Morphismes de Lyndon

G. Richomme

29 avril 2002

Résumé

Nous donnons une caractérisation des morphismes qui préservent les mots de Lyndon finis (morphismes que nous appelons morphismes de Lyndon) ainsi que des morphismes qui préservent l'ordre lexicographique sur les mots finis. Nous montrons que les monoïdes formés de ces morphismes ne sont pas finiment engendrés. A titre d'exemple d'utilisation des morphismes de Lyndon, nous les employons pour obtenir les décompositions de deux mots sturmiens. Nous terminons en caractérisant les morphismes sturmiens qui sont des morphismes de Lyndon et ceux qui préservent l'ordre lexicographique.

Soit P une propriété des mots finis. Pour générer des mots arbitrairement grands ayant la propriété P (voire des mots infinis), un moyen commode consiste à utiliser des morphismes. De tels morphismes existent par exemple pour les mots primitifs [7], pour les mots sans carré ou sans chevauchement [13], pour les mots sturmiens (voir [5] pour une présentation générale), pour les mots équilibrés [8]. Bien que cela ne soit pas nécessaire, il suffit de considérer des morphismes qui préservent les mots ayant la propriété P (et qui accroissent la longueur des mots). Ici nous considérons les mots de Lyndon.

Un mot (fini) de Lyndon est un mot qui est plus petit (dans l'ordre lexicographique) que tous ces suffixes propres non vides (voir [4] pour une présentation de ces mots). Un mot infini de Lyndon est un mot ayant une infinité de préfixes qui sont des mots de Lyndon. Le théorème de factorisation de Lyndon [2] établit que tout mot fini peut se décomposer de manière unique en un produit décroissant (pour l'ordre lexicographique) de mots de Lyndon. Ce théorème a été étendu pour les mots infinis par Siromoney et al. [12]. Pour obtenir la décomposition en mots de Lyndon du mot infini appelé mot de Thue-Morse, les auteurs de [3] montrent que certains morphismes préservent les mots de Lyndon. Une autre utilisation de tels morphismes peut être trouvée dans [5, chapitre 4]. Notre but est d'étudier ces morphismes.

On suppose le lecteur familier avec les notions de base des langages formels et nous suivons les notations de [5]. Nous ne considérons dans la suite que des alphabets finis d'au moins deux lettres (sur un alphabet $\{a\}$ constitué d'une seule lettre, seuls le mot vide ε et a sont des mots de Lyndon). Nous supposons les alphabets considérés totalement ordonnés et nous notons $\prec$ l'ordre lexicographique sur les mots.

Etant donnés deux alphabets A et B, un morphisme f de A^* vers B^* (si B n'a pas d'importance, on dira morphisme sur A) est une application de A^* vers B^* telle que $f(uv) =$

$f(u)f(v)$ pour tous mots u et v sur A. Un morphisme est entièrement défini par l'image des lettres. Un endomorphisme sur A est un morphisme de A^* vers A^*.

Nous appelons *morphisme de Lyndon* tout morphisme qui préserve les mots de Lyndon. Plus précisément, si f est un morphisme sur A, alors f est un morphisme de Lyndon si $f(w)$ est un mot de Lyndon pour tout mot de Lyndon w sur A. Une condition nécessaire pour qu'un morphisme sur A soit de Lyndon est qu'il *préserve l'ordre lexicographique* sur les mots finis, i.e., pour tous mots u, v dans A^* avec $u \prec v$, on a $f(u) \prec f(v)$.

Pour n un entier, notons A_n un alphabet ordonné à n lettres : $A_n = \{a_1, \ldots, a_n\}$ avec $a_1 \prec a_2 \prec \ldots \prec a_n$. La proposition suivante fournit une caractérisation testable des morphismes qui préserve l'ordre lexicographique.

Proposition 1 *Un morphisme f sur A_n préserve l'ordre lexicographique si et seulement si pour tout entier i avec $1 \leq i < n$, $f(a_i a_n^{k_i}) \prec f(a_{i+1})$ où k_i est le plus petit entier tel que $|f(a_i a_n^{k_i})| \geq |f(a_{i+1})|$.*

Les couples de mots $(a_i a_n^{k_i}, a_{i+1})$ impliqués dans cette proposition dépendent de f. On peut se demander si cette dépendance est nécessaire. Nous donnons une famille de morphismes qui montre que oui. Toutefois, certaines situations ne la nécessitent pas comme le montre les corollaires suivants de la proposition 1.

Corollaire 2 *Un morphisme sur $\{a, b\}$ avec $a \prec b$ préserve l'ordre lexicographique si et seulement si $f(ab) \prec f(b)$.*

Corollaire 3 *Un morphisme* uniforme *(resp.* préfixe*) f sur A_n préserve l'ordre lexicographique si et seulement si $f(a_i) \prec f(a_{i+1})$ pour tout entier i avec $1 \leq i < n$.*

Dans [11], nous pouvons noter l'utilisation d'un morphisme qui préserve l'ordre lexicographique sans être un morphisme de Lyndon. Le théorème suivant fournit une caractérisation des morphismes non effaçant de Lyndon (un morphisme effaçant est de Lyndon si et seulement s'il est le morphisme vide, c'est-à-dire, que l'image de n'importe quel mot est le mot vide).

Théorème 4 *Un morphisme non-effaçant f défini sur un alphabet ordonné A est un morphisme de Lyndon si et seulement si*

1. *pour toute lettre a de A, $f(a)$ est un mot de lyndon, et,*
2. *f est un morphisme qui préserve l'ordre lexicographique sur les mots finis.*

La proposition 1 et le théorème précédent fournissent un moyen pour tester si un morphisme est de Lyndon. On peut chercher à utiliser une autre méthode basée sur la notion d'ensemble fini de test (Cette notion d'ensemble de test existe par exemple pour les morphismes sans-chevauchement [9] et les morphismes sans puissance k [10, 14]). Un *ensemble de test* pour les morphismes de Lyndon est un ensemble S de mots de Lyndon tel qu'un morphisme est de Lyndon si et seulement si tous les mots de $f(S)$ sont des mots de Lyndon. En conséquence des résultats sur les morphismes qui préservent l'ordre, on peut voir que :

– Pour $n \geq 3$, il n'existe pas d'ensemble fini de test pour les morphismes de Lyndon sur A_n.

– L'ensemble $\{a, b, ab\}$ est un ensemble de test pour les morphismes non effaçants de Lyndon définis sur $\{a, b\}$ avec $a \prec b$.
– L'ensemble $A_n \cup \{a_i a_{i+1} \mid 1 \leq i < n\}$ est un ensemble de tests pour les morphismes uniformes de Lyndon et pour les morphismes préfixes de Lyndon.

Etant donné un alphabet ordonné A, l'ensemble des endomorphismes de Lyndon sur A (morphismes de A^* vers A^*) est un monoïde. Quand un monoïde d'endomorphismes est finiment engendré, on a un moyen naturel pour déterminer si un endomorphisme appartient à ce monoïde en essayant de le décomposer sur la base du monoïde. C'est le cas pour les endomorphismes binaires sans-chevauchement [13] et pour les morphismes sturmiens [5, chapitre 2]. Nous montrons :

Proposition 5 *Etant donné un alphabet ordonné A, le monoïde des endomorphismes (uniformes) de Lyndon sur A et le monoïde des endomorphismes (uniformes) qui préservent l'ordre lexicographique sur A sont des monoïdes non finiment engendrés.*

La preuve de cette proposition consiste à fournir une famille infinie de morphismes uniformes de Lyndon (et donc préservant l'ordre lexicographique) qui ne peuvent pas s'écrire gh avec g, h deux morphismes différents de l'identité (la composition des morphismes est notée par juxtaposition).

En dernière partie, nous étudions des liens entre les morphismes de Lyndon et les morphismes sturmiens (morphismes sur l'alphabet $\{a, b\}$ qui préservent les mots infinis sturmiens, voir [5, chapitre 2] pour plus d'informations sur ces morphismes). Rappelons qu'un morphisme f est sturmien si et seulement si $f \in St = \{\varphi, \tilde{\varphi}, E\}^*$ où φ, $\tilde{\varphi}$ et E sont les morphismes définis par $\varphi(a) = ab$, $\varphi(b) = a$, $\tilde{\varphi}(a) = ba$, $\tilde{\varphi}(b) = a$, $E(a) = b$, $E(b) = a$.

Tout d'abord, comme exemple d'utilisation des morphismes de Lyndon, on obtient une nouvelle preuve de la factorisation en mots de Lyndon du mot $\varphi^\omega(a)$ [6], et du fait que le mot infini $a\varphi^\omega(a)$ est un mot de Lyndon infini [1] ($\varphi^\omega(a)$ est le mot infini obtenu en appliquant infiniment le morphisme φ à partir de a).

Nous montrons ensuite :

Proposition 6 *Les morphismes sturmiens qui préservent l'ordre lexicographique sont les éléments de $\{E\varphi, \varphi E, E\tilde{\varphi}, \tilde{\varphi}E\}^* \setminus (St\ \varphi(\varphi E)^*))$.*

Par exemple, le morphisme $E\varphi E\tilde{\varphi}$ ne préserve pas l'ordre car $\varphi E\tilde{\varphi} = \tilde{\varphi}E\varphi$.

Proposition 7 *Les morphismes sturmiens qui sont des morphismes de Lyndon sont les éléments de $\{\varphi E, E\tilde{\varphi}\}^*$.*

Ces deux propositions peuvent être rapprochées du fait que $St = \{E\varphi, \varphi E, E\tilde{\varphi}, \tilde{\varphi}E, E\}^*$. Par ailleurs elles admettent en corollaire : un morphisme standard (i.e. élément de $\{E, \varphi\}^*$) préserve l'ordre si et seulement si il est un morphisme de Lyndon si et seulement si il appartient à $(E\varphi)^*$.

Références

[1] J.P. Borel and F. Laubie. Quelques mots sur la droite projective réelle. *Journal de Théorie des Nombres de Bordeaux*, 5:23–51, 1993.

[2] K.T. Chen, R.H. Fox, and R.C. Lyndon. Free differential calculus IV – the quotient groupes of the lower central series. *Ann. Math. 68*, 68:81–95, 1958.

[3] A. Ido and G. Melançon. Lyndon factorization of the Thue-Morse word and its relatives. *Discret Math. and Theoret. Comput. Sci.*, 1:43–52, 1997.

[4] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, 1983. reprinted in 1997 by Cambridge University Press in the Cambridge Mathematical Library.

[5] M. Lothaire. Algebraic combinatorics on words. *Cambridge University Press*, To appear. See http://www-igm.univ-mlv.fr/~berstel.

[6] G. Melançon. Viennot factorization of infinite words. *Information Processing Letters*, 60:53–57, 1996.

[7] V. Mitrana. Primitive morphisms. *Information Processing Letters*, 64:277–281, 1997.

[8] G. Richomme. Test-words for Sturmian morphisms. *Bulletin of the Belgian Mathematical Society*, 6:481–489, 1999.

[9] G. Richomme and F. Wlazinski. Overlap-free morphisms and finite test-sets. Technical Report 2002-05, LaRIA, Avril 2002.

[10] G. Richomme and F. Wlazinski. Some results on k-power-free morphisms. *Theoretical Computer Science*, 273:119–142, 2002.

[11] P. Séébold. Lyndon factorization of the Prouhet words. Technical report, LaRIA, March 2002.

[12] R. Siromoney, L. Mahtew, V.R. Dare, and K.G. Subramanian. Infinite Lyndon words. *Information Processing Letters*, 50:101–104, 1994.

[13] A. Thue. Uber die gegenseitige Lage gleigher Teile gewisser Zeichenreihen. *Kristiania Videnskapsselskapets Skrifter Klasse I. Mat.-naturv*, 1:1–67, 1912.

[14] F. Wlazinski. A test-set for k-power-free binary morphisms. Technical Report 2001-05, LaRIA, July 2001. To appear in Theoretical Informatics and Applications.