

Récritures suffixes de mots

Didier Caucal

▶ To cite this version:

Didier Caucal. Récritures suffixes de mots. [Research Report] RR-0871, 1988. <inria-00075683>

HAL Id: inria-00075683 https://hal.inria.fr/inria-00075683

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNITÉ DE RECHERCHE INRIA-RENNES

> Institut National de Recherche en Informatique et en Automatique

Domaine de Voluceau Rocquencourt B.P. 105 78153 Le Chesnay Cedex France Tél. (1) 39 63 55 11

Rapports de Recherche

N°871

RECRITURES SUFFIXES DE MOTS

Didier CAUCAL

JUILLET 1988





INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu 35042-RENNES CÉDEX **FRANCE**

Téléphone: 99 36 20 00 Télex: UNIRISA 950 473 F Télécopie: 99 38 38 32

> Publication Interne n° 413 Juin 1988 12 Pages

Récritures suffixes de mots

Didier CAUCAL

IRISA , Campus de Beaulieu , F35042 Rennes-Cédex , France

Résumé. Etant donné un alphabet X et une relation R finie binaire sur X^{\star} , le plus petit pré-ordre contenant R et compatible à gauche avec la concaténation de tout mot, est une relation rationnelle dont un transducteur est effectivement constructible à partir de R.

Rewriting right factors

Abstract. Given R, a finite binary relation over X^* , the least preorder relation extending R and preserved by left product is rational, and we give an effective construction for the corresponding transducer.

Introduction

Un mot u sur X se récrit par suffixe [resp. préfixe] en un mot v selon une relation R binaire sur X^* s'il existe une suite finie u_0,\ldots,u_n telle que $u=u_0$, $v=u_n$ et pour tout entier i compris entre 1 et n, $u_{i-1}=w_ig_i$, $u_i=w_id_i$ [resp. $u_{i-1}=g_iw_i$, $u_i=d_iw_i$] et g_i R d_i . Aussi toute récriture selon un système (fini) R de couples de mots sur X correspond à la récriture suffixe selon le système (rationnel) { $(gu,du) \mid g \ R \ d$ et $u \in X^*$ }. Les récritures rationnelles de [Ch 82] sont des récritures préfixes de systèmes particuliers. Les récritures suffixes de systèmes finis interviennent notamment pour l'étude des centres des langages algébriques [Bo-Ni 84], et s'étendent aux récritures de termes de base (sans variable). Concernant les systèmes finis de récritures suffixes de mots, on a les résultats suivants :

- a) l'ensemble des mots ayant une dérivation infinie est régulier et effectif (au sens où le langage est reconnu par un automate fini, constructible effectivement à partir du système) [Bo-Ni 84]
- b) le problème des mots est décidable, comme cas particulier de l'égalité de deux termes modulo un système fini de termes de base [Ne-Op 80]
- c) les propriétés de confluence et de nœthériennité sont décidables [Da-... 87] et [Hu-La 78]
- d) le langage des configurations (états concaténés avec mots de la pile) accessibles d'un automate à pile est rationnel et effectif [Au 87] (problème 14).

Ces propriétés se déduisent de celle établie dans cet article, à savoir : la récriture suffixe de tout système fini est une transduction rationnelle dont un transducteur s'obtient effectivement à partir du système.



1. Préliminaires

Pour toute relation $R \subseteq ExF$, on pourra noter x R y au lieu de $(x,y) \in R$, et l'inverse de R est $R^{-1} = \{ (y,x) \mid x R y \}$. L'image R(A) d'un ensemble A par R est $\{ y \mid \exists \ x \in A \ , x R y \}$. On note $Dom(R) = R^{-1}(F)$ le domaine de R et Im(R) = R(E) l'image de R. La relation identité sur R est notée R est

Dans un monoïde (M,.) d'élément neutre 1, le produit d'une partie A par une partie B de M est $A.B = \{uv \mid u \in A \text{ et } v \in B\}$ et l'étoile d'une partie A est $A^* = U \{A^i \mid i \geq 0\}$ avec $A^0 = \{1\}$ et $A^{i+1} = A^i.A$. Dans le cas d'un ensemble réduit à un élément a et pout tout entier a, on a a and a are a

$$_{B} \mid A = \{ u \mid A \cap Bu \neq \emptyset \} \text{ et } A \mid_{B} = \{ u \mid A \cap uB \neq \emptyset \}.$$

Considérons le monoïde $(2^{E\times E}, \circ)$ des parties de $E\times E$, muni du produit de composition \circ , défini par $R\circ S = \{ (u,v) \mid \exists \ w\in E \ , \ u\ R\ w\ et\ w\ S\ v \} \ .$

Une relation R est un pré-ordre si elle est réflexive ($\iota_E \subseteq R$) et est transitive ($R \circ R \subseteq R$).

Soient X un alphabet et $(X^*, .)$ le monoïde libre engendré par X. La longueur de tout mot u est notée |u|. On considère le monoïde produit $(X^*xX^*, .)$ défini par

$$(u,v) \cdot (x,y) = (ux,vy)$$
.

Ainsi pour R , S \subseteq X*xX* , R.S = { (ux,vy) | u R v et x S y } est le produit composante à composante, qu'il faut distinguer de R \circ S . De même, l'étoile de R est R* = U { R^i | i \geq 0 } avec R^0 = {(1,1)} et R^{i+1} = R^i.R , qu'il faut distinguer de l'étoile R(*) de {R} pour le monoïde $(2^{X*xX*}, \circ)$, c'est-à-dire R(*) = U { R(i) | i \geq 0 } avec R(0) = ι_{X*} noté Δ et R(i+1) = R(i) \circ R . La relation R est nœthérienne s'il n'existe pas de suite infinie $(u_n)_{n\geq0}$ telle que u_n R u_{n+1} . La relation R est confluente si $(R^{-1})^{(*)} \circ R^{(*)} \subseteq R^{(*)} \circ (R^{-1})^{(*)}$. On rappelle [Be 79] que Rat(X*xX*) est fermée pour \circ , et pour R \in Rat(X*xX*) et A \in Rat(X*) , R(A) \in Rat(X*). De plus Rat(X*xX*) contient la famille Rec(X*xX*) des parties reconnaissables de X*xX*, c'est-à-dire (d'après le théorème de Mezei) l'ensemble des unions finies de produits de parties rationnelles de X*. Une partie R de X*xX* est compatible à gauche (avec la concaténation de tout mot) si Δ .R = R .

On pose

$$\frac{1}{R} = \Delta R$$
 un pas de récriture suffixe selon R $\frac{1}{R} = (\frac{1}{R})^{(*)}$ la récriture suffixe selon R, c'est-à-dire le plus petit pré-ordre contenant R et compatible à gauche.

Exemple 1: [Ch 82] posons
$$X = \{a,b\}$$
 et $R = (ab,a^2b^2).(b,b)^* \in Rat(X^*xX^*)$, alors $\xrightarrow{*}_{R} (ab) = \{a^nb^n \mid n \ge 1\}$ donc $\xrightarrow{*}_{R} \notin Rat(X^*xX^*)$.

Exemple 2: [Ne-Op 80] posons
$$X = \{a,b\}$$
 et $R = S \cup S^{-1}$ avec $S = \{(a^3b,b), (a^5b,b)\}$, alors $\xrightarrow{*}_{R} = \{(a,a),(b,b)\}^* \cdot ((a,1)^* \cdot (1,a)^* \cdot (b,b) \cup \{(1,1)\}) \in Rat(X^* \times X^*)$.

2. Automate d'une récriture suffixe selon un système fini

Dorénavant, R est un ensemble fini de couples de mots sur un alphabet X et S est l'ensemble $Dom(R) \cup \{1\}$. Comme indiqué à l'introduction, on va établir que $\frac{\star}{R}$ \in Rat(X*xX*) et qu'à partir de R, on peut construire un transducteur réalisant $\frac{\star}{R}$. Préalablement, on construit un automate fini G(R) sur X^* , d'ensemble des états S, d'unique état terminal 1, et tel que le langage reconnu par G(R) à partir de tout état s soit égal à $\frac{\star}{R}$ (s).

Considérons le monoïde $(2^{SxX*xS}, \circ)$ des graphes à arcs étiquetés sur X^* et d'ensemble de sommets S, muni du produit de composition \circ défini par

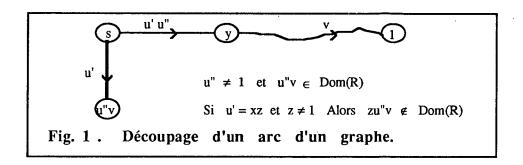
$$A \circ B = \{ (s,uv,t) \mid \exists r, (s,u,r) \in A \text{ et } (r,v,t) \in B \}.$$

Son élément neutre, noté 1_R , est $\{(s,1,s) \mid s \in S\}$. On définit $G^* = U\{G^i \mid i \geq 0\} = U\{G\}^*$ le graphe infini de G, d'ensemble de sommets S et d'ensemble d'arcs les chemins finis de G. Pour tout graphe \underline{fini} A, on définit le découpage A de A par R comme suit :

$$<\!\!A\!\!>\ =\ \{\ (s,u',u''v)\not\in A^*\mid (s,u,y)\in A\ ,\ (y,v,1)\in A^*\ ,\ u''\ \ \text{plus grand suffixe de}\ u\ \ \text{tel que}$$

$$u''\ne 1\ \ \text{et}\ \ u''v\in Dom(R)\ ,\ \ u'u''=u\ \}$$

La construction de <A> par découpage des arcs de A est illustrée à la figure 1.



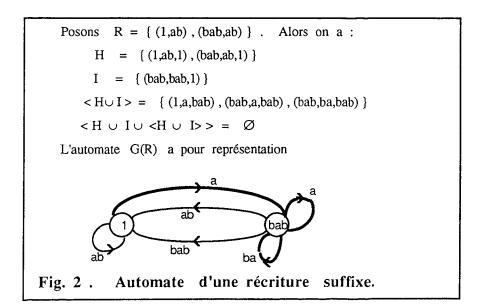
On définit récursivement le complété \overline{A} de A (fini) par R comme suit : si $\langle A \rangle = \emptyset$ alors $\overline{A} = A$ sinon $\overline{A} = \overline{A \cup \langle A \rangle}$. Le complété \overline{A} de A existe et est fini parce que Dom(R) est fini et qu'il n'existe qu'un nombre fini de préfixes d'arcs. On note Suff(u) le plus grand suffixe de u élément de S, et Pref(u) le mot de X^* tel que Pref(u).Suff(u) = u. On définit l'automate G(R) en complétant un graphe de base. Ce graphe de base est l'union d'un graphe H_R des règles de R, et d'un graphe I_R de découpage des mots du domaine de R:

$$G(R) = \overline{H_R \cup I_R} \quad \text{avec}$$

$$H_R = \{ (x, \text{Pref}(y), \text{Suff}(y)) \mid x (R-\Delta) y \}$$

$$I_R = \{ (ay, a. \text{Pref}(y), \text{Suff}(y)) \mid ay \in \text{Dom}(R) \text{ et } a \in X \}$$

On remarquera que $\, 1_R \cap (H_R \cup I_R) = \varnothing \,$. La figure 2 donne un exemple d'un tel automate .



Le graphe G(R) est fini parce que H_R et I_R le sont. On va établir (proposition 1) que l'ensemble des mots récrits par suffixes selon R à partir du mot u de X^* est égal à l'ensemble des mots étiquetant les chemins de $G(R \cup \{(u,u)\})$ allant de u à 1. Pour se faire, on établit deux résultats préalables. Montrons I_R^* est le graphe de découpage de $S = Dom(R) \cup \{1\}$ par suffixe de S.

Lemme 1.
$$I_R^* = \{ (uv, u, v) \mid uv, v \in S \}$$
.

Preuve. \subseteq : $I_R^0 = 1_R = \{(v,1,v) \mid v \in S \}$. De plus, A.B vérifie l'inclusion si A et B la vérifient. Aussi par récurrence sur $n \ge 0$, l'inclusion est satisfaite pour le graphe I_R^n ; et par union, l'inclusion directe est vérifiée.

 \supseteq : Montrons pour tout uv et v de S, $(uv,u,v) \in I_R^*$. Si u=1 alors $(uv,u,v)=(v,1,v) \in I_R^0$. Si $u \neq 1$ alors l'appartenance se montre par récurrence sur la cardinalité de $\{xv \in S \mid x \text{ suffixe de } u \text{ et } x \neq u \}$.

On dit qu'un graphe A est S-complet si tout chemin de A se découpe par suffixe de S en un chemin de A, i.e.

pour tout $(x,u,y) \in A^*$ et pour vz = uy avec $z \in S$, on a $(x,v,z) \in A^*$ Montrons que le complété d'un graphe contenant I_R est S-complet.

Lemme 2. Si $<A>=\emptyset$ et $I_R\subseteq A$ alors A est S-complet.

Preuve. Considérons un graphe A contenant I_R et tel que $\langle A \rangle = \emptyset$, et soient $(x,u,y) \in A^*$, vz = uy et $z \in S$. On distingue les 2 cas suivants :

<u>Cas 1</u>: $|v| \ge |u|$ i.e. v = uw et wz = y. Par le lemme 1, $(y,w,z) \in I_R^*$; d'où $(x,v,z) = (x,uw,z) \in A^* \circ I_R^* \subseteq A^*$

<u>Cas 2</u>: |v| < |u| i.e. vw = u, z = wy, w ≠ 1. Soit w" le plus grand suffixe de u = vw tel que w"y ∈ S. Aussi w est suffixe de w", et on note w" = sw. Comme w"y, wy ∈ S, on a par le lemme 1 (w"y,s,wy) ∈ I_R^* . Comme (x,u,y) ∈ A^* et u ≠ 1, \exists (x,u₁,y₁), (y₂,u₃,y) ∈ A^* , (y₁,u₂,y₂) ∈ A tels que u₁u₂u₃ = u et |u₃| < |w"| ≤ |u₂u₃|. Mais <A> = Ø et par le lemme 1 (y,y,1) ∈ I_R^* ⊆ A^* , donc par définitions de w" et <A>, (y₁,w',w"y) ∈ A^* pour

$$w'w''=u_2u_3\;.\;\;\mathrm{D'où}\;\;(x,v,z)=(x,u_1w's,z)\;\in\;\;\mathrm{A}^*\circ\mathrm{A}^*\circ\mathrm{I_R}^*\;\subseteq\;\mathrm{A}^*\;\;.$$

Dans les deux cas, $(x,v,z) \in A^*$. Ainsi A est S-complet.

Le langage accepté par l'automate fini $G(R \cup \{(u,u)\})$ à partir de l'état $u \in X^*$ et d'état final 1 est égal à $\xrightarrow{*}_R (u)$

Proposition 1.
$$\forall u \in X^*$$
, $\xrightarrow{*}_R (u) = \{ v \mid (u,v,1) \in G(R \cup \{(u,u)\})^* \}$.

Preuve de la proposition 1 : comme $\frac{1}{R} = \frac{1}{S}$ pour $S = R \cup \{(u,u)\}$ et $u \in X^*$, il revient au même de montrer que

$$\forall u \in Dom(R)$$
, $\xrightarrow{*}_{R} (u) = \{ v \mid (u,v,1) \in G(R)^* \}$

 \subseteq : il suffit d'établir par récurrence sur $n \ge 0$ et pour $u \in Dom(R)$ que

$$\left(\begin{array}{c} \longrightarrow \end{array}\right)^{(n)} (u) \subseteq \left\{\begin{array}{ccc} v \mid (u,v,1) \in G(R)^* \end{array}\right\}$$

 $\underline{n=0}$: $(-\frac{1}{R})^{(0)}$ (u) = {u} et par le lemme 1, (u,u,1) $\in I_R^* \subseteq G(R)^*$

 $n \Rightarrow n+1$: $u \xrightarrow{R}^{(n)} x \xrightarrow{R} y$. Par hypothèse de récurrence, $(u,x,1) \in G(R)^*$. Par définition de \xrightarrow{R} , on a x = ws, y = wt et s R t. On considère les deux cas suivants :

Cas 1: s = t alors x = y donc $(u,y,1) \in G(R)^*$.

Cas $2: s \neq t$. Comme $(u,x,1) \in G(R)^*$, x = ws et par le lemme 2, G(R) est S-complet, d'où $(u,w,s) \in G(R)^*$. De plus $s(R-\Delta)t$ donc $(s,t_1,t_2) \in H_R \subseteq G(R)$ avec $t_1t_2 = t$. Par le lemme 1, $(t_2,t_2,1) \in I_R^* \subseteq G(R)^*$, d'où $(u,y,1) = (u,wt_1t_2,1) \in G(R)^*$.

La récurrence est terminée ainsi que la preuve de l'inclusion directe.

⊇ : on considère l'ensemble E des graphes suivants :

 $A \in E$ si et seulement si pour tout $(x,u,y) \in A$, $x \xrightarrow{R} uy$.

Soit A un graphe de E. Pour tout entier n, $A^n \in E$ (par récurrence sur n); d'où $A^* \in E$,

donc <A> et \overline{A} sont des graphes de E. Ainsi, si $A \in E$ alors $\overline{A}^* \in E$.

Par définitions, H_R et I_R appartiennent à E et comme $G(R) = \overline{H_R \cup I_R}$, le graphe $G(R)^*$ est élément de E; en particulier, pour $(u,v,1) \in G(R)^*$, on a $u \xrightarrow{R} v$.

L'inclusion inverse est démontrée, ce qui termine la preuve de la proposition 1.

Exemple 3: pour $X = \{a,b\}$ et $R = \{(1,ab), (bab,ab)\}$ et par application de la proposition 1 au graphe G(R) de la figure 2, $\frac{*}{R}$ $(1) = (a^+b)^*$.

De la construction de G(R) et de la proposition 1, on déduit

$$\forall u \in X^*$$
, $\frac{\star}{R}$ (u) $\in Rat(X^*)$ et est effectif; d'où $\frac{\star}{R}$ est décidable.

Le cœur de R [Bo-Ni 84] est l'ensemble $C(R) = \{ u \in X^* \mid \forall n, \xrightarrow{R} {}^{(n)}(u) \neq \emptyset \}$. Aussi

$$C(R) = X^* \cdot (\xrightarrow{R} (Dom(R) \cap C(R)))$$
.

A partir de G(R) on détermine l'ensemble fini $A = Dom(R) \cap C(R)$, et à partir de $G(R^{-1} \cup \iota_A)$ on déduit que C(R) est rationnel et effectif. De plus \xrightarrow{R} est nœthérienne si et seulement si $Dom(R) \cap C(R) = \emptyset$. Ainsi la nœthériennité de tout système fini de récritures suffixes de mots est décidable.

3. Transducteur d'une récriture suffixe selon un système fini

La construction effective de G(R) va nous permettre d'extraire un transducteur réalisant $\frac{*}{R}$.

Cette extraction s'appuie sur le lemme ci-dessous donnant une détermination de $\frac{*}{R}$ en fonction de la récriture directe et de la récriture inverse de $S = Dom(R) \cup \{1\}$.

Lemme 3.
$$u \xrightarrow{\kappa} v \Leftrightarrow (u = wx, v = wy, x \xrightarrow{\kappa} z \xrightarrow{\kappa} y \land z \in S).$$

Preuve. La condition suffisante résulte de la compatibilité à gauche de $\xrightarrow{*}_{R}$. Montrons la condition nécessaire par récurrence sur n pour $u \xrightarrow{R}^{(n)} v$.

$$\underline{n=0}$$
: $\underline{u} \xrightarrow{R} {0 \choose v}$ i.e. $\underline{u} = v$ donc $\underline{w} = \underline{u}$ et $\underline{x} = \underline{y} = \underline{z} = 1$ conviennent.

$$n \Rightarrow n+1$$
: supposons $u \xrightarrow{R} ^{(n)} v \xrightarrow{R} w$. Par hypothèse de récurrence, on a $u = sx$, $v = sy$ et $x \xrightarrow{*} r \xrightarrow{*} y$ avec $r \in Dom(R) \cup \{1\}$

Comme $v \longrightarrow_R w$, on a v = tp, w = tq et p R q. Aussi sy = tp et on considère les deux cas suivants :

 $Cas\ où\ |t| \ge |s|$: alors t = sh et y = hp. Aussi u = sx, w = shq et $x \xrightarrow{*}_{R} r \xrightarrow{*}_{R} hq$.

Cas où |t| < |s|: alors s = th et p = hy. Aussi u = thx, w = tq et $hx \xrightarrow{*} hy = p \xrightarrow{R} q$.

La récurrence est terminée, d'où la condition nécessaire.

On peut maintenant établir le résultat annoncé à l'introduction.

Théorème. Soit R une relation finie et binaire sur X^* . La récriture suffixe \xrightarrow{R} selon R est une partie rationnelle de X^*xX^* , et on peut de façon effective définir à partir de R, un transducteur réalisant \xrightarrow{R} .

Preuve: Notons $T = U\{(\frac{x}{R}, x) \mid (z)\} \times (\frac{x}{R}, x) \mid z \in S\}$.

D'après la proposition 1 , T est une union finie de produits de langages rationnels sur X, i.e. T est une partie reconnaissable donc rationnelle de X^*xX^* . De plus, de la construction effective de G(R) et $G(R^{-1} \cup \iota_S)$, on définit effectivement un transducteur rationnel réalisant T. Enfin, le lemme 3 signifie que

$$\frac{\bullet}{\mathsf{R}}\mathsf{H} = \Delta.\mathsf{T} \tag{1}$$

d'où le théorème.

De l'égalité (1), on déduit que la confluence de \xrightarrow{R} est décidable. En effet, considérons l'ensemble $\Re = \{ \Delta.T \mid T \in \operatorname{Rec}(X^*xX^*) \}$. Cette famille de relations rationnelles sur X^* est fermée par union et par composition, parce que

i) la composition et le produit sont distributifs (à gauche et à droite) par rapport à l'union

ii)
$$\Delta.(A \times B) \circ \Delta.(C \times D) = \Delta.(A \times (B \mid_{C}).D) \cup \Delta.((C \mid_{B}).A \times D)$$

Pour toute relation R binaire sur X^* , on définit l'ensemble [R] des paires de mots obtenues par suppression du plus grand préfixe commun de chaque paire de R, i.e.

$$[R] = \int_{\Lambda} |R - \iota_{X}(X^* x X^*).$$

Donc [R] est reconnaissable si R l'est. De plus, on a Δ .[R] = Δ .R et $[\Delta$.R] = [R], d'où Δ .R $\subseteq \Delta$.S si et seulement si $[R] \subseteq [S]$.

Comme l'inclusion est décidable sur $\operatorname{Rec}(X^*xX^*)$, il en est de même pour l'inclusion sur \Re , et la confluence de tout système fini de récritures suffixes de mots est décidable.

Conclusion

On conclut par une question et une invitation. La question est de savoir si la récriture suffixe selon un système reconnaissable reste une transduction rationnelle et, dans l'affirmative, peut-on construire un transducteur de la récriture à partir d'un automate du système ? L'invitation est d'étendre cet article aux récritures suffixes de termes.

Remerciement

Je remercie Jean-Claude Raoult pour ses multiples remarques.

Références

Au 87	JM. Autebert "langages algébriques" Ed. Masson.				
Be 79	J. Berstel "Transductions and Context-Free Languages" Ed. Teubner.				
Bo-Ni	L. Boasson, M. Nivat "Centers of context-free languages"				
84	Rapport LITP 84-44.				
Ch 82	L. Chottin "Langages algébriques et systèmes de réécritures rationnels"				
	Rairo 16, pp 93-112.				
Da-Ti	M. Dauchet, S. Tison, T. Heuillard, P. Lescanne "The confluence of ground				
He-Le	term rewriting systems is decidable" Preliminary proceeding of the				
87	conference on resolution of equations in algebraic structures. Austin.				
Hu-La	G. Huet, D. Lankford "On the uniform halting problem for term rewriting				
. 78	systems" Rapport INRIA 283.				
Ne-Op	G. Nelson, D. Oppen "Fast Decision Procedures Based on Congruence				
80					

ε			
1			
,			
à.			
•			
₹		υ,	