JOURNAL DE THÉORIE DES NOMBRES DE BORDEAUX

JEAN-ÉRIC PIN

Topologie p-adique sur les mots

Journal de Théorie des Nombres de Bordeaux, tome 5, nº 2 (1993), p. 263-281

<http://www.numdam.org/item?id=JTNB_1993__5_2_263_0>

© Université Bordeaux 1, 1993, tous droits réservés.

L'accès aux archives de la revue « Journal de Théorie des Nombres de Bordeaux » (http://jtnb.cedram.org/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

Journal de Théorie des Nombres
de Bordeaux **5** *(1993), 263–281*

Topologie p-adique sur les mots

par JEAN-ÉRIC PIN

RÉSUMÉ — Cet article est une introduction aux aspects combinatoires de la distance p-adique et de la topologie p-adique sur les mots. On donne plusieurs définitions équivalentes de ces notions, illustrées par divers exemples et propriétés. Après avoir décrit de façon détaillée les ouverts, on démontre que la distance p-adique est uniformément équivalente à une distance obtenue à partir des coefficients binomiaux définis sur les mots. On donne également deux exemples de suites convergentes dans la topologie p-adique. Le premier exemple est constitué par la suite des puissances d'ordre p^n d'un mot fixé, qui converge vers le mot vide. Le second est formé par la suite des préfixes du mot de Prouhet-Thue-Morse : pour chaque nombre premier p, on peut extraire de cette suite une sous-suite qui converge vers le mot vide dans la topologie p-adique. La plupart des démonstrations sont omises, à l'exception de celles qui tiennent en quelques lignes.

1. Introduction

Cet article est un article de synthèse sur les aspects combinatoires de la distance p-adique et de la topologie p-adique sur les mots. En revanche, nous n'abordons pas les aspects analytiques et les applications à la théorie des groupes. Nous invitons le lecteur intéressé par ces développements mathématiques à consulter le récent livre [4]. Cet article ne contient pas de résultats nouveaux, sauf peut-être le théorème 4.4. On ne trouvera pas non plus de démonstrations sauf lorsqu'elles s'écrivent en quelques lignes.

La topologie p-adique sur les mots peut être définie de diverses façons, qui sont décrites dans la section 2 : distances p-adiques, topologie initiale, ou encore topologie trace de la topologie p-adique sur le groupe libre. La description précise des ouverts nécessite quelques outils de théorie des automates, qui sont présentés dans la section 3. La section 4 est de nature plus combinatoire : on y démontre que les coefficients binomiaux sur les

Manuscrit reçu le 17 octobre 1991, version définitive le 7 décembre 1993.
Travail réalisé dans le cadre du PRC Mathématiques et Informatique.
Les résultats contenus dans cet article ont été exposés au colloque "Thémate", (CIRM, Luminy, Mai 1991).

mots fournissent une définition purement combinatoire de la topologie p-adique. On en déduit une série de propriétés remarquables des ouverts, qui sont détaillées dans la section 5. On évoque aussi dans cette section une conjecture sur la caractérisation des ouverts qui sont reconnaissables au sens de la théorie des automates. On présente dans la dernière section un curieux lien entre la suite des préfixes du mot infini de Prouhet-Thue-Morse et la topologie p-adique : pour chaque nombre premier p, on peut extraire de cette suite une sous-suite qui converge vers le mot vide.

2. Topologie et structure uniforme p-adique sur les mots

Soit p un nombre premier. Pour tout entier n, on note $v_p(n)$ la *valuation p-adique* de n, c'est-à-dire l'exposant de p dans la décomposition de n en facteurs premiers. Les nombres p-adiques peuvent être définis comme le complété de $\mathbb{Z}$ pour la *distance p-adique*, définie, pour tout couple d'entiers relatifs (x, y), en posant

$$d_p(x,y) = \begin{cases} p^{-v_p(|x-y|)} & \text{si } x \neq y, \\ 0 & \text{si } x = y. \end{cases}$$

On se propose d'étendre la construction précédente de $\mathbb{Z}$ aux groupes libres de base finie. Ce qui va jouer maintenant le rôle de valuation p-adique, ce sont les morphismes de groupe du groupe libre dans un p-groupe (i. e. un groupe fini dont l'ordre est une puissance de p). Cette construction est en fait un cas particulier des topologies *profinies* introduites par M. Hall [7].

Soit $F(A)$ le groupe libre de base (finie) A. On dit qu'un groupe G *sépare* deux éléments u et v du groupe libre s'il existe un morphisme de groupe φ de $F(A)$ dans G tel que $\varphi(u) \neq \varphi(v)$. On pose alors, pour tout $u, v \in F(A)$,

$$v_p(u) = \min \{n \mid \text{il existe un } p\text{-groupe d'ordre } p^n \text{ qui sépare } u \text{ et } 1\}$$

et

$$d_p(u,v) = p^{-v_p(uv^{-1})}.$$

Nous démontrerons plus loin que d_p est une distance sur $F(A)$, qui est *ultramétrique* et *invariante par translation*, ce qui signifie que, pour tout triplet de (u, v, w) de $F(A)$, on a

$$d_p(u,w) \leq \max(d_p(u,v), d_p(v,w)),$$
$$d_p(wu, wv) = d_p(u,v) = d_p(uw, vw).$$

On appellera d_p la *distance p-adique* sur $F(A)$. Cette distance définit une topologie (la *topologie p-adique*) et une structure uniforme(*), (la *structure uniforme p-adique*).

Muni de cet distance, le groupe libre devient un groupe *topologique*. Plus précisément, on a la propriété suivante.

PROPOSITION 2.1. *La multiplication du groupe libre est uniformément continue pour la distance p-adique.*

Il est temps de rappeler la construction du monoïde et du groupe libres. On part d'un ensemble A, baptisé *alphabet*, dont les éléments sont des *lettres*. Un *mot* sur l'alphabet A est une suite finie de lettres, que l'on note par simple juxtaposition :

$$a_1 a_2 \cdots a_n.$$

La *longueur* d'un mot u, notée $|u|$, est le nombre d'éléments de la suite de lettres qui compose le mot. Par exemple $|abaab| = 5$. Le produit (ou concaténation) de deux mots $a_1 a_2 \cdots a_r$ et $b_1 b_2 \cdots b_s$ est le mot $a_1 a_2 \cdots a_r b_1 b_2 \cdots b_s$ obtenu en les plaçant bout à bout. La concaténation est une opération associative, qui munit l'ensemble A^* des mots d'une structure de monoïde dont l'élément neutre est le mot vide, que l'on note 1 pour cette raison. Le produit s'étend aux parties de A^* en posant, pour tout $X, Y \subset A^*$,

$$XY = \{xy \mid x \in X,\ y \in Y\}$$

et munit à son tour l'ensemble $\mathcal{P}(A^*)$ des parties de A^* d'une structure de monoïde.

Le monoïde A^* est le monoïde *libre* sur l'ensemble A, car il possède la propriété universelle qui caractérise les structures libres : toute application de A dans un monoïde M se prolonge de façon unique en un morphisme de monoïde de A^* dans M. Il en résulte en particulier que tout monoïde engendré par un ensemble A est quotient de A^*.

La construction du groupe libre de base A est un peu plus compliquée. On considère d'abord l'ensemble des inverses formels des éléments de A

$$\bar{A} = \{\bar{a} \mid a \in A\}$$

(*) La notion d'espace uniforme est peut-être moins familière que celle de topologie. Très grossièrement, les topologies permettent une définition abstraite des suites convergentes, alors que les structures uniformes permettent de définir suites de Cauchy et espaces complets, cf. [3].

et on prolonge l'application $a \mapsto \bar{a}$ (de A dans $\bar{A}$) en une involution de l'ensemble $\tilde{A} = A \cup \bar{A}$. Il suffit pour cela de poser $\bar{\bar{a}} = a$ pour toute lettre $\bar{a}$ de $\bar{A}$. On considère ensuite le monoïde présenté sur $\tilde{A}$ par les relations $a\bar{a} = \bar{a}a = 1$ pour tout $a \in A$. Par construction, le monoïde ainsi obtenu est un groupe, et c'est en fait le groupe libre sur A. On le note $F(A)$ et on note $\pi : \tilde{A}^* \to F(A)$ le morphisme de monoïde qui définit le quotient.

Il existe une autre représentation du groupe libre qui est plus facile à utiliser. On dit qu'un mot de $\tilde{A}^*$ est *réduit* s'il ne contient aucun facteur de la forme $a\bar{a}$ ou $\bar{a}a$. On peut démontrer que tout mot u de $\tilde{A}^*$ est équivalent à un unique mot réduit $\delta(u)$. Ce mot est obtenu à partir de u en supprimant les occurrences de tous les facteurs de la forme $a\bar{a}$ et $\bar{a}a$, et en itérant ce procédé (on démontre que le résultat est indépendant de l'ordre de suppression des facteurs). Par exemple, $\delta(aa\bar{b}\bar{a}b\bar{b}ab\bar{b}b\bar{a}b) = ab$. Il est clair que u et $\delta(u)$ ont la même image par π. De plus, la restriction de π à l'ensemble R des mots réduits est une bijection, notée β, de R sur $F(A)$, qui permet d'identifier le groupe libre à R. La situation est résumée sur la figure suivante :

$$\begin{tikzcd} \tilde{A}^* \arrow[rr, "\pi"] \arrow[dr, "\delta"'] & & F(A) \\ & R \arrow[ur, "\beta"'] & \end{tikzcd}$$

Figure 2.1. Un diagramme commutatif

Les opérations de groupe s'interprètent facilement dans R. L'inverse d'un mot s'obtient en «retournant» le mot et en remplaçant chaque lettre par son inverse formel. Par exemple, si $u = ab\bar{a}\bar{b}aa$, l'inverse de u est le mot $\bar{a}\bar{a}ba\bar{b}\bar{a}$. Pour calculer le produit dans le groupe libre, on effectue la concaténation habituelle et on réduit le résultat obtenu. Ainsi, si $u = ab\bar{a}\bar{b}aa$ et si $v = \bar{a}\bar{a}bab$, le produit dans le groupe libre est égal à $\delta(ab\bar{a}\bar{b}aa\bar{a}\bar{a}bab) = abb$.

Il est important de remarquer que R n'est pas un sous-monoïde de $\tilde{A}^*$ et que β n'est donc pas un morphisme de monoïdes. En revanche, puisqu'un mot ne contenant aucune lettre barrée est nécessairement réduit, A^* est un sous-ensemble de R, et β induit une application injective du monoïde libre A^* dans le groupe libre $F(A)$, ce qui permet d'identifier le monoïde libre à une partie du groupe libre.

En particulier, on appellera *topologie p-adique* (resp. *structure uniforme p-adique*) *sur* A^* la restriction à A^* de la topologie p-adique et de la

structure uniforme p-adique sur $F(A)$. Nous allons voir que l'on peut donner plusieurs définitions équivalentes de ces deux notions.

Une première façon de procéder consiste à utiliser la notion de structure initiale. Rappelons que si E est un ensemble et si $\mathcal{F} = (\varphi_i)_{i\in I}$ est une famille d'applications $\varphi_i : E \to (F_i, \mathcal{T}_i)$ de E dans un espace topologique (resp. uniforme) $(F_i, \mathcal{T}_i)$, la topologie (resp. structure uniforme) *initiale* définie par $\mathcal{F}$ est la topologie (resp. structure uniforme) la moins fine qui rend continue (resp. uniformément continue) chacune des applications φ_i. Une base d'ouverts pour cette topologie est obtenue en prenant les intersections finies d'ensembles la forme $\omega_i\varphi_i^{-1}$, où ω_i est un ouvert de $\mathcal{T}_i$. On peut alors énoncer

PROPOSITION 2.2.

(1) *La topologie p-adique (resp. la structure uniforme p-adique) sur $F(A)$ est la topologie (resp. structure uniforme) initiale définie par les morphismes de groupe de $F(A)$ dans un p-groupe muni de la topologie discrète (resp. la structure uniforme discrète).*

(2) *La topologie p-adique (resp. la structure uniforme p-adique) sur A^* est la topologie (resp. structure uniforme) initiale définie par les morphismes de monoïde de $F(A)$ dans un p-groupe muni de la topologie discrète (resp. la structure uniforme discrète).*

Ce résultat permet de donner une caractérisation simple des suites convergentes et des suites de Cauchy. Rappelons qu'une suite $(u_n)_{n\geq 0}$ d'éléments d'un ensemble E est dite *ultimement constante* (resp. *ultimement égale* à u) s'il existe un entier n_0 tel que, pour tout $n \geq n_0$, $u_n = u_{n_0}$ (resp. $u_n = u$).

PROPOSITION 2.3.

(1) *Une suite d'éléments $(u_n)_{n\geq 0}$ de $F(A)$ est une suite de Cauchy si et seulement si, pour tout morphisme de groupe φ de $F(A)$ dans un p-groupe, la suite $(\varphi(u_n))_{n\geq 0}$ est ultimement constante.*

(2) *La suite $(u_n)_{n\geq 0}$ converge vers un élément u de $F(A)$ si et seulement si, pour tout morphisme de groupe φ de $F(A)$ dans un p-groupe, la suite $(\varphi(u_n))_{n\geq 0}$ est ultimement égale à $\varphi(u)$.*

On en déduit un premier exemple de suite convergente :

COROLLAIRE 2.4. *Pour tout $u \in F(A)$, on a $\lim_{n\to\infty} u^{p^n} = 1$.*

Démonstration. Soit φ un morphisme de groupe de $F(A)$ dans un p-groupe G d'ordre p^k. Pour tout $n \geq k$, on a $\varphi(u^{p^n}) = (\varphi(u))^{p^n} = 1$ puisque p^n est un multiple de l'ordre de G. □

Comme la multiplication est continue, on en déduit, pour tout $x, y, u \in F(A)$,

$$\lim_{n\to\infty} xu^{p^n} y = xy \tag{2.1}$$

En particulier, en prenant $x = u^{-1}$ et $y = 1$, on obtient la formule

$$\lim_{n\to\infty} u^{p^n-1} = u^{-1} \tag{2.2}$$

Cette dernière formule permet de «remplacer les inverses par des limites», et permet de démontrer que le monoïde libre est dense dans le groupe libre.

PROPOSITION 2.5. *Le complété du groupe libre pour la distance d_p est un groupe compact. Le monoïde libre est dense dans le groupe libre et son complété pour la distance d_p est donc aussi un groupe compact.*

Citons encore une propriété élémentaire des morphismes entre structures libres. Rappelons qu'une application φ d'un espace métrique (E, d) dans un espace métrique (E', d') est contractante si, pour tout $u, v \in E$, $d'(\varphi(u), \varphi(v)) \leq d(u, v)$.

PROPOSITION 2.6. *Tout morphisme de monoïde (resp. groupe) entre deux monoïdes (resp. groupes) libres est une application contractante pour la distance p-adique.*

Pour décrire les ouverts de la topologie p-adique de A^*, nous aurons besoin de quelques définitions empruntées à la théorie des automates.

3. Parties reconnaissables

On dit qu'une partie L de A^* est *reconnue* par un morphisme de monoïde $\varphi : A^* \to M$ si $L = \varphi^{-1}(\varphi(L))$. Notons qu'il est équivalent de dire qu'il existe une partie P de M telle que $L = \varphi^{-1}(P)$ car si $L = \varphi^{-1}(P)$, il vient $\varphi^{-1}\varphi(L) = \varphi^{-1}(\varphi\varphi^{-1}(P)) = \varphi^{-1}(P) = L$. Par extension, on dit qu'un monoïde M reconnaît L s'il existe un morphisme de monoïde $\varphi : A^* \to M$ qui reconnaît L. Une partie L de A^* est dite *reconnaissable* si elle est reconnue par un monoïde fini.

Exemple 3.1. L'ensemble L des mots de longueur multiple de n est reconnu par le morphisme de monoïde $\varphi : A^* \to \mathbb{Z}/n\mathbb{Z}$ défini par

$$\varphi(u) = |u| \bmod n.$$

En effet, on a $L = \varphi^{-1}(0)$.

Exemple 3.2. Soit M le monoïde formé des six matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

muni de la multiplication usuelle des matrices. Le morphisme $\varphi : \{a, b\}^* \to M$ défini par

$$\varphi(a) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \varphi(b) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

reconnaît l'ensemble $(ab)^+$ des puissances non nulles du mot ab, puisque

$$(ab)^+ = \varphi^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

On peut aussi définir les parties reconnaissables en utilisant des automates. Un *automate déterministe fini* ou simplement *automate* est un quintuplet $\mathcal{A} = (Q, A, \cdot, q_0, F)$ où Q est un ensemble fini (l'ensemble des *états*), A est un ensemble fini (*l'alphabet*) dont les éléments sont des *lettres*, $\cdot$ est une application de $Q \times A$ dans Q, qui à un état q et une lettre a associe l'état $q \cdot a$, q_0 est un état de Q, appelé état *initial* et F est un ensemble fini d'états, appelé l'ensemble des états *finaux*(*). L'application $(q, a) \mapsto q \cdot a$ se prolonge de façon naturelle en une action de A^* sur Q en posant, pour tout $q \in Q$, $u \in A^*$ et $a \in A$,

$$q \cdot 1 = q$$
$$q \cdot (ua) = (q \cdot u) \cdot a$$

ce qui revient à poser, pour tout mot $u = a_1 a_2 \cdots a_n$,

$$q \cdot u = (\cdots((q \cdot a_1) \cdot a_2) \cdots a_n).$$

(*) Faut-il dire « finals » ou « finaux » ? Les deux sont acceptables, mais Grévisse indique que « finaux » se répand de plus en plus, notamment chez les grammairiens. On le trouve aussi chez Alexis Carrel, Raymond Aron et Jean-Pierre Chevènement...

Chaque mot u de A^* définit donc une transformation $\rho(u)$ de Q (i. e. une application de Q dans lui-même). L'application ρ ainsi définie est un morphisme de monoïde de A^* dans le monoïde $\mathcal{T}(Q)$ des transformations de Q, muni de l'opération $(f,g) \mapsto g \circ f$. L'ensemble $\rho(A^*)$ est un sous-monoïde de $\mathcal{T}(Q)$, noté $M(\mathcal{A})$, et appelé le *monoïde de* $\mathcal{A}$.

Les automates peuvent être représentés par un graphe étiqueté dont les états forment les sommets et les arêtes sont les triplets de la forme $(q, a, q.a)$ où q est un état et a une lettre. L'état initial est représenté par une flèche entrante et les états finaux par des flèches sortantes. Par exemple, l'automate $\mathcal{A} = (\{0,1,2\}, \{a,b\}, \cdot, 1, \{0,2\})$, avec

$$0 \cdot a = 0 \qquad 1 \cdot a = 2 \qquad 2 \cdot a = 0 \qquad\qquad 0 \cdot b = 0 \qquad 1 \cdot b = 0 \qquad 2 \cdot b = 1$$

est représenté sur la figure 3.1.

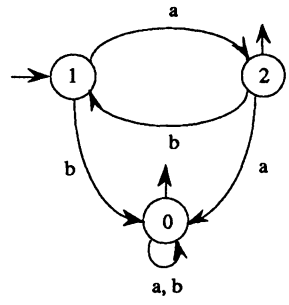

Figure 3.1. Un automate.

Soient q un état et u un mot. Pour «lire» $q \cdot u$ sur ce graphe, il suffit de suivre le chemin d'origine q et d'étiquette u : l'état d'arrivée donne le résultat. Par exemple $1 \cdot abaaba = 0$.

Le monoïde d'un automate est un groupe si et seulement si chaque lettre induit une permutation de l'ensemble des états. Par exemple, le monoïde associé à l'automate représenté sur la figure 3.2 est un groupe.

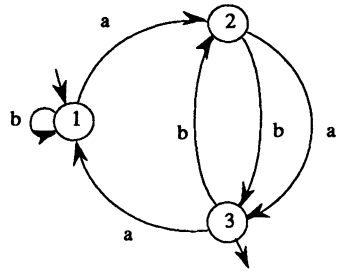

Figure 3.2. Un automate dont le monoïde associé est un groupe.

L'ensemble des mots reconnus par $\mathcal{A}$ est l'ensemble, noté $L(\mathcal{A})$, des étiquettes des chemins issus de l'état initial et qui arrivent dans un état final :

$$L(\mathcal{A}) = \{u \in A^* \mid q_0 \cdot u \in F\}.$$

Il reste à établir l'équivalence des deux définitions des parties reconnaissables. Si $\mathcal{A}$ est un automate, l'ensemble $L(\mathcal{A})$ est reconnu par le morphisme $\rho : A^* \to M(\mathcal{A})$. En effet, si on pose $P = \{f \in \mathcal{T}(Q) \mid f(q_0) \in F\}$, on a $L(\mathcal{A}) = \rho^{-1}(P)$. Réciproquement, soit $\varphi : A^* \to M$ un morphisme de monoïde de A^* dans un monoïde fini M et soit P une partie de M. L'ensemble $L = \varphi^{-1}(P)$ est alors reconnu par l'automate $(M, A, \cdot, 1, P)$, où l'action de A sur M est donnée, pour tout $m \in M$ et tout $a \in A$, par les formules :

$$m \cdot a = m\varphi(a).$$

Si on a, avec les notations précédentes, $L = \varphi^{-1}(P)$, on a aussi $A^* \setminus L = \varphi^{-1}(M \setminus P)$. Par conséquent, si L est reconnu par M, il en est de même de $A^* \setminus L$. On démontre également facilement que si L_1 et L_2 sont reconnus respectivement par les monoïdes M_1 et M_2, alors $L_1 \cap L_2$ et $L_1 \cup L_2$ sont reconnus par le produit direct $M_1 \times M_2$. On en déduit en particulier que l'ensemble des parties de A^* reconnues par des p-groupes forme une algèbre de Boole (pour l'union, l'intersection et la complémentation dans A^*). Nos donnerons dans la section 4 une description plus explicite de ces parties.

On déduit des remarques précédentes et de la proposition 2.2 que la topologie p-adique sur A^* admet pour base d'ouverts les parties reconnues par un p-groupe, ce qui fournit la description suivante des ouverts.

PROPOSITION 3.1. *Les ouverts de la topologie p-adique de A^* sont les unions (finies ou infinies) de parties de A^* reconnues par des p-groupes.*

Notons enfin que les parties reconnues par un p-groupe sont à la fois ouvertes et fermées.

4. Topologie p-adique et coefficients binomiaux

Les coefficients binomiaux définis sur les entiers peuvent être eux aussi étendus aux mots. Leur définition, due à S. Eilenberg [5], repose sur la notion de *sous-mot*. On dit qu'un mot $u \in A^*$ est un sous-mot d'un mot $v \in A^*$ si $u = a_1 a_2 \cdots a_n$ (où $a_1, a_2, \ldots, a_n \in A$) et si v se factorise sous la forme $v = v_0 a_1 v_1 a_2 v_2 \cdots v_{n-1} a_n v_n$ avec $v_0, v_1, \ldots, v_n \in A^*$. Par exemple, *abca* est un sous mot de *accbacabac* ; pour le voir, il suffit de faire apparaître en gras le mot *abca* : **acc**b**aca**b*ac*. Le coefficient binomial $\binom{v}{u}$ est le nombre de façons distinctes d'écrire u comme sous-mot de v. De façon plus formelle, si $u = a_1 a_2 \cdots a_n$ comme ci-dessus,

$$\binom{v}{u} = \text{Card } \{(v_0, v_1, \ldots, v_n) \in (A^*)^{n+1} \mid v = v_0 a_1 v_1 a_2 v_2 \cdots v_{n-1} a_n v_n\}.$$

On a par exemple

$$\binom{abab}{a} = 2 \qquad \binom{abab}{ab} = 3 \qquad \binom{abab}{ba} = 1 \qquad \binom{abab}{aba} = 1$$

et, si n et m sont des entiers,

$$\binom{a^n}{a^m} = \binom{n}{m}$$

ce qui montre bien qu'il s'agit d'une généralisation des coefficients binomiaux habituels ! On peut aussi définir les coefficients binomiaux à partir des formules de récurrence suivantes, où $u, v \in A^*$ et $a, b \in A$:

$$\text{(4.1)} \qquad \begin{cases} \binom{u}{1} = 1, \\ \binom{1}{u} = 0 \quad \text{si } u \neq 1, \\ \binom{va}{ub} = \begin{cases} \binom{v}{ub} & \text{si } a \neq b, \\ \binom{v}{ub} + \binom{v}{u} & \text{si } a = b. \end{cases} \end{cases}$$

Une troisième méthode consiste à utiliser l'automorphisme de Magnus μ de l'algèbre $\mathbb{Z}[A^*]$ défini par $\mu(a) = 1 + a$ pour tout $a \in A$. On démontre que pour tout mot $v \in A^*$,

$$\text{(4.2)} \qquad \mu(v) = \sum_{u \in A^*} \binom{v}{u} u$$

d'où l'on déduit la formule

$$\text{(4.3)} \qquad \binom{uv}{x} = \sum_{x_1 x_2 = x} \binom{u}{x_1} \binom{v}{x_2}.$$

Enfin, on peut utiliser le fait que pour tout mot $a_1a_2\cdots a_n \in A^*$, l'application $\tau : A^* \to \mathcal{M}_{n+1}(\mathbb{Z})$ du monoïde libre dans le monoïde des matrices carrées d'ordre n à coefficients entiers définie par

$$\tau(u) = \begin{pmatrix} 1 & \binom{u}{a_1} & \binom{u}{a_1a_2} & \binom{u}{a_1a_2a_3} & \cdots & \binom{u}{a_1a_2\cdots a_n} \\ 0 & 1 & \binom{u}{a_2} & \binom{u}{a_2a_3} & \cdots & \binom{u}{a_2\cdots a_n} \\ 0 & 0 & 1 & \binom{u}{a_3} & \cdots & \binom{u}{a_3\cdots a_n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \binom{u}{a_n} \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

est un morphisme de monoïdes.

Le lien entre la topologie p-adique et les coefficients binomiaux provient essentiellement du résultat suivant :

PROPOSITION 4.1. [5]. *Soit $x \in A^*$, soit p un nombre premier et soit r un entier. Alors l'ensemble*

$$L(x,r,p) = \left\{u \in A^* \mid \binom{u}{x} \equiv r \mod p\right\}$$

est reconnu par un p-groupe.

Démonstration. Posons $x = a_1a_2\cdots a_n \in A^*$, et notons $\tau_p : A^* \to \mathcal{M}_{n+1}(\mathbb{Z}/p\mathbb{Z})$ le morphisme de monoïdes obtenu en réduisant τ modulo p. L'image de τ_p est un monoïde de matrices *unitriangulaires* (c'est-à-dire triangulaires supérieures et ayant des 1 sur la diagonale) à coefficients dans $\mathbb{Z}/p\mathbb{Z}$, et on peut vérifier que ce monoïde est en réalité un p-groupe. La formule

$$L(x,r,p) = \{u \in A^* \mid \tau(u)_{1,n+1} \equiv r \mod p\}$$

montre que τ reconnaît $L(x,r,p)$. □

La proposition 4.1 permet de vérifier que d_p est bien une distance. En effet, la principale difficulté consiste à démontrer que si $u \neq v$ alors $d_p(u,v) \neq 0$. Or si $u \neq v$, on peut supposer, quitte à intervertir u et v, que u n'est pas un sous-mot de v. Or d'après la proposition 4.1, l'ensemble

$$L = \left\{w \in A^* \mid \binom{w}{u} \equiv 1 \mod p\right\}$$

est reconnu par un p-groupe G. Comme $u \in L$ puisque $\binom{u}{u} = 1$ et $v \notin L$ puisque $\binom{v}{u} = 0$, G sépare u et v, et $d_p(u,v) > 0$. Ceci démontre en particulier que le groupe libre est résiduellement fini.

Il est facile de construire un automate qui reconnaît $L(x,r,p)$. Prenons par exemple $A = \{a,b,c\}$, $x = abc$, $p = 2$ et $r = 1$. Ce qui précède montre qu'il suffit de mémoriser $\binom{u}{a}$, $\binom{u}{ab}$ et $\binom{u}{abc}$ modulo 2. On prendra donc pour ensemble d'états $\{0,1\}^3$, chaque état représentant une valeur possible du triplet $\left(\binom{u}{a}, \binom{u}{ab}, \binom{u}{abc}\right)$. L'action des lettres se calcule facilement en utilisant la formule 4.1. L'automate obtenu est représenté sur la figure 4.1.

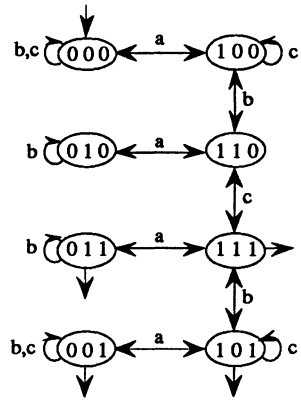

Figure 4.1. Un automate reconnaissant $L(abc,1,2)$.

La proposition 4.1 donne des exemples de langages reconnus par un p-groupe. Ces exemples sont en fait génériques, comme le montre le résultat qui suit.

PROPOSITION 4.2. (Eilenberg et Schützenberger [5]). *L'ensemble des parties de A^* reconnues par un p-groupe est l'algèbre de Boole engendrée par les parties de la forme $L(x,r,p)$ avec $x \in A^*$ et $0 \leq r < p$.*

La démonstration de ce résultat utilise une caractérisation remarquable des p-groupes. Définissons une congruence $\sim_k$ sur A^* en posant $u \sim_k v$ si et seulement si

$$\binom{u}{x} \equiv \binom{v}{x} \bmod p \quad \text{pour tout mot } x \text{ tel que } 0 < |x| < k.$$

Le monoïde quotient $G_k(A,p) = A^*/\sim_k$ est un p-groupe qui possède la propriété suivante :

PROPOSITION 4.3. *Tout p-groupe d'ordre p^n engendré par A est quotient de $G_{p^n}(A,p)$.*

Posons maintenant, pour tout mot $u, v \in A^*$,

$$\delta'_p(u,v) = \min\left\{|x| \mid x \in A^* \text{ et } \binom{u}{x} \not\equiv \binom{v}{x} \bmod p\right\}$$

(on fait la convention habituelle $\min \emptyset = -\infty$) et

$$d'_p(u,v) = p^{-\delta'_p(u,v)}.$$

On vérifie que d'_p est une distance linéaire ultramétrique.

THÉORÈME 4.4. *Les distances d_p et d'_p sont uniformément équivalentes.*

Démonstration. Si $\binom{u}{x} \not\equiv \binom{v}{x} \bmod p$, le p-groupe qui reconnaît l'ensemble $L(x, \binom{u}{x}, p)$ sépare u et v. L'ordre de ce groupe est majoré par le nombre de matrices unitriangulaires supérieures d'ordre $|x|+1$ à coefficients dans $\mathbb{Z}/p\mathbb{Z}$, soit $p^{|x|(|x|+1)/2}$. On en déduit que

$$\delta_p(u,v) \leq \frac{1}{2}\,\delta'_p(u,v)\big(\delta'_p(u,v)+1\big).$$

Réciproquement, si u et v sont séparés par un p-groupe G d'ordre p^n, la proposition 4.3 montre que G est quotient de G_{p^n}. Il en résulte que $u \not\sim_{p^n} v$ et donc

$$\delta'_p(u,v) \leq \delta_p(u,v).$$

Par conséquent, les distances d_p et d'_p définissent la même structure uniforme. □

5. Propriétés des ouverts

Les propositions 3.1 et 4.2 fournissent une description des ouverts de A^* pour la topologie p-adique. Nous allons compléter cette description en montrant que les ouverts sont stables par diverses opérations. Soient L_0, L_1, ..., L_k des parties de A^*, a_1, a_2, ..., a_k des lettres et r et n deux entiers. On pose

$$L_0a_1L_1\cdots a_kL_k = \{u \in A^* \mid u = u_0a_1u_1\cdots a_ku_k \text{ avec } u_0 \in L_0, u_1 \in L_1, \ldots, u_k \in L_k\}$$

et on note $(L_0a_1L_1\cdots a_kL_k)_{r,n}$ l'ensemble des mots u de A^* tels que le nombre de factorisations de u de la forme $u = u_0a_1u_1\cdots a_ku_k$ avec $u_0 \in L_0$, $u_1 \in L_1$, ..., $u_k \in L_k$, soit un entier congru à r modulo n. Par exemple, on a $L(a_1a_2\cdots a_r, r, p) = (A^*a_1A^*a_2A^*\cdots a_kA^*)_{r,p}$. Le résultat qui suit explique l'intérêt de cette opération pour l'étude de la topologie p-adique.

PROPOSITION 5.1. [11]. *Soit p un nombre premier et k, r et n des entiers. Si L_0, L_1, ..., L_k sont des parties de A^* reconnues par des p-groupes, alors l'ensemble $(L_0a_1L_1\cdots a_kL_k)_{r,p^n}$ est également reconnu par un p-groupe.*

On en déduit immédiatement deux corollaires, également tirés de [11].

COROLLAIRE 5.2. *Si L_0, L_1, ..., L_k sont des ouverts de A^* pour la topologie p-adique, alors l'ensemble $(L_0a_1L_1\cdots a_kL_k)_{r,p^n}$ est également ouvert.*

Démonstration. D'après la proposition 3.1, chacun des L_i est union (finie ou infinie) de parties reconnues par des p-groupes. Comme l'opération $(L_0, L_1, \ldots, L_k) \mapsto (L_0a_1L_1\cdots a_kL_k)_{r,p^n}$ est distributive par rapport à l'union, $L = (L_0a_1L_1\cdots a_kL_k)_{r,p^n}$ est union de parties élémentaires de la forme $(L_0a_1L_1\cdots a_kL_k)_{r,p^n}$ où chacun des L_i est reconnu par un p-groupe. Mais d'après la proposition 5.1, ces parties élémentaires sont elles-mêmes reconnues par des p-groupes. Par conséquent, L est ouvert. □

COROLLAIRE 5.3. *Si L_0, L_1, ..., L_k sont des ouverts de A^* pour la topologie p-adique, et si a_1, a_2, ...a_k sont des lettres, alors l'ensemble $L_0a_1L_1\cdots a_kL_k$ est également ouvert.*

Démonstration. Cela résulte du corollaire 5.2 et de la formule

$$L_0a_1L_1\cdots a_kL_k = \bigcup_{\substack{n>0\\0<r<p^n}} (L_0a_1L_1\cdots a_kL_k)_{r,p^n}. \qquad \square$$

Sous les mêmes hypothèses, on démontre que l'ensemble $L_0L_1\cdots L_k$ est également ouvert.

Exemple 5.1. Soient a_1, a_2, ..., a_k des lettres de A. Alors l'ensemble $A^*a_1A^*a_2A^*\cdots a_kA^*$ des mots ayant $a_1a_2\cdots a_k$ comme sous-mot est un ouvert de la topologie p-adique.

Un problème qui n'est pas encore résolu à l'heure actuelle, mais qui semble en bonne voie, est la caractérisation des ouverts *reconnaissables* de la topologie p-adique de A^*. Ce qui précède montre que les parties de A^* qui sont union finies de parties de la forme

$$L_0a_1L_1\cdots a_kL_k$$

où les a_i sont des lettres et les L_i sont des parties reconnues par des p-groupes, sont des parties reconnaissables ouvertes pour la topologie p-adique. On conjecture que réciproquement, toute partie reconnaissable ouverte est de cette forme. On trouvera dans [11] une discussion approfondie de cette conjecture et de conjectures connexes.

On appelle *substitution*(*) de A^* dans B^* un morphisme de monoïde de A^* dans $\mathcal{P}(B^*)$. C'est donc une application σ de A^* dans $\mathcal{P}(B^*)$ telle que $\sigma(1) = \{1\}$ et $\sigma(a_1a_2\cdots a_n) = \sigma(a_1)\sigma(a_2)\cdots\sigma(a_n)$ pour tout mot $a_1a_2\ldots a_n$. En réalité, on considére σ comme une relation sur $A^*\times B^*$. En particulier, la notation σ^{-1} désigne la relation inverse de σ [1]. Si L est une partie de A^*, et K est une partie de B^*, on pose

$$\sigma(L) = \bigcup_{u\in L}\sigma(u), \qquad \sigma^{-1}(K) = \{u\in A^* \mid \sigma(u)\cap K\neq\emptyset\}.$$

Ainsi $\sigma^{-1}(K)$ est l'ensemble des mots de A^* qui sont en relation avec un mot de K par la relation σ^{-1}.

THÉORÈME 5.4. [11]. *Soit σ une substitution de A^* dans B^*. Si K est un ouvert de B^* pour la topologie p-adique, $\sigma^{-1}(K)$ est un ouvert de A^*.*

La démonstration de ce résultat repose sur le fait (non trivial) que si K est une partie reconnue par un p-groupe, alors $\sigma^{-1}(K)$ est union finie de parties de la forme $L_0a_1L_1\cdots a_kL_k$, où les L_i sont des parties reconnues par des p-groupes.

(*) Cette terminologie provient de la théorie des langages [1].

COROLLAIRE 5.5. *Soit φ une application surjective d'un alphabet A sur un alphabet B. Alors le morphisme de monoïde de A^* dans B^* défini par φ est une application ouverte.*

Démonstration. Il suffit d'observer que φ^{-1} est une substitution de B^* dans A^* et d'appliquer le théorème 5.4. □

Nous pouvons à présent en déduire les propriétés topologiques des applications π, δ et β de la figure 2.1, R étant muni de la topologie induite par la distance d_p.

PROPOSITION 5.6. *Relativement à la distance p-adique, on a les propriétés suivantes :*
(1) l'application $\pi : \tilde{A}^* \to F(A)$ est continue et ouverte,
(2) l'application $\delta : \tilde{A}^* \to R$ est ouverte mais non continue,
(3) l'application $\beta : R \to F(A)$ est un isomorphisme d'espaces uniformes, mais pas une isométrie.

Démonstration. (1) résulte de la proposition 2.6 et du corollaire 5.5.

6. Le mot de Prouhet-Thue-Morse et la topologie p-adique

Le mot de Prouhet-Thue-Morse est curieusement relié à la topologie p-adique. Rappelons qu'il s'agit du mot infini t obtenu à partir du mot a par itération du morphisme de monoïde $\tau : A^* \to A^*$ défini par $\tau(a) = ab$ et $\tau(b) = ba$. On a ainsi

$$\begin{aligned}
\tau(a) &= ab \\
\tau^2(a) &= abba \\
\tau^3(a) &= abbabaab \\
&\vdots \\
t = \tau^\infty(a) &= abbabaabbaababbabaababbaabbabaab\cdots
\end{aligned}$$

On notera t_n le préfixe de t de longueur n. Par exemple, $t_5 = abbab$. On

s'intéresse aux coefficients binomiaux de la forme $\binom{t_m}{x}$, dont le tableau ci-dessous donne les premières valeurs :

$x \setminus m$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
a	1	1	1	2	2	3	4	4	4	5	6	6	7	7	7	8
b	0	1	2	2	3	3	3	4	5	5	5	6	6	7	8	8
aa	0	0	0	1	1	3	6	6	6	10	15	15	21	21	21	28
ab	0	1	2	2	4	4	4	8	12	12	12	18	18	25	32	32
ba	0	0	0	2	2	5	8	8	8	13	18	18	24	24	24	32
bb	0	0	1	1	3	3	3	6	10	10	10	15	15	21	28	28

On constate expérimentalement que, pour certaines valeurs de m, les coefficients binomiaux $\binom{t_m}{x}$ du tableau sont tous pairs, et que pour $m = 12$, ils sont tous multiples de 3. L'explication est fournie par le résultat suivant.

THÉORÈME 6.1. [2]. *Pour tout nombre premier p et pour tout entier n, il existe un entier $m = f(p,n)$ tel que, pour tout mot x tel que $0 < |x| \leq n$, on ait $\binom{t_m}{x} \equiv 0 \bmod p$.*

Lorsque $p \neq 2$, on peut prendre pour f la fonction

$$f(p,n) = 2^n p^{1+\lfloor \log_p n \rfloor}$$

où $\lfloor x \rfloor$ désigne le plus grand entier inférieur ou égal à x. Pour $p = 2$, on peut prendre $f(p,n) = 2^k$ si $F_{k-1} \leq n < F_k$, où F_k est le k-ième élément de la suite de Fibonacci définie par $F_0 = 1$, $F_1 = 1$ et $F_{n+2} = F_{n+1} + F_n$ pour tout $n \geq 0$. Les premières valeurs de $f(p,n)$ sont données dans le tableau ci-dessous.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
$f(2,n)$	4	8	16	16	32	32	32	64	64	64	64	64	128
$f(3,n)$	6	12	72	144	288	576	1152	2304	13824	27648	55296	110592	221184
$f(5,n)$	10	20	40	80	800	1600	3200	6400	12800	128000	256000	512000	1024000

Ces propriétés combinatoires conduisent au résultat suivant.

THÉORÈME 6.2. [2]. *Pour tout nombre premier p, il existe une sous-suite de la suite $(t_n)_{n\geq 0}$ qui converge vers 1 dans la topologie p-adique.*

Démonstration. D'après le théorème 4.4, une suite u_n converge vers le mot vide si et seulement si $\delta'_p(u_n, 1)$ tend vers l'infini, c'est-à-dire si et seulement si, pour tout k, il existe un entier n_k tel que, pour tout $n \geq n_0$ et pour tout mot x tel que $0 < |x| < k$, on ait $\binom{u_n}{x} \equiv 0 \bmod p$. Le théorème 6.1 montre alors que la suite $(t_{f(p,n)})_{n\geq 0}$ converge vers 1 dans la topologie p-adique. □

Le théorème 6.2 fournit des exemples de suites convergentes pour la topologie p-adique qui ne sont pas conséquences du corollaire 2.4. On sait en effet que la suite de Prouhet-Thue-Morse ne contient aucun facteur de la forme x^3 (avec x mot non vide). Elle ne contient donc a fortiori aucun facteur de la forme x^{p^n} pour $n > 1$ et le corollaire 2.4 ne peut être utilisé.

BIBLIOGRAPHIE

[1] J. Berstel, *Transductions and Context-free Languages*, Teubner, Stuttgart (1979).

[2] J. Berstel, M. Crochemore et J.-E. Pin, *Thue-Morse sequence and p-adic topology for the free monoid*, Discrete Math. **76** (1989), 89–94.

[3] N. Bourbaki, *Eléments de Mathématique, Topologie générale*, chapitres I à IV (1971).

[4] J. D. Dixon, M. P. F. Sautoy, A. Mann et D. Segal *Analytic pro-p groups*, London Math. Society Lecture Note Series **157**, Cambridge University Press, Cambridge, Grande-Bretagne (1991).

[5] S. Eilenberg, *Automata, languages and machines*, Academic Press, New York, Vol. A (1974), Vol. B (1976).

[6] M. D. Fried and M. Jarden, *Field arithmetic*, Springer, Berlin (1986).

[7] M. Hall Jr, *A topology for free groups and related groups*, Ann. Math. **52** (1950), 127–139.

[8] M. Lothaire, *Combinatorics on Words*, Encyclopedia of Mathematics **17**, Addison Wesley, New-York (1983).

[9] P. Ochsenschlager, *Binomialkoeffizenten und Shuffle-Zahlen*, Technischer Bericht, Fachbereicht Informatik, T. H. Darmstadt (1981).

[10] J.-E. Pin, *Finite group topology and p-adic topology for free monoids*, 12th ICALP, Lecture Notes in Computer Science **194** (1985), 445–455.

[11] J.-E. Pin, *Topologies for the free monoids*, Journal of Algebra **137** (1991), 297–337.

[12] M. E. Prouhet, *Mémoire sur quelques relations entre les puissances des nombres*, C. R. Acad. Sc. **33**, N 8 (1851), 225.

[13] C. Reutenauer, *Une topologie du monoïde libre*, Semigroup Forum **18** (1979), 33–49.

[14] C. Reutenauer, *Sur mon article « Une topologie du monoïde libre »*, Semigroup Forum **22** (1981), 93–95.

J.-E. Pin
LITP-IBP, Université Paris VI et CNRS, Tour 55-65
4 Place Jussieu
F-75252 Paris Cedex 05
France
e-mail : pin@litp.ibp.fr