SUR LA DENSITE ASYMPTOTIQUE DE LANGAGES FORMELS.

JEAN BERSTEL
Centre de Calcul , UER de Mathématique
Université Louis Pasteur, Strasbourg, France.

Résumé : On introduit une notion de densité asymptotique d'un langage formel et on en étudie les propriétés. Une application aux ensembles acceptables de nombres est donnée.

INTRODUCTION.

La présente étude a été motivée par les propriétés concernant les ensembles acceptables de nombres qui ont été obtenus grâce à l'examen du comportement asymptotique des ensembles de nombres en question (Cobham (1972), Minsky, Papert (1966)). En particulier, c'est l'extension de ces résultats aux ensembles de nombres correspondant à des langages algébriques, comme déjà suggérée dans Minsky, Papert (1966) , qui a paru intéressante. Il s'est alors avéré que l'on pouvait dépasser le cadre initial des ensembles acceptables de nombres et définir une notion générale de densité asymptotique d'un langage formel, Le résultat le plus frappant obtenu est que la densité asymptotique, si elle existe, d'un langage algébrique non ambigu (resp. rationnel) est un nombre algébrique, resp. rationnel. Transposé aux ensembles acceptables de nombres, il donne, pour des ensembles de nombres correspondant à des langages algébriques non ambigus, un critère analogue à celui de Cobham (1972) pour les ensembles de nombres rationnellement acceptables. La première partie introduit la définition de la densité asymptotique et donne quelques propriétés générales. Au deuxième paragraphe, nous démontrons le théorème principal mentionné plus haut (Th. 2.1) et en tirons quelques corollaires, Le paragraphe suivant est consacré à l'étude du comportement de la suite de densité d'un langage rationnel ne possédant pas de densité asymptotique. Cette étude est rendue possible grâce à une propriété particulière des pôles des séries $\mathfrak{R}$-rationnelles en une variable, où $\mathfrak{R}$ est un sous-semianneau de $\mathbb{R}_+$ (th. 3.2). Nous examinons au quatrième paragraphe la notion de densité relative de deux langages qui semble pouvoir être utile dans l'étude de l'approximation d'un langage par un autre, avant d'exposer l'application aux ensembles acceptables de nombres dans la dernière partie.

1. DEFINITIONS.

Tous les monoïdes libres considérés dans la suite sont supposés engendrés par un alphabet fini non vide. Soit X^* un monoïde libre, et $H(X^*)$ l'ensembledes morphismes de X^* dans le monoïde multiplicatif des nombres réels strictement positif. Etant donné un langage $L \subset X^*$, et un morphisme $\eta \in H(X^*)$, on pose, pour tout entier $r \geq 0$,

$$\eta_r(L) = \Sigma\{\eta f : f \in L \backslash X^* X^{r+1}\} .$$

<u>Définition</u> : La densité asymptotique $d^{\eta}(L)$ du langage $L \subset X^*$ relativement au morphisme η est la limite, si elle existe, définie par

$$d^{\eta}(L) = \lim_{r\to\infty} \frac{\eta_r(L)}{\eta_r(X^*)} .$$

La terminologie est justifiée par la remarque suivante : Soit L un langage sur un alphabet à une seule lettre x, et soit $L^o = \{n \in \mathbb{N} : x^n \in L\}$. Si $\eta \in H(x^*)$ est défini par $\eta(x) = 1$, alors $d^{\eta}(L)$ n'est autre que la densité asymptotique, au sens habituel du terme, de la partie $L^o \subset \mathbb{N}$.

On appelle morphisme <u>naturel</u> l'élément $\eta \in H(X^*)$ vérifiant $\eta x = 1$ pour tout $x \in X$. La densité correspondante, appelée densité naturelle, et notée $d(L)$, est définie par

$$d(L) = \lim_{r\to\infty} \frac{\mathrm{Card}\,(L\backslash X^r X^*)}{\mathrm{Card}\,(X^*\backslash X^r X^*)} .$$

Enfin, le nombre $|\eta| = \sum_{x\in X} \eta x$ sera appelé la <u>norme</u> du morphisme η. La norme du morphisme naturel est $\mathrm{Card}\, X$. Si $|\eta| = 1$, alors η définit une distribution de probabilité sur X. Ces morphismes ont été étudiés dans Schützenberger (1965).

<u>Exemples</u> : 1. Soit $f \in X^*$, et $L = fX^*$. On a, pour $\eta \in H(X^*)$, $d^{\eta}(L) = \eta f$ si $|\eta| \le 1$, $d^{\eta}(L) = \eta f\, |\eta|^{-|f|}$ sinon.

2. Soit $L = (X^p)^*$, où $p \ge 2$, et $\eta \in H(X^*)$ avec $q = |\eta| > 1$. Si $r = mp + s$ ($m,s \in \mathbb{N}$ et $0 \le s < p$), alors $\eta_r(L) = \frac{q^{(m+1)p}-1}{q^p-1}$. Il s'ensuit que pour $s = 0,\dots,p-1$,

$$\lim_{\substack{r\to\infty \\ r\equiv s \pmod p}} \frac{\eta_r(L)}{\eta_r(X^*)} = q^{p-s-1}\,\frac{q-1}{q^p-1} ,$$

donc que L ne possède pas de densité asymptotique relativement au morphisme η.

Dans la suite on utilisera plusieurs fois sans mention explicite la fait suivant découlant d'une propriété classique sur la sommation des séries :

Si la suite $\dfrac{\Sigma\{\eta f : f \in L \cap X^r\}}{\Sigma\{\eta f : f \in L' \cap X^r\}}$ a une limite s, et si la série de terme général $\Sigma\{\eta f : f \in L' \cap X^r\}$ diverge, alors $s = \lim_{r\to\infty} \dfrac{\eta_r(L)}{\eta_r(L')}$.

L'exemple précédent montre qu'un langage, même rationnel, ne possède pas toujours une densité. On a par contre le résultat suivant :

<u>Proposition 1.1.</u> : <u>Soient</u> $L \subset X^*$ <u>et</u> $\eta \in H(X^*)$; <u>si</u> $|\eta| < 1$, <u>alors la densité</u> $d^{\eta}(L)$ <u>existe</u>.

<u>Preuve</u> : La suite $(\eta_r(L))_{r\geq 0}$ est non décroissante et bornée supérieurement par $\frac{1}{1-|\eta|}$; elle possède donc une limite. Il en est de même de la suite $\left(\frac{\eta_r(L)}{\eta_r(X^*)}\right)_{r\geq 0}$

Avant d'examiner la nature de la densité, nous établissons un lien entre les densités d'un même langage relativement à des morphismes différents.

<u>Proposition 1.2.</u> : <u>Soient</u> $L \in X^*$, <u>et</u> $\eta, \eta' \in H(X^*)$ <u>deux morphismes linéairement dépendant vérifiant</u> $|\eta| > |\eta'|$. <u>Si</u> $d^{\eta}(L)$ <u>existe, alors</u> $d^{\eta'}(L)$ <u>existe, et de plus,</u> $d^{\eta}(L) = d^{\eta'}(L)$ <u>ssi</u> $|\eta'| \geq 1$.

<u>Preuve</u> : Comme η et η' sont linéairement dépendants, il existe un nombre réel k tel que $\eta = k\eta'$,et par hypothèse on a $k > 1$. Posons, pour tout entier $n \geq 0$

$$p_{ni} = \frac{k-1}{k^{i+1}} \frac{\sum_{j=0}^{i} |\eta|^j}{\sum_{j=0}^{n} |\eta'|^j} \quad (i=0,\dots,n-1) \; ; \quad p_{nn} = \frac{\sum_{j=0}^{n} |\eta|^j}{k^n \sum_{j=0}^{n} |\eta'|^j} .$$

Par construction, on a $p_{n,i} \geq 0$ pour tout $n \geq i \geq 0$, et $\sum_{i=0}^{n} p_{n,i} = 1$.

De plus, $\frac{\eta'_r(L)}{\eta'_r(X^*)} = \sum_{j=0}^{r} p_{r,j} \frac{\eta_r(L)}{\eta_r(X^*)}$ pour tout $r \geq 0$. Enfin, pour tout entier i fixé,

$$\lim_{n\to\infty} p_{ni} = \begin{cases} 0 & \text{ssi } |\eta'| \geq 1 \; ; \\ \frac{(k-1)(1-|\eta'|)}{k^{i+1}} \sum_{j=0}^{i} |\eta|^j & \text{sinon.} \end{cases}$$

La proposition résulte alors de l'application d'un théorème de Toeplitz (cf. Pólya, Szegö (1964), I. Abschn. N° 66 et III. Abschn. N° 43).

2. DENSITE D'UN LANGAGE ALGEBRIQUE NON AMBIGU.

Nous en venons maintenant au premier résultat principal concernant la nature de la densité des langages formels, si elle existe, Pour cela, soit $\eta \in H(X^*)$; on note, pour simplifier, $\mathbb{Q}(\eta)$ l'extension du corps $\mathbb{Q}$ engendrée par les nombres ηx , où x parcourt l'alphabet X .

<u>Théorème 2.1.</u> : <u>Soit</u> $L \subset X^*$ <u>un langage algébrique non ambigu (resp. rationnel), et soit</u> $\eta \in H^*(X)$, <u>de norme</u> $|\eta| \neq 1$, <u>tel que</u> $d^\eta(L)$ <u>existe ; alors</u> $d^\eta(L)$ <u>est un nombre algébrique sur</u> $Q(\eta)$ <u>(resp.</u> $d^\eta(L) \in Q(\eta)$).

<u>Preuve</u> : Soit y une lettre qui n'est pas dans X , posons $Y = X \cup \{y\}$, et considérons le langage $L' = y^* L \subset Y^*$. Etant donné un entier $r \geq 0$, à tout mot $f \in L$ de longueur $k \leq r$ correspond le mot unique $g = y^{r-k}f \in L' \cap Y^r$, et réciproquement, on en tire $\mathrm{Card}\,(L' \cap Y^r) = \mathrm{Card}\,(L \cap X^* \backslash X^{r+1}X^*)$. Définissons alors un morphisme $\bar\eta \in H(Y^*)$ par $\bar\eta(y) = 1$ et $\bar\eta|_X = \eta$. Il en résulte que $\eta_r(L) = \Sigma\{\bar\eta g : g \in L' \cap Y^r\}$.

D'autre part, le langage L' est algébrique non ambigu (resp. rationnel) ssi L l'est. La série formelle $\mathrm{car}\, L' = \Sigma\{f : f \in L\} \in \mathbb{R}\langle\langle Y\rangle\rangle$ est donc une série algébrique (resp. rationnelle). Soit t une nouvelle lettre. Le morphisme $\varphi : \mathbb{R}\langle\langle Y\rangle\rangle \to \mathbb{R}[[t]]$, défini par $\varphi z = (\bar\eta z)t$ pour tout $z \in Y$, transforme la série $\mathrm{car}\, L'$ en une série $b(t) = \sum_{r\geq 0} b_r t^r$ de $\mathbb{R}[[t]]$ qui est algébrique (resp. rationnelle). Par ailleurs, on a $b_r = \Sigma\{\bar\eta g : g \in L' \cap Y^r\} = \eta_r(L)$. Comme enfin $\dfrac{\eta_r(L)}{\eta_r(X^*)} = \dfrac{(1-|\eta|)}{(1-|\eta|^{r+1})}\, b_r$, en posant :

$$a_r = \begin{cases} b_r & \text{si } |\eta| < 1 \\ \dfrac{b_r}{|\eta|^{r+1}} & \text{si } |\eta| > 1 \end{cases}$$

la série $a(t) = \Sigma\, a_r t^r$ est algébrique (resp. rationnelle) à coefficients dans $\mathbb{Q}(\eta)$ avec la série $b(t)$ et par construction

$$d^\eta(L) = |1-|\eta||\ \lim_{r\to\infty} a_r\ .$$

La preuve du théorème est donc achevée avec celle du lemme suivant :

<u>Lemme 2.2.</u> : <u>Soit</u> $a(t) = \sum_{n\geq 0} a_n t^n$ <u>une série algébrique (resp. rationnelle) à coefficients</u> a_n <u>dans un sous-corps</u> K <u>de</u> $\mathbb{R}$; <u>si</u> $d = \lim_{n\to\infty} a_n$ <u>existe, alors</u> d <u>est algébrique sur</u> K <u>(resp.</u> $d \in K$).

<u>Preuve</u> : Si $d = 0$, le lemme est vrai. Supposons donc $d \neq 0$; par conséquent le rayon de convergence $\rho(a)$ de la série $a(t)$ est 1 . Si $b(t) = \sum_{n\geq 0} b_n t^n = (1-t)a(t)$, alors $\sum_{i=0}^{n} b_i = a_n$, donc $b_n \in K$ pour tout $i \in \mathbb{N}$; d'autre part $b(t)$ est algébrique ou rationnelle avec $a(t)$, et le rayon de convergence $\rho(b)$ est ≥ 1 . Par hypothèse $d = \lim_{n\to\infty} \sum_{i=0}^{n} b_i$, donc $d = \lim_{t\to 1-0} b(t)$. D'autre part, il existe des polynômes $P_i(t) \in K[t]$ $(0 \leq i \leq k)$ (avec $k = 1$ si $b(t)$ est rationnelle) que l'on

peut supposer sans facteur commun, et tels que l'on ait, pour $|t| < 1$

$$\sum_{i=0}^{k} P_i(t)(b(t))^i = 0 .$$

Lorsque t tend vers 1 par valeurs réelles inférieures à 1, on obtient :

$$\sum_{i=0}^{k} P_i(1)d^i = 0 ,$$

et comme on ne peut avoir $P_i(1) = 0$ pour tout $i \in [k]$, le nombre d est bien algébrique sur K (resp. $d \in K$), ce qui prouve le lemme.

<u>Exemple</u> : Soit $X = \{a,b,c\}$, $\not{L}$ le langage de Lukasiewicz sur $\{a,b\}$ défini par $\not{L} = a\not{L}\not{L} \cup b$, et soit $L = X^* \subset \not{L}$. On voit que, pour $r \geq 0$

$$3 \frac{\operatorname{Card}(L \cap X^r)}{\operatorname{Card}(X^r)} = \sum_{j=0}^{r-1} \frac{\operatorname{Card}(\not{L} \cap X^j)}{\operatorname{Card}(X^j)} .$$

Comme par ailleurs $\sum_{j\geq 0} t^j \operatorname{Card}(\not{L} \cap X^j) = \frac{1-\sqrt{1-4t^2}}{2t}$, on trouve que la densité naturelle $d(L)$ du langage L est égale à $\frac{3-\sqrt{5}}{6}$.

Le théorème ci-dessus peut être utilisé, dans certains cas, pour prouver qu'un langage n'est pas rationnel. Je ne connais pas d'exemple d'un langage algébrique inhéremment ambigu ayant une densité non algébrique. Par ailleurs, je n'ai pas pu prouver pour les morphismes de norme 1 , une propriété analogue au résultat suivant de M.P. Schützenberger (1965) qui étend également la proposition 1.2. :

<u>Proposition 2.3.</u> : <u>Si</u> $L \subset X^*$ <u>est rationnel et si</u> $\eta \in H(X^*)$ <u>est de norme</u> 1 , <u>alors</u> $d^\eta(L)$ <u>existe et appartient à</u> $\mathbb{Q}(\eta)$.

Il n'y a pas de réciproque au théorème 2.1. Plus précisément, notons $g^\eta(L) = \sum_{r\geq 0} t^r \Sigma\{\eta f : f \in L \cap X^r\}$ la <u>fonction génératrice relativement à</u> η du langage $L \subset X^*$. Si η est le morphisme naturel, on retrouve la notion habituelle de fonction génératrice.

<u>Corollaire 2.4.</u> : <u>Soit</u> $L \subset X^*$ <u>et</u> $\eta \in H(X^*)$ <u>de norme</u> $|\eta| \neq 1$ <u>tel que</u> $d^\eta(L)$ <u>existe ; pour que</u> $d^\eta(L) \in \mathbb{Q}(\eta)$, <u>il suffit que la série</u> $g^\eta(L)$ <u>soit une série rationnelle.</u>

<u>Preuve</u> : Si la série $g^\eta(L)$ est rationnelle, il en est de même de la série $g^\eta(L)(1-t)^{-1} = \sum_{r\geq 0} \eta_r(L)t^r$. Le corollaire résulte alors de la preuve du théorème 2.1.

Soit par exemple L le langage sur $X = \{x,y\}$ défini par $L = L_1 \cup XL_1 \cup \{1\}$, où $L = \{f \in X(X^2)^* : |f|_x > |f|_y\}$. On a $\operatorname{Card}(L_1 \cap X^{2r+1}) = 2^{2r}$, et la fonction génératrice naturelle du langage L

est $g(L) = \frac{1-t}{1-2t}$. La densité naturelle $d(L)$ de L , si elle existe, est donc un nombre rationnel. En fait, $d(L) = 1/2$.

3. DENSITE DE LANGAGES RATIONNELS.

Dans ce paragraphe, nous étudions la nature et le nombre de points d'accumulation de la suite $\left(\frac{\eta_r(L)}{\eta_r(X^*)}\right)_{r\geq 0}$, que nous appelons *suite de densité*, lorsque L est un langage rationnel ne possédant pas de densité relativement au morphisme η . Ceci nous permet en particulier d'étendre le théorème 2.1. à ce cas en remplaçant la densité par la densité supérieure ou inférieure.

Soit $\mathcal{R}$ un sous-semi-anneau unitiare du semianneau des nombres réels non négatifs. Le semianneau $Rat_{\mathcal{R}}(t)$ des *séries* *$\mathcal{R}$-rationnelles* en la variable t est le plus petit semianneau $\mathcal{K}$ de séries formelles contenant les polynômes et tel que $a(t) \in \mathcal{K}$ et $a(0) = 0$ impliquent $a^+(t) = \sum_{n\geq 0} (a(t))^n = a(t)(1-a(t))^{-1} \in \mathcal{K}$.

On peut démontrer le résultat suivant (Berstel (1972)) :

Proposition 3.1. : *Une série* $a(t)$ *vérifiant* $a(0) = 0$ *appartient à* $Rat_{\mathcal{R}}(t)$ *ssi il existe un alphabet* X , *un langage rationnel* $L \subset X^+$, *et un morphisme* $\eta \in H(X^*)$ *à valeurs dans* $\mathcal{R}$ *tels que* $a(t) = g^{\eta}(L)$.

On ne dispose pas, pour l'instant, de caractérisation analytique des séries $\mathcal{R}$-rationnelles. Une condition nécessaire pour qu'une série rationnelle à coefficients dans $\mathcal{R}$ soit $\mathcal{R}$-rationnelle est donnée dans le théorème suivant (Berstel (1971)), dont nous rappelons la démonstration par souci de complétude.

Théorème 3.2. : *Soit* $a(t)$ *une série rationnelle à coefficients dans* $\mathcal{R}$ *qui n'est pas un polynôme ; pour que* $a(t) \in Rat_{\mathcal{R}}(t)$, *il faut que l'argument de tout pôle de* $a(t)$ *situé sur le cercle de convergence de la série soit multiple rationnel de* 2π .

Preuve : Notons $\mathcal{K}$ la famille des séries rationnelles à coefficients dans $\mathcal{R}$ qui sont des polynômes ou dont les pôles situés sur le cercle de convergence ont leurs arguments multiples rationnels de 2π . Pour démontrer le théorème, donc que $Rat_{\mathcal{R}}(t) \subseteq \mathcal{K}$, vérifions d'abord que $\mathcal{K}$ est un semianneau. Pour cela, soient $a(t)$, $b(t) \in \mathcal{K}$, et posons $c(t) = a(t)+b(t)$ (resp. $c(t) = a(t)b(t)$).

Si $a(t)$ et $b(t)$ sont des polynômes, alors $c(t) \in \mathcal{K}$. Sinon, les pôles de la série $c(t)$ figurent parmi les pôles des séries $a(t)$ et $b(t)$. Par ailleurs, $\rho(c) = \min(\rho(a),\rho(b))$, où $\rho(s)$ est le rayon de convergence de la série s , et les pôles de $c(t)$ de module minimum figurent parmi les pôles de module minimum de la série $a(t)$ si $\rho(a) < \rho(b)$, des séries $a(t)$ et $b(t)$, si $\rho(a) = \rho(b)$. Dans les deux cas, les pôles de $c(t)$ de module minimum ont leurs arguments multiples rationnels de 2π, donc $c(t) \in \mathcal{K}$.

Si $a(t)$ est une série rationnelle non nulle, à coefficients non négatifs, vérifiant $a(0) = 0$, alors il existe un nombre réel positif unique $r < \rho(a)$ tel que $a(r) = 1$, et r est pôle de la série $c(t) = (1-a(t))^{-1}$. Cette série est à coefficients non négatifs et possède, par un théorème de Pringsheim (cf. par exemple Titchmarsh (1968)) un pôle réel positif sur son cercle de convergence, donc $r = \rho(c)$. Si z est un pôle de $c(t)$ de module r, alors $1 = \sum_{n\geq 1} a_n z^n = \sum_{n\geq 1} a_n r^n$, d'où l'on tire $a_n z^n = a_n r^n$ pour tout entier $n \geq 0$ puisque $a_n \geq 0$. Comme la série $a(t)$ est non nulle, ceci implique l'existence d'un entier $p \geq 1$ tel que $z^p = r^p$, et en choisissant p minimal, on a $a_n = 0$ pour tout entier $n \not\equiv 0 \pmod p$. L'argument de z est bien multiple rationnel de 2π, ce qui achève la preuve du théorème.

On peut donner des exemples de séries rationnelles à coefficients dans $\mathbb{R}$ dont tous les pôles sur le cercle de convergence ont leurs arguments multiples rationnels de 2π, sans que ces séries soient $\mathbb{R}$-rationnelles (voir Berstel (1972)).

Nous utilisons maintenant ce théorème pour étudier le comportement de la suite $\left(\frac{\eta_r(R)}{\eta_r(X^*)}\right)_{r\geq 0}$ pour un langage rationnel $R \subset X^*$.

__Théorème 3.3.__ : __Soit__ $R \subset X^*$ __un langage rationnel et__ $\eta \in H(X^*)$; __il existe un entier__ $p \geq 1$ __tel que, pour tout__ k $(0 \leq k \leq p-1)$, __la suite__ $\left(\frac{\eta_{np+k}(R)}{\eta_{np+k}(X^*)}\right)_{n\geq 0}$ __converge vers un élément de__ $\underset{\sim}{Q}(\eta)$.

__Preuve__ : Le résultat est vrai avec $p = 1$ si $|\eta| \leq 1$ par le théorème 2.1. et la proposition 2.3. Supposons donc $q = |\eta| > 1$. Soit y une lettre qui n'est pas dans X, $Y = \{y\} \cup X$, et R' la partie rationnelle de Y^* définie par $R' = y^* R$. Si $\bar{\eta} \in H(Y^*)$ est définie par $\bar{\eta} y = 1$, $\bar{\eta}|_X = \eta$, la série $g^{\bar{\eta}}(R')$ est $\mathbb{R}$-rationnelle par la proposition 3.1, où $\mathbb{R} \subset \underset{\sim}{Q}(\eta)$ est le sous-semianneau de $\mathbb{R}$ engendré par $\underset{\sim}{Q}$ et les ηx $(x \in X)$. Par ailleurs $g^{\bar{\eta}}(R') = a(t) = \sum_{r\geq 0} \eta_r(R) t^r$.

Soit $\mathfrak{Z}$ l'ensemble des pôles de la série $a(t)$ dont les arguments sont multiples rationnels de 2π, et soit p le plus petit entier tel que $\rho \in \mathfrak{Z}$ implique $\rho^p \in R_+$. Pour chaque entier $k \in \{0,\ldots,p-1\}$, la série $a_k(t) = \sum_{n\geq 0} \eta_{np+k} t^n$ est $\mathbb{R}$-rationnelle comme image de la série $a(t)$ par la transduction $\mathbb{R}$-rationnelle $r = (t^k,1)(1-(t^n,t))^{-1}$.

Fixons alors $k \in \{0,\dots,p-1\}$, et supposons que la série $a_k(t)$ n'est pas un polynôme ; ses pôles étant les puissances p-ièmes des pôles de la série $a(t)$, la série $a_k(t)$ a un pôle unique réel positif σ de module minimum r en vertu du théorème précédent, module qui peut d'ailleurs être plus grand que $\rho(a)$. Pour n assez grand, $\eta_{np+k}(R)$ s'écrit :

$$\eta_{np+k}(R) = P(n)\alpha^n + \sum_{i=1}^{s} P_i(n)\alpha_i^n \quad ,$$

où $P(t)$, $P_i(t)$ sont des polynômes, $\alpha\sigma = 1$, et les α_i sont les inverses des pôles de $a_k(t)$ autres que σ ; donc $\alpha > |\alpha_i|$ $(i \in [s])$. Il s'ensuit que :

$$d_k = \lim_{n\to\infty} \frac{\eta_{np+k}(R)}{\eta_{np+k}(X^*)} = \frac{q-1}{q} \lim_{n\to\infty} \frac{P(n)\alpha^n + \sum_{i=1}^{r} P_i(n)\alpha_i^n}{q^{np+k} - 1}$$

existe. Pour que ce nombre soit non nul, il faut que $\alpha = q^p$, et que le polynôme $P(n)$ se réduise à une constante λ, auquel cas $d_k = \frac{(q-1)\lambda}{q^{k+1}}$. Notant $A(t)/B(t)$ $(A(t), B(t) \in \mathbb{R}[t])$ la fraction rationnelle dont la série $a_k(t)$ est le développement, il est classique que $\lambda = -\alpha \frac{A(\sigma)}{B'(\sigma)}$, où B' est la dérivée de B. Il en résulte que $d_k \in \mathbb{Q}(\eta)$. Si enfin $a_k(t)$ est un polynôme, alors évidemment $d_k = 0$. La preuve du théorème est achevée.

Appelons <u>densité supérieure (inférieure)</u> du langage $R \subset X^*$ relativement au morphisme $\eta \in H(X^*)$ le nombre

$$d_+^{\eta}(R) = \limsup_{r\to\infty} \left(\frac{\eta_r(R)}{\eta_r(X^*)}\right) \quad \left(\text{resp. } d_-^{\eta}(R) = \liminf_{r\to\infty} \left(\frac{\eta_r(R)}{\eta_r(X^*)}\right)\right).$$

On a donc le :

<u>Corollaire 3.4.</u> : <u>Pour tout langage rationnel</u> $R \subset X^*$, <u>et tout morphisme</u> $\eta \in H^*(X^*)$, <u>les nombres</u> $d_+^{\eta}(R)$ <u>et</u> $d_-^{\eta}(R)$ <u>appartiennent à</u> $\mathbb{Q}(\eta)$.

L'exemple 2 du paragraphe 1 montre que le nombre de points d'accumulation de la suite de densité d'un langage rationnel R peut être arbitrairement grand. Par contre, il est possible d'en donner une borne supérieure en fonction du nombre d'états d'un automate reconnaissant R dans le cas où le morphisme considéré est le morphisme naturel. Pour cela, soit ψ la fonction définie par :

$$\psi(N) = \mathrm{Max}\,\{p : p = \mathrm{ppcm}(p_1, p_2, \dots, p_l), l \geq 1, \sum_{j=1}^{l} \varphi(p_j) \leq N\} \quad ,$$

où φ désigne la fonction d'Euler. On a le :

Théorème 3.5. : Soit $R \subset X^*$ un langage rationnel et $\eta \in H(X^*)$ le morphisme naturel ; la suite $\left(\frac{\eta_r(R)}{\eta_r(X^*)}\right)_{r\geq 0}$ a au plus $\psi(N)$ points d'accumulations, où N est le nombre d'états de l'automate minimal reconnaissant R ; de plus, si R n'a pas une densité naturelle nulle, le rayon de convergence de la série génératrice de R est $(\mathrm{Card}\ X)^{-1}$, et tous ses pôles sur le cercle de convergence sont simples.

Preuve : Si $\mathrm{Card}\ X = |\eta| = 1$, la densité $d(R)$ existe par la proposition 2.3. Elle est non nulle ssi R est infini, et alors la série génératrice de R est de la forme $g(R) = P(t)(1-t^m)^{-1}$ pour un polynôme $P(t)$ et un entier $m \geq 1$. D'où le résultat.

Supposons donc $q = \mathrm{Card}\ X > 1$, et que la suite de densité possède au moins un point d'accumulation non nul. Soient $\alpha_1, \alpha_2, \ldots, \alpha_r$ $(\alpha_1 = |\alpha_i|$, $i = 2,\ldots,r)$ les inverses des pôles de module minimum de la série $a(t) = \sum_{r\geq 0} \eta_r(R)t^r$; soient $a_{r+1},\ldots,a_s$ les inverses des autres pôles, de sorte que pour n assez grand, on ait :

$$\eta_n(R) = \sum_{i=1}^{s} P_i(n)\alpha_i^n = \sum_{i=1}^{r} P_i(n)\alpha_i^n + \beta_n ,$$

où $\beta_n = \sum_{i=r+1}^{n} P_i(n)\alpha_i^n$, et $P_i(t) \in C(t)$ pour tout $i \in [s]$. Soit p le plus petit entier, qui existe par le théorème 3.2, tel que $\alpha_i^p = \alpha_1^p$ pour tout $i \in \{2,\ldots,r\}$. Nous vérifions d'abord la deuxième partie de l'énoncé. Pour $k = 0,\ldots,p-1$, on a $\eta_{np+k}(R) = Q_k(n)(\alpha_1^p)^n + \beta_{np+k}$, et le polynôme $Q_k(t) = \sum_{i=1}^{r} P_i(tp+k)\alpha_i^k$ n'est pas identiquement nul pour un k au moins. On entire $\alpha_1 = q$, et que $Q_k(t)$ est une constante μ_k pour tout $k \in \{0,\ldots,p-1\}$.

Nous vérifions maintenant que les polynômes $P_i(t)$ $(i \in [r])$ sont des constantes. Soit en effet υ le maximum de leurs degrés, posons $P_i(t) = \sum_{l=0}^{\upsilon} P_{il}t^l$. Le fait que $Q_k(t)$ soit une constante se traduit par :

$$\sum_{i=1}^{r} \Big(\sum_{l=j}^{\upsilon} P_{il}\binom{l}{j}k^{l-j}\Big)\alpha_i^k = 0 \qquad j=1,\ldots,\upsilon\ ;\ k=0,\ldots,p-1 .$$

Comme les α_i $(i \in [r])$ sont distincts, on a $p \geq r$. Il en découle, en considérant le déterminant de Vandermonde (α_i^k) $(i \in [r]$, $k = 0,\ldots,r-1)$ que $\sum_{l=j}^{\upsilon} P_{il}\binom{l}{j}k^{l-j} = 0$ pour tout $j \in [\upsilon]$ et $k \in \{0,\ldots,p-1\}$, d'où enfin $P_{il} = 0$ pour tout $i \in [r]$ et tout $l \in [\upsilon]$. Ceci montre que les polynômes $P_i(t)$ $(i \in [r])$ sont des constantes, soit λ_i , donc que les pôles α_i $(i \in [r])$ sont simples. Par le théorème 3.3, on sait que le nombre de points d'accumulation de la suite de densité est $\leq p$, nombre que nous allons majorer. Pour cela, soit $A(t)/B(t) = a(t)$, où $A(t), B(t)$ sont deux polynômes sans facteur commun, et soit $w = $ degré $B(t)$. Posons $B'(t) = B(\frac{t}{q})$. Les r racines de l'unité $\frac{q}{\alpha_i}$ $(i \in [r])$ figurent parmi les racines de $B'(t)$. Or si ces racines sont d'ordre p_i , alors $p = \mathrm{ppcm}(p_1, \ldots, p_r)$, avec d'ailleurs $p_1 = 1$. Autrement dit, le polynôme $B'(t)$ contient comme facteurs les polynômes cyclotomiques des racines p_i-ièmes de l'unité qui sont respectivement de degré $\varphi(p_i)$. Il en résulte que le degré w de $B'(t)$ vérifie $w-1 \geq \sum_{i=2}^{r} \varphi(p_i)$, et que le nombre p appartient à l'ensemble des nombres p' de la forme $p' = (p'_2, p'_3, \ldots, p'_m)$ pour lesquels $\sum_{i=2}^{m} \varphi(p'_i) \leq w-1$. D'où $p \leq \psi(w-1)$.

Soit maintenant $A = \langle S, s_o, F, \delta\rangle$ l'automate minimal sur X reconnaissant R , et $N = \mathrm{Card}\, S$. L'automate $A' = \langle S', s_o, F, \delta'\rangle$ sur $Y = \{y\} \cup X$ $(y \notin X)$, avec $S' = \{\bar{s}\} \cup S$ $(\bar{s} \notin S)$, et la fonction de transition δ' définie par : $\delta' \mid S \times X = \delta$, $\delta'(s_o, y) = s_o$, $\delta'(s,y) = \bar{s}$ $\delta'(\bar{s}, z) = \bar{s}$ $(z \in Y)$, reconnait le langage $R' = y^* R$ et a $N+1$ états, Soit $M = (m_{s,s'})$ la $S' \times S'$-matrice définie par $m_{s,s'} = \mathrm{Card}\,\{z \in Y :$ $\delta'(s,z) = s'\}$, posons $M^r = (m^{(r)}_{s,s'})$, de sorte que $m^{(r)}_{s,s'} = \{f \in Y^r :$ $\delta'(s,f) = s'\}$ et

$$\sum_{s \in F} m^{(r)}_{s_o,s} = \mathrm{Card}\,(R' \cap Y^r) = \eta_r(R) .$$

Chacune des suites $(m^{(r)}_{s,s'})_{r \geq 0}$ vérifie l'équation de récurrence défini par le polynôme caractéristique de la matrice M , et il en est de même de la suite $(\eta_r(R))_{r \geq 0}$. Comme $B(t)$ divise ce polynôme, on a $w \leq N+1$. Ceci achève la preuve du théorème.

4. DENSITE RELATIVE DE DEUX LANGAGES.

Lorsqu'un langage $R \subset X^*$ est inclus dans un autre langage $R' \subset X^*$, il peut être intéressant de savoir à quel point R approche de R' , en comparant les densités de R et R' . Ceci conduit à la définition suivante :

<u>Définition</u> : Soient $R \subset R' \subset X^*$ deux langages et $\eta \in H(X^*)$. La <u>densité relative</u> $d^{\eta}(R/R')$ de R à R' par rapport à η est la limite, si elle existe, définie par :

$$d^{\eta}(R/R') = \lim_{r \to \infty} \frac{\eta_r(R)}{\eta_r(R')} .$$

On définit de même la densité relative supérieure $d^{\eta}_{+}(R/R')$ et inférieure $d^{\eta}_{-}(R/R')$. La densité (absolue) d'un langage $R \subset X^*$ n'est autre que la densité relative de R par rapport à X^*.

Exemple : Soit $X = \{a,b\}$, $A = \{a^2,b\}^*$, $B = bA$. Pour calculer la densité relative naturelle $d(B/A)$, on constate que $\mathrm{Card}\,(A \cap X^r) = \mathrm{Card}\,(A \cap X^{r-1}) + \mathrm{Card}\,(A \cap X^{r-2}) = \mathrm{Card}\,(B \cap X^{r+2})$ $(r \geq 2)$. Notant f_r ce nombre, on a $f_o = 1$, $f_1 = 1$, donc $(f_n)_{n\geq 0}$ est la suite des nombres de Fibonacci. Il est facile de voir que $d(B/A) = \lim\limits_{r\to\infty} \frac{f_{r-1}}{f_r} = \frac{2}{1+\sqrt{5}}$.

Cet exemple montre que la densité relative, même naturelle, de deux langages rationnels n'est pas en général un nombre rationnel. Par contre, on a le :

Théorème 4.1. : Soient $R \subset R' \subset X^*$ deux langages rationnels, et $\eta \in H(X^*)$. La densité relative $d^{\eta}(R/R')$, si elle existe, est un nombre algébrique sur $\mathbb{Q}(\eta)$.

Nous démontrons le résultat plus général suivant :

Proposition 4.2. : Soient $R \subset R' \subset X^*$ deux langages rationnels et $\eta \in H(X^*)$; il existe un entier $p \geq 0$ tel que, pour tout k $(0 \leq k \leq p-1)$, la suite $\left(\frac{\eta_{np+k}(R)}{\eta_{np+k}(R')}\right)_{n\geq 0}$ converge vers un nombre algébrique sur $\mathbb{Q}(\eta)$.

Preuve : Soit $\mathcal{L}$ le semianneau des nombres de $\mathbb{Q}(\eta)$ non négatifs. Les séries $\delta(t) = \sum\limits_{r\geq 0} \eta_r(R)t^r$ et $\delta'(t) = \sum\limits_{r\geq 0} \eta_r(R')t^r$ sont $\mathcal{L}$-rationnelles, et on peut donc appliquer le théorème 3.2. Soit p un entier tel que pour tout pôle σ de $\delta(T)$ et tout pôle σ' de $\delta'(t)$ d'argument multiple rationnel de π, les nombres σ^p et $\sigma'^{p'}$ soient rationnels positifs, et désignons par $\delta_k(t)$ et $\delta'_k(t)$ les séries $\mathcal{L}$-rationnelles $\delta_k(t) = \sum\limits_{r\geq 0} \delta_{rp+k}(R)t^r$, $\delta'_k(t) = \sum\limits_{r\geq 0} \delta_{rp+k}(R')t^r$. Les séries $\delta_k(t)$ et $\delta'_k(t)$ ont un pôle unique de module minimum qui est réel positif par le théorème 3.2. Notons α_k et α'_k les inverses respectifs de ces pôles ; il existe deux polynômes $P_k(t) = a_k t^m + \ldots$, et $P'_k(t) = a'_k t^{m'} + \ldots$, à coefficients algébriques sur $Q(\eta)$ tels que

$$\lim_{r\to\infty} \frac{\eta_{rp+k}(R)}{\eta_{rp+k}(R')} = \lim_{r\to\infty} \frac{a_k r^m \alpha_k^r}{a'_k r^{m'} \alpha'^r_k}.$$

Cette limite existe puisque $\eta_{rp+k}(R) \leq \eta_{rp+k}(R')$, et si elle est non nulle, alors $\alpha_k = \alpha'_k$, $m = m'$, et elle vaut a_k/a'_k qui est algébrique sur $\mathbb{Q}(\eta)$, ce qu'il fallait prouver.

Corollaire 4.3. : La densité relative supérieure $d_+^{\eta}(R/R')$ (resp. inférieure $d_-^{\eta}(R/R')$) de deux langages rationnels $R \subset R' \subset X^*$ par rapport à un morphisme $\eta \in H(X^*)$ est un nombre algébrique sur $Q(\eta)$.

Comme exemple d'application, considérons le langage $D_1'^* \subset X^*$, où $X = \{x,y\}$ et $D_1'^* = xD_1'^* yD_1'^* \cup \{1\}$. On a le résultat suivant (Berstel (1972)) :

Corollaire 4.4. : Pour tout langage rationnel $R \subset D_1'^*$, la densité relative naturelle $d(R/D_1'^*)$ est nulle.

Ceci montre que $D_1'^*$ ne peut pas être "bien" approché par un langage rationnel contenu en lui.

5. APPLICATION AUX ENSEMBLES ACCEPTABLES DE NOMBRES.

Soit $k \geq 2$ un entier, $k = \{0,1,\dots,k-1\}$ et e le mot vide de k^*. On définit une application ρ_k de k^* dans $\mathbb{N}$ en posant, pour $f = x_0 x_1 \dots x_n \in k^*$ $(x_i \in k)$:

$$\rho_k f = \sum_{i=0}^{n} x_i k^{n-i} .$$

L'entier $\rho_k f$ est le nombre $m \geq 0$ qui s'écrit, en numérotation de base k et éventuellement précédé de lettres 0, sous la forme $f = x_0 x_1 \dots x_n$. Pour tout entier $m \geq 0$, on note $\sigma_k m$ le mot $f \in k^*$ le plus court tel que $\rho_k f = m$. En particulier $\sigma_k 0 = e$, et $\rho_k^{-1}(m) = 0^* \sigma_k m$.

Une partie $A \subset \mathbb{N}$ est rationnellement (resp. algébriquement, algébriquement et non ambigument) acceptable en base $k \geq 2$ ssi le langage $\sigma_k A$, ou de façon équivalente le langage $\rho_k^{-1} A = 0^* \sigma_k A$ est un langage rationnel (resp. algébrique, algébrique non ambigu) de k^*. On note respectivement $\mathrm{Rat}(k)$, $\mathrm{Alg}(k)$, $\mathrm{UAlg}(k)$ l'ensemble des parties de $\mathbb{N}$ de cette nature.

Exemples : 1. Pour $k \geq 2$, l'ensemble $A_k = \{k^n : n \in \mathbb{N}\}$ appartient à $\mathrm{Rat}(k)$ puisque $\sigma_k A = e \cup 10^*$.

2. L'ensemble $B_k = \{(k^n+1)^2 : n \geq 1\}$ appartient à $\mathrm{UAlg}(k)$ pour tout $k \geq 2$. On a en effet :

$$\sigma_k B = \begin{cases} \{10^n 20^n 1 : n \geq 0\} & \text{si } k \geq 3 ; \\ \{10^n 10^{n+2} 1 : n \geq 0\} \cup \{10^2 1\} & \text{si } k = 2 . \end{cases}$$

La propriété fondamentale découlant des considérations antérieures est la suivante :

Propriété 5.1. : Soit $A \subset \mathbb{N}$ un ensemble possédant une densité asymptotique $\delta(A)$ $\left(= \lim_{n\to\infty} \frac{\mathrm{Card}\, A \cap [n]}{n}\right)$; si $A \in \mathrm{Rat}(k)$, alors $\delta(A)$ est un nombre rationnel ; si $A \in \mathrm{UAlg}(k)$, alors $\delta(A)$ est un nombre algébrique.

Dans le cas des ensembles rationnellement acceptables de nombres, cette propriété a été établie directement par Cobham (1972). Nous la déduisons du théorème 2.1. par la :

<u>Proposition 5.2.</u> : <u>Si</u> $A \subset \mathbb{N}$ <u>a une densité asymptotique</u> $\delta(A)$, <u>alors le langage</u> $L = \rho_k^{-1} A \subset k^*$ <u>possède, pour tout</u> $k \geq 2$, <u>une densité naturelle</u> $d(L)$, <u>et</u> $d(L) = \delta(A)$.

<u>Preuve</u> : On a $\text{Card}\,\{L \cap X^r\} = \text{Card}\,A \cap [k^r - 1]$, donc

$$\delta(A) = \lim_{r\to\infty} \frac{\text{Card}\,A \cap [k^r-1]}{[k^r-1]} = \lim_{r\to\infty} \frac{\text{Card}\,(L \cap X^r)}{\text{Card}\,(X^r)} = d(L) .$$

<u>Exemple 1</u> : L'ensemble F des nombres quadrat-frei a une densité asymptotique $\delta(F) = \frac{6}{\pi^2}$ (cf. Hardy, Wright (1965)). Il en résulte pour tout $k \geq 2$, on a $F \notin \text{UAlg}(k)$.

La proposition suivante donne un critère plus souple :

<u>Proposition 5.3.</u> : <u>Soit</u> $A \subset \mathbb{N}$, <u>et soit</u> $b(t) = \Sigma\, b_r t^r$ <u>une série rationnelle à coefficients rationnels telle que pour un entier</u> $k \geq 2$, <u>la limite</u>

$$c = \lim_{r\to\infty} b_r \,\text{Card}\,(A \cap [k^r])$$

<u>existe ; si</u> $A \in \text{UAlg}(k)$ <u>(resp.</u> $A \in \text{Rat}(k)$), <u>alors</u> c <u>est un nombre algébrique (resp. rationnel).</u>

<u>Preuve</u> : La série $a(t) = \sum_{r\geq 0} \text{Card}\,(A \cap [k^r - 1])t^r$ est algébrique (resp. rationnelle) en tant que série génératrice du langage $\rho_k^{-1}A$. Il en est de même de la série $u(t) = \Sigma\, u_r t^r$, où $u_r = b_r\,\text{Card}\,(A \cap [k^r-1])$ en tant que produit de Hadamard de $a(t)$ par $b(t)$. Comme $\lim_{r\to\infty} u_r = c$ existe, la proposition résulte du lemme 2.2.

Cette proposition peut s'appliquer comme suit :

<u>Exemple 2</u> : Soit $d(n)$ le nombre de diviseurs de l'entier n , et $D = \{d(1) + d(2) + \ldots + d(n) : n \geq 1\}$. On sait (Hardy, Wright (1965)) que $\text{Card}\,(D \cap [m]) \sim \frac{m}{\log m}$, donc $\text{Card}\,D \cap [k^r] \sim \frac{k^r}{r \log k}$. La série $b(t)$ de terme général $b_r = \frac{r}{k^r}$ est rationnelle, de sorte que le nombre c de la proposition précédente vaut $c = \frac{1}{\log k}$ qui n'est pas algébrique. Il en résulte que $D \notin \text{UAlg}(k)$, quel que soit l'entier $k \geq 2$.

REFERENCES.

J. Berstel (1971), Sur les pôles et le quotient de Hadamard de séries $\mathbb{N}$-rationnelles, C.R. Acad. Sci. Paris, <u>272</u>, 1079-1081.

J. Berstel (1972), Contribution à l'étude des propriétés arithmétiques des langages formels, Thèse, Paris.

A. Cobham (1972), Uniform Tag Sequences, Math. System Theory, <u>6</u>, 164-192.

G.H. Hardy, E.M. Wright (1965), The Theory of Numbers, Oxford, 4th ed.

M. Minsky, S. Papert (1966), Unrecognizable sets of numbers, Journal ACM, 13, 281-286.

G. Pólya, G. Szegö (1964), Aufgaben und Lehrsätze aus der Analysis, Springer, 3. Aufl.

M.P. Schützenberger (1965), Sur certains sous-monoïdes libres, Bull. Soc. Math. France, 93, 209-223.

E.C. Titchmarsh (1968), The Theory of Functions, Oxford, 2nd ed.