confondre avec la longueur d'un mot) on a :

$$q(n) = | F_{n+1} | - | F_n | = \left(\sum_{k \geq 1} | F_n^{(k)} | \right) - | F_n | = \sum_{k \geq 1} (k - 1) | F_n^{(k)} | .$$

Soit L_0 la constante du Théorème 1 et soit

$$E := \cup \{ F_n^{(k)} ; \ 1 \leq n \leq L_0 \}$$

Nous allons étudier les ensembles $F_n^{(k)}$, pour $n > L_0$, par une méthode récursive.

Décomposition des entiers.

Notations .- Soit X une partie de $\mathcal{A}$. On pose

$$\lambda_X := \text{Max}\{ s ; \ \exists \ m \in \mathcal{A}^s , \ \forall x \in X , \ m = p_s(f(x)) \}$$

et

$$p(X,n) := \text{Min} \ \{ k \in \mathbb{N} ; \ k \geq (n - \lambda_X)/\rho \ \} = p.$$

LEMME 5 .- *Les éléments de* $F_n^{(k)}$, *pour* $n > L_0$ *sont factorisés en suffixes de longueur* $(n - \lambda_X)$ *des images des éléments de* $F_p^{(k+\ell)}$ *et préfixes qui sont les plus grands facteurs préfixes commun aux images des éléments d'un sous-ensemble* X *de* $\mathcal{A}$ *de cardinal* $(k + \ell)$, $\ell \geq 0$, *tel*

que $k = \text{card}\{ p_{\lambda_X + 1}(f(a)) ;\ a \in X \}$.

Démonstration. Soit B facteur de u de longueur $> L_0$. D'après le Théorème 2, B admet une factorisation unique $B = B_1 f(A) B_2$ avec $|B_i| < \rho$, B_1 et B_2 respectivement dans $\mathcal{S}$ et $\mathcal{P}$, où $\mathcal{S}$ (resp. $\mathcal{P}$) est l'ensemble des suffixes (resp. préfixe) de longueur au plus ρ , des mots dans $f(\mathcal{A})$.

Le mot B dans $F_n^{(k)}$ signifie qu'il existe k lettres distincte $a_1,\ldots, a_k$ telles que

$$\forall\ i \in \{1,\ldots,k\} ;\ B_1 f(A) B_2 a_i \in F.$$

Chacun des $B_2 a_i$ est donc dans $\mathcal{P}$, ce qui implique non seulement que B_2 est préfixe commun aux images d'au moins k lettres de $\mathcal{A}$, mais qu'il en est le plus grand. Désignons alors par X le sous-ensemble de $\mathcal{A}$ formé par ces lettres; on a $\text{card}(X) \geq k$ et nous noterons $\text{card}(X) = k + \ell$, $\ell \geq 0$.

Le mot B dans $F_n^{(k)}$ signifie aussi que

$$\text{card}\{ p_{|B_2|+1}(f(a)) ;\ a \in X \} = k.$$

Comme par ailleurs

$$|B| = n = |B_1| + \rho|A| + \lambda_X ,$$

en posant $|A| = p$ et $|B_1| = r$, on obtient :

$$n - \lambda_X = \rho p + r \quad (0 \leq r < \rho).$$

Deux cas sont à distinguer :

Premier cas: $r > 0$.

Soit α le nombre de lettres a telles que B_1 soit suffixe de $f(a)$. Alors :

$$B \in F_n^{(k)} \iff \exists\ \{a_1,\ldots, a_\alpha\} \subset \mathcal{A} ;\ a_i A \in F_p^{(k+\ell)} .$$

Deuxième cas: $r = 0$.

On a maintenant $B = f(A)B_2$ et

$$B \in F_n^{(k)} \iff \mathcal{A} \in F_{p-1}^{(k+\ell)} .$$

Nous écrirons alors $n - \lambda_X = \rho p = \rho(p - 1) + \rho$, pour "uniformiser".

La décomposition de n sous l'unique forme $n - \lambda_X = \rho p + r$, avec cette fois-ci $0 < r \leq \rho$, démontre le Lemme 5.

◆

Soient C_1 , C_2 dans $F_{p+1}^{(k+\ell)}$; par construction, ils donnerons deux mots différents de $F_n^{(k)}$ si $s_p(C_1) \neq s_p(C_2)$ ou dans le cas contraire, si le plus grand suffixe commun à $f(p_1(C_1))$ et $f(p_1(C_2))$ est de longueur strictement inférieur à r . Il vient donc :

$$| F_n^{(k)}| = \sum_{\ell \geq 0} \sum_i \text{ card } \{ s_r(f(a)) ; a = p_1(C) , C \in F_{p+1}^{(k+\ell)} \text{ et } s_p(C) = C_i \} .$$

L'entier $q(n)$ est donc parfaitement déterminé dès que $q(p+1)$ et r sont connus. Si $(p+1)$ est superieur à L_0 , le Lemme 5 peut lui être appliqué. Ainsi de suite, nous déduisons la première partie du Théorème 3, puisque pour tout $n > L_0$ $q(n)$ sera dans l'ensemble Q (parfaitement défini au préalable) des valeurs prises par $q(n)$ pour les n inférieurs à L_0 .

En désignant par σ_r l'application de Q dans Q qui à $q(p+1)$ fait correspondre $q(n)$ lorsque $n - \lambda_X = \rho p + r$ $(0 < r \leq \rho)$ on a alors le résultat suivant

LEMME 6. *Pour tout* $n > L_0$, *il existe un* ρ *automate (dont un programme formel indique ci-dessous) qui donne la valeur de* $q(n)$ *à partir de la décomposition de* n *sous la forme* $n - \lambda_X = \rho p + r$ $(0 < r \leq \rho)$.

$n := n_{\text{initials}}$

$\sigma :=$ Application identité dans Q .

Tant que $n > L_0$, faire

$p \leftarrow \begin{cases} (n - \lambda_X)\rho^{-1} \text{ si entier} , \\ [(n - \lambda_X)\rho^{-1}] + 1 \text{ sinon} \end{cases}$

$r \leftarrow n - \rho p + \rho + 1$

$\sigma \leftarrow \sigma o \sigma_r$

$n \leftarrow p$

$$q(n_{initial}) = \sigma(q(n)) .$$

IV.3.- FIN DE LA DEMONSTRATION DU THEOREME 3.

Une légère transformation du programme formel du Lemme 6 nous donne le Lemme 7 suivant qui termine la démonstration de la seconde partie du Théorème 3.

LEMME.-7 : *Pour tout* $n > \mathbb{N}_0$, *il existe un* ρ-*automate (dont on a un programme formel ci-dessous) qui donne la valeur de* $q(n)$ *à partir de l'écriture de* n *en base* ρ .

```
n : = n initials
σ : = Application identité dans Q .
retenue : = 0
Tant que (n + retenue ) > N0 , faire
    m       ←   [nρ^-1]
    s       ←   n - [nρ^-1]
    Si (retenue + s - λ) > 0 :
                r           ←   retenue + s - λ
                retenue     ←   1
    Sinon
                r           ←   retenue + s - λ + ρ
                retenue     ←   0

    σ       ←   σoσ_r
    n       ←   m
```

$$q(n_{initial}) := \sigma(q(n)) .$$

IV.- COMPLEXITE POLYNOMIALE .

Le Théorème 3 donne :

PROPOSITION.- *Pour toute suite minimale non périodique, point fixe d'une substitution uniforme injective,* $p(n)$ *est de forme polynomiale si et seulement si* $q(n)$ *est une constante. De plus, le polynôme est de degré au plus égal à un et son terme de degré un est la constante* $q(n)$.

Démonstation. Supposons $p(n) = \alpha_k n^k + \ldots + \alpha_1 + \alpha_0$, alors

$$q(n) = \alpha_1 + \sum_{i=0}^{k} \alpha_i ((n + 1)^i - n^i) .$$

Comme $q(n)$ prend un nombre fini de valeurs, on a $\alpha_i = 0$, pour $2 \leq i \leq k$. Ainsi $q(n) = \alpha_1$ et $p(n) = \alpha_0 + \alpha_1 n$.

♦

V.- EXEMPLES

V.1.- SUITE DE MORSE

Soit μ la suite de Morse donnée dans l'introduction. On a vu que les mots de longueur ≥ 4 sont rythmés et un dénombrement direct donne $q(1) = q(2) = 2$ et $q(3) = 4$.

Notons que $f_1(1)$ et $f_1(2)$ n'ont ni suffixe commun, ni préfixe commun ; ainsi pour $n > 4$, $q(n) = |F_n^{(2)}| = |F_{p+1}^{(2)}|$ où $2(p+1)+r = n$ et $r = 1$ ou 2 . Donc $\sigma_1 = \sigma_2 =$ identité dans Q.

Pour tout $n \geq 4$, il existe un seul entier k tel que

$$2^{k+1} < n \leq 2^{k+2}$$

Nous distinguons deux cas :

Premier cas : $2^{k+1} + 2^k < n \leq 2^{k+2}$.

Nous avons alors $3 < n2^{-k} \leq 4$, donc $4 = n2^{-k}$ si $n2^{-k}$ est entier et $4 = \lceil n2^{-k} \rceil + 1$ sinon et par suite $q(n) = q(4) = 2$.

Deuxieme cas : $2^{k+1} < n \leq 2^{k+1} + 2^k$

Nous avons maintenant $2 < n2^{-k} \leq 3$ et un raisonnement analogue au cas précédent donne $q(n) = q(3) = 4$.

Dans tous les cas, en remarquant que $3.2^{-2} < 1 \leq 4.2^{-2}$; $3.2^{-1} < 2 \leq 4.2^{-1}$ et $3.2^0 < 4 \leq 4.2^0$, on obtient en définitive $q(n) = 2$ pour tou entier n pour lequel il existe un entier relatif k tel que : $3.2^k < n \leq 4.2^k$ et $q(n) = 4$ sinon .

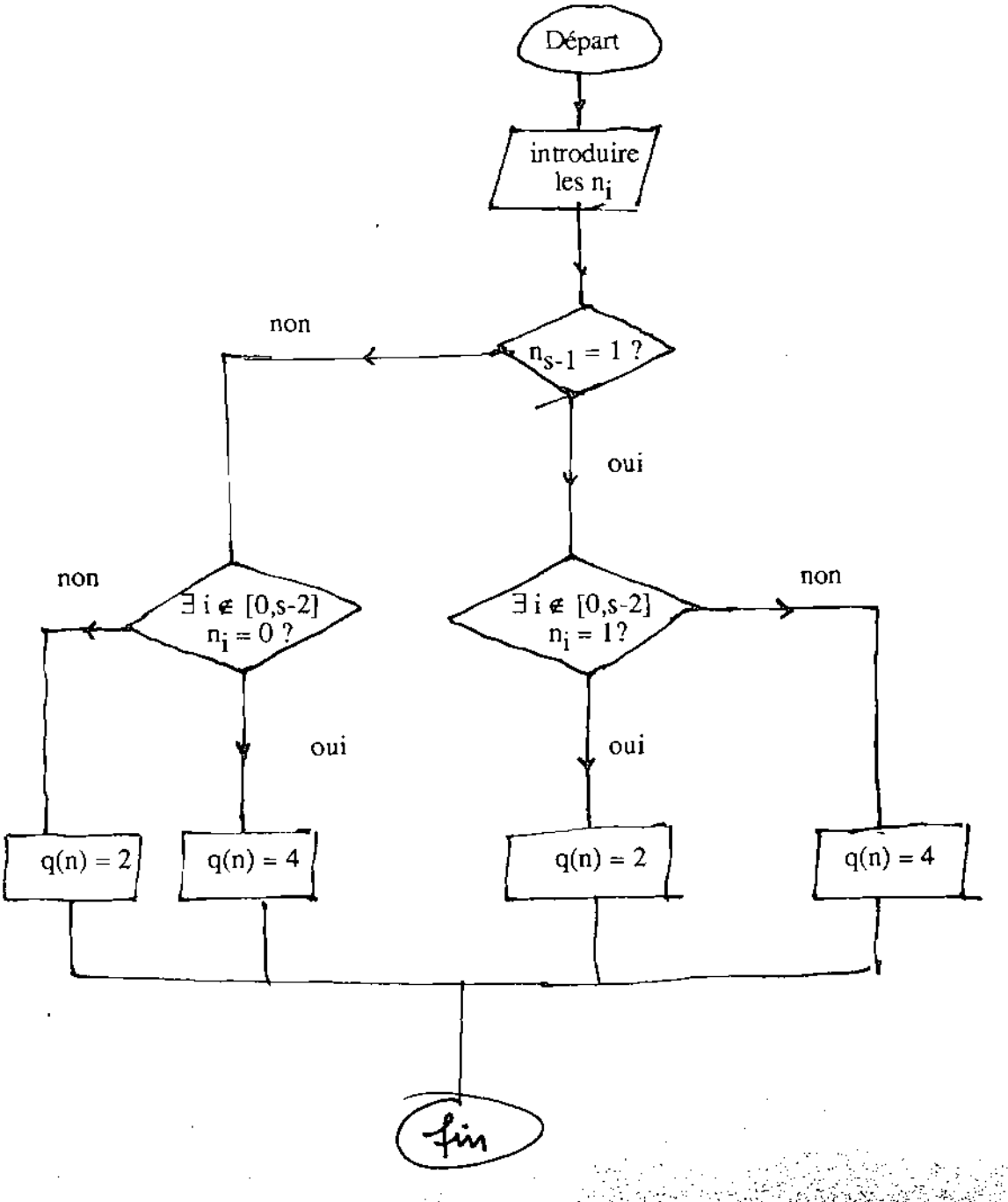
Montrons à présent comment est obtenu $q(n)$ à partir de l'écriture de n en base 2 .Soit $n = n_s 2^s + \ldots + n_1.2 + n_0$. avec $n_s \neq 0$; on a donc $n_s = 1$. Supposons $q(n) = 2$. Alors il existe $k \in \mathbb{Z}$, tel que $3.2^k < n \leq 4.2^k$, c'est - à - dire :

$$2^{k+1} + 2^k < 2^s + n_{s-1}\,2^{s-1} + n_{s-2}\,2^{s-2} + \ldots + n_0 \leq 2^{k+2}.$$

Si $n_{s-1} = 1$, alors $k = s-1$ et $n_{s-1}\,2^{s-1} + n_{s-2}\,2^{s-2} + \ldots + n_0 \neq 0$, donc il existe i, $0 \leq i \leq s-2$, tel que $n_i = 1$.

Si $n_{s-1} = 0$; nous avons maintenant $k = s-2$ et pour tout i , $2 \leq i \leq s-2$, on a $n_i = 0$.

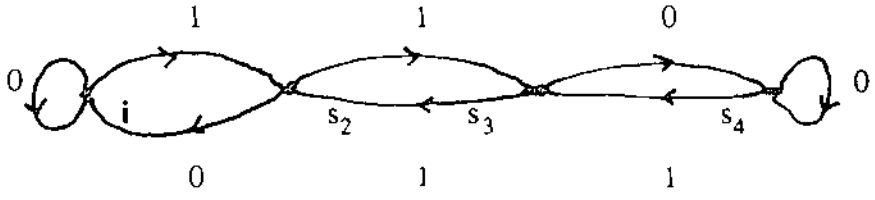
Dans tous les cas, $q(n) = 2$ si $n_{s-1} = 1$ et s'il existe $i < s-1$ tel que $n_i = 1$, ou si pour tout $i < s$, $n_i = 0$; autrement $q(n) = 4$. Le diagramme résume ce calcul :



Fin

V.2.- SUITE DE RUDIN-SHAPIRO

La suite de Rudin-Shapiro compte les "11" molulo 2 dans l'écriture de n en base 2 . Elle est reconnaissable par l'automate suivant :



Soit u le point fixe de la substitution τ associée à l'automate ci-dessus (cf. I. 3) et de préfixe 1 . Ici les ensembles $\mathcal{S}$ et $\mathcal{P}$ sont disjoints, par conséquent, les facteurs de deux lettres sont rythmés. Alors pour tout n

$$F_n^{(4)} = F_n^{(3)} = \emptyset$$

donc

$$q(n) = |F_n^{(2)}| = \sum_i \text{card } \{ s_r(\tau(a)) ; a = p_1(C), C \notin F_{p+1}^{(2)} \text{ et } s_p(C) = C_i \}.$$

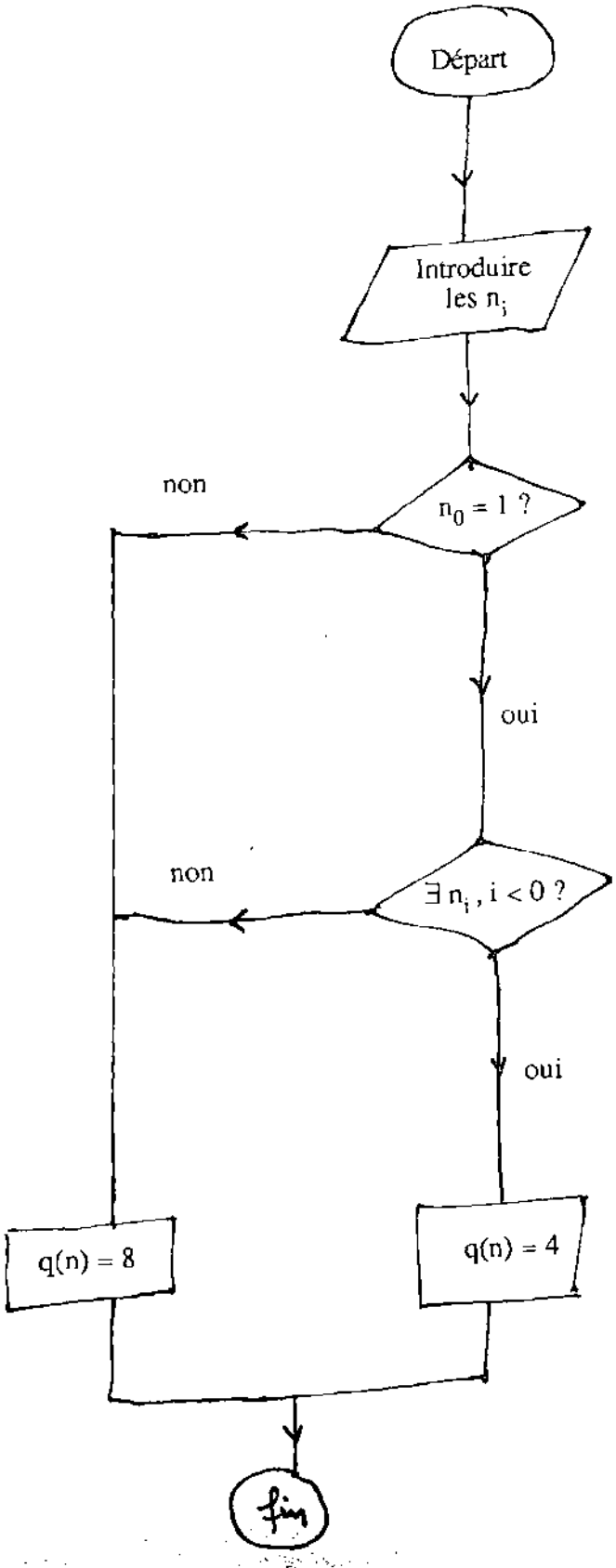
Soit $\{C, C'\} \subset F_{p+1}^{(2)}$ tel que $s_p(C) = s_p(C') = C_1$. On a toujours

$$s_2(\tau(p_1(C))) \neq s_2(\tau(p_1(C'),$$

donc pour $r = 2$, $|F_n^{(2)}| = |F_{p+1}^{(2)}|$ et σ_2 est l'identité.

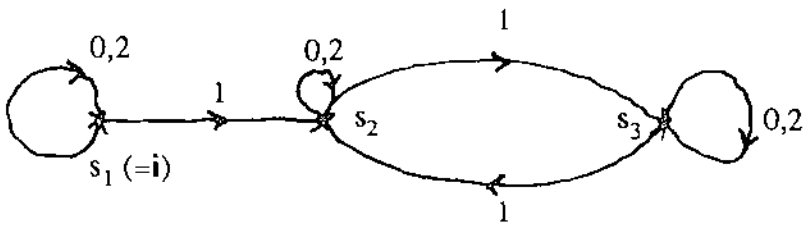
D'autre part on a $s_1(\tau(p_1(C))) \neq s_1(\tau(p_1(C')))$. En effet dans le cas contraire, le couple $(p_1(C), (p_1(C'))$ vaut (1,3) ou (2,4) ; $C = 1C_1$ ou $C = 2C_1$ et $C' = 3C_1$ ou $C' = 4C_1$. Alors 1 et 3 (resp. 2 et 4) se prolongent par une même lèttre, ce qui n'est pas le cas. Donc σ_1 est l'identité.

Ainsi, $q(1) = 4$ et pour $n \geq 2$, $q(n) = 8$. Le calcul automatique est donné par le diagramme suivant :


Départ
Introduire les n_i
$n_0 = 1$?
non
oui
$\exists n_i , i < 0$?
non
oui
q(n) = 8
q(n) = 4
fin

VI.- UN EXEMPLE NON MINIMAL

Soit u le point fixe dans $1\mathcal{A}^*$ de la substitution f_3 définie sur l'alphabet $\mathcal{A} = \{1,2,3\}$ par $f_3(1) = 121$, $f_3(2) = 232$, $f_3(3) = 323$. La suite u est reconnaissable par l'automate $(S,\mathcal{A},I,i)$ donné par le graphe orienté suivant :



avec l'application $\tau : S \to \mathcal{A}$ donnée par $\tau(s_k) = k$. Remarquons que l'on a la propriété arithmétique suivante : l'entier n ne contient pas de chiffre 1 dans son écriture en base trois si et seulement si $u_n = 1$.

Quel que soit l'entier k , 1 ne figure pas dans $f_3^k(2)$. Donc u n'est pas minimale d'après le Théorème B. Montrons ce pendant que q(n) prend un nombre fini de valeurs et qu'il existe un 3-automate tel que q(n) en est la sortie lorsqu'on entre l'écriture de n en base trois.

Il suffit évidemment de montrer que le Lemme 5 (pour $L_0 = 1$ par exemple) est vérifié. Déterminons donc $F_n^{(k)}$ pour k = 2 ou 3 . $F_n^{(3)} = \emptyset$ car u ne contient aucun terme $a^{(2)}$, a = 1, 2 ou 3 , et pour toute partie X de {1,2,3} , on a

$$\text{card } X > 1 \Rightarrow \lambda = 0 .$$

On est donc ramené à montrer que les éléments de $F_n^{(2)}$ sont les suffixes de longueur n des images des éléments de $F_{p+1}^{(2)}$ où p = n/3 si n est un multiple de 3 et $p = [n/3] + 1$ sinon.

Soit $a_{q+1}a_{q+2}\ldots a_{q+n}$ dans $F_n^{(2)}$; alors, il existe $\{x_1,x_2\}$ dans $\{1,2,3\}$ tel que $a_{q+1}a_{q+2}\ldots a_{q+n}x_i$ soit dans F_{n+1} , pour tout $i \in [1,2]$. Ainsi, il existe A dans F et j dans $\{0,1,2\}$ tels que $f_3(A) = a_{q-j}\, a_{q-j+1}\ldots a_{q+1}\ldots a_{q+n}$; autrement, on aurait $a_{q+n-1}\, a_{q+n}$ ou a_{q+n} préfixe commun aux images de deux lettres différentes, contradiction. Les lettres x_1 et x_2 sont donc respectivement préfixe des images de deux lettres différentes t et z . At et Az sont alors dans $F_{|A|+1}$, donc $a \in F_{|A|}^{(2)}$.

Par ailleurs, $n < |f_3(A)| \leq n+3$, donc $|A| = (n/3) + 1$ si n est un multiple de 3 et $|A| = [n/3] + 1$ sinon.

Réciproquement, soit $a_1a_2\ldots a_pa_{p+1}$ dans $F_{p+1}^{(2)}$; il existe alors $\{x_1,x_2\}$ dans $\{1,2,3\}$ tel que $a_1a_2\ldots a_{p+1}x_i$ soit dans F_{p+2p}, pour i = 1, 2. Ainsi, $f_3(a_1a_2\ldots a_{p+1}x_1)$ et $f_3(a_1a_2\ldots a_{p+1}px_2)$ sont dans $F_{3(p+1)+3}$. Comme $p_1(f_3(x_1)) \neq p_1(f_3(x_2))$, $f(a_1a_2\ldots a_{p+1}) \in F_{3(p+1)}^{(2)}$. Il en est de même pour tout suffixe de ce mot et a fortiori pour les suffixes dont la longueur n est comprise entre 3p et 3(p+1) , borne inférieure incluse.

$$3p \leq n < 3(p+1) \Rightarrow \begin{cases} p = n/3 & \text{si n est multiple de 3} \\ p = [n/3] + 1 & \text{si n n'est pas multiple de 3} \end{cases}$$

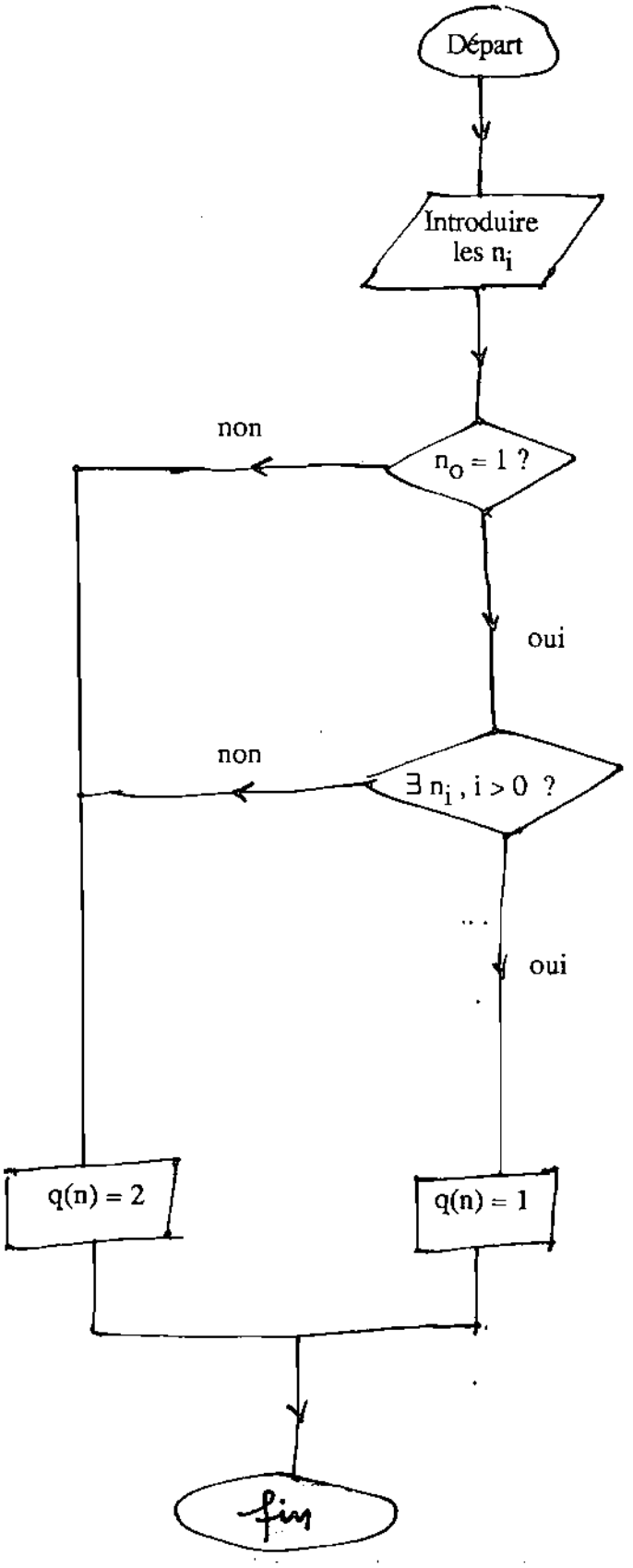
Le suffixe de longueur n prolongeable de deux manières est donc suffixe de l'image du mot $a_1a_2\ldots a_{p+1}$ de $F_{p+1}^{(2)}$. On a

$$q(n) = |F_n^{(2)}| = \sum_i \text{card}\,\{s_r(f_3(a)) \ ; \ a = p_1(C)\,,\, C \in F_{p+1}^{(2)} \text{ et } s_p(C) = C_i\}.$$

Soit $\{C,C'\} \subset F_{p+1}^{(2)}$ tel que $s_p(C) = s_p(C') = C_1$; pour r = 1, 2 ou 3 on a $s_r(f_3(p_1(C))) \neq s_r(f_3(p_1(C')))$ donc $|F_n^{(2)}| = |F_{p+1}^{(2)}|$ et σ_r est l'identité. Ainsi

$$q(1) = 1 \text{ et pour tout } n \geq 2 \ , \ q(n) = 2 \,.$$

D'où le diagramme suivant :


Départ
Introduire les n_i
non
$n_o = 1$?
oui
non
$\exists\, n_i , i > 0$?
oui
q(n) = 2
q(n) = 1
fin

Les conclusions du Théorème 3 restent donc vraies bien que u_β ne soit pas minimale; la minimalité n'est donc pas une condition nécessaire.

Notons que les suites périodiques vérifient également les conclusions du Théorème 3 , puisque pour de telles suites p(n) est constant à partir d'un certain rang.

Problème :

Comment caractériser les suites pour lesquelles les conclusions du Théorème 3 sont vraies ?

BIBLIOGRAPHIE

1.- ARSON S. : Démonstration de l'existence de suites asymétriques infinies.

MAT. Sb. 44 (1937) : 769-777.

2.- BLEUZEN-GUERNALEC N. : Suites points fixes de transductions uniformes.

CRAS, Paris, T. 300, Série I, n° 3 (1985) : 85-88.

3.- COBHAM A. : Uniform Tag Sequences.

Mathematical Systems Theory 6 (1972) : 164-192.

4.- GOTTSCHALK W.H. and HEDLUND G.A. : Topological dynamics.

Am. Math. Soc. Colloq. Publ. Vol 36. Providence R.I. (1968).

5.- GOTTSCHALK W.H. and HEDLUND G.A. : A Characterization of the Morse Minimal Set.

Proc. Am. Math. Society 15 (1964) : 70-74 .

6.- HEDLUND G.A. and MORSE M. : Symbolic Dynamics.

American J. Math.,60,(1938) : 815-866.

7.- KEANE M. : Generalized Morse sequences.

Z. Wahrscheinlichkeitstheorie Verw. Gebiete,10(1968) : 335-353.

8.- LOTHAIRE : Combinatorics on words.

Addison Wesley Reading MA (1982) : chapter 12.

9.- MARTIN J.C. : Substitution minimal flows.

Amer. J. Math 93 (1971) : 503-526.

10.- MARTIN J.C. : Minimal flows arising from substitutions of non-constant length.

Math. Systems Th. 7 n° 1 (1973) : 73-82.

11.- MORSE M. : Recurrent geodesic on a surface of negative curvature.

Trans. Amer. Math Soc. 22 (1921) : 84-100.

12.- PANSIOT J.J. : A propos d'une conjecture de F. Dejeun sur les répétitions dans les mots.

(A paraître dans Discrete Applied Mathematics).

13.- QUEFFELEC M.: Contribution à l'étude Spectrale de suites arithmétiques

Paris-Nord Thèse d'Etat, 1984.

14.- RAUZY G. : Suites à termes dans un alphabet fini.

Séminaire de théorie des nombres. Bordeaux exposé n° 25 (1982-1983) : 1-16.

15.- RAUZY G. : Des mots en Arithmétique.

Journée Algorithmes et Complexités. Avignon 1973.

16.- THUE A. : Über unendliche Zeichenreihen.

Norske. vid. Selsk I. Mat. Nat Kl (Christiana) N° 7 (1906) : 1-22.

Résumé:

On montre que dans une suite u minimale obtenue par substitution injective uniforme, tout mot f_u dans u et de longueur assez longue admet une factorisation unique.

Ce résultat permet de reconnaître par un automate $n \to p(n+1)-p(n)$ où $p(n)$ désigne le nombre de mots lus dans u, de longueur n.