USTL
FLANDRES ARTOIS

UNIVERSITE DES SCIENCES ET TECHNIQUES
DE LILLE FLANDRES ARTOIS

N° d'ordre : 208

THÈSE

Nouveau Régime

présentée à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

pour obtenir le titre de

DOCTEUR EN INFORMATIQUE

par

Alain TERLUTTE

CONTRIBUTION A L'ETUDE DES LANGAGES ENGENDRES PAR DES MORPHISMES ITERES

Thèse soutenue le 15 Mars 1988, devant la Commission d'Examen

Membres du Jury

Président	G. JACOB
Rapporteurs	J. BERSTEL
	C. CHOFFRUT
Directeur de thèse	M. LATTEUX
Examinateur	M. DAUCHET

UNIVERSITE DES SCIENCES
ET TECHNIQUES DE LILLE
FLANDRES ARTOIS

DOYENS HONORAIRES DE L'ANCIENNE FACULTE DES SCIENCES

M. H. LEFEBVRE, M. PARREAU.

PROFESSEURS HONORAIRES DES ANCIENNES FACULTES DE DROIT ET SCIENCES ECONOMIQUES, DES SCIENCES ET DES LETTRES

MM. ARNOULT, BONTE, BROCHARD, CHAPPELON, CHAUDRON, CORDONNIER, DECUYPER, DEHEUVELS, DEHORS, DION, FAUVEL, FLEURY, GERMAIN, GLACET, GONTIER, KOURGANOFF, LAMOTTE, LASSERRE, LELONG, LHOMME, LIEBAERT, MARTINOT-LAGARDE, MAZET, MICHEL, PEREZ, ROIG, ROSEAU, ROUELLE, SCHILTZ, SAVARD, ZAMANSKI, Mes BEAUJEU, LELONG.

PROFESSEUR EMERITE

M. A. LEBRUN

ANCIENS PRESIDENTS DE L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

MM. M. PARREAU, J. LOMBARD, M. MIGEON, J. CORTOIS.

PRESIDENT DE L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE FLANDRES ARTOIS

M. A. DUBRULLE.

PROFESSEURS - CLASSE EXCEPTIONNELLE

M. CONSTANT Eugène	Electronique
M. FOURET René	Physique du solide
M. GABILLARD Robert	Electronique
M. MONTREUIL Jean	Biochimie
M. PARREAU Michel	Analyse
M. TRIDOT Gabriel	Chimie appliquée

PROFESSEURS - 1ère CLASSE

M. BACCHUS Pierre	Astronomie
M. BIAYS Pierre	Géographie
M. BILLARD Jean	Physique du solide
M. BOILLY Bénoni	Biologie
M. BONNELLE Jean Pierre	Chimie-Physique
M. BOSCQ Denis	Probabilités
M. BOUGHON Pierre	Algèbre
M. BOURIQUET Robert	Biologie végétale
M. BREZINSKI Claude	Analyse numérique
M. BRIDOUX Michel	Chimie-Physique
M. CARREZ Christian	Informatique
M. CELET Paul	Géologie générale
M. CHAMLEY Hervé	Géotechnique
M. COEURE Gérard	Analyse
M. CORDONNIER Vincent	Informatique
M. DEBOURSE Jean Pierre	Gestion des entreprises
M. DHAINAUT André	Biologie animale
M. DOUKHAN Jean Claude	Physique du solide
M. DYMENT Arthur	Mécanique
M. ESCAIG Bertrand	Physique du solide
M. FAURE Robert	Mécanique
M. FOCT Jacques	Métallurgie
M. FRONTIER Serge	Ecologie numérique
M. GRANELLE Jean jacques	Sciences Economiques
M. GRUSON Laurent	Algèbre
M. GUILLAUME Jean	Microbiologie
M. HECTOR Joseph	Géométrie
M. LABLACHE-COMBIER Alain	Chimie organique
M. LACOSTE Louis	Biologie végétale
M. LAVEINE Jean Pierre	Paléontologie
M. LEHMANN Daniel	Géométrie
Mme LENOBLE Jacqueline	Physique atomique et moléculaire
M. LEROY Jean Marie	Spectrochimie
M. LHOMME Jean	Chimie organique biologique
M. LOMBARD Jacques	Sociologie
M. LOUCHEUX Claude	Chimie physique
M. LUCQUIN Michel	Chimie physique
M. MACKE Bruno	Physique moléculaire et rayonnements atmosphériqu
M. MIGEON Michel	E.U.D.I.L.
M. PAQUET Jacques	Géologie générale
M. PETIT Francis	Chimie organique
M. POUZET Pierre	Modélisation - Calcul scientifique
M. PROUVOST Jean	Minéralogie
M. RACZY Ladislas	Electronique
M. SALMER Georges	Electronique
M. SCHAMPS Joel	Spectroscopie moléculaire
M. SEGUIER Guy	Electrotechnique
M. SIMON Michel	Sociologie
Mle SPIK Geneviève	Biochimie
M. STANKIEWICZ François	Sciences Economiques
M. TILLIEU Jacques	Physique théorique
M. TOULOTTE Jean Marc	Automatique
M. VIDAL Pierre	Automatique
M. ZEYTOUNIAN Radyadour	Mécanique

PROFESSEURS - 2ème CLASSE

M. ALLAMANDO Etienne	Composants électroniques
M. ANDRIES Jean Claude	Biologie des organismes
M. ANTOINE Philippe	Analyse
M. BART André	Biologie animale
M. BASSERY Louis	Génie des procédés et réactions chimiques
Mme BATTIAU Yvonne	Géographie
M. BEGUIN Paul	Mécanique
M. BELLET Jean	Physique atomique et moléculaire
M. BERTRAND Hugues	Sciences Economiques et Sociales
M. BERZIN Robert	Analyse
M. BKOUCHE Rudolphe	Algèbre
M. BODARD Marcel	Biologie végétale
M. BOIS Pierre	Mécanique
M. BOISSIER Daniel	Génie civil
M. BOIVIN Jean Claude	Spectrochimie
M. BOUQUELET Stéphane	Biologie appliquée aux enzymes
M. BOUQUIN Henri	Gestion
M. BRASSELET Jean Paul	Géométrie et topologie
M. BRUYELLE Pierre	Géographie
M. CAPURON Alfred	Biologie animale
M. CATTEAU Jean pierre	Chimie organique
M. CAYATTE Jean Louis	Sciences Economiques
M. CHAPOTON Alain	Electronique
M. CHARET Pierre	Biochimie structurale
M. CHIVE Maurice	Composants électroniques optiques
M. COMYN Gérard	Informatique théorique
M. COQUERY Jean Marie	Psychophysiologie
M. CORIAT Benjamin	Sciences Economiques et Sociales
Mme CORSIN Paule	Paléontologie
M. CORTOIS Jean	Physique nucléaire et corpusculaire
M. COUTURIER Daniel	Chimie organique
M. CRAMPON Norbert	Tectolique Géodynamique
M. CROSNIER Yves	Electronique
M. CURGY Jean jacques	Biologie
Mle DACHARRY Monique	Géographie
M. DAUCHET Max	Informatique
M. DEBRABANT Pierre	Géologie appliquée
M. DEGAUQUE Pierre	Electronique
M. DEJAEGER Roger	Electrochimie et Cinétique
M. DELORME Pierre	Physiologie animale
M. DELORME Robert	Sciences Economiques
M. DEMUNTER Paul	Sociologie
M. DENEL Jacques	Informatique
M. DE PARIS Jean Claude	Analyse
M. DEPREZ Gilbert	Physique du solide - Cristallographie
M. DERIEUX Jean Claude	Microbiologie
Mle DESSAUX Odile	Spectroscopie de la réactivité chimique
M. DEVRAINNE Pierre	Chimie minérale
Mme DHAINAUT Nicole	Biologie animale
M. DHAMELINCOURT Paul	Chimie physique
M. DORMARD Serge	Sciences Economiques
M. DUBOIS Henri	Spectroscopie hertzienne
M. DUBRULLE Alain	Spectroscopie hertzienne
M. DUBUS Jean Paul	Spectrométrie des solides

M. DUPONT Christophe	Vie de la firme (I.A.E.)
Mme EVRARD Micheline	Génie des procédés et réactions chimiques
M. FAKIR Sabah	Algèbre
M. FAUQUEMBERGUE Renaud	Composants électroniques
M. FONTAINE Hubert	Dynamique des cristaux
M. FOUQUART Yves	Optique atmosphérique
M. FOURNET Bernard	Biochimie structurale
M. GAMBLIN André	Géographie urbaine, industrielle et démographie
M. GLORIEUX Pierre	Physique moléculaire et rayonnements atmosphérique
M. GOBLOT Rémi	Algèbre
M. GOSSELIN Gabriel	Sociologie
M. GOUDMAND Pierre	Chimie physique
M. GOURIEROUX Christian	Probabilités et statistiques
M. GREGORY Pierre	I.A.E.
M. GREMY Jean Paul	Sociologie
M. GREVET Patrice	Sciences Economiques
M. GRIMBLOT Jean	Chimie organique
M. GUILBAULT Pierre	Physiologie animale
M. HENRY Jean Pierre	Génie mécanique
M. HERMAN Maurice	Physique spatiale
M. HOUDART René	Physique atomique
M. JACOB Gérard	Informatique
M. JACOB Pierre	Probabilités et statistiques
M. JEAN Raymond	Biologie des populations végétales
M. JOFFRE Patrick	Vie de la firme (I.A.E.)
M. JOURNEL Gérard	Spectroscopie hertzienne
M. KREMBEL Jean	Biochimie
M. LANGRAND Claude	Probabilités et statistiques
M. LATTEUX Michel	Informatique
Mme LECLERCQ Ginette	Catalyse
M. LEFEBVRE Jacques	Physique
M. LEFEVRE Christian	Pétrologie
Mle LEGRAND Denise	Algèbre
Mle LEGRAND Solange	Algèbre
M. LEGRAND Pierre	Chimie
Mme LEHMANN Josiane	Analyse
M. LEMAIRE Jean	Spectroscopie hertzienne
M. LE MAROIS Henri	Vie de la firme (I.A.E.)
M. LEROY Yves	Composants électroniques
M. LESENNE Jacques	Systèmes electroniques
M. LHENAFF René	Géographie
M. LOCQUENEUX Robert	Physique théorique
M. LOSFELD Joseph	Informatique
M. LOUAGE Francis	Electronique
M. MAHIEU Jean Marie	Optique - Physique atomique
M. MAIZIERES Christian	Automatique
M. MAURISSON Patrick	Sciences Economiques et Sociales
M. MESMACQUE Gérard	Génie Mécanique
M. MESSELYN Jean	Physique atomique et moléculaire
M. MONTEL Marc	Physique du solide
M. MORCELLET Michel	Chimie organique
M. MORTREUX André	Chimie organique
Mme MOUNIER Yvonne	Physiologie des structures contractiles
M. NICOLE Jacques	Spectrochimie
M. NOTELET Francis	Systèmes électroniques
M. PARSY Fernand	Mécanique
M. PECQUE Marcel	Chimie organique
M. PERROT Pierre	Chimie appliquée

M. PERTUZON Emile	Physiologie animale
M. PONSOLLE Louis	Chimie physique
M. PORCHET Maurice	Biologie animale
M. POSTAIRE Jack	Informatique industrielle
M. POVY Lucien	Automatique
M. RICHARD Alain	Biologie animale
M. RIETSCH François	Physique des polymères
M. ROBINET Jean CLaude	EUDIL
M. ROGALSKI Marc	Analyse
M. ROY Jean Claude	Psychophysiologie
Mme SCHWARZBACH Yvette	Géométrie
M. SLIWA Henri	Chimie organique
M. SOMME Jean	Géographie
M. STAROSWIECKI Marcel	Informatique
M. STERBOUL François	Informatique
M. TAILLIEZ Roger	Génie alimentaire
M. THERY Pierre	Systèmes électroniques
M. THIEBAULT François	Sciences de la terre
M. THUMERELLE Pierre	Démographie - Géographie Humaine
Mme TJOTTA Jacqueline	Mathématiques
M. TOURSEL Bernard	Informatique
M. TREANTON Jean rené	Sociologie du Travail
M. TURREL Georges	Spectrochimie infrarouge et Raman
M. VANDORPE Bernard	Chimie minérale
M. VASSEUR Christian	Automatique
M. VAST Pierre	Chimie inorganique
M. VERBERT André	Biochimie
M. VERNET Philippe	Génétique
M. WACRENIER Jean Marie	Electronique
M. WALLART Francis	Spectrochimie infrarouge et Raman
M. WARTEL Michel	Chimie inorganique
M. WATERLOT Michel	Géologie générale
M. WEINSTEIN Olivier	Analyse économique de la recherche et développement
M. WERNER Georges	Informatique théorique
M. WOZNIAK Michel	Spectrochimie
Mme ZINN JUSTIN Nicole	Algèbre

Je tiens à remercier Gérard Jacob qui me fait l'honneur de présider le jury.

Je tiens à exprimer ma reconnaissance à Jean Berstel et à Christian Choffrut qui ont accepté la tache ingrate de rapporteurs.

Je ne peux qu'exprimer toute ma sympathie et ma gratitude envers Michel Latteux qui dirigea mes recherches ; de nombreuses discussions ont mené à l'achévement de cette thèse.

Je remercie Max Dauchet d'avoir accepté de participer au jury de cette thèse.

Je remercie le laboratoire d'informatique de l'Université de Lille en tant qu'entité abstraite, mais dont toute les composantes ont fait que ce travail ne soit pas une peine. Je tiens à remercier tout particulièrement ceux qui m'ont initié aux plaisirs complexes de l'informatique théorique.

Je remercie Henri Gland pour la rapidité et la qualité de la reprographie.

CONTRIBUTION A L'ÉTUDE DES LANGAGES

ENGENDRÉS PAR DES MORPHISMES ITÉRÉS

TABLE DES MATIÈRES

Chapitre I

INTRODUCTION

Les L-systèmes ont été introduits par A.Lindenmayer pour décrire des développements d'organismes en biologie. La comparaison avec la théorie des langages formels met en évidence de nombreuses analogies. A l'origine, le modèle mathématique utilisait des tableaux d'automates finis. Puis la description fit appel à des constructions plus proche des grammaires telles qu'elles sont définies dans la hiérarchie de Chomsky.

On retrouve, comme dans la hiérarchie de Chomsky, la notion d'indépendance ou de dépendance vis à vis du contexte. La notation des L-systèmes utilise le suffixe $(m, n)L$ pour indiquer que, dans un tel système, la réécriture d'une lettre dépend de m lettres à sa gauche et de n lettres à sa droite ; la terminaison $(0,0)L$ est abrégée en $0L$.

La différence principale entre la hiérarchie de Chomsky et les familles de L-langages provient de la notion de simultanéité. Dans les L-systèmes, à chaque étape, toute la chaine est dérivée et il n'y a pas de symboles non terminaux au sens des grammaires de Chomsky. Cela conduit, pour les systèmes à contexte indépendant, à des langages obtenus par morphismes ou par substitutions finies, itérés à partir d'un axiome. En fait, le langage est engendré par l'application itérative de tables de production, ce qui est symbolisé par la lettre T dans le nom de la classe de systèmes (ce symbole est absent lorsqu'il n'y a qu'une seule table) : $T0L$-systèmes, $0L$-systèmes. Dans le cas des systèmes à contexte indépendant, chaque table contient des règles qui correspondent à une substitution finie. Le L-système est déterministe quand chacune de ses tables de production est déterministe, il s'agit alors de morphismes itérés : $DT0L$-systèmes, $D0L$-systèmes. Les L-systèmes peuvent être enrichis par la notion d'alphabet-cible en ne conservant dans le langage que les mots écrits sur un alphabet donné. On parle alors de systèmes étendus : $ET0L$-systèmes, $E0L$-systèmes, $EDT0L$-systèmes, $ED0L$-systèmes. Une autre notion qui a un intérêt évident en biologie est celle de langage adulte ; le langage adulte d'un L-système représente l'ensemble des mots engendrés par le système qui ne peuvent se dériver qu'en eux-mêmes.

Exceptée la notion de langage adulte, les notions rappelées ci-dessus sont connus en théorie des langages et, tout en conservant ses rapports avec la biologie, l'étude des L-langages a apporté de nombreux résultats et de nombreux problèmes à la théorie des langages formels. Citons un des liens entre les L-langages et la hiérarchie de Chomsky : un langage est algébrique si et seulement s'il est le langage adulte d'un $0L$-système.

Parmi les L-systèmes ,nous nous intéresserons plus particulièrement aux $D0L$-systèmes (un seul morphisme itéré) et aux $DT0L$-systèmes (plusieurs morphismes itérés).

Les $D0L$-systèmes représentent la construction de base des L-systèmes. Le principe des dérivations peut s'envisager selon deux approches : analyse de la séquence des dérivations, analyse du langage engendré.

Dans le chapitre III, nous examinerons principalement les mots infinis engendrés par des $D0L$-systèmes. Deux notions importantes permettent l'étude de la structure infinie des langages : la limite et l'adhérence qui sont toutes deux des ensembles de mots infinis. Un mot infini appartient à l'adhérence d'un langage si ses facteurs gauches, ou préfixes, sont aussi des facteurs gauches du langage. Il appartiendra à la limite du langage si un nombre infini de ses facteurs gauches appartiennent au langage. La limite d'un langage est toujours incluse dans l'adhérence de ce langage. Le mot de Thue [66] est l'adhérence et la limite du $D0L$-langage engendré, à partir de l'axiome a, par le morphisme h de $\{a,b\}^*$ dans $\{a,b\}^*$ défini par $h(a) = ab$ et $h(b) = ba$.

Un mot est sans k-répétition s'il ne contient aucun facteur de la forme w^k excepté le mot vide (sans carré pour $k = 2$). L'étude des mots infinis ne contenant pas de carré est liée à celle des $D0L$-langages. En effet, la plupart des constructions de mots infinis sans carré fait intervenir un morphisme itéré [9]. L'un des mots sans répétition les plus célèbres est certainement le mot de Thue qui, sur un alphabet de deux lettres, est sans cube. D'un autre côté, le rapport entre mots infinis et langages algébriques est mis en évidence par le langage de Goldstine dont le complémentaire est l'ensemble des facteurs gauches d'un mot infini. Nous remarquerons aussi que les facteurs gauches d'un ensemble fini de mots infinis constituent un langage algébrique seulement si ce langage est rationnel.

Un langage dont le complémentaire est algébrique est appelé co-algébrique. Le fait que l'ensemble des facteurs gauches du mot de Thue soit co-algébrique a permis à W. Bucher, D. Haussler et M.G. Main [15] d'apporter une réponse à certaines conjectures concernant les mots sans carré, conjectures émises dans [2,3]. J. Berstel a généralisé ce résultat en montrant que le centre, ensemble des facteurs gauches de l'adhérence, de tout langage engendré par un morphisme prolongeable itéré était co-algébrique [10].

Nous poursuivons cette étude du rapport entre le centre d'un langage obtenu par un procédé itératif et la famille des langages algébriques. Nous verrons, en particulier, que le centre est co-algébrique quand il s'agit de $D0L$-systèmes, de gsm prolongeables itérés ou de Tag-systèmes uniformes. Récemment J.M. Autebert et J. Gabarró ont prouvé l'existence de mots infinis qui n'étaient engendrés par aucun gsm prolongeable itéré [6], répondant négativement à une conjecture de J. Berstel [11]. Signalons aussi leurs études du rapport entre facteurs gauches de mots infinis et langages algébriques ambigus [4,5].

La décidabilité de l'équivalence des séquences de deux $D0L$-systèmes est longtemps restée un problème ouvert. Il a été résolu de façon positive dans [20]. En rapport avec ce résultat, il a été démontré que l'équivalence des limites de deux $D0L$-langages est décidable [21] et que, pour un $D0L$-langage donné, on peut décider si sa limite est un mot ultimement périodique donné [53].

Un $D0L$-langage infini se décompose en une union finie de $D0L$-langages dont l'adhérence contient un seul mot. En se référant aux résultats que nous venons d'énoncer, cela permet de montrer la décidabilité de l'équivalence des adhérences de deux $D0L$-langages. Ce résultat nous permet aussi de démontrer facilement quelques propriétés des centres algébriques de $D0L$-langages, c'est-à-dire quand l'adhérence contient uniquement des mots ultimement périodiques. Nous conclurons l'étude des $D0L$-langages en montrant que le problème de savoir si un mot ultimement périodique est limite d'un $D0L$-langage est décidable. Dans le cas d'un alphabet quelconque, nous donnerons des conditions suffisantes pour décider de ce problème ; pour un alphabet de deux lettres, nous établirons une caractérisation des mots infinis ultimement périodiques qui sont les limites de $D0L$-langages, retrouvant les résultats de P. Séébold [63].

Dans le chapitre IV, nous nous intéresserons à l'étude des $EDT0L$-langages.

La famille des $ET0L$-langages (substitutions itérées avec un alphabet-cible), notée $ET0L$, semble être la plus riche en propriétés parmi les L-langages à contexte indépendant. $ET0L$ est un cône rationnel principal. La sous-famille des $EDT0L$-langages (morphismes itérés avec un alphabet-cible), notée $EDT0L$, présente un domaine de recherche très investi. La structure mathématique des $EDT0L$-systèmes permet une meilleure comprèhension du parallélisme dans les dérivations. Rappelons quelques résultats importants.

Grâce à des théorèmes de duplication, la recherche de langages n'appartenant pas à $ET0L$ peut se réduire à celle de langages n'appartenant pas à $EDT0L$ [65].

Alors que les langages algébriques sont tous des $ET0L$-langages, il existe des langages algébriques qui n'appartiennent pas à $EDT0L$ [24]. En particulier, aucun générateur algébrique n'appartient à $EDT0L$ [43].

N'étant pas fermée par morphisme inverse, $EDT0L$ n'est pas un cône rationnel.

La famille des $EDT0L$-langages d'index fini, notée $EDT0L_{FIN}$, admet plusieurs définitions équivalentes [61] : $ET0L$-langages d'index fini [60], langages matriciels d'index fini [7,31,54] ou langages totalement parallèles [56].

Nous montrerons que tout langage de cette famille est Parikh-borné c'est-à-dire qu'il contient un langage borné qui lui est commutativement équivalent. Cette notion introduite par M. Latteux et J. Leguy [44] a été étudiée dans différents articles [12,38,50,57,58]. En particulier, il a été montré que les langages algébriques sont Parikh-bornés [12]. Pour les $EDT0L$-langages d'index fini, la question a été soulevée par G. Păun [55].

Les principales propriétés de clôture de la famille des $EDT0L$-langages ont été établies, notamment par A. Ehrenfeucht, G. Rozenberg [23] et J. Engelfriet [29]. Elle est fermée par union, concaténation, morphisme, intersection avec les langages rationnels, fonction rationnelle mais ne l'est pas par morphisme inverse. En particulier, si c est une nouvelle lettre, $L \sqcup\!\sqcup\, c^* \in EDT0L$ implique $L \in EDT0L_{FIN}$ [30,41]. Ce résultat entraine que $EDT0L_{FIN}$ est le plus grand cône rationnel contenu dans $EDT0L$. A. Ehrenfeucht, G. Rozenberg [23] poursuivent l'investigation de la famille des $EDT0L$-langages en amorçant un début de hiérarchie, $EDT0L \subsetneq \mathbf{H}^{-1}(EDT0L) \subsetneq \mathbf{H} \circ \mathbf{H}^{-1}(EDT0L)$ qu'ils conjecturent infinie.

M. Latteux et P. Turakainen [46] ont établi que toute composition de morphismes et de morphismes inverses pouvait se réduire à une composition dans $\mathbf{H}^{-1} \circ \mathbf{H} \circ \mathbf{H}^{-1} \circ \mathbf{H}$. Ce résultat suffisait pour apporter une réponse négative à la conjecture. En utilisant les travaux de M. Latteux et J. Leguy [45], J. Karhumaki et M. Linna [40] sur les compositions de morphismes et de morphismes inverses, nous démontrons que la hiérarchie s'arrête à $\mathbf{H} \circ \mathbf{H}^{-1}(EDT0L) = \mathbf{C}(EDT0L)$, le cône rationnel engendré par $EDT0L$.

Si un morphisme est effaçant, le morphisme inverse est d'image infinie. Si un morphisme non effaçant transforme deux lettres différentes en un même mot, le morphisme inverse est d'image finie mais le nombre d'images d'un mot exprimé en fonction de la longueur de ce mot peut être exponentiel. En appliquant des morphismes inverses de ce type à des $EDT0L$-langages, on trouve des exemples de langages qui n'appartiennent plus à $EDT0L$.

Il semblait donc intéressant de se demander si $EDT0L$ restait close par des transductions rationnelles qui, à tout mot, associeraient un ensemble d'images dont le cardinal serait polynômialement borné en fonction de la longueur du mot de départ. Les polynômes de fonctions rationnelles vérifient ce critère ; la famille des $EDT0L$-langages est fermée par polynôme de fonctions rationnelles. Mais savoir si $EDT0L$ est fermée par transduction rationnelle polynômialement bornée reste un problème ouvert.

Le chapitre V est consacré à l'étude des transductions rationnelles polynômialement bornées. La recherche de propriétés de clôture des familles de langages conduit tout naturellement à l'étude des transductions rationnelles. Introduites par C.C. Elgot et G. Mezei [28], leur caractérisation en termes de bimorphismes, due à M. Nivat, reste un résultat fondamental [51].

On peut distinguer des classes particulières dans l'ensemble des transductions rationnelles en considérant la longueur des images, leur nombre,... . Dans les chapitres précédents, nous avons utilisé des propriétés des transductions rationnelles décroissantes [47], des transductions rationnelles étoilées [45], des fonctions rationnelles [29],... . Une autre classe , celle des transductions rationnelles cycliques d'image finie sera étudiée dans ce chapitre. Une transduction est cyclique si l'image du domaine est un sous-ensemble du monoïde libre engendré par un unique mot. Elle sera dite d'image finie si l'image de chaque mot est un sous-ensemble fini.

Le fait que *EDT0L* soit fermée par polynôme de fonctions rationnelles nous a amenés à étudier les transductions rationnelles polynômialement bornées. M.P. Schützenberger a montré que l'ensemble de ces dernières étaient exactement l'ensemble des polynômes en des transductions rationnelles fonctionnelles ou cycliques d'image finie [64]. Nous établirons l'inclusion stricte de l'ensemble des polynômes de fonctions rationnelles dans l'ensemble des transductions rationnelles polynômialement bornées, apportant ainsi une réponse à une interrogation de M.P. Schützenberger. Nous montrerons qu'il existe dans $\mathbf{H} \circ \mathbf{H}^{-1}$ des transductions rationnelles cycliques d'image finie qui ne sont pas des polynômes de fonctions rationnelles. En utilisant les propriétés du produit de Hadamard sur les fonctions cycliques [16], nous donnons une méthode qui permet de trouver d'autres exemples de telles transductions cycliques.

Chapitre II

PRÉLIMINAIRES

Nous présentons dans ce chapitre les notations et les définitions qui seront utilisées par la suite.

N désignera l'ensemble des entiers naturels et $\mathbf{N}_+ = \mathbf{N} \setminus \{0\}$ l'ensemble des entiers strictement positifs. Pour n et $p \in \mathbf{N}$, l'intervalle des entiers compris entre n et p s'écrit $[n,p] = \{i \in \mathbf{N} \ / \ n \leq i \leq p\}$ et $[n,p[= \{i \in \mathbf{N} \ / \ n \leq i < p\}$.

Soit X un alphabet fini. X^* représente le monoïde libre engendré par X, c'est-à-dire l'ensemble des mots finis construits sur X, avec le mot vide noté ε. On définit $X^+ = X^* \setminus \{\varepsilon\}$. Un langage L est un sous-ensemble de X^*. Si w est un mot de X^*, on note $|w|$ la longueur de w. Soit Z un sous-ensemble de X. On note $|w|_Z$ le nombre d'occurrences des lettres de Z dans le mot w. Nous noterons $ab(u)$ l'ensemble des lettres apparaissant dans u. Le cardinal d'un ensemble A sera noté $||A||$.

Soient u et v deux mots de X^*. Nous noterons $\sqcup\!\sqcup$, l'opérateur de shuffle défini par

$$u \sqcup\!\sqcup v = \{u_1 v_1 u_2 v_2 \ldots u_n v_n \ / \ n \in \mathbf{N}, \ \ u_i, \ v_i \in X^*, \ \ u_1 \ldots u_n = u \ \ et \ \ v_1 \ldots v_n = v\}.$$

Il se définit sur les langages par

$$L_1 \sqcup\!\sqcup L_2 = \bigcup_{\substack{u \in L_1 \\ v \in L_2}} u \sqcup\!\sqcup v.$$

Si $X = \{a_1, a_2, \ldots, a_k\}$ est un alphabet ordonné, la fonction de Parikh, notée Ψ, associe, à tout mot $w \in X^*$, le k-uple $\Psi(w) = (|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_k})$. Deux langages L et L' sont commutativement équivalent si $\Psi(L) = \Psi(L')$. Un langage est borné s'il existe un n-uple de mots $(w_1, w_2, \ldots, w_n)$ tel que $L \subseteq w_1^* w_2^* \ldots w_n^*$. Un langage L est Parikh-borné s'il existe un langage borné L' inclus dans L et commutativement équivalent à L ou, de façon équivalente, s'il existe des mots $u_1, \ldots, u_k \in X^*$ tels que $\Psi(L) = \Psi(L \cap u_1^* u_2^* \ldots u_k^*)$.

Un mot infini α est une application de $\mathbf{N}_+$ dans X. Pour $n \in \mathbf{N}_+$, $p \in \mathbf{N}$, $\alpha_{(n)}$ désigne la n-ième lettre de α, $\alpha_{[n]}$ représente le mot fini $\alpha_{(1)}\alpha_{(2)}\ldots\alpha_{(n)}$, $\alpha_{[n,n+p]}$ le mot $\alpha_{(n)}\alpha_{(n+1)}\ldots\alpha_{(n+p)}$ et $\alpha_{[n,n+p[}$ le mot $\alpha_{(n)}\alpha_{(n+1)}\ldots\alpha_{(n+p-1)}$. On posera $\alpha_{[0]} = \varepsilon$ et $\alpha_{[n,n[} = \varepsilon$. Nous utiliserons les mêmes notations pour les mots finis en respectant la longueur du mot, c'est-à-dire en posant, pour tout mot $w \in X^*$, $w_{(n)} = \varepsilon$ si $n > |w|$. L'ensemble des mots infinis sur X est noté X^ω et l'ensemble des mots finis ou infinis, $X^\infty = X^* \cup X^\omega$.

Un mot α appartenant à X^ω sera dit ultimement périodique s'il existe $u \in X^*$ et $v \in X^+$ tels que $\alpha = uv^\omega$. Un mot u de X^* est primitif si $u \in w^+$ implique $u = w$. Pour tout mot ultimement périodique $\alpha \in X^\omega$, il existe un couple unique $(u, v) \in X^* \times X^+$ tel que v est primitif et soit $u = \varepsilon$, soit la dernière lettre de u est différente de la dernière lettre de v. On dira alors que uv^ω est sous forme primitive.

L'ensemble des facteurs gauches et l'ensemble des facteurs d'un mot u de X^∞ sont définis par

$$FG(u) = \{w \in X^* \ / \ \exists\, v \in X^\infty \ tel\, que\ wv = u\} = \{u_{[n]} \ / \ n \in \mathbb{N}\}$$

$$F(u) = \{w \in X^* \ / \ \exists\, v_1 \in X^* \ et\ v_2 \in X^\infty \ tel\, que\ v_1 w v_2 = u\}.$$

On définit de la même façon l'ensemble des facteurs droits d'un mot u de X^* par

$$FD(u) = \{w \in X^* \ / \ \exists\, v \in X^* \ tel\, que\ vw = u\}.$$

Pour un langage $L \subseteq X^\infty$, on pose

$$FG(L) = \{FG(u) \ / \ u \in L\} \quad \text{et} \quad F(L) = \{F(u) \ / \ u \in L\}.$$

Pour un langage $L \subseteq X^*$, on pose $FD(L) = \{FD(u) \ / \ u \in L\}$.

A tout langage L inclus dans X^*, on peut associer son adhérence et sa limite, ensembles de mots infinis notés $Adh(L)$ et $Lim(L)$, qui sont définies par

$$\begin{aligned} Adh(L) &= \{\alpha \in X^\omega \ / \ \forall\, n \in \mathbb{N}_+ \ \ \alpha_{[n]} \in FG(L)\} \\ &= \{\alpha \in X^\omega \ / \ \forall\, n \in \mathbb{N}_+ \ \exists\, v \in X^* \ \ tel\, que\ \ \alpha_{[n]} v \in L\} \\ Lim(L) &= \{\alpha \in X^\omega \ / \ \forall\, n \in \mathbb{N}_+ \ \exists\, p \in \mathbb{N} \ \ tel\, que\ \ \alpha_{[n+p]} \in L\}. \end{aligned}$$

Il est clair que $Lim(L) \subseteq Adh(L)$, pour tout langage L, et que $Adh(L)$ est vide si et seulement si L est fini. Par contre, $Lim(a^* b) = \emptyset$.

Le centre d'un langage $L \subseteq X^*$, noté L^c, est défini par

$$L^c = FG(Adh(L)) = \{w \in X^* \ / \ wX^* \cap L \ est\ infini\}.$$

L'ensemble des langages rationnels (resp. algébriques) est noté *Rat* (resp. *Alg*). Le langage $\overline{L}$ représente le complémentaire de L et le langage L est co-algébrique si son complémentaire est algébrique.

Soient X et Y deux alphabets finis. Un morphisme h de X^* dans Y^* est dit alphabétique (resp. strictement alphabétique, non effaçant, uniforme de module k) si $h(X)$ est inclus dans $Y \cup \{\varepsilon\}$ (resp. dans Y, dans Y^+, dans Y^k).

$\mathbf{H}$ (resp $\mathbf{H}_\alpha$, $\mathbf{H}_{s\alpha}$, $\mathbf{H}_\varepsilon$) représentera l'ensemble des morphismes (resp. morphismes alphabétiques, strictement alphabétique, non effaçant) et $\mathbf{H}^{-1}$ l'ensemble des morphismes inverses.

Nous rappelons ci-dessous les principales définitions concernant les transductions rationnelles. Pour plus de détails, le lecteur pourra consulter [8].

Une transduction rationnelle τ de X^* dans Y^* est une application de X^* dans 2^{Y^*} dont le graphe $\hat{\tau}$ est une partie rationnelle du produit $X^* \times Y^*$, définie par

$$\hat{\tau} = \{(u,v) \in X^* \times Y^* \ / \ v \in \tau(u)\}.$$

Le domaine d'une transduction τ, noté $Dom(\tau)$, est l'ensemble des mots qui possédent une image, c'est-à-dire $Dom(\tau) = \{u \in X^* \ / \ \exists\, v \in Y^* \ tel\ que\ (u,v) \in \hat{\tau}\}$.

Théorème [51]

Une transduction τ de X^* dans Y^* est rationnelle si et seulement s'il existe un alphabet Z, un langage rationnel $R \subseteq Z^*$ et deux morphismes (alphabétiques) h et g, respectivement de Z^* dans X^* et de Z^* dans Y^*, tels que

$$\hat{\tau} = \{(h(w), g(w)) \ / \ w \in R\},$$

ce qui s'écrit en utilisant les notations de S. Eilenberg [27]

$$\tau = g \circ \cap R \circ h^{-1}.$$

Si nous notons $\wedge Rat = \{\cap R \ / \ R \in Rat\}$, l'ensemble des intersections avec un langage rationnel, et $\wedge Rat^* = \{\cap R^* \ / \ R \in Rat\}$, une transduction est rationnelle si et seulement si elle appartient à $\mathbf{C} = \mathbf{H} \circ \wedge Rat \circ \mathbf{H}^{-1}$; elle est rationnelle étoilée si et seulement si elle appartient à $\mathbf{C}_* = \mathbf{H} \circ \wedge Rat^* \circ \mathbf{H}^{-1}$. Pour un langage L, le cône rationnel $\mathbf{C}(L) = \mathbf{H} \circ \wedge Rat \circ \mathbf{H}^{-1}(L)$ est la clôture de L par transductions rationnelles. Pour une famille de langages $\mathbf{L}$, on a $\mathbf{C}(\mathbf{L}) = \{\mathbf{C}(L) \ / \ L \in \mathbf{L}\}$. De la même façon, on définit le cône rationnel étoilé engendré par un langage L, $\mathbf{C}_*(L) = \mathbf{H} \circ \wedge Rat^* \circ \mathbf{H}^{-1}(L)$, et le cône engendré par une famille $\mathbf{L}$ de langages, $\mathbf{C}_*(\mathbf{L}) = \{\mathbf{C}_*(L) \ / \ L \in \mathbf{L}\}$.

La projection sur X^* sera notée π_X. Nous utiliserons aussi la caractérisation suivante des transductions rationnelles :

Théorème [27]

Soient X et Y deux alphabets disjoints. Une transduction τ de X^* dans Y^* est rationnelle si et seulement s'il existe un langage rationnel $R \subseteq (X \cup Y)^*$ tel que

$$\hat{\tau} = \{(\pi_X(w), \pi_Y(w)) \ / \ w \in R\},$$

ce qui s'écrit en termes de bimorphismes

$$\tau = \pi_Y \circ \cap R \circ \pi_X^{-1}.$$

Une transduction est cyclique si son graphe est inclus dans $X^* \times v^*$ où v est un mot de Y^*.

Une transduction τ de X^* dans Y^* est d'image finie si, pour tout mot w de X^*, $\|\tau(w)\|$ est finie. Elle sera fonctionnelle si, pour tout mot w de X^*, $\|\tau(w)\| \leq 1$.

Une transduction τ sera dite polynômialement bornée s'il existe un polynôme P tel que, pour tout mot $w \in X^*$, on ait $\|\tau(w)\| \leq P(|w|)$. Elle sera dite linéairement bornée si le polynôme P est du premier degré. Les transductions rationnelles cycliques d'image finie sont donc linéairement bornées.

On définit, pour des transductions rationnelles τ et τ' de X^* dans Y^*, la somme $\Sigma = \tau + \tau'$ et le produit $\Pi = \tau \times \tau'$ par

$$\Sigma(w) = \tau(w) \cup \tau'(w) \quad \text{pour tout } \ w \in X^*$$
$$\Pi(w) = \bigcup_{\substack{w_1 w_2 = w \\ w_1, w_2 \in X^*}} \tau(w_1)\tau(w_2) \quad \text{pour tout } \ w \in X^*.$$

Une fonction séquentielle généralisée, qui sera appelée gsm par référence au transducteur qui la réalise (generalized sequential machine), est définie par un sextuple $\langle X, Y, Q, q_0, \delta, \lambda \rangle$ où X et Y sont les alphabets d'entrée et de sortie, Q est un ensemble d'états dans lequel q_0 repère l'état initial, δ est la fonction de sortie de $Q \times X$ dans Y^* et λ la fonction de transition de $Q \times X$ dans Q. Les fonctions δ et λ sont étendues en des fonctions de $Q \times X^*$ dans Y^* et de $Q \times X^*$ dans Q en posant :

$$\lambda(q, \varepsilon) = q \quad \text{pour tout q} \in \text{Q}$$
$$\lambda(q, wx) = \lambda(\lambda(q, w), x) \quad \text{pour tout q} \in \text{Q}, \ \text{w} \in \text{X}^*, \ \text{x} \in \text{X}$$
$$\delta(q, \varepsilon) = \varepsilon \quad \text{pour tout q} \in \text{Q}$$
$$\delta(q, wx) = \delta(q, w)\delta(\lambda(q, w), x) \quad \text{pour tout q} \in \text{Q}, \ \text{w} \in \text{X}^*, \ \text{x} \in \text{X}$$

Une fonction rationnelle f de X^* dans X^* est prolongeable en un mot $w \in X^+$ s'il existe un mot $z \in X^+$ tel que $f(w) = wz$.

Avant d'aborder les différentes familles de L-langages, définissons la famille des $D0L$-langages, langages obtenus par itération d'un morphisme, ce qui nous permettra de présenter les Tag-systèmes et les gsm itérés.

Un $D0L$-système est un triplet $\langle X, h, u\rangle$ où X est un alphabet fini, h est un morphisme de X^* dans X^* et l'axiome u appartient à X^*. Le $D0L$-langage $L(G)$ contient tous les mots de la forme $h^n(u)$, avec $n \in \mathbf{N}$. On notera

$$L(G) = h^*(u) = \{h^n(u) \ / \ n \in \mathbf{N}\}.$$

Nous dirons qu'un $D0L$-système est prolongeable si le morphisme qui le définit est prolongeable en l'axiome et nous parlerons alors de $D0L$-langage prolongeable.

Exemple :

Le mot de Thue, noté ici $\mathbf{t}$, est engendré par le $D0L$-système prolongeable $\langle X, h, a\rangle$ où $X = \{a, b\}$, $h(a) = ab$ et $h(b) = ba$:

$$\mathbf{t} = abbabaabbaababbabaababbaabbaba\ldots$$

Nous distinguons dans la suite, pour un $D0L$-système $G = \langle X, h, u\rangle$, deux sous-ensembles disjoints de l'alphabet X : X_I, l'ensemble des lettres engendrant, par itérations successives du morphisme, des langages infinis et X_F, l'ensemble de celles engendrant des langages finis

$$X_F = \{x \in X \ / \ h^*(x) \quad fini\} \qquad \text{et} \qquad X_I = \{x \in X \ / \ h^*(x) \quad infini\}.$$

Un $CD0L$-langage est l'image par un morphisme strictement alphabétique, parfois appelé codage, d'un $D0L$-langage. Un $CD0L$-système est donc un triplet $\langle G, g, Y\rangle$ où $G = \langle X, h, u\rangle$ est un $D0L$-système et g un morphisme strictement alphabétique de X^* dans Y^*.

Les Tag-systèmes ont été introduits par A. Cobham [18]. Bien qu'il y ait une comparaison possible avec les $CD0L$-systèmes, les Tag-systèmes ont surtout été étudiés pour les mots infinis qu'ils engendrent. Un Tag-système est un quintuplet $\langle X, h, u, g, Y\rangle$ où X et Y sont deux alphabets finis, h est un morphisme de X^* dans X^* prolongeable en u et g est un morphisme strictement alphabétique de X^* dans Y^*. Alors $\langle X, h, u\rangle$ est un $D0L$-système prolongeable ; on appelle séquence interne, notée ***intseq***, l'adhérence du $D0L$-langage qu'il engendre et séquence externe, notée ***extseq***, l'image de ***intseq*** par g. Remarquons que ***intseq*** est aussi la limite du $D0L$-langage. Le Tag-système sera dit uniforme si le morphisme h est uniforme de module k pour un certain k.

Exemple :

Le mot de S. Aršon, noté **a**, est engendré par le Tag-système uniforme de module 3 : $\langle X, h, a, g, Y\rangle$ où $X = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$, $Y = \{a, b, c\}$,

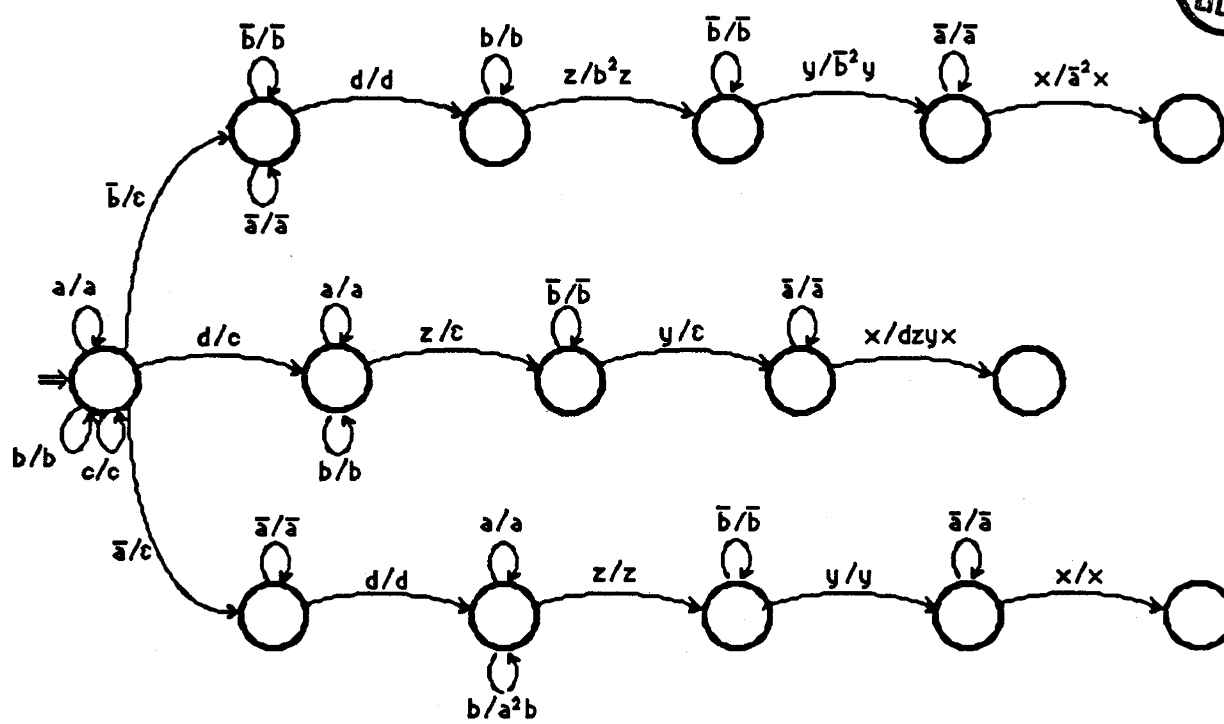
$$h(a) = a\bar{b}c \qquad h(\bar{a}) = \bar{c}b\bar{a}$$
$$h(b) = b\bar{c}a \qquad h(\bar{b}) = \bar{a}c\bar{b}$$
$$h(c) = c\bar{a}b \qquad h(\bar{c}) = \bar{b}a\bar{c}$$
$$g(x) = g(\bar{x}) = x, \text{ pour } x \in \{a, b, c\}$$
$$\mathbf{a} = abcacbcabcbacabacbcabcbabcabacbcacbacabcbabca\ldots$$

S. Aršon a démontré que ce mot est sans carré [1]. J. Berstel [9] a montré qu'il n'était la limite d'aucun $D0L$-langage.

Nous considérons ici la notion de gsm itéré à partir d'un axiome comme une extension des $D0L$-systèmes. Soit $G = \langle X, Y, Q, q_0, \delta, \lambda\rangle$ un gsm. Nous supposerons que $X = Y$. Pour tout mot $w \in X^*$, nous définissons $\delta^0(q_0, w) = w$ et $\delta^{n+1}(q_0, w) = \delta(q_0, \delta^n(q_0, w))$, pour tout $n \in \mathbb{N}$. Le langage engendré par itération de G à partir d'un axiome u est alors $\{\delta^n(q_0, u) \ / \ n \in \mathbb{N}\}$.

Exemple :

Considérons le gsm défini ci-dessous



L'adhérence du langage engendré en itérant ce gsm à partir de l'axiome $ab\bar{b}\bar{a}dzyx$ est

$$\alpha = abca^2ba^2bca^4ba^4ba^4ba^4bc\ldots(a^{2^n}b)^{2^n}c\ldots$$

Il a été montré dans [6] que ce mot infini α ne peut être engendré par aucun gsm prolongeable itéré. Par contre le centre du langage est co-algébrique.

Partant de la définition des $T0L$-systèmes, nous introduisons différentes classes de L-systèmes et de L-langages à contexte indépendant. Pour plus d'informations, le lecteur se référera à [36,59].

Un $T0L$-système est un triplet $G = \langle X, P, \varpi\rangle$ où

(1) X est un alphabet fini,

(2) P est un ensemble fini de tables de production ;
pour tout t de P, on a $t \subseteq X \times X^*$, t étant un ensemble fini
et, pour tout $x \in X$, il existe $w \in X^*$ tel que $(x, w) \in t$,

(3) ϖ est l'axiome appartenant à X.

Soit $x = x_{(1)}x_{(2)}\ldots x_{(n)} \in X^*$ avec $n = |x| \geq 1$ et $y \in X^*$. Nous dirons que x se dérive en y, représenté par la relation $x \Longrightarrow y$, s'il existe une table $t \in P$ telle que $(x_{(1)}, w_1)$, $(x_{(2)}, w_2)$,...,$(x_{(n)}, w_n)$ appartiennent à t et $y = w_1w_2\ldots w_n$. La clôture transitive de cette relation étant notée $\overset{*}{\Longrightarrow}$, le langage engendré par le système G est défini par $L(G) = \{y \in X^* \mid \varpi \overset{*}{\Longrightarrow} y\}$. On remarquera que t est le graphe d'une substitution finie.

Si $||P|| = 1$, le système est appelé $0L$-système.

En introduisant la notion d'alphabet-cible, on définit les $T0L$-systèmes étendus, notés $ET0L$-systèmes. Un $ET0L$-système est un quadruplet $G = \langle X, P, \varpi, Y\rangle$ où $G' = \langle X, P, \varpi\rangle$ est un $T0L$-système engendrant un langage $S(G)$. Le langage engendré par G ne contient que les mots dérivés à partir de l'axiome qui appartiennent à Y^*, c'est-à-dire $L(G) = S(G) \cap Y^*$.

Si $||P|| = 1$, le système est appelé $E0L$-système.

Si, pour tout t de P et pour tout x de X, il existe un et un seul $w \in X^*$ tel que $(x, w) \in t$, le système est déterministe. On définit ainsi les $D0L$-systèmes, les $ED0L$-systèmes, les $DT0L$-systèmes et les $EDT0L$-systèmes.

Si, pour tout t de P, on a $t \in X \times X^+$, le système est dit non effaçant (propagating). On définit de cette façon les $EPDT0L$-systèmes ...

Pour les L-systèmes que nous venons de définir, nous adopterons la convention suivante : X étant un préfixe appartenant à $\{\varepsilon, D, CD, DT, ET, EDT, EPDT\}$, la famille des $X0L$-langages, engendrés par des $X0L$-systèmes, sera notée $X0L$.

Pour illustrer ces notions, donnons quelques exemples :

Exemple :

Soient $X = \{a, b\}$, $P = \{t_1, t_2\}$, $t_1 = \{(a, a^2), (b, b^2)\}$, $t_2 = \{(a, a^3), (b, b^3)\}$ et $\varpi = ab$. Alors $G = \langle X, P, \varpi\rangle$ est un $DT0L$-système qui engendre le langage $L(G) = \{a^{2^n 3^m} b^{2^n 3^m} \ / \ n, m \in \mathbf{N}\}$.

Exemple :

Soient $X = \{A, B, a\}$, $P = \{t_1, t_2\}$, $t_1 = \{(A, A^2), (B, B^3), (a, a)\}$, $t_2 = \{(A, a), (B, a), (a, a)\}$, $\varpi = AB$ et $Y = \{a\}$. Alors $G = \langle X, P, \varpi\rangle$ est un $EDT0L$-système. L'ensemble des mots dérivables est $S(G) = \{A^{2^n} B^{3^n} \ / \ n \in \mathbf{N}\} \cup \{a^{2^n + 3^n} \ / \ n \in \mathbf{N}\}$ et le langage engendré est $L(G) = \{a^{2^n+3^n} \ / \ n \in \mathbf{N}\}$.

Une autre définition des $ET0L$-systèmes fait apparaître des non-terminaux au sens des grammaires de Chomsky :

Un $ET0L$-système est un quintuplet $\langle X, N, H, \phi, \varpi\rangle$ où

(1) X et N sont des alphabets finis disjoints, $V = X \cup N$, X est l'alphabet terminal, N l'alphabet non-terminal,

(2) H représente l'alphabet fini des règles de productions,

(3) ϕ est une application de H dans l'ensemble des substitutions finies de V^* dans V^* dont la restriction à X est l'identité,

(4) ϖ est l'axiome appartenant à N.

L'application ϕ est étendue à H^* en posant :

(i) $\phi(\varepsilon)$ est l'identité sur V

(ii) $\phi(\delta_1\delta_2) = \phi(\delta_2) \circ \phi(\delta_1)$ pour tout δ_1, δ_2 de H^*.

Nous utiliserons surtout cette définition pour les $EDT0L$-systèmes. Dans ce cas, ϕ est une application dans l'ensemble des morphismes de V^* dans V^* dont la restriction à X est l'identité.

L'ensemble des mots, terminaux ou non, dérivés par un système $G = \langle X, N, H, \phi, \varpi\rangle$ est noté $S(G) = \{(\phi(\delta))(\varpi) \ / \ \delta \in H^*\}$ et le langage engendré est $L(G) = S(G) \cap X^*$.

Exemple :

Soient $X = \{a, b\}$, $N = \{A, B, C, D\}$, $H = \{\delta_1, \delta_2, \delta_3\}$, $\varpi = CD$ et ϕ défini par

avec

$\phi(\delta_1) = h_1$	$\phi(\delta_2) = h_2$	$\phi(\delta_3) = h_3$
$h_1(A) = A$	$h_2(A) = A$	$h_3(A) = a$
$h_1(B) = B$	$h_2(B) = B$	$h_3(B) = b$
$h_1(C) = ACB$	$h_2(C) = CB$	$h_3(C) = \varepsilon$
$h_1(D) = DA$	$h_2(D) = D$	$h_3(D) = \varepsilon$
$h_1(a) = a$	$h_2(a) = a$	$h_3(a) = a$
$h_1(b) = b$	$h_2(b) = b$	$h_3(b) = b$

Alors $G = \langle X, N, H, \phi, \varpi \rangle$ est un $EDT0L$-système engendrant le langage $L(G) = \{a^n b^m a^n \ / \ n, m \in \mathbb{N} \ et \ m \geq n\}$.

Un $EDT0L$-système est dit d'index fini s'il existe un entier k tel que tout mot de $S(G)$ contient au plus k occurrences de lettres de N : $S(G) \subseteq (X^* N X^*)^{\leq k}$. La famille des $EDT0L$-langages d'index fini, engendrés par ces systèmes, est notée $EDT0L_{FIN}$.

Chapitre III

ITÉRATION D'UN SEUL MORPHISME :

$D0L$-SYSTÈME

A. Centre de langages

1) présentation

L'itération d'un morphisme prolongeable permet de générer une suite de mots w_i dont chaque élément est préfixe de son image par le morphisme : $h(w_i) = w_{i+1} = w_i u_{i+1}$. On voit tout de suite l'interêt d'un tel procédé pour générer un mot infini unique qui est donc le seul élément de l'adhérence du langage correspondant à la suite de mots.

Le centre d'un langage L, noté L^c, est l'ensemble des facteurs gauches de son adhérence, $L^c = FG(Adh(L))$, c'est-à-dire l'ensemble des mots qui peuvent être prolongés en une infinité de mots de L. Le fait que l'ensemble des facteurs gauches du mot de Thue est co-algébrique a permis d'apporter une réponse négative à la conjecture suivante : si L est un langage algébrique alors soit le complémentaire de L est fini, soit l'intersection entre le complémentaire de L et l'ensemble des mots qui contiennent un carré est infinie [2,3]. Il est connu que ce mot s'obtient par morphisme itéré. J. Berstel a démontré que le centre d'un langage engendré par un morphisme prolongeable itéré est co-algébrique [10]. En respectant l'esprit de cette démonstration, nous allons étudier d'autres procédés itératifs.

La première généralisation à examiner est celle des $D0L$-systèmes. Un $D0L$-système correspond à un morphisme itéré, sans restriction sur le morphisme. Nous verrons que le centre d'un $D0L$-langage est co-algébrique.

Puis nous nous intéresserons à une extension des morphismes fournie par les fonctions séquentielles généralisées, appelées aussi gsm (generalized sequential machine). Un gsm correspond à un transducteur déterministe dont tous les états seraient finals. Dans le cas d'un gsm prolongeable itéré, le centre du langage sera co-algébrique. Mais contrairement aux morphismes, le résultat n'est plus vrai pour les gsm quelconques.

Nous terminerons cette section en examinant le cas des Tag-systèmes. Les Tag-systèmes constituent un mécanisme légèrement plus puissant que les morphismes prolongeables itérés. J. Berstel a montré que le mot d'Aršon, qui pouvait être obtenu par un Tag-système uniforme, ne l'était par un $D0L$-système [9]. La séquence interne d'un Tag-système représente l'adhérence d'un $D0L$-langage prolongeable en un axiome $x_0 \in X$. La séquence externe est un codage de la séquence interne par un morphisme strictement alphabétique. Ce codage masque une partie de l'information qui était contenue dans la séquence interne. Nous verrons que, pour les Tag-systèmes uniformes, l'ensemble des facteurs gauches de la séquence externe est co-algébrique. Par contre, pour les Tag-systèmes quelconques, le problème reste ouvert. L'étude d'exemples semble indiquer qu'ils vérifient la même propriété.

2) différents procédés itératifs

La preuve de J. Berstel s'appuie sur la remarque suivante : quand un morphisme prolongeable h de X^* dans X^* engendre par itération un mot infini α, la place de l'image de la k-ième lettre est déterminée par la longueur de l'image des $k-1$ lettres précédentes. Ce qui se traduit par la propriété des morphismes prolongeables en $\alpha_{[l_0]}$:

$$
\text{(i)} \qquad \begin{array}{l} h(\alpha_{[l_0]}) = \alpha_{[l_1]} \\ \forall \ \ k > l_0 \ \ h(\alpha_{(k)}) = \alpha_{[r+1,r+l]} \text{ où } r = |h(\alpha_{[k-1]})| \text{ et } l = |h(\alpha_{(k)})| \end{array}
$$

Dans le complémentaire, soit le mot ne commence pas par un facteur gauche de $\alpha_{[l_1]}$, soit la condition (i) n'est pas vérifiée pour au moins une lettre. Ce qui s'énonce :

$$
\text{(ii)} \qquad w \in \overline{FG(\alpha)} \iff \begin{array}{l} \text{soit } w \in (\overline{FG(\alpha_{[l_1]})} \cap X^{\leq l_1})X^* \\ \text{soit } \exists\, k \quad l_0 < k < |w| \ = n \text{ tel que} \\ \qquad w_{[r+1,n]} \in (\overline{FG(h(w_{(k)}))} \cap X^{\leq l})X^* \\ \qquad r = |h(w_{[k-1]})| \text{ et } l = |h(w_{(k)})| \end{array}
$$

Pour obtenir le complémentaire, il faut donc rompre le lien entre une lettre et son image. Ce principe réapparaitra dans les démonstrations concernant la généralisation aux gsm prolongeables itérés et aux Tag-systèmes uniformes.

Un problème important reste à résoudre : contrôler la longueur séparant une lettre de son image. J. Berstel traite la question en utilisant un automate à pile. Nous emploierons des langages obtenus par transductions du langage de Dyck sur une lettre D_1^*.

Théorème III.A.1 [10]

Le centre d'un langage engendré à partir de $w_0 \in X^*$ en itérant un morphisme h de X^* dans X^* prolongeable en w_0 est co-algébrique.

Preuve :

Si le langage est fini, l'adhérence est vide et le centre est rationnel.

Si le langage est infini, l'adhérence contient un mot infini α unique avec $\alpha_{[l_0]} = w_0$. Sachant qu'il faut vérifier la propriété (ii), nous nous intéressons au langage

$$L = \{u \bigtriangledown \#^m \bigtriangleup \ / \ u \in \alpha_{[l_0]}X^* \quad et \quad m = |h(u)| - |u| - 1\}$$

où $Z = \{\#, \bigtriangledown, \bigtriangleup\}$ est un ensemble disjoint de X.

Le nombre de $\#$ représente la longueur nécessaire pour qu'une lettre qui remplacerait $\bigtriangledown$ ait son image à la place de $\bigtriangleup$. Le langage L s'obtient par une transduction sur le langage de Dyck D_1^*. Pour cela, nous définissons le morphisme φ de $(X \cup Z)^*$ dans $\{a,b\}^*$ par :

$$\begin{array}{l}
\varphi(x) = a^{l-1} \quad \forall\, x \in X \quad \text{et} \quad l = |h(x)| \geq 1 \\
\varphi(x) = b \quad \forall\, x \in X \quad \text{et} \quad l = |h(x)| = 0 \\
\varphi(\#) = b \\
\varphi(\bigtriangledown) = b \\
\varphi(\bigtriangleup) = \varepsilon
\end{array}$$

et le langage rationnel $R = \alpha_{[l_0]}X^* \bigtriangledown \#^* \bigtriangleup$.

Montrons que $L = \varphi^{-1}(D_1^*) \cap R$ qui est donc un langage algébrique. Définissons l'ensemble des lettres mortelles $X_M = \{x \in X \ / \ h(x) = \varepsilon\}$ et $X_{NM} = X \setminus X_M$. Soit $w = u \bigtriangledown \#^m \bigtriangleup$ un mot de L. Nous avons $\varphi(w) = \varphi(u)b^{m+1}$ avec

$$|\varphi(u)|_a = |\varphi(\pi_{X_{NM}}(u))| = |h(\pi_{X_{NM}}(u))| - |\pi_{X_{NM}}(u)|$$

$$|\varphi(u)|_b = |\varphi(\pi_{X_M}(u))| = |\pi_{X_M}(u)|.$$

Nous avons bien l'égalité

$$m + 1 + |\pi_{X_M}(u)| = |h(\pi_{X_{NM}}(u))| - |\pi_{X_{NM}}(u)|$$

puisque

$$m = |h(u)| - |u| - 1 = |h(\pi_{X_M}(u))| - |\pi_{X_M}(u)| - |\pi_{X_{NM}}(u)| - 1.$$

Donc $\varphi(w) \in D_1^*$ et, comme $w \in \alpha_{[l_0]}X^* \bigtriangledown \#^* \bigtriangleup$, nous en déduisons que $w \in \varphi^{-1}(D_1^*) \cap R$.

Prenons maintenant un mot v de D_1^*. L'ensemble

$$\begin{aligned}
\varphi^{-1}(v) &= \{w \in (X \cup Z)^* \ / \ |w|_{\{\#,\bigtriangledown\}} + |\pi_{X_M}(w)| = |h(\pi_{X_{NM}}(w)| - |\pi_{X_{NM}}(w)|\} \\
&= \{w \in (X \cup Z)^* \ / \ |w|_{\{\#,\bigtriangledown\}} = |h(\pi_X(w)| - |\pi_X(w)|\} \\
&= \{w \in (X \cup Z)^* \ / \ |w|_{\{\#\}} = |h(\pi_X(w)| - |\pi_X(w)| - |w|_{\{\bigtriangledown\}}\}.
\end{aligned}$$

Nous obtenons bien $\varphi^{-1}(v) \cap R \subseteq L$, pour tout v de D_1^*, et donc $\varphi^{-1}(D_1^*) \cap R \subseteq L$.

Il suffit maintenant de remplacer ∇ et Δ comme l'indique la propriété (ii). Pour cela, nous définissons, pour chaque lettre x de X, la substitution rationnelle s_x de $(X \cup Z)^*$ dans X^* par

$$\begin{aligned} &s_x(y) = y \quad \forall\, y \in X \\ &s_x(\#) = X \\ &s_x(\nabla) = x \\ &s_x(\Delta) = (\overline{FG(h(x))} \cap X^{\leq l})X^* \text{ où } l = |h(x)|. \end{aligned}$$

On vérifie que

$$(\star) \qquad \overline{FG(\alpha)} = (\overline{FG(\alpha_{[l_1]})} \cap X^{\leq l_1})X^* \bigcup (\bigcup_{x \in X} s_x(L)).$$

Il est évident que

$$(\overline{FG(\alpha_{[l_1]})} \cap X^{\leq l_1})X^* \bigcup (\bigcup_{x \in X} s_x(L)) \subseteq \overline{FG(\alpha)} \qquad \text{propriété(ii)}.$$

Inversement, tous les mots ne commençant pas par un facteur gauche de $\alpha_{[l_1]}$ appartiennent à $(\overline{FG(\alpha_{[l_1]})} \cap X^{\leq l_1})X^*$. Vérifions simplement que tous les mots de $\overline{FG(\alpha)}$ qui commencent par $\alpha_{[l_1]}$ sont obtenus. Prenons $w \in \overline{FG(\alpha)} \cap \alpha_{[l_1]}X^*$. Il existe alors une lettre $w_{(p)}$ différente de $\alpha_{(p)}$. Nous pouvons supposer que $w_{[p-1]} = \alpha_{[p-1]}$ avec $p-1 \geq l_1$. Puisque le langage est prolongeable infini, tout facteur gauche $\alpha_{[k-1]}$ a une image de longueur supérieure à $k-1$. Alors il existe une lettre $\alpha_{(k)}$ avec $p-1 \geq k > l_0$ telle que $|h(\alpha_{[k-1]})| = r$, $|h(\alpha_{(k)})| = l$ et $r+1 \leq p \leq r+l$.D'où

$$\begin{aligned} w &\in \alpha_{[k-1]}\alpha_{(k)}X^{r-k}(\overline{FG(h(\alpha_{(k)}))} \cap X^{\leq l})X^* \\ &\in s_{\alpha_{(k)}}(\alpha_{[k-1]} \nabla \#^{r-k}\Delta) \\ &\in s_{\alpha_{(k)}}(L). \end{aligned}$$

L'égalité $(\star)$ est obtenue.

La famille des langages algébriques étant fermée par union et par substitution rationnelle, le centre du langage est co-algébrique. □

Remarquons que, si le morphisme prolongeable h est non effaçant, le complémentaire du centre appartient à $\mathbf{C}(\{a^nb^n \;/\; n \in \mathbf{N}\})$.

En effet, l'inclusion $\varphi^{-1}(\{a^nb^n \;/\; n \in \mathbf{N}\}) \subseteq \varphi^{-1}(D_1^*)$ est vérifiée quelque soit le morphisme h. Dans l'autre sens, si le morphisme h est non effaçant, nous avons, pour tout mot $w = u \nabla \#^m \Delta$, $\varphi(w) \in \{a^nb^n \;/\; n \in \mathbf{N}\}$ car

$$|\varphi(w)|_a = |\varphi(u)|_a = |h(u)| - |u| \geq 0 \quad \text{et} \quad |\varphi(w)|_b = |\varphi(u)|_b + m + 1 = m + 1.$$

Donc, si le morphisme est non effaçant, $L = \varphi^{-1}(D_1^*) \cap R = \varphi^{-1}(\{a^nb^n \;/\; n \in \mathbf{N}\}) \cap R$.

Nous pouvons aussi constater que le complémentaire du centre d'un $D0L$-langage prolongeable peut être obtenu par les mêmes transductions rationnelles appliquées sur le langage de Dyck restreint $D_1'^*$.Nous avons alors un langage

$$L' = \{u \bigtriangledown \#^m \triangle \ / \ u \in \alpha_{[l_0]}X^* \ \ et \ \ m = |h(u)| - |u| - 1 \ \ et \ \ \forall\, v \in FG(u),\ |h(v)| \geq |v|\}.$$

Le reste de la preuve n'est pas modifié. Dans un sens l'inclusion est évidente

$$\forall\, x \in X \qquad s_x(\varphi^{-1}(D_1'^*) \cap R) \subseteq s_x(\varphi^{-1}(D_1^*) \cap R) \subseteq \overline{FG(\alpha)}.$$

Dans l'autre sens, en conservant les mêmes notations, si $w \in s_{\alpha_{(k)}}(\alpha_{[k-1]} \bigtriangledown \#^{r-k} \triangle)$, comme $|h(\alpha_{[k-1]})| > k-1$, on a $w \in s_{\alpha_{(k)}}(L')$.

Pour généraliser ce résultat aux $D0L$-langages, nous introduisons la notion de $D0L$-système et de $D0L$-langage unitaires, construits à partir de morphismes particuliers. Les propriétés des $D0L$-systèmes unitaires et la décomposition d'un $D0L$-système quelconque en $D0L$-systèmes unitaires nous permettront de démontrer que le centre d'un $D0L$-langage est co-algébrique.

L'adhérence d'une union finie de langages est égale à l'union des adhérences de ces langages [13]. De plus, l'adhérence d'un langage est vide si et seulement si ce langage est fini. Nous en déduisons immédiatement une propriété que nous utiliserons par la suite.

Propriété III.A.2

Soient $G = \langle X, h, u\rangle$ un $D0L$-système, n_0 un entier positif et n un entier strictement positif. Alors

$$Adh(L(G)) = \bigcup_{i=0}^{n-1} Adh(L(G_i))$$

où $\forall\, i \in [0, n-1], G_i$ désigne le $D0L$-système $\langle X, h^n, h^{n_0+i}(u)\rangle$.

Nous utiliserons le lemme suivant démontré par T. Head dans [35] :

Lemme III.A.3 [35]

Soient $G_1 = \langle X, h, u_1\rangle$ et $G_2 = \langle X, h, u_1u_2\rangle$ deux $D0L$-systèmes. Alors, si $Adh(L(G_1))$ est non vide, $L(G_1)$ et $L(G_2)$ ont la même adhérence.

Nous distinguons dans la suite, pour un $D0L$-système $G = \langle X, h, u\rangle$, deux sous-ensembles disjoints de l'alphabet X : X_I, l'ensemble des lettres engendrant, par itérations successives du morphisme, des langages infinis et X_F, l'ensemble de celles engendrant des langages finis. Le lemme III.A.3 nous permet d'affirmer que, parmi les lettres de X_I apparaissant dans l'axiome, seule celle située le plus à gauche détermine l'adhérence.

Il est clair que

$$X_F = \{x \in X \ / \ \exists\, i \neq j \in \mathbf{N} \quad \textit{tels que } h^i(x) = h^j(x)\}$$
$$X_I = \{x \in X \ / \ \forall\, i \neq j \in \mathbf{N},\ h^i(x) \neq h^j(x)\} \quad \text{et} \quad X = X_F \cup X_I.$$

Remarquons aussi que $w \in X_F^*$ implique $h^n(w) \in X_F^*$, pour tout $n \in \mathbf{N}$. De plus, il existe deux entiers $r_0 \in \mathbf{N}$ et $r \in \mathbf{N}_+$ vérifiant, pour tout $w \in X_F^*$, $h^{r_0}(w) = h^{r_0+r}(w)$. Enfin, $x \in X_I$ implique que, pour tout entier $n \in \mathbf{N}$, $h^n(x)$ appartient à $X^*X_IX^* = X_F^*X_IX^*$.

Définition

Un $D0L$-système $G = \langle X, h, u\rangle$ est unitaire s'il existe une factorisation $u = u_1u_2u_3$ avec u_1, u_2, $u_3 \in X^*$, $h(u_1) = u_1$, $u_2 \in X_F^*X_IX^*$, $h(u_2) = w_1u_2w_3$ avec w_1, $w_3 \in X^*$ et $h(w_1) = w_1$.

Nous appellerons $D0L$-langage unitaire, un langage engendré par un $D0L$-système unitaire.

Lemme III.A.4

Soit $G = \langle X, h, u\rangle$ un $D0L$-système unitaire engendrant le langage L. Alors, $Adh(L)$ contient exactement un élément et L^c est co-algébrique.

Preuve :

Deux cas peuvent se présenter :

1. $w_1 = \varepsilon$

D'après l'hypothèse, u se factorise en $u_1u_2u_3$ avec u_2 engendrant un langage infini. Donc, $G = \langle X, h, u_1u_2u_3\rangle$ et $G' = \langle X, h, u_1u_2\rangle$ engendrent deux langages L et L' qui ont même adhérence (lemme III.A.3), donc même centre.

Puisque le morphisme h est prolongeable en u_1u_2, le centre $L'^c = L^c$ est co-algébrique.

2. $w_1 \neq \varepsilon$

Alors, $Adh(L) = \{u_1w_1^\omega\}$ et $L^c = FG(u_1v_1^*)$ est rationnel. Donc $\overline{L^c}$ est rationnel. □

Montrons maintenant que tout $D0L$-langage infini est égal à l'union d'un langage fini et d'une union finie de $D0L$-langages unitaires.

Proposition III.A.5

Soit $G = \langle X, h, u\rangle$ un D0L-système. Si $L = L(G)$ est infini, alors il existe deux entiers $n \in \mathbf{N}_+$ et $n_0 \in \mathbf{N}$ tels que, pour tout $i \in \mathbf{N}$, $G_i = \langle X, h^n, h^{n_0+i}(u)\rangle$ est unitaire.

Preuve :

Dans un premier temps, on établit le résultat suivant : il existe $x \in X_I, t_0 \in \mathbf{N}$ et $t \in \mathbf{N}_+$ tels que $h^{t_0}(u) = v_1 x v_3$ avec $v_1 \in X_F^*$, $v_3 \in X^*$, $h^t(x) = w_1' x w_3'$, $w_1' \in X_F^*$ et $w_3' \in X^*$.

En effet, L infini implique que, quelle que soit l'itération i, $h^i(u)$ appartient à $X_F^* X_I X^*$. Soit k le nombre d'éléments de X_I. Pour $i \in [0, k]$, $h^i(u)$ représente $k+1$ itérations. Donc un élément x de X_I a été utilisé deux fois en tant que première lettre de X_I apparaissant dans $h^i(u)$. Il existe donc deux entiers t et t_0, avec $0 \leq t_0 < t_0 + t \leq k$ pour lesquels x est facteur de $h^{t_0}(u)$ et de $h^t(x)$ et, dans les deux cas, n'est précédé que de lettres de X_F.

Nous avons $v_1 \in X_F^*$ et $w_1' \in X_F^*$; il existe donc p_0, $q_0 \in \mathbf{N}$ et p, $q \in \mathbf{N}_+$ tels que

$$h^{p_0}(v_1) = h^{p_0+p}(v_1) \quad \text{et} \quad h^{q_0}(w_1') = h^{q_0+q}(w_1').$$

Posons $n_0 = p_0 + q_0 + t_0$ et $n = ptq$. Pour tout $i \in \mathbf{N}$, nous avons

$$h^{n_0+i}(u) = h^{p_0+q_0+i}(h^{t_0}(u)) = h^{p_0+q_0+i}(v_1 x v_3).$$

Posons

$$u_1 = h^{p_0+q_0+i}(v_1), \quad u_2 = h^{p_0+q_0+i}(x) \quad \text{et} \quad u_3 = h^{p_0+q_0+i}(v_3).$$

Nous vérifions que :

(a) $h^n(u_1) = h^{ptq}(u_1) = h^{ptq+p_0+q_0+i}(v_1) = h^{q_0+i}(h^{p_0+ptq}(v_1)) = h^{q_0+i}(h^{p_0}(v_1)) = u_1$;

(b) puisque $x \in X_I$, nous avons $u_2 \in X_F^* X_I X^*$;

(c)
$$\begin{aligned}
h^n(u_2) &= h^{ptq}(h^{p_0+q_0+i}(x)) \\
&= h^{p_0+q_0+i+(pq-1)t}(w_1' x w_3') \\
&= h^{p_0+q_0+i+(pq-1)t}(w_1') h^{p_0+q_0+i+(pq-2)t}(w_1') \ldots h^{p_0+q_0+i}(w_1') h^{p_0+q_0+i}(x) \\
&\qquad h^{p_0+q_0+i}(w_3') h^{p_0+q_0+i+t}(w_3') \ldots h^{p_0+q_0+i+(pq-1)t}(w_3') \\
&= w_1 u_2 w_3
\end{aligned}$$

en posant

$$w_1 = h^{p_0+q_0+i+(pq-1)t}(w_1') \ldots h^{p_0+q_0+i}(w_1')$$
$$w_3 = h^{p_0+q_0+i}(w_3') \ldots h^{p_0+q_0+i+(pq-1)t}(w_3') \;;$$

(d) puisque, $w_1' \in X_F^*$, nous avons $h^{p_0+q_0+i+(pq-j)t}(w_1') \in X_F^*$, pour tout $j \in [1,pq]$.
De plus, pour tout $j \in [1,pq]$, on a :

$$\begin{aligned} h^{ptq}(h^{p_0+q_0+i+(pq-j)t}(w_1')) &= h^{p_0+i+(pq-j)t}(h^{q_0+ptq}(w_1')) \\ &= h^{p_0+i+(pq-j)t}(h^{q_0}(w_1')) \\ &= h^{p_0+q_0+i+(pq-j)t}(w_1') \ ; \end{aligned}$$

donc $h^n(w_1) = h^{ptq}(w_1) = h^{ptq}(h^{p_0+q_0+i+(pq-1)t}(w_1') \dots h^{p_0+q_0+i}(w_1')) = w_1$.

Pour tout $i \in \mathbf{N}$, $G_i = \langle X, h^n, h^{n_0+i}(u) \rangle$ est donc unitaire. □

Des propositions III.A.4 et III.A.5 , nous pouvons déduire que l'adhérence d'un $D0L$-langage est toujours fini, résultat qui avait été annoncé dans [22].

En effet, si le $D0L$-langage est fini, son adhérence est vide. Dans le cas contraire, le $D0L$-langage est égal à l'union d'un langage fini, dont l'adhérence est vide, et de l'union, pour $i \in [0, n-1]$, des $D0L$-langages unitaires $L(G_i)$ définis par la proposition III.A.5, dont l'adhérence contient un seul élément. Nous avons donc

Proposition III.A.6

L'adhérence d'un $D0L$-langage est fini.

Pour démontrer que le centre d'un $D0L$-langage est co-algébrique, nous utilisons le lemme suivant :

Lemme III.A.7

Soient $\alpha_1, \alpha_2, \dots, \alpha_n \in X^\omega$. Si, pour tout $i \in [1,n]$, $L_i = \overline{FG(\alpha_i)}$ est algébrique, alors $L_1 \cap L_2 \cap \dots \cap L_n$ est encore algébrique.

Preuve :

On peut supposer, sans nuire à la généralité, que tous les α_i sont différents. Il existe alors un entier $k \in \mathbf{N}_+$ tel que les préfixes de longueur k des α_i sont différents. Posons $\alpha_i = \alpha_{i[k]}\beta_i$, pour tout $i \in [1,n]$. Nous avons $i \neq j$ implique $\alpha_{i[k]} \neq \alpha_{j[k]}$.

X^* peut s'écrire :

$$X^* = \left(\bigcup_{i=1}^{n} FG(\alpha_{i[k]})\right) \bigcup \left(\bigcup_{i=1}^{n} \alpha_{i[k]} X^*\right) \bigcup \left(\bigcap_{i=1}^{n} \overline{FG(\alpha_{i[k]}X^*)}\right)$$

en fonction des différents préfixes $\alpha_{i[k]}$. Le langage

$$L = \bigcap_{i=1}^{n} L_i$$

est examiné par rapport à cette décomposition.

$$L \cap FG(\alpha_{i[k]}) \subset L_i \cap FG(\alpha_{i[k]}) = \emptyset \quad \text{pour tout } i \in [1,n].$$

$$L \cap \alpha_{i[k]}X^* = L_i \cap \alpha_{i[k]}X^* \quad \text{pour tout } i \in [1,n].$$

$$L \bigcap \left(\bigcap_{i=1}^{n} \overline{FG(\alpha_{i[k]}X^*)} \right) = \bigcap_{i=1}^{n} \overline{FG(\alpha_{i[k]}X^*)}.$$

Donc nous pouvons écrire

$$L = \left(\bigcup_{i=1}^{n} (L_i \cap \alpha_{i[k]}X^*) \right) \bigcup \left(\bigcap_{i=1}^{n} \overline{FG(\alpha_{i[k]}X^*)} \right).$$

L_i étant algébrique, $\alpha_{i[k]}X^*$ et $\overline{FG(\alpha_{i[k]}X^*)}$ étant rationnels, pour tout $i \in [1,n]$, nous déduisons que

$$L = \bigcap_{i=1}^{n} L_i$$

est algébrique. □

Nous avons vu que si l'adhérence n'était pas vide, elle pouvait être déterminée par une union finie de $D0L$-langages unitaires (propositions III.A.2 et III.A.5):

$$Adh(L(G)) = \bigcup_{i=0}^{n-1} Adh(L(G_i)) = \{\alpha_0, \alpha_1, \ldots, \alpha_m\} \text{ avec } m \leq n-1.$$

Chaque α_j est l'adhérence d'un $L(G_i)$ et, G_i étant unitaire, $\overline{L(G_i)^c} = \overline{FG(\alpha_j)}$ est algébrique (lemme III.A.4). Le complémentaire d'une union finie étant l'intersection des complémentaires,

$$\overline{L(G)^c} = \overline{\bigcup_{j=0}^{m} FG(\alpha_j)} = \bigcap_{j=0}^{m} \overline{FG(\alpha_j)},$$

le lemme III.A.7 nous permet de conclure par la proposition suivante :

Proposition III.A.8

Le centre d'un $D0L$-langage est co-algébrique.

Nous pouvons préciser ce résultat. Le complémentaire du centre d'un $D0L$-langage prolongeable appartient à *Rocl*, la famille des langages à un compteur restreint. Il en va de même pour le complémentaire du centre d'un $D0L$-langage unitaire. *Rocl* étant un cône rationnel principal est fermé par union ; les preuves du lemme et de la proposition précédents nous permettent de conclure que le complémentaire du centre d'un $D0L$-langage appartient à $Rocl = \mathbf{C}(D_1'^*)$ et au cône $\mathbf{C}(D_1^*)$.

Nous trouvons pour les gsm prolongeables itérés la proposition annoncée par J. Berstel [11].

Proposition III.A.9

Le centre d'un langage engendré à partir de $w_0 \in X^+$ en itérant un gsm prolongeable en w_0 est co-algébrique.

Preuve :

Soit $M = \langle X, X, Q, q_0, \delta, \lambda \rangle$ un gsm prolongeable en w_0.

Si le langage L engendré par itération de M à partir de w_0 est fini, l'adhérence est vide et le centre est rationnel.

Si le langage est infini, l'adhérence contient, comme pour les morphismes prolongeables, un seul mot infini α qui vérifie :

$$\begin{array}{ll} & \alpha_{[l_0]} = w_0 \qquad \delta(q_0, \alpha_{[l_0]}) = \alpha_{[l_1]} \\ & \forall\, k > 0 \quad \lambda(q_{j_{k-1}}, \alpha_{(k)}) = q_{j_k} \text{ avec } q_{j_0} = q_0 \\ \text{(i)} & \forall\, k > l_0 \quad \delta(q_{j_{k-1}}, \alpha_{(k)}) = \alpha_{[r+1, r+l]} \text{ où } r = |\delta(q_0, \alpha_{[k-1]})| \text{ et } l = |\delta(q_{j_{k-1}}, \alpha_{(k)})| \end{array}$$

Nous pouvons séparer le complémentaire du centre en deux sous-ensembles. Le premier contient les mots qui ne commencent pas par un facteur gauche de $\alpha_{[l_1]}$; ce sous-ensemble est rationnel. Le deuxième sous-ensemble contient les mots qui commencent par $\alpha_{[l_0]}$ mais qui ne vérifient pas la contrainte (i) pour au moins une de leurs lettres.

$$\begin{array}{lll} & w \in \overline{FG(\alpha)} \iff & \text{soit } w \in (\overline{FG(\alpha_{[l_1]})} \cap X^{\leq l_1})X^* \\ \text{(ii)} & & \text{soit } \exists\, k \quad l_0 < k < |w| = n \text{ tel que} \\ & & \qquad w_{[r+1,n]} \in (\overline{FG(\delta(q_{j_{k-1}}, w_{(k)}))} \cap X^{\leq l})X^* \\ & & \qquad \text{où } q_{j_{k-1}} = \lambda(q_0, w_{[k-1]}), \\ & & \qquad\quad r = |\delta(q_0, w_{[k-1]})| \text{ et } l = |\delta(q_{j_{k-1}}, w_{(k)})| \end{array}$$

Remarquons que le mot $w_{[k]}$ appartient à $Dom(gsm)$ puisque la lettre qui vérifie la propriété (ii) peut être trouvée dans un facteur gauche de w qui appartient à $FG(\alpha)$.

Comme pour les morphismes, nous construisons le langage qui permet de réserver la place de l'image du début de mot. Mais il faut tenir compte des états. Nous construisons donc

$$L = \{u \nabla_q \#^m \Delta \ / \ u \in \alpha_{[l_0]} X^* \cap Dom(gsm), \ \ q = \lambda(q_0, u) \ \ et \ \ m = |\delta(q_0, u)| - |u| - 1\}$$

où $\nabla_Q = \{\nabla_q \ / \ q \in Q\}$ et $Z = \{\#, \Delta\}$ sont des ensembles disjoints de X. Nous pouvons obtenir L par transduction du langage de Dyck D_1^*. Pour cela, posons :

$$X' = X \cup \{\nabla\}$$

$$R_1 = \{(q_i, x, q_j) \ / \ x \in X \quad q_j = \lambda(q_i, x)\}^* \{(q, \nabla, q') \ / \ q, \ q' \in Q\}$$

$$R_2 = \bigcap_{q_i \neq q_j} \overline{(Q \times X \times Q)^*(Q \times X \times \{q_i\})(\{q_j\} \times X' \times Q)(Q \times X' \times Q)^*}$$

$$R_3 = (q_0, \alpha_{(1)}, \lambda(q_0, \alpha_{(1)})) \ldots (\lambda(q_0, \alpha_{[l_0-1]}), \alpha_{[l_0]}, \lambda(q_0, \alpha_{[l_0]}))(Q \times X' \times Q)^*$$

Le langage $R = R_1 \cap R_2 \cap R_3$ contient les mots représentant des suites de transitions autorisées par le gsm, commençant par les transitions dues à $\alpha_{[l_1]}$ et terminées par une transition (q, ∇, q') où q est l'état atteint par les transitions précédentes.

Nous pourrions vérifier que $L = \mu \circ \cap R\#^*\Delta \circ \varphi(D_1^*)$ où φ et μ sont deux morphismes de $(Q{\times}X'{\times}Q \ \cup \ Z)^*$ dans $\{a, b\}^*$ et dans $(X \cup \nabla_Q \cup Z)^*$ definis par :

$\varphi((q_i, x, q_j)) = a^{l-1}$ si $l = |\delta(q_i, x)| > 0$ et $q_j = \lambda(q_i, x)$
$\varphi((q_i, x, q_j)) = b$ si $l = |\delta(q_i, x)| = 0$ et $q_j = \lambda(q_i, x)$
$\varphi((q, \nabla, q')) = b$ pour tout $q, q' \in Q$ $l = |\delta(q_i, x)| = 0$
$\varphi(\Delta) = \varepsilon$
$\varphi(\#) = b$
$\mu((q_i, x, q_j)) = x$
$\mu((q, \nabla, q')) = \nabla_q$
$\mu(\#) = \#$
$\mu(\Delta) = \Delta$

Le langage L est donc algébrique.

Nous définissons alors, pour chaque lettre x de X et pour chaque état q de Q la substitution s_{xq} de $(X \cup \{\nabla_Q\} \cup Z)^*$ dans X^*

$$\begin{array}{lll} s_{xq}(y) & = y & \forall y \in X \\ s_{xq}(\#) & = X & \\ s_{xq}(\nabla_{q'}) & = x & \text{si } q' = q \text{ et } \lambda(q,x) \text{ est défini} \\ & = \emptyset & \text{sinon} \\ s_{xq}(\Delta) & = (\overline{FG(\delta(q,x))} \cap X^{\leq l})X^* & \text{où } l = |\delta(q,x)| \end{array}$$

Nous vérifions que

$$\overline{FG(\alpha)} = (\overline{FG(\alpha_{[l_1]})} \cap X^{\leq l_1})X^* \bigcup (\bigcup_{\substack{x \in X \\ q \in Q}} s_{xq}(L)).$$

De façon évidente,

$$(\overline{FG(\alpha_{[l_1]})} \cap X^{\leq l_1})X^* \bigcup (\bigcup_{\substack{x \in X \\ q \in Q}} s_{xq}(L)) \subseteq \overline{FG(\alpha)}.$$

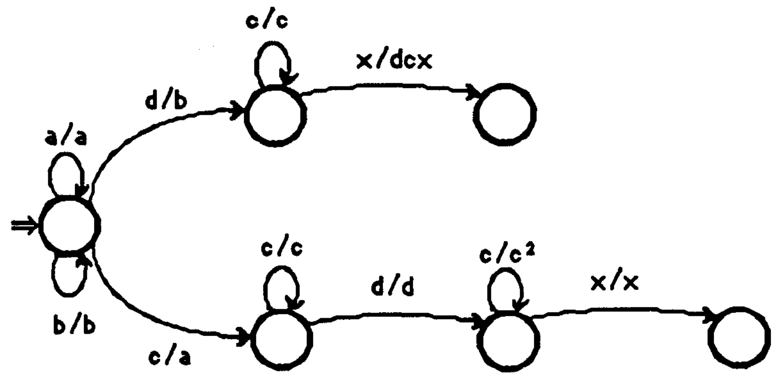
Dans l'autre sens, si un mot w appartient $\overline{FG(\alpha)}$, soit il ne commence pas par un facteur gauche de $\alpha_{[l_1]}$, soit il existe une lettre $w_{(p)}$ différente de $\alpha_{(p)}$. Dans ce dernier cas, nous pouvons supposer que $w_{[p-1]} = \alpha_{[p-1]}$ avec $p - 1 \geq l_1 > l_0$. On constatera que $w_{[p-1]} \in Dom(gsm)$. Puisque le gsm engendre un langage prolongeable infini, il existe une lettre $w_{(k)} = \alpha_{(k)}$ où $l_0 < k \leq p-1$ telle que $\lambda(q_0, w_{[k-1]}) = q_{k-1}$, $|\delta(q_0, w_{[k-1]})| = r$, $|\delta(q_{k-1}, w_{(k)})| = l$ et $r + 1 \leq p \leq r + l$. D'où

$$\begin{aligned} w &\in w_{[k-1]} w_{(k)} X^{r-k} (\overline{FG(\delta(q_{k-1}, w_k))} \cap X^{\leq l})X^* \\ &\in s_{w_{(k)} q_{k-1}}(w_{[k-1]} \nabla_{q_{k-1}} \#^{r-k} \Delta) \\ &\in s_{w_{(k)} q_{k-1}}(L) \end{aligned}$$

La famille des langages algébriques étant fermée par union finie et substitution rationnelle, $\overline{FG(\alpha)}$ est un langage algébrique. □

Alors que la généralisation d'un morphisme prolongeable à un morphisme quelconque ne pose pas de difficulté, la généralisation à un gsm quelconque ne peut pas être faite.

Le gsm ci-dessous fournit un exemple de fonction séquentielle engendrant par itération à partir de l'axiome $w_0 = cdcx$ un langage dont le centre n'est pas co-algébrique.



Les premières dérivations sont

$$cdcx \Rightarrow adc^2x \Rightarrow abc^2dcx \Rightarrow abacdc^2x \Rightarrow aba^2dc^4x \Rightarrow aba^2bc^4dcx \overset{*}{\Longrightarrow} aba^2ba^4bc^{16}dcx \Rightarrow \dots$$

Par induction, on vérifie que le gsm engendre par itération un seul mot infini

$$\alpha = Adh(abA^* \cap A^*a^+b) \quad \text{où } A = \{a^nba^{2^n}b \;/\; n \in \mathbb{N}\}.$$

En appliquant le résultat suivant, démontré par A. Grazon [33], on vérifie que $\overline{FG(\alpha)}$ n'est pas algébrique.

Proposition III.A.10 [33]

Soit $\alpha = a^{s_1}ba^{s_2}b\dots a^{s_n}b$ *un mot infini sur* $\{a,b\}^\omega$. *Si, pour tout* $K \in \mathbb{N}$, *il existe* $z \in \mathbb{N}$ *tel que*

$$s_{z+1} \geq K! + K + (K-1)z + (K-1)\sum_{i=1}^{z} s_i,$$

alors $\overline{FG(\alpha)}$ *n'est pas un langage algébrique.*

Nous terminons notre examen des résultats positifs en montrant que l'ensemble des facteurs gauches de la séquence externe engendrée par un Tag-système uniforme est co-algébrique.

La démonstration ne pose pas de difficulté majeure. Ce que le morphisme cache en passant de la séquence interne à la séquence externe, le caractére uniforme nous permet de le découvrir en partie.

La proposition III.A.11 résulte directement du théorème 1 dans [18]. Il s'ensuit une progression de propriétés simples vers la propriété III.A.14 qui nous sera utile.

Proposition III.A.11 [18]

Soit $T = \langle X, h, \alpha_1, g, Y\rangle$ un Tag-système uniforme. Il existe n, $p \in \mathbf{N}$ avec $0 \leq n < p$ tels que, pour tout x, $y \in X$, $g(h^n(x)) = g(h^n(y))$ implique $g(h^p(x)) = g(h^p(y))$.

Cette propriété reste vraie si nous remplaçons les lettres x et y par des mots u et v, puisque $h^n(u) = h^n(v)$ signifie $h^n(u_{(1)}) = h^n(v_{(1)})$, $h^n(u_{(2)}) = h^n(v_{(2)})$...quand il s'agit de morphismes uniformes.

Propriété III.A.12

Soit $T = \langle X, h, \alpha_1, g, Y\rangle$ un Tag-système uniforme. Il existe n, $p \in \mathbf{N}$ avec $0 \leq n < p$ tels que, pour tout u, $v \in X^+$, $g(h^n(u)) = g(h^n(v))$ implique $g(h^p(u)) = g(h^p(v))$.

Comme $p > n$, on peut appliquer de nouveau la propriété III.A.12 en prenant $u_1 = h^{p-n}(u)$ et $v_1 = h^{p-n}(v)$. Par induction, on en déduit une suite d'égalités. Donc pour une période assez grande $q = k(p-n) \geq n$, on obtient la propriété suivante :

Propriété III.A.13

Soit $T = \langle X, h, \alpha_1, g, Y\rangle$ un Tag-système uniforme. Il existe $q \in \mathbf{N}_+$ tel que, pour tout u, $v \in X^+$, $g(h^q(u)) = g(h^q(v))$ implique $g(h^{\lambda q}(u)) = g(h^{\lambda q}(v))$, pour tout $\lambda \in \mathbf{N}_+$.

La séquence *intseq* restant la même si on remplace h par $h_1 = h^q$, nous pouvons écrire la propriété:

Propriété III.A.14

Soit $T = \langle X, h, \alpha_1, g, Y\rangle$ un Tag-système uniforme. Il existe un Tag-système uniforme $T_1 = \langle X, h_1, \alpha_1, g, Y\rangle$ engendrant les mêmes séquences interne et donc externe et vérifiant , pour tout $u, v \in X^+$, $g(h_1(u)) = g(h_1(v))$ implique $g(h_1^\lambda(u)) = g(h_1^\lambda(v))$, pour tout $\lambda \in \mathbf{N}_+$.

Cette propriété nous permet de retrouver une construction similaire à celle des propositions III.A.1 et III.A.9.

Proposition III.A.15

L'ensemble des facteurs gauches de la séquence externe engendrée par un Tag-système uniforme est co-algébrique.

Preuve :

Soit $T = \langle X, h, x_1, g, Y\rangle$ un Tag-système uniforme de module l engendrant les séquences $intseq$ et $extseq$. Nous pouvons supposer que T vérifie l'implication de la propriété précédente.

Considérons le langage L défini par

$$L = \{u \bigtriangledown \#^m \bigtriangleup \ / \ u \in extseq_{[l]} Y^{kl-2l}, \quad k \geq 2 \quad et \quad m = (k-1)l^2 - kl\}$$

où $Z = \{\#, \bigtriangledown, \bigtriangleup\}$.

Ce langage est construit pour que $\bigtriangledown$ soit remplacé par l'image de la k-ième lettre et $\bigtriangleup$ par l'image de cette image.

Nous vérifions aisément que L est algébrique, puisque

$$L = \mu \circ \cap R \circ \varphi^{-1}(\{a^n b^n \ / \ n \in \mathbb{N}\})$$

où μ, φ sont des morphismes de $(Y \cup Z)^*$ dans $\{a, b\}^*$ et dans $(Y \cup Z)^*$ définis par

$$\begin{array}{ll}
\varphi(y) = a^{l-1} \quad \forall\, y \in Y & \mu(y) = y \quad \forall\, y \in Y \\
\varphi(\#) = b & \mu(\#) = \# \\
\varphi(\flat) = b & \mu(\flat) = \varepsilon \\
\varphi(\bigtriangledown) = b & \mu(\bigtriangledown) = \bigtriangledown \\
\varphi(\bigtriangleup) = \varepsilon & \mu(\bigtriangleup) = \bigtriangleup
\end{array}$$

et le langage rationnel $R = extseq_{[l]}(Y^l)^* \bigtriangledown \flat^{l-1}\#^*\bigtriangleup$.

Définissons, pour chaque x de X, la substitution s_x de $(Y \cup Z)^*$ dans Y^*, par

$$\begin{aligned}
s_x(y) &= y \quad \text{pour tout } y \in Y \\
s_x(\bigtriangledown) &= g(h(x)) \\
s_x(\#) &= Y \\
s_x(\bigtriangleup) &= (\overline{FG(g(h^2(x)))} \cap Y^{\leq l^2})Y^*
\end{aligned}$$

On démontre facilement que

$$\overline{FG(extseq)} = (\overline{FG(extseq_{[l^2]})} \cap Y^{\leq l^2})Y^* \bigcup (\bigcup_{x \in X} s_x(L))$$

prouvant ainsi que $FG(extseq)$ est co-algébrique.

En effet, la propriété III.A.14 nous indique que, si la substitution $s_x(\nabla)$, pour $x \in X$, correspond au facteur de la séquence externe $s_{intseq_{(k)}}(\nabla)$ pour un certain k, alors $s_x(\Delta)$ correspond à $s_{intseq_{(k)}}(\Delta)$. On en déduit que $s_x(L) \subseteq \overline{FG(extseq)}$ et

$$(\overline{FG(extseq_{[l^2]})} \cap Y^{\leq l^2})Y^* \bigcup (\bigcup_{x \in X} s_x(L)) \subseteq \overline{FG(extseq)}.$$

Dans l'autre sens, prenons un mot w dans $\overline{FG(extseq)}$ et plus particulièrement w appartenant à $\overline{FG(extseq)} \cap extseq_{[l^2]}Y^*$. Il existe alors un entier $p > l^2$ tel que $w_{[p-1]} = extseq_{[p-1]}$ et $w_{(p)} \neq extseq_{(p)}$. Comme $p > l^2$, l'entier k vérifiant $(k-1)l^2 < p \leq kl^2$ est ≥ 2, nous avons donc $w_{[kl-l+1,kl]} = extseq_{[kl-l+1,kl]}$ et $w_{[kl^2-l^2+1,kl^2]} \neq extseq_{[kl^2-l^2+1,kl^2]}$. On en déduit que $w \in s_{extseq_{[k]}}(L)$. □

Remarquons que les gsm itérés sont plus puissants que les Tag-systèmes.

Proposition III.A.16

Toute séquence externe engendrée par un Tag-système est l'adhérence d'un langage engendré par un gsm itéré.

Preuve :

Soit $T = \langle X, h, x_1, g, Y \rangle$ un Tag-système engendrant les séquences *intseq* et *extseq*. On peut supposer $X \cap Y = \emptyset$.

Construisons le gsm suivant $M = \langle X \cup Y \cup \{\nabla\}, X \cup Y \cup \{\nabla\}, Q, q_0, \delta, \lambda \rangle$ où

$$\begin{array}{lll}
Q = \{q_0\} \cup \{q_x \ / \ x \in X\} \cup \{q_f\} & & \\
\delta(q_0, y) = y & \lambda(q_0, y) = q_0 & \forall\, y \in Y \\
\delta(q_0, x) = g(x) & \lambda(q_0, x) = q_x & \forall\, x \in X \\
\delta(q_x, x') = x' & \lambda(q_x, x') = q_x & \forall\, x, x' \in X \\
\delta(q_x, \nabla) = h(x)\nabla & \lambda(q_x, \nabla) = q_f & \forall\, x \in X
\end{array}$$

Les itérations commenceront sur l'axiome $extseq_{[1]}intseq_{[2,l_1]}\nabla$ où $l_1 = |h(intseq_{(1)})|$. On vérifie aisément qu'à chaque étape, la séquence croît sous la forme $extseq_{[k]}intseq_{[k+1,l]}\nabla$ où $l = |h(intseq_{[k]})|$. En effet, les alphabets X et Y étant disjoints, la lecture de $extseq_{[k]}$ conduit à une réécriture du mot et maintient dans l'état q_0. La première lettre n'appartenant pas à Y est $intseq_{(k+1)}$ qui est transformée en $extseq_{(k+1)}$ et provoque la transition à l'état indicé par la lettre $intseq_{(k+1)}$. Les autres lettres de la séquence $intseq$ sont lues et recopiées sans changement d'état. Finalement, le marqueur de fin ∇ engendre $h(intseq_{(k+1)})\nabla$. □

3) conclusion

L'analyse d'autres familles de L-systèmes fournit des exemples de langages dont le centre n'est pas co-algébrique. Prenons par exemple le $DT0L$-système $G = \langle\{a,b\},h_1,h_2,ab\rangle$ avec h_1 et h_2 définis par

$$\begin{array}{ll} h_1(a) = a^2 & h_2(a) = a \\ h_1(b) = b & h_2(b) = b^2 \end{array}$$

Nous avons $Adh(L(G)) = \{a^\omega\} \cup \{a^{2^n}b^\omega \ / \ n \in \mathbf{N}\}$ dont l'ensemble des facteurs gauches n'est pas co-algébrique.

Nous pouvons aussi signaler que le complémentaire des facteurs de l'adhérence d'un $D0L$-langage n'est pas toujours algébrique. Il suffit de prendre le $D0L$-système suivant pour s'en convaincre : $G = \langle\{a,b\},h,a\rangle$ avec $h(a) = aba$ et $h(b) = b^2$. On vérifiera que

$$\overline{F(Adh(L(G)))} \cap ab^*a = \{ab^m a \ / \ \nexists n \in \mathbf{N} \quad tel\ que\ m = 2^n\}$$

qui n'est pas un langage algébrique.

Il serait intéressant de comparer la famille des langages co-algébriques avec des familles de langages engendrés par d'autres procédés itératifs. Il existe en particulier toute une hiérarchie de Tag-systèmes [52] engendrant des séquences externes qui sont toutes des adhérences de $CD0L$-langages.

Donnons un exemple de séquence externe engendrée par un Tag-système non uniforme:

$$G = \langle X,h,e,g,Y\rangle \quad \text{avec} \quad X = \{a,b,c,d,e\} \quad \text{et} \quad Y = \{a,d,e\}$$

$$\begin{array}{ll} h(a) = a & g(a) = a \\ h(b) = ba & g(b) = a \\ h(c) = cba & g(c) = a \\ h(d) = dcba & g(d) = d \\ h(e) = edcba & g(e) = e \end{array}$$

Nous obtenons les séquences suivantes :

$$intseq = edcbadcbacbabaadcbacbabaacbabaabaaa\ldots$$
$$extseq = eda^3da^9da^{19}da^{34}da^{59}\ \ldots$$

Si nous notons $extseq = eda^{s_1}da^{s_2}da^{s_3}\ldots$, l'équation suivante détermine la valeur de s_n en fonction de n et on en déduit la relation qui unit un facteur a^{s_n} aux deux facteurs

$a^{s_{n-1}}$ et $a^{s_{n-2}}$.

$$s_n = \sum_{j=1}^{n-1} \sum_{i=1}^{j} i + 2\sum_{i=1}^{t-1} i + 3(t-1) + 3 \qquad \text{et} \qquad s_n = 2s_{n-1} - s_{n-2} + n + 1$$

Cette dernière relation ne doit plus être vérifiée dans le complémentaire

$$\overline{FG(extseq)} = (\overline{FG(eda^3da^9d)} \cap Y^{\leq 16})Y^*$$
$$\cup\{euda^p da^q dw \ / \ u \in Y^* \ \ w \in (\overline{FG(a^m d)} \cap Y^{\leq m+1})Y^* \ et \ m = 2q - p + |u|_d + 4\}.$$

$\overline{FG(extseq)}$ est un langage algébrique qui appartient à *Rocl*.

Des exemples de mots infinis dont le complémentaire des facteurs gauches n'est pas algébrique ont été fournis par A. Grazon [32]. La structure des séquences externes engendrées par des Tag-systèmes comparée à celle de ces exemples nous incite à émettre la conjecture suivante :

Conjecture

Le complémentaire des facteurs gauches d'une séquence externe engendrée par un Tag-système est algébrique.

A l'inverse, J.M. Autebert et J. Gabarró [6] ont montré récemment qu'il existait des mots infinis, dont l'ensemble des facteurs gauches était co-algébrique, et qui n'étaient engendrés par aucun gsm prolongeable itéré. Ils émettent la conjecture suivante :

Conjecture [6]

Il existe un gsm prolongeable G et un morphisme strictement alphabétique h tels que l'ensemble des facteurs gauches de l'image par h du mot infini engendré par itération du gsm G n'est pas co-algébrique.

B.Décidabilité dans les problèmes d'adhérence et de limite de $D0L$-langages

1) présentation

Nous avons vu qu'un $D0L$-système unitaire engendrait un seul mot infini. Nous pouvons aussi affirmer que l'adhérence d'un $D0L$-langage unitaire est soit un mot ultimement périodique, soit engendrée par un $D0L$-système prolongeable. Les travaux de K. Čulik II et A. Salomaa [22] nous indiquent que, dans le cas des $D0L$-systèmes prolongeables, l'adhérence et la limite sont égales. Dès lors, la décomposition d'un $D0L$-langage en une union finie de langages finis et de $D0L$-langages unitaires nous permet d'obtenir quelques propriétés de décidabilité.

L'égalité de deux mots ultimement périodiques, l'égalité de deux limites de $D0L$-langages prolongeables [21] et l'égalité d'une limite de $D0L$-langage prolongeable et d'un mot ultimement périodique [48] sont trois problèmes décidables. On peut donc décider si deux $D0L$-langages unitaires ont même adhérence et donc même centre.

Si deux $D0L$-langages sont infinis, chaque adhérence est une union finie d'adhérences de $D0L$-langages unitaires. Cette union étant effectivement constructible, il suffit de comparer deux à deux les adhérences des $D0L$-langages unitaires pour décider si deux $D0L$-langages quelconques ont même adhérence. Nous retrouvons donc le théorème de T. Head.

Théorème III.B.1 [34]

On peut décider si deux $D0L$-langages ont même adhérence et donc même centre.

Après avoir montré que les centres de $D0L$-langages étaient co-algébriques, nous pouvons nous interroger sur les centres de $D0L$-langages qui sont algébriques. Nous remarquons alors que, si le centre est algébrique, l'adhérence ne contient que des mots ultimement périodiques.

Lemme III.B.2

Soit $L_\omega = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ un ensemble fini de mots infinis. Les trois propriétés suivantes sont équivalentes :

(1) *$FG(L_\omega)$ est algébrique.*

(2) *$FG(L_\omega)$ est rationnel.*

(3) *L_ω est un ensemble de mots ultimement périodiques.*

Preuve :

Nous avons, de façon triviale, (3) $\Longrightarrow$ (2) $\Longrightarrow$ (1).

Il ne reste à montrer que (1) $\Longrightarrow$ (3).

Remarquons d'abord que les facteurs gauches de chaque mot infini forment un langage algébrique. En effet, les mots infinis étant supposés différents, il existe une longueur K qui les distinguent : $i \neq j$ implique $\alpha_{i[K]} \neq \alpha_{j[K]}$. Pour tout mot α_i dans L_ω, nous avons

$$FG(\alpha_i) = (FG(L_\omega) \cap \alpha_{i[K]}X^*) \cup FG(\alpha_{i[k]}).$$

Donc, si $FG(L_\omega)$ est un langage algébrique, $FG(\alpha_i)$ l'est aussi.

Puisque $FG(\alpha_i)$ est un langage algébrique qui a la propriété de préfixe, $FG(\alpha_i)$ est un langage rationnel [14]. Nous pouvons en déduire que α_i est ultimement périodique. □

L'adhérence d'un $D0L$-langage unitaire est soit ultimement périodique, soit engendrée par un $D0L$-système prolongeable. Dans ce dernier cas, l'adhérence et la limite sont égales. Nous pouvons donc formuler le résultat de J.J. Pansiot de la manière suivante :

Théorème III.B.3 [53]

On peut décider si l'adhérence d'un D0L-langage unitaire est un mot ultimement périodique.

En appliquant toujours le même raisonnement, nous en déduisons la propriété suivante

Proposition III.B.4

On peut décider si le centre d'un D0L-langage est algébrique.

Plusieurs études ont montré qu'on pouvait décider si un $D0L$-système engendrait un mot ultimement périodique [48,53]. Analysons le problème dans l'autre sens. Un mot ultimement périodique est toujours l'adhérence d'un $D0L$-langage. Par contre, il n'en est pas toujours la limite.

2) décidabilité du problème de synthèse

Nous nous intéréssons maintenant au problème de synthèse suivant : un mot ultimement périodique donné peut-il être la limite d'un $D0L$-langage ? Nous avons besoin de quelques propriétés très simples pour montrer que ce problème est décidable.

Lemme III.B.5

Soit $G = \langle X, h, w_1 \rangle$ un $D0L$-système tel que $Lim(L(G)) = \{\alpha\}$ où α est un mot ultimement périodique qui s'écrit uv^ω sous forme primitive.

Alors, le $D0L$-système $G' = \langle X', h', uv \rangle$ engendre un langage qui possède la même limite ; le morphisme h' est la restriction de h à l'alphabet $X' = \alpha b(uv)$.

Preuve :

Par définition de la limite, il existe un facteur gauche de uv^ω, plus long que uv, appartenant à $L(G)$. Notons $w_k = h^{k-1}(w_1)$ ce facteur. La limite du langage engendré par le $D0L$-système $G_0 = \langle X, h, w_k \rangle$ est égale à celle du langage $L(G)$.

Nous pouvons aussi constater que uv engendre, par itérations successives de h, un langage infini. Si ce n'était pas le cas, $L(G_0)$ serait fini et la limite serait vide.

La limite de $L(G_0)$ contient un seul mot qui vérifie donc l'équation $h(uv^\omega) = uv^\omega$. Nous en déduisons que $h^*(uv^\omega) = uv^\omega$ et $h^*(FG(uv^\omega)) \subseteq FG(uv^\omega)$. Donc, le langage engendré par $G' = \langle X, h, uv \rangle$ est inclus dans $FG(uv^\omega)$ et il est infini ; sa limite est uv^ω. Il est alors évident que l'on peut restreindre le morphisme à l'alphabet $X' = \alpha b(uv)$. □

Lemme III.B.6

Soit uv^ω un mot infini mis sous forme primitive. Le $D0L$-langage engendré par $G = \langle X, h, uv \rangle$ a pour limite uv^ω si et seulement si $h^*(uv)$ est infini et s'il existe v_1, $v_2 \in X^*$ tels que $v = v_1 v_2$, $h(u) \in uv^* v_1$ et $h(v) \in v_2 v^* v_1$.

Preuve :

En utilisant les lemmes de T. Head [35] sur les mots ultimement périodiques, on vérifie facilement cette propriété. Remarquons que le $D0L$-système est prolongeable , puisque la limite contient un et un seul mot infini, et engendre un langage $h^*(uv)$ infini.

Supposons que $|h(u)| < |u|$. Nous avons alors $h(u) = u_1$ et $h(v) = u_2 v^i v_1$ avec $u_2 \neq \varepsilon$, $i \in \mathbb{N}$ et $|v_1| < |v|$. Le mot uv^ω ne serait pas sous forme primitive $uv^\omega = h(uv^\omega) = u_1(u_2 v^i v_1)^\omega = u_1 u_2 (v^i v_1 u_2)^\omega$. Donc, $h(u) = uv_1$ avec $v_1 \in FG(v)$.

Supposons que $|h(v)| < |v|$. Nous aurions alors $h(v) = v_2$ et $v^\omega = v_1 v_2^\omega$. Le fait de mettre $v_1 v_2^\omega$ sous forme primitive $v_1' v_2'^\omega$ entraine $|v_2'| \leq |v_2| < |v|$, ce qui contredit le lemme 6.1 de T. Head [35]. Donc $|h(v)| \geq |v|$ et $h(v) \in v_2 v^* v_1''$ pour un certain $v_1'' \in FG(v)$ et $v_1 v_2 = v$. Alors $h(v^\omega) = (v_2 v^* v_1'')^\omega = v_2 v^\omega$ entraine $v_1'' v_2 = v$.

Réciproquement, nous avons $h(uv) \in uv^+v_1$ et $h^*(uv)$ est un langage infini ; le $D0L$-système $G = \langle X, h, uv \rangle$ est donc prolongeable engendrant un langage infini. Pour vérifier que ce mot infini est bien uv^ω, il suffit de constater que l'ensemble des facteurs gauches de uv^ω reste stable par le morphisme h : l'image de uv^kv' où $v' \in FG(v)$ appartient à $uv^+v_1FG(v_2v^*v_1)$. □

Proposition III.B.7

Un mot ultimement périodique étant donné, l'existence d'un D0L-langage ayant ce mot pour limite est décidable.

Preuve :

Supposons acquis le mot uv^ω sous forme primitive. D'après les propriétés précédentes, si un $D0L$-langage a pour limite uv^ω, nous pouvons rechercher le $D0L$-système en prenant uv pour axiome et $X = \alpha b(uv)$ pour alphabet.

Il suffit de montrer que le morphisme h, s'il existe, pourra être trouvé dans

$$\mathbf{H}_k = \{h \in \mathbf{H} \ / \ \forall x \in X \ \ |h(x)| \leq |uv^2|\}$$

qui est un ensemble fini.

Supposons qu'il existe un morphisme g qui convienne avec, pour une certaine lettre x de X, $g(x) = w_1vvw_2$ et $|w_1| \geq |u|$. Alors, il existe un morphisme h qui convient avec $h(y) = g(y)$, pour tout $y \neq x$ et $h(x) = w_1vw_2$. La propriété précédente nous indique qu'il existe v_1, $v_2 \in X^*$ tels que $g(u) \in uv^*v_1$ et $g(v) \in v_2v^*v_1$ avec $v_1v_2 = v$.

Si x appartient à $\alpha b(u)$, comme nous avons $|w_1| \geq |u|$ et uv^ω sous forme primitive, le lemme 6.2 de T. Head [35] nous permet de déduire que $h(u)$ appartient encore à uv^*v_1. De même, si $x \in \alpha b(v)$, nous conservons $h(v) \in v_2v^*v_1$.

Nous vérifions aussi que le langage engendré est toujours infini. Si x appartient à $\alpha b(u)$, nous avons $g(u) \in uv^2v^*v_1$ et donc $h(u) \in uv^+v_1$, $h(v) \in v_2v^*v_1$. Si x appartient à $\alpha b(v)$, nous avons $g(v) \in v_2v^2v^*v_1$ et $h(v) \in v_2v^+v_1$, $h(u) \in uv^*v_1$. Dans les deux cas, $h(uv) \in uv^2v^*v_1$ et $h(v) \in v_2v^*v_1$, ce qui est suffisant pour garantir que le langage est infini.

Nous pouvons décider si $Lim(h^*(uv)) = \{uv^\omega\}$ pour chacun des morphismes de $\mathbf{H}_k$ [48]. □

Nous terminons cette partie en donnant des conditions suffisantes pour la décision du problème de synthèse. Ces conditions s'obtiennent à l'aide de la propriété suivante. Elles permettent de traiter la plupart des cas de mots infinis ultimement périodiques sur un alphabet de deux lettres. Nous retrouvons alors des résultats de P. Séébold [63].

Lemme III.B.8

Soit uv^ω un mot ultimement périodique sous forme primitive. Soit $G = \langle X, h, uv\rangle$ un $D0L$-système tel que $Lim(L(G)) = \{uv^\omega\}$ et $X = \alpha b(uv)$. Pour toute lettre infinie a de X_I, on a $|u|_a|uv|_a < 2$.

Preuve :

Supposons que les hypothèses soient vérifiées et qu'il existe une lettre infinie a telle que $|u|_a|uv|_a \geq 2$. Nous considérons a la première lettre de X_I apparaissant dans u. Deux cas sont à envisager :

1. $|u|_a \geq 1 \quad |v|_a \geq 1$

Posons $u = u_1au_2$ et $v = v_1av_2$ avec $|u_1|_a = 0$. Nous avons $h(uv^\omega) = uv^\omega$. Nous en déduisons l'existence d'un entier n tel que $h^n(a) = u_1'au_2v^iw_1$ avec $i > 0$ et $h^n(u_1)u_1' = u_1$. Ceci implique

$$h^{2n}(a) = h^n(u_1')u_1'au_2v^iw_1h^n(u_2)h^n(v_1)u_1'au_2v^iw_1h^n(v_2)h^n(v^{i-1})h^n(w_1).$$

L'occurrence de u_2 entre deux mots v conduit à des contradictions. Soit $u_2 = \varepsilon$ et le mot v se termine, comme le mot u, par a ; soit $u_2 \neq \varepsilon$ et les mots u, v se terminent par la même lettre, la dernière lettre de u_2. Ces deux possibilités contredisent le fait que uv^ω est sous forme primitive.

2. $|u|_a \geq 2$ et $|v|_a = 0$

Posons $u = u_1au_2au_3$ avec $|u_1|_a = |u_2|_a = 0$. Il existe un entier n tel que $h^n(a) = u_1'au_2au_3v^iw_1$ avec $i > 0$ et $h^n(u_1)u_1' = u_1$ et donc

$$h^{2n}(a) = h^n(u_1')u_1'au_2au_3v^iw_1h^n(u_2)u_1'au_2au_3v^iw_1h^n(u_3v^iw_1)$$

ce qui contredit le fait que le mot v ne contienne pas a. □

Dans le lemme III.B.5, nous avons vu que l'existence d'un $D0L$-langage ayant pour limite un mot uv^ω coïncidait avec l'existence d'un morphisme de $(\alpha b(uv))^*$ sur $(\alpha b(uv))^*$ qui sera itéré à partir de l'axiome uv. La propriété précédente nous permet donc de conclure qu'il n'existe pas de $D0L$-langage ayant α pour limite dans les cas suivants ; si α s'écrit uv^ω sous forme primitive :

-$\alpha b(v) = \alpha b(u)$

-$\alpha b(v) \subset \alpha b(u)$ et $\forall\, a \in \alpha b(u) \quad |uv|_a \geq 2$

Par contre, nous pouvons répondre affirmativement dans les cas suivants :

-$\alpha b(u) \cap \alpha b(v) = \emptyset$ on définit alors $h(x) = x \quad \forall\, x \in \alpha b(u)$
$h(x) = v^2 \quad \forall\, x \in \alpha b(v)$

-$\alpha b(v) \subset \alpha b(u)$ et $\exists\, a \in X$ tel que $u = u_1 a u_2$ avec $|u_1 u_2 v|_a = 0$ et
-soit $\alpha b(v) \setminus \alpha b(u_1) \neq \emptyset$ on définit alors $h(x) = \varepsilon \quad \forall\, x \in \alpha b(u_1)$
$h(a) = u$
$h(x) = v \quad \forall\, x \in \alpha b(u_2) \setminus \alpha b(u_1)$

-soit $\alpha b(u_1) \cap \alpha b(u_2) = \emptyset$ on définit alors $h(x) = \varepsilon \quad \forall\, x \in \alpha b(u_1)$
$h(a) = u$
$h(x) = v \quad \forall\, x \in \alpha b(u_2)$

-$\alpha b(v) \setminus \alpha b(u) \neq \emptyset$ et $\exists\, a \in X$ tel que $|uv|_a = 1$
si $|u|_a = 0 \quad v = v_1 a v_2$ on définit alors $h(x) = x \quad \forall\, x \neq a$
$h(a) = a v_2 v_1 a$

si $|v|_a = 0$ on définit alors $h(x) = \varepsilon \quad \forall\, x \in \alpha b(u) \setminus \{a\}$
$h(a) = u$
$h(x) = v^2 \quad \forall\, x \in \alpha b(v) \setminus \alpha b(u)$

Les cas suivants ne présentent pas de critéres simples :

(i) $\alpha b(v) \setminus \alpha b(u) \neq \emptyset$ et $\alpha b(v) \cap \alpha b(u) \neq \emptyset$ et $\forall\, a \in X \quad |uv|_a \geq 2$

(ii) $\alpha b(v) \subset \alpha b(u)$ et $\exists\, a \in X$ tel que $u = u_1 a u_2$ avec $|u_1 u_2 v|_a = 0$
et $\alpha b(v) \subset \alpha b(u_1)$ et $\alpha b(u_1) \cap \alpha b(u_2) \neq \emptyset$.

Le cas (i) a pour caractéristique le fait que chaque lettre de $\alpha b(uv)$ apparait deux fois.

Vu la propriété III.B.8, nous pourrions penser qu'il sera généralement difficile de trouver un $D0L$-système engendrant de tels mots. Donnons cependant quelques exemples pour lesquels la réponse est affirmative :

cas (i)	$u = dbb$ $v = accadd$	$h(a) = acca$ $h(b) = b$ $h(c) = d$ $h(d) = d$
cas (i)	$u = ecb$ $v = baeecbaec$	$h(a) = aeecba$ $h(b) = b$ $h(c) = ec$ $h(e) = \varepsilon$
cas (ii)	$u = ecac$ $v = ce$	$h(a) = acc$ $h(c) = ec$ $h(e) = \varepsilon$

Par contre on constatera que, cas (i), $u_1 = b$, $v_1 = cbc$ et, cas (ii), $u_2 = dbcad$, $v_2 = cb$ définissent deux mots infinis $u_1 v_1^\omega$ et $u_2 v_2^\omega$ qui ne sont pas des limites de $D0L$-langages.

Pour clore le problème sur un alphabet de deux lettres, $X = \{a, b\}$, nous devons encore regarder le cas particulier : $\alpha b(u) = \{a\}$ et $\alpha b(v) = \{a, b\}$.

Lemme III.B.9

Soit uv^ω un mot infini ultimement périodique sous forme primitive tel que $\alpha b(u) = \{a\}$, $\alpha b(v) = \{a, b\}$ et $|v|_b \geq 2$. Ce mot n'est pas la limite d'un $D0L$-langage.

Preuve :

Posons $u = a^n$ avec $n > 0$.

Nous déduisons de la propriété précédente que la lettre a n'appartient pas à l'ensemble X_I des lettres engendrant un langage infini. S'il existe un $D0L$-système, $G = \langle X, h, uv \rangle$ qui engendre un langage ayant uv^ω pour limite, il faut que b appartienne à X_I. Pour la lettre a, il n'y a donc que deux images possibles : $h(a) = a$ ou $h(a) = \varepsilon$. Nous pouvons éliminer la possibilité $h(a) = \varepsilon$ qui impliquerait $h(u) = \varepsilon$. Donc $h(u) = u = a^n$.

Comme uv^ω est sous forme primitive, nous pouvons poser $v = a^{m_1} b a^{m_2} b \ldots a^{m_k} b$ où

$k \in \mathbf{N}_+$ et, pour tout $j \in [1,k]$, $m_j \in \mathbf{N}$. Comme $h(u) = u$, nous aurions $h(v) = v^i$ avec $i > 0$; donc $h(b) \in bX^*b$. Il existerait alors un entier n tel que $h^n(b) \in v_1 v^+ v_2$ avec $|v_1| < |v|$ et $|v_2| < |v|$. Mais $h^n(v) \in v^+$ implique $a^{m_1} v_1 v^+ v_2 a^{m_2} v_1 v^+ v_2 \ldots a^{m_k} v_1 v^+ v_2 \subset v^+$. on en déduit que $v_2 = \varepsilon$ et que les m_j sont égaux ; ce qui contredit la forme primitive de uv^ω. $\square$

Des lemmes III.B.8 et III.B.9, on déduit la proposition suivante :

Proposition III.B.10

Soit $uv^\omega \in \{a,b\}^\omega$ un mot infini ultimement périodique mis sous forme primitive. Ce mot est la limite d'un D0L-langage si et seulement si

$$\text{soit } u \in a^*b \qquad v = a$$
$$\text{soit } u \in a^+ \qquad v \in a^*b$$

ou symétriquement

$$\text{soit } u \in b^*a \qquad v = b$$
$$\text{soit } u \in b^+ \qquad v \in b^*a$$

ou encore

$$u = \varepsilon \qquad v \in \{a,b\}^*.$$

Cette dernière propriété est une conséquence des résultats obtenus par P. Séébold qui a établi une caractérisation complète des morphismes engendrant des mots infinis ultimement périodique sur un alphabet de deux lettres [63].

Chapitre IV

ITÉRATION DE PLUSIEURS MORPHISMES :

$EDT0L$-SYSTÈMES

A. Sur le plus grand cône rationnel contenu dans $EDT0L$

Dans la définition des $EDT0L$-systèmes que nous utilisons, les non-terminaux sont les lettres actives du système. Une des restrictions sur la structure des $EDT0L$-systèmes consiste à borner le nombre de non-terminaux apparaissant à chaque étape de la dérivation. Cela revient à borner le parallélisme. Les $EDT0L$-systèmes ainsi obtenus sont appelés $EDT0L$-systèmes d'index fini. La famille des langages qu'ils engendrent, notée $EDT0L_{FIN}$, est le plus grand cône rationnel contenu dans $EDT0L$ [41]. Remarquons aussi que $ET0L_{FIN}$ et $EDT0L_{FIN}$ ne représentent qu'une seule famille de langage [60].

La fonction de Parikh fournit des renseignements sur le nombre d'occurences de chaque lettre dans un mot. Un langage est dit Parikh-borné s'il contient un langage borné qui lui est commutativement équivalent. Contrairement aux langages algébriques, les $D0L$-langages ne sont pas forcément Parikh-bornés [12].

Exemple : le $D0L$-système $G = \langle X, h, a\rangle$ avec $X = \{a, b, c\}$, $h(a) = abc$, $h(b) = bc$, $h(c) = c$ engendre un langage qui n'est pas Parikh-borné.

Dans [55], G. Păun démontre l'existence de langages matriciels qui ne sont pas Parikh-bornés. Il termine son article en s'interrogeant sur le cas de la famille des langages matriciels d'index fini qui est égale à $EDT0L_{FIN}$. Il conjecture que les langages de cette famille sont Parikh-bornés. Nous apportons une réponse positive à cette conjecture.

Nous allons définir les $D0L$-langages d'index fini ; chaque mot d'un tel langage possédera une dérivation dans laquelle le nombre de lettres actives à chaque étape est borné. Ces $D0L$-langages d'index fini permettront de montrer que les $EDT0L$-langages d'index fini sont Parikh-bornés.

Définition

Soit $G = \langle X, h, u\rangle$ un $D0L$-système. Une lettre a de X est active si $h(a) \neq a$. Le $D0L$-système est d'index fini s'il existe un entier k tel que tout mot du langage $L(G)$ contienne au plus k occurences de lettres actives. Un $D0L$-langage est d'index fini s'il existe un $D0L$-système d'index fini qui l'engendre.

Lemme IV.A.1

Tout $D0L$-langage d'index fini est borné.

Preuve :

Soit $G = \langle X, h, u \rangle$ un $D0L$-système d'index k engendrant le langage considéré. Soit $A \subseteq X$ l'ensemble des lettres actives. Nous allons raisonner par induction sur le cardinal $\|X\|$.

Si $\|X\| = 1$, l'alphabet est réduit à un certain $\{a\}$ et le langage engendré $L(G)$ est inclus dans a^* ; ce langage est donc borné.

Supposons que $\|X\| > 1$. Le langage $L(G) = h^*(u) = h^*(u_{(1)}u_{(2)} \ldots u_{(n)})$ est inclus dans $h^*(u_{(1)})h^*(u_{(2)}) \ldots h^*(u_{(n)})$ où $n = |u| \geq 1$. Etant donné que la famille des langages bornés est fermée par concaténation , nous pouvons raisonner en supposant $u \in X$.

Distinguons alors deux cas :

- $L(G) \setminus \{u\} \subseteq {X_1}^*$ où $X_1 = X \setminus \{u\}$. Le $D0L$-système $G_1 = \langle X_1, h, h(u) \rangle$ est d'index fini avec $\|X_1\| < \|X\|$. Nous en déduisons, par hypothèse de récurrence, que $L(G_1)$ est borné et $L(G) = \{u\} \cup L(G_1)$ est aussi borné.

- Si nous ne sommes pas dans le cas précédent, il existe un entier $i \in \mathbf{N}_+$ tel que $h^i(u) = xuy$ avec x, $y \in X^*$. Posons $g = h^i$ et montrons que $g^{k-1}(xy) \in T^*$ où $T = X \setminus A$. Si cela n'était pas vrai, nous aurions $|g^j(xy)|_A \geq 1$ pour tout $j \in [1, k-1]$. Cela impliquerait que $|g^k(u)|_A > k$ car $g^k(u) = g^{k-1}(x) \ldots g(x)xuyg(y) \ldots g^{k-1}(y)$. Ceci est en contradiction avec l'appartenance de $g^k(u)$ au langage $L(G)$.

Posons $x_1 = g^{k-1}(x)$, $y_1 = g^{k-1}(y)$ et $z = g^k(u) = g^{k-1}(xuy) = x_1 g^{k-1}(u) y_1$. Puisque x_1, $y_1 \in T^*$, nous avons $g(x_1) = h^i(x_1) = x_1$; de même, $g(y_1) = y_1$. Alors $g(z) = g(x_1)g^k(u)g(y_1) = x_1 z y_1$ et $g^*(z) = \{x_1^n z y_1^n \ / \ n \in \mathbf{N}\}$ est un langage borné. La famille des langages bornés est fermée par union, morphisme et contient les langages finis. Nous en déduisons que

$$h^*(z) = \bigcup_{j=0}^{i-1} h^j(g^*(z)) \quad \text{et} \quad L(G) = \{h^j(u) \ / \ 0 \leq j \leq ki\} \cup h^*(z)$$

sont des langages bornés. □

Nous allons montrer que le lemme précédent peut être généralisé à certains sous-ensembles des $EDT0L$-langages d'index fini.

Lemme IV.A.2

Soit $G = \langle X, N, H, \phi, \varpi\rangle$ un $EDT0L$-système d'index fini. Pour tout langage borné $K \subseteq H^*$, le langage $(\phi(K))(\varpi) = \{(\phi(\delta))(\varpi) \ / \ \delta \in K\}$ est un langage borné.

Preuve :

Puisque $K \subseteq \delta_1^* \ldots \delta_n^*$ avec $\delta_1,\ldots, \delta_n \in H^*$, l'ensemble $(\phi(K))(\varpi)$ est inclus dans

$$L = (\phi(\delta_1^* \ldots \delta_n^*))(\varpi).$$

Il suffit donc de prouver que L est borné.

Nous raisonnons par récurrence sur n.

Si $n = 0$, le langage $L = \{\varpi\}$ est borné.

Si $n > 0$, nous avons $L = (\phi(\delta_n^*))(L')$ où $L' = (\phi(\delta_1^* \ldots \delta_{n-1}^*))(\varpi) \subseteq S(G)$. Par hypothèse de récurrence, L' est borné. Pour tout $B \in N$, considérons le $D0L$-système $G_B = \langle V, h, B\rangle$ où $h = \phi(\delta_n)$. Le $D0L$-système G_B est d'index fini et d'après le lemme précédent, $L_B = h^*(B)$ est un langage borné. Définissons la substitution s par $s(x) = \{x\}$, pour tout $x \in X$, et $s(B) = L_B$, pour tout B de N. Nous avons l'inclusion $L \subseteq s(L')$. Il reste à prouver que $s(L')$ est borné. Puisque L' est un langage borné, il existe $u_1,\ldots, u_t \in V^*$ tels que $L' \subseteq u_1^* \ldots u_t^*$. Le $D0L$-système G étant d'index fini, il existe un entier k tel que $L' \subseteq S(G) \subseteq (X^*(N \cup \{\varepsilon\})X^*)^k$. Nous en déduisons que L' est inclus dans $R = u_1^* \ldots u_t^* \cap (X^*(N \cup \{\varepsilon\})X^*)^k$. Ce langage rationnel R est une union finie de langages de la forme $L_1B_1L_2 \ldots B_pL_{p+1}$ avec $0 \le p \le k$, $B_1, B_2,\ldots, B_p \in N$ et où $L_1,\ldots, L_{p+1}$ sont des langages rationnels bornés inclus dans X^*. Puisque la famille des langages bornés est fermée par concaténation et union, les langages $s(L_1B_1L_2 \ldots B_pL_{p+1}) = L_1L_{B_1}L_2 \ldots L_{B_p}L_{p+1}$ et $s(R)$ sont bornés. Donc $L \subseteq s(L') \subseteq s(R)$ est un langage borné. □

Pour prouver le résultat recherché, nous utiliserons la notion d' $EDT0L$-système d'index fini mis sous forme normale [41,42].

Définition

Un $EDT0L$-système d'index fini $G = \langle X, N, H, \phi, \varpi\rangle$ est sous forme normale si

1) $N = M \times [1, k]$ pour un certain entier k et un certain ensemble M.
   Pour tout $B \in M$, nous noterons $\bar{B} = (B, 1)(B, 2)\ldots(B, k)$.
2) $\varpi = \bar{A}$ pour un certain $A \in M$.
3) Pour tout $h \in H$, il existe un seul $B \in M$, noté $ntg(h)$,
   tel que $\phi(h)$ est l'identité sur $(M \setminus \{B\}) \times [1, k]$.
   De plus, soit $(\phi(h))(\bar{B}) \in X^*$ et on note $ntd(h) = \varepsilon$,
   soit il existe un seul $C \in M$ tel que $(\phi(h))(\bar{B}) = u_1(C, 1)u_2 \ldots u_k(C, k)u_{k+1}$
   avec $u_1,\ldots, u_{k+1} \in X^*$ et on note $ntd(h) = C$.

Alors, si $y \in S(G) \backslash X^*$, il existe $B \in M$, noté $nt(y)$, tel que $y = y_1(B,1)y_2 \ldots (B,k)y_{k+1}$ avec $y_1, \ldots, y_{k+1} \in X^*$. Si $y \in X^*$, on note $nt(y) = \varepsilon$.

Proposition IV.A.3 [42]

Tout $ET0L$-langage d'index fini peut être généré par un $EDT0L$-système d'index fini mis sous forme normale.

Nous pouvons maintenant démontrer le principal résultat de cette section.

Proposition IV.A.4

Tout $ET0L$-langage d'index fini est Parikh-borné.

Preuve :

La proposition précédente nous permet de considérer le langage engendré par un $EDT0L$-système d'index fini sous forme normale $G = \langle X, N, H, \phi, \varpi \rangle$ avec $N = M \times [1, k]$ et $\varpi = \bar{A}$.

Nous allons d'abord construire un langage rationnel $K \subseteq H^*$ tel que $(\phi(K))(\bar{A}) = L(G)$ et $\forall\, \delta \in K$, le vecteur de Parikh $\Psi((\phi(\delta))(\bar{A}))$ dépend seulement de $\Psi(\delta)$. Puis nous montrerons que l'on peut trouver un langage rationnel $R \subseteq K$ tel que $\Psi(R) = \Psi(K)$ et nous montrerons que $\Psi((\phi(R))(\bar{A})) = \Psi((\phi(K))(\bar{A}))$. Le lemme IV.A.2 nous permet de déduire que $L = (\phi(R))(\bar{A})$ est un langage borné. Alors L est un langage borné inclus dans $L(G)$ et commutativement équivalent à $L(G)$.

Le langage rationnel K est défini par une grammaire linéaire à droite $G' = \langle H, M, P, A \rangle$ où P est l'ensemble des règles $P = \{B \rightarrow hC \;/\; h \in H \;,\; B = ntg(h) \text{ et } C = ntd(h)\}$.

Par induction sur la longueur de $\delta \in H^*$, on obtient facilement la propriété suivante :

(1) Si $A \xrightarrow{*} \delta B$ avec $B \in M \cup \{\varepsilon\}$ alors $B = nt((\phi(\delta))(\bar{A}))$.

Montrons maintenant que tout mot du langage $L(G)$ peut être obtenu en utilisant une séquence de morphismes correspondant à un mot de K. Plus précisément, montrons la propriété :

(2) $L(G) = (\phi(K))(\bar{A})$.

Puisque $K \subseteq H^*$, nous avons $(\phi(K))(\bar{A}) \subseteq (\phi(H^*))(\bar{A}) = S(G)$. Nous déduisons de la propriété (1) que $nt((\phi(\delta))(\bar{A})) = \varepsilon$, pour tout $\delta \in K$; donc $(\phi(K))(\bar{A}) \subseteq X^*$ et $(\phi(K))(\bar{A}) \subseteq S(G) \cap X^* = L(G)$.

Pour l'inclusion inverse, nous prouvons par récurrence sur la longueur de δ que, pour tout $\delta \in H^*$, il existe $\lambda \in H^*$ tel que $(\phi(\lambda))(\bar{A}) = (\phi(\delta))(\bar{A})$ et $A \stackrel{*}{\Longrightarrow} \lambda B$ avec $B = nt((\phi(\delta))(\bar{A}))$.

Si $\delta = \varepsilon$, il suffit de prendre $\lambda = \varepsilon$.

Sinon $\delta = \delta' h$ avec $h \in H$. Soit $w' = (\phi(\delta'))(\bar{A})$, $B' = nt(w')$, $w = (\phi(\delta))(\bar{A})$ et $B = nt(w)$. Par hypothèse de récurrence, il existe $\lambda' \in H^*$ tel que $(\phi(\lambda'))(\bar{A}) = w'$ et $A \stackrel{*}{\longrightarrow} \lambda' B'$. Si $w' = w$, alors $B' = B$ et il suffit de prendre $\lambda = \lambda'$. Si $w' \neq w$, puisque $w = (\phi(h))(w')$ on peut en déduire que $ntg(h) = nt(w') = B'$ et $ntd(h) = nt(w) = B$; donc $B' \rightarrow hB$ est une règle de P et $A \stackrel{*}{\longrightarrow} \lambda' hB$. Puisque $(\phi(\lambda' h))(\bar{A}) = (\phi(h))(w') = w = (\phi(\delta))(\bar{A})$, nous pouvons choisir $\lambda = \lambda' h$.

Maintenant, si $w \in L(G)$, alors il existe $\delta \in H^*$ tel que $w = (\phi(\delta))(\bar{A})$ et, comme nous venons de le montrer, $w = (\phi(\lambda))(\bar{A})$ pour un certain $\lambda \in H^*$ tel que $A \stackrel{*}{\longrightarrow} \lambda B$ où $B = nt(w) = \varepsilon$. Donc $\lambda \in K$ et $w \in (\phi(K))(\bar{A})$; on a donc l'inclusion $L(G) \subseteq (\phi(K))(\bar{A})$.

Nous allons maintenant montrer que, pour $\delta \in K$, le vecteur $\Psi((\phi(\delta))(\bar{A}))$ dépend seulement de $\Psi(\delta)$. Pour cela, nous définissons le morphisme θ de H^* dans X^* : pour tout $h \in H$, $\theta(h) = \pi_X((\phi(h))(\bar{B}))$ où $\bar{B} = ntg(h)$. Alors, si $\lambda \in FG(K)$, nous pouvons prouver l'égalité

(3) $\Psi \circ \pi_X((\phi(\lambda))(\bar{A})) = \Psi \circ \theta(\lambda)$.

La preuve se fait par récurrence sur la longueur de λ.

Si $\lambda = \varepsilon$, nous obtenons immédiatement $\pi_X((\phi(\varepsilon))(\bar{A})) = \pi_X(\bar{A}) = \varepsilon = \theta(\varepsilon)$.

Autrement, $\lambda = \lambda' h$ avec $h \in H$. Posons $w' = (\phi(\lambda'))(\bar{A})$ et $w = (\phi(\lambda))(\bar{A}) = (\phi(h))(w')$. Puisque $\lambda \in FG(K)$, nous avons $A \stackrel{*}{\longrightarrow} \lambda C$ pour un certain $C \in M \cup \{\varepsilon\}$ et, de la propriété (1), $nt((\phi(\lambda))(\bar{A})) = C$. Alors il existe $B \in M$ tel que $A \stackrel{*}{\longrightarrow} \lambda' B$ et $B \rightarrow hC$ est une règle de P, ce qui implique $B = ntg(h)$ et $C = ntd(h)$. De la propriété (1), on déduit que $B = nt(w')$. Donc on peut écrire $w' = w_0(B,1)w_1 \ldots (B,k)w_k$ et $w = w_0 y_1 w_1 \ldots y_k w_k$

avec $w_0, \ldots, w_k \in X^*$ et $\pi_X(y_1 \ldots y_k) = \pi_X((\phi(h))(\bar{B})) = \theta(h)$. Par hypothèse de récurrence, nous obtenons $\Psi \circ \pi_X(w') = \Psi \circ \theta(\lambda')$. Donc

$$
\begin{aligned}
\Psi \circ \pi_X(w) &= \Psi \circ \pi_X(w') + \Psi \circ \pi_X(y_1 \ldots y_k) \\
&= \Psi \circ \theta(\lambda') + \Psi \circ \theta(h) \\
&= \Psi \circ \theta(\lambda' h) \\
&= \Psi \circ \theta(\lambda)
\end{aligned}
$$

ce qui achève la preuve de la propriété (3).

Pour tout langage rationnel K, il existe un langage rationnel borné $R \subseteq K$ et commutativement équivalent à K [44]. Par le lemme IV.A.2, on sait que $L = (\phi(R))(\bar{A})$ est un langage borné. De la propriété (2), on déduit que $L(G) = (\phi(K))(\bar{A}) \supseteq (\phi(R))(\bar{A}) = L$. Il reste à prouver que L et $L(G)$ sont commutativement équivalents. Puisque $L \subseteq L(G)$, il suffit d'établir que $\Psi(L(G)) \subseteq \Psi(L)$. Prenons $w \in L(G)$. Il existe $\delta \in K$ tel que $w = (\phi(\delta))(\bar{A})$ et il existe $\delta' \in R$ commutativement équivalent à δ. Donc $\theta(\delta)$ et $\theta(\delta')$ sont commutativement équivalent ; c'est à dire $\Psi \circ \theta(\delta) = \Psi \circ \theta(\delta')$. Puisque δ et δ' appartiennent à K, nous pouvons conclure en appliquant la propriété (3) :

$$
\begin{aligned}
\Psi(w) &= \Psi((\phi(\delta))(\bar{A})) \\
&= \Psi \circ \pi_X((\phi(\delta))(\bar{A})) \\
&= \Psi \circ \theta(\delta) \\
&= \Psi \circ \theta(\delta') \\
&= \Psi \circ \pi_X((\phi(\delta'))(\bar{A})) \\
&= \Psi((\phi(\delta'))(\bar{A})) \\
&\subseteq \Psi((\phi(R))(\bar{A})) \ = \Psi(L)
\end{aligned}
$$

□

B. Sur le plus petit cône rationnel contenant $EDT0L$

Les principales propriétés de clôture de la famille $EDT0L$ sont connues : union, concaténation, étoile, morphisme, intersection avec les langages rationnels, fonction séquentielle [23,29]. Cette dernière propriété peut être généralisée aux fonctions rationnelles en utilisant le fait que toute fonction rationnelle s'obtient par composition d'une fonction séquentielle droite et d'une fonction séquentielle gauche [43].

Mais $EDT0L$ n'est pas fermée par morphisme inverse. L'exemple le plus connu est fourni par le langage $L = \{a^{2^n} \ / \ n \in \mathbf{N}_+\}$ et le morphisme h de $\{a,b\}^*$ dans $\{a,b\}^*$ défini par $h(a) = h(b) = a$; le langage $h^{-1}(L)$ n'est pas un $EDT0L$-langage. Donc, $EDT0L$ n'est pas un cône rationnel.

A. Ehrenfeucht et G. Rozenberg [23] établissent une méthode qui, partant de langages qui appartiennent à $\mathbf{H}^{-1}(EDT0L) \setminus EDT0L$, permet d'obtenir des langages qui ne sont pas dans $\mathbf{H}^{-1}(EDT0L)$. Cette méthode consiste à appliquer un morphisme d'un type particulier. Cela leur permet donc de démontrer que $EDT0L \subsetneq \mathbf{H}^{-1}(EDT0L) \subsetneq \mathbf{H} \circ \mathbf{H}^{-1}(EDT0L)$, début d'une hiérarchie qu'ils conjecturent infinie :

Conjecture [23]

Soient $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{L}_3$,... *la suite infinie de familles de langages, définies par :*

$$\mathbf{L}_1 = EDT0L$$
$$\mathbf{L}_{2n} = \mathbf{H}^{-1}(\mathbf{L}_{2n-1}) \quad \forall\, n \geq 1$$
$$\mathbf{L}_{2n+1} = \mathbf{H}(\mathbf{L}_{2n}) \quad \forall\, n \geq 1.$$

Alors $\mathbf{L}_1 \subsetneq \mathbf{L}_2 \subsetneq \mathbf{L}_3 \subsetneq \ldots \subsetneq \mathbf{L}_i \subsetneq \ldots \subseteq ET0L$.

Les différents travaux sur les compositions de morphismes et de morphismes inverses [40,45,46] ont permis d'élaborer des formes normales de ces compositions. En particulier, M. Latteux et P. Turakainen [46] ont démontré que toute composition de morphismes et de morphismes inverses pouvait se réduire à une composition dans $\mathbf{H} \circ \mathbf{H}^{-1} \circ \mathbf{H} \circ \mathbf{H}^{-1} = \mathbf{H}^{-1} \circ \mathbf{H} \circ \mathbf{H}^{-1} \circ \mathbf{H}$. Cela permet donc de contredire la conjecture. Nous allons préciser ce résultat en montrant que la hiérarchie s'arrête à $\mathbf{H} \circ \mathbf{H}^{-1}(EDT0L) = \mathbf{C}(EDT0L)$.

Nous utiliserons surtout les propriétés des transductions rationnelles étoilées.

Définitions

Une transduction est rationnelle étoilée si et seulement si elle appartient à $\mathbf{C}_* = \mathbf{H} \circ \wedge Rat^* \circ \mathbf{H}^{-1}$. Pour un langage L, le cône rationnel étoilé $\mathbf{C}_*(L) = \mathbf{H} \circ \wedge Rat^* \circ \mathbf{H}^{-1}(L)$

est la clôture de L par transductions rationnelles étoilées ; pour une famille $\mathbf{L}$ de langages, $\mathbf{C}_*(\mathbf{L}) = \{\mathbf{C}_*(L) \ / \ L \in \mathbf{L}\}$

Nous montrerons l'égalité du cône rationnel et du cône rationnel étoilé engendrés par les $EDT0L$-langages. Mieux encore, la structure des $EDT0L$-langages permet d'engendrer le cône rationnel en utilisant uniquement des transductions rationnelles étoilées décroissantes.

En utilisant le fait que toute transduction rationnelle est équivalente à une transduction rationnelle étoilée précédée d'une application qui ajoute une marque à la fin de chaque mot [40,67], nous pouvons énoncer le lemme suivant :

Lemme IV.B.1

Le cône rationnel et le cône rationnel étoilé engendrés par les $EDT0L$-langages sont égaux :

$$\mathbf{C}(EDT0L) = \mathbf{C}_*(EDT0L).$$

Preuve :

Soit L un langage sur X^*. Notons $L\flat$, le langage constitué des mots de L suivis d'une marque $\flat$ n'appartenant pas à X. Le langage $L\flat$ s'obtient en appliquant une fonction rationnelle au langage L. La famille $EDT0L$ étant fermée par fonction rationnelle [29,43], pour tout $EDT0L$-langage L, le langage $L\flat$ appartient encore à $EDT0L$. Pour toute transduction rationnelle τ, il existe une transduction rationnelle étoilée τ_* telle que $\tau(L) = \tau_*(L\flat)$ [40,67]. Donc le cône rationnel et le cône rationnel étoilé engendrés par $EDT0L$ sont égaux. □

Pour démontrer l'égalité $\mathbf{C}_*(EDT0L) = \mathbf{H} \circ \mathbf{H}^{-1}(EDT0L)$, nous utiliserons le shuffle d'un $EDT0L$-langage et de $\sharp^*$, où $\sharp$ est une nouvelle lettre. Or ce langage n'appartient pas toujours à $EDT0L$. Le lemme suivant nous permettra de démontrer le résultat désiré.

Lemme IV.B.2

Soit L un $EDT0L$-langage. Soit $\sharp$ une nouvelle lettre terminale.

Le langage $L(\sharp) = \{\sharp^n x_1 \sharp^n ... x_p \sharp^n \ / \ x_i \in X \ , \ x_1 x_2 ... x_p \in L \ , \ n \in \mathbf{N}\}$ appartient encore à $EDT0L$.

Preuve :

Nous allons modifier le système engendrant L pour qu'apparaisse, entre chaque lettre terminale, un nouveau non-terminal qui ne pourra se dériver qu'en dernier lieu, générant ainsi parallèlement les blocs $\sharp^n$.

Soit $G = \langle X, N, H, \phi, \varpi\rangle$ un $EDT0L$-système engendrant le langage L. Soit S une nouvelle lettre non-terminale. Soit f le morphisme défini par $f(x) = Sx$, pour tout $x \in X$, et $f(A) = A$, pour tout $A \in N$.

Posons $G_\sharp = \langle X', N', H', \phi', \varpi'\rangle$ avec $X' = X \cup \{\sharp\}$, $N' = N \cup \{A_p \ / \ A \in N\} \cup \{S\}$, $H' = H \cup \{r_1, r_2\}$, $\varpi' = \varpi S$ et où ϕ' est défini par :

pour tout h de H	$(\phi'(h))(A) = f((\phi(h))(A))$	
	$(\phi'(h))(A_p) = A_p$	pour tout A de N
	$(\phi'(h))(S) = S$	
pour r_1	$(\phi'(r_1))(A) = A_p$	
	$(\phi'(r_1))(A_p) = A_p$	pour tout A de N
	$(\phi'(r_1))(S) = \sharp S$	
pour r_2	$(\phi'(r_2))(A) = A_p$	
	$(\phi'(r_2))(A_p) = A_p$	pour tout A de N
	$(\phi'(r_2))(S) = \varepsilon$	

L'ensemble $\{A_p \ / \ A \in N\}$ contient les lettres non-terminales qui, si elles apparaissent dans une dérivation, ne pourront être éliminées. Cela oblige l'utilisation des règles r_1 et r_2 après avoir effacé toutes les lettres de N ; c'est-à-dire sur un mot $f(w)S$ avec w appartenant à L. On vérifie aisément que $L(G_\sharp) = L(\sharp)$. $\square$

Proposition IV.B.3

Tout langage appartenant au cône rationnel engendré par $EDT0L$ peut être obtenu à partir de $EDT0L$ par application d'un morphisme (non effaçant) inverse suivi d'un morphisme (alphabétique) :

$$EDT0L \subsetneq \mathbf{H}^{-1}(EDT0L) \subsetneq \mathbf{H}_\alpha \circ \mathbf{H}_\varepsilon^{-1}(EDT0L) = \mathbf{C}(EDT0L).$$

Preuve :

Etant donné l'égalité du cône rationnel et du cône rationnel etoilé pour $EDT0L$ (lemme IV.B.1), nous allons vérifier, dans un premier temps, que le cône rationnel étoilé peut être obtenu en utilisant uniquement des transductions décroissantes.

Nous utilisons la caractérisation du cône rationnel étoilé par des morphismes alphabétiques [45]. Soit $\tau_* = g \circ \cap R^* \circ h^{-1}$ une transduction rationnelle étoilée appartenant à $\mathbf{H}_\alpha \circ \wedge Rat^* \circ \mathbf{H}_\alpha^{-1}$. Pour tout $EDT0L$-langage L, nous avons $\tau_*(L) = \tau'_*(L \sqcup\!\sqcup \sharp^*)$ avec $\tau'_* = g \circ \cap R^* \circ h'^{-1}$ où h' est défini par $h'(x) = \sharp$ si $h(x) = \varepsilon$ et $h'(x) = h(x)$ sinon. La transduction τ'_* appartient à $\mathbf{H}_\alpha \circ \wedge Rat^* \circ \mathbf{H}_{s\alpha}^{-1}$ et est décroissante.

Or, le langage $L \sqcup\!\sqcup \sharp^*$ s'obtient en appliquant une substitution au langage $L(\sharp)$. Soit σ la substitution de $(X \cup \{\sharp\})^*$ dans $(X \cup \{\sharp\})^*$ définie par $\sigma(x) = x$, pour tout x de X, et $\sigma(\sharp) = \{\sharp, \varepsilon\}$. Nous avons $L \sqcup\!\sqcup \sharp^* = \sigma(L(\sharp))$. Par composition, nous avons $\tau_*(L) = \tau'_* \circ \sigma(L(\sharp))$. Cette substitution est aussi une transduction rationnelle étoilée décroissante et l'ensemble des transductions rationnelles étoilées décroissante est fermée par composition.

Donc, pour un $EDT0L$-langage quelconque L et une transduction rationnelle étoilée quelconque τ_*, nous avons réalisé $\tau_*(L)$ à partir de $L(\sharp)$, qui appartient à $EDT0L$ (lemme IV.B.2), et de $\tau'_* \circ \sigma$, qui est une transduction rationnelle étoilée décroissante.

Toute transduction rationnelle étoilée décroissante appartient à $\mathbf{H}_\alpha \circ \mathbf{H}_\varepsilon^{-1} \circ \mathbf{H}_u$, propriété démontrée dans [45]. La famille $EDT0L$ étant close par morphisme, nous en concluons que :

$$\mathbf{C}(EDT0L) = \mathbf{C}_*(EDT0L) = \mathbf{H}_\alpha \circ \mathbf{H}_\varepsilon^{-1} \circ \mathbf{H}_u(EDT0L) = \mathbf{H}_\alpha \circ \mathbf{H}_\varepsilon^{-1}(EDT0L). \qquad \square$$

Quelques problèmes restent à résoudre.

Le cône rationnel engendré par les $EDT0L$-langages est inclus dans la famille des $ET0L$-langages. L'inclusion stricte reste encore à démontrer. Nous savons que, pour tout entier $n \geq 1$, le langage de Dyck $D_n'^*$ n'appartient pas à $EDT0L$ [62]; on peut s'interroger sur leur appartenance au cône $\mathbf{C}(EDT0L)$. Il est aisé de montrer que $D_1'^* \notin \mathbf{H}^{-1}(EDT0L)$ et nous conjecturons que $D_1'^* \notin \mathbf{H} \circ \mathbf{H}^{-1}(EDT0L)$.

La famille $ET0L$ est un cône rationnel principal [19]. Dans le cas où l'inclusion serait stricte, il faudrait aussi s'interroger sur la principalité de $\mathbf{C}(EDT0L)$.

Le cône rationnel engendré par $EDT0L$ peut-il l'être fidélement ?

C. $EDT0L$ est fermée par polynôme de fonctions rationnelles

La famille des $EDT0L$-langages n'est pas fermée par morphisme inverse. On peut même préciser dans le cas d'un morphisme inverse particulier :

Proposition IV.C.1 [30,41]

Soit $\sharp$ *une lettre n'appartenant pas à* X. *Soient* $L \subseteq X^*$ *et* π_X *la projection de* $X \cup \{\sharp\}$ *sur* X. *Alors,* $L \sqcup\!\sqcup \sharp^* = \pi_X^{-1}(L) \in EDT0L$ *implique* $L \in EDT0L_{FIN}$.

Généralement, l'application d'un morphisme inverse associe à un mot de départ un ensemble de mots dont le cardinal peut être grand. Par contre, $EDT0L$ est fermée par fonctions rationnelles ; les fonctions rationnelles à un mot associent au plus un mot. On peut donc s'interroger sur le comportement de la famille des $EDT0L$-langages, quand on lui applique des transductions rationnelles d'image finie d'un type particulier. Nous remarquons alors que $EDT0L$ est fermée par polynôme de fonctions rationnelles. Pour la preuve de ce résultat, établissons d'abord qu'à l'inverse de la proposition précédente,

Lemme IV.C.2

Soit L un $EDT0L$-langage inclus dans X^*. Soit $\sharp$ une lettre n'appartenant pas à X. Pour tout $n \in \mathbb{N}$, le langage $L \sqcup\!\sqcup \sharp^n$ appartient à $EDT0L$.

Preuve :

Montrons que $L \sqcup\!\sqcup \sharp$ appartient à $EDT0L$.

Il existe un $EPDT0L$-système $G_\varepsilon = \langle X, N, H, \phi, \varpi\rangle$ qui engendre $L_\varepsilon = L \setminus \{\varepsilon\}$. Nous pouvons, de plus, supposer que $\varpi \in N$.

On construit un $EPDT0L$-système

$$G_\sharp = \langle X \cup X_\sharp, N \cup N_\sharp \cup N_p, H', \phi', \varpi_\sharp\rangle$$

où $X_\sharp = \{x_\sharp \ / \ x \in X\}$, $N_\sharp = \{A_\sharp \ / \ A \in N\}$, $N_p = \{A_p \ / \ A \in N\}$,
$H' = \{h_{A,i} \ / \ A \in N \ \ et \ \ 1 \leq i \leq |(\phi(h))(A)|\}$,
et où ϕ' est défini par :
$(\phi'(h_{A,i}))(B) = (\phi(h))(B) \qquad \forall\, B \in N$
$(\phi'(h_{A,i}))(A_\sharp) = u_{[i-1]} z_\sharp u_{[i+1,n[} \qquad$ où $u = (\phi(h))(A)$, $n = |u| \geq 1$ et $z = u_{(i)}$
$(\phi'(h_{A,i}))(B_\sharp) = B_p \qquad \forall\, B \in N$ et $B \neq A$
$(\phi'(h_{A,i}))(B_p) = B_p \qquad \forall\, B \in N$.

Dans une dérivation, le symbole $\sharp$ marque le chemin de l'axiome à une lettre terminale quelconque.

Notons $V = X \cup N$, $V_\sharp = X \cup X_\sharp \cup N \cup N_\sharp$, $V_{\sharp p} = X \cup X_\sharp \cup N \cup N_\sharp \cup N_p$.

Nous vérifions que

$$(\star) \qquad S(G_\sharp) \cap V_\sharp^* = \{w_1 z_\sharp w_2 \ / \ w_1 z w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^* \ et \ z \in V\}.$$

Par induction sur la longueur de $\delta \in H'^*$, nous montrons que

$$(\phi'(\delta))(\varpi_\sharp) \cap V_\sharp^* \subseteq \{w_1 z_\sharp w_2 \ / \ w_1 z w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^* \ et \ z \in V\}.$$

Si $\delta = \varepsilon$, nous avons $(\phi'(\delta))(\varpi_\sharp) = \varpi_\sharp$ et $\varpi = (\phi(f(\delta)))(\varpi)$.

Sinon $\delta = \delta_1 h_{A,i}$ avec $\delta_1 \in H'^*$ et $h_{A,i} \in H'$.

$$\begin{aligned}
(\phi'(\delta))(\varpi_\sharp) \cap V_\sharp^* &= (\phi'(h_{A,i}))((\phi'(\delta_1))(\varpi_\sharp) \cap V_\sharp^*) \cap V_\sharp^* \\
&\quad \cup (\phi'(h_{A,i}))((\phi'(\delta_1))(\varpi_\sharp) \cap V_{\sharp p}^* N_p V_{\sharp p}^*) \cap V_\sharp^* \\
&= (\phi'(h_{A,i}))((\phi'(\delta_1))(\varpi_\sharp) \cap V_\sharp^*) \cap V_\sharp^* \\
&\subseteq (\phi'(h_{A,i}))(\{w_1 z_\sharp w_2 \ / \ w_1 z w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^* \ et \ z \in V\}) \cap V_\sharp^* \\
&\subseteq (\phi'(h_{A,i}))(\{w_1 z_\sharp w_2 \ / \ w_1 z w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^* \ et \ z \in X\}) \cap V_\sharp^* \\
&\quad \cup (\phi'(h_{A,i}))(\{w_1 z_\sharp w_2 \ / \ w_1 z w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^* \ et \ z \in N\}) \cap V_\sharp^* \\
&\subseteq \{(\phi'(h_{A,i}))(w_1 z_\sharp w_2) \ / \ w_1 z w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^* \ et \ z \in X\} \\
&\quad \cup \{(\phi'(h_{A,i}))(w_1 A_\sharp w_2) \ / \ w_1 A w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^*\} \\
&\subseteq \{(\phi(h))(w_1) z_\sharp (\phi(h))(w_2) \ / \ w_1 z w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^* \ et \ z \in X\} \\
&\quad \cup \{(\phi(h))(w_1)(\phi'(h_{A,i}))(A_\sharp)(\phi(h))(w_2) \ / \ w_1 A w_2 \in S(G_\varepsilon) \ avec \ w_1, w_2 \in V^*\} \\
&\subseteq \{w_1' z_\sharp' w_2' \ / \ w_1' z' w_2' \in S(G_\varepsilon) \ avec \ w_1', w_2' \in V^* \ et \ z' \in V\}
\end{aligned}$$

Inversement, nous montrons, par induction sur la longueur de λ, que

$$\forall \lambda \in H^*, \quad \text{si } w_1 z w_2 \in (\phi(\lambda))(\varpi) \quad \text{avec} \quad z \in V \qquad \text{alors } w_1 z_\sharp w_2 \in S(G_\sharp)$$

Si $\lambda = \varepsilon$, nous avons $z = \varpi$ et $\varpi_\sharp \in S(G_\sharp)$.

Si $\lambda = \lambda_1 h$, pour tout mot $w = w_1 z w_2 \in (\phi(\lambda))(\varpi)$, il existe un mot $w' \in S(G_\varepsilon)$ tel que $w = (\phi(h))(w')$. Il existe alors une factorisation $w' = w_1' z w_2'$ tel que

$$(\phi(h))(w'_1) = v_1 \in FG(w_1)$$
$$(\phi(h))(w'_2) = v_2 \in FD(w_2)$$
$$\text{et } z \in F((\phi(h))(z'))$$

Si $z' \in X$, alors $z = z'$ et $(\phi(h_{A,i}))(w'_1 z_\sharp w'_2) = w_1 z_\sharp w_2$ pour A et i quelconques. Si $z' \in N$, alors $(\phi(h_{z',i}))(w'_1 z_\sharp w'_2) = w_1 z_\sharp w_2$ pour $i = |w_1| - |v_1| + 1$.

De l'égalité $(\star)$, nous pouvons déduire que

$$L(G_\sharp) = S(G_\sharp) \cap (X \cup X_\sharp)^* = \{w_1 z_\sharp w_2 \ / \ w_1 z w_2 \in L(G_\varepsilon) \ avec \ w_1, w_2 \in X^* \ et \ z \in X\}.$$

Nous définissons alors le morphisme φ de $(X \cup X_\sharp)^*$ dans $(X \cup \{\sharp\})^*$ par $\varphi(x) = x$, pour tout $x \in X$, et $\varphi(x_\sharp) = \sharp x$, pour tout $x_\sharp \in X_\sharp$. Comme $L\sharp$, l'ensemble des mots de L suivis de la marque $\sharp$, s'obtient en appliquant une fonction rationnelle à L et que $EDT0L$ est fermée par union et fonction rationnelle, nous pouvons conclure que $L ⧢ \sharp = L\sharp \cup \varphi(L(G_\sharp))$ est un $EDT0L$-langage.

Par induction sur le nombre n, nous pouvons affirmer que, pour tout $EDT0L$-langage L et pour tout $n \in \mathbf{N}$, le langage $L ⧢ \sharp^n$ appartient à $EDT0L$. □

$EDT0L$ étant fermée par fonction rationnelle et par union, on peut déduire du lemme précédent que $EDT0L$ est fermée par shuffle avec les langages finis.

Lemme IV.C.3

$EDT0L$ est fermée par monôme de fonctions rationnelles.

Preuve :

Soit $M = \prod_{i=1}^{n} \rho_i$ un monôme de fonctions rationnelles de X^* dans Y^*.

Soit L un $EDT0L$-langage inclus dans X^*.

Dans un premier temps, nous allons transformer le langage L pour obtenir des factorisations de chaque mots sur n alphabets disjoints $X_{\{i\}}$ dont l'union sera notée Z.

Le langage $L' = L ⧢ \sharp^{n-1}$ est un $EDT0L$-langage (lemme IV.C.2).

Le langage $L'' = \{w_{1\{1\}} w_{2\{2\}} ... w_{n\{n\}} \ / \ w_1 w_2 ... w_n \in L \ \ et \ \ w_{i\{i\}} \in X^*_{\{i\}}\}$ s'obtient à partir de L' en appliquant une fonction rationnelle. Donc, L'' est un $EDT0L$-langage.

Considérons maintenant notre monôme $M = \prod_{i=1}^{n} \rho_i$.

Il suffit de transformer chaque fonction ρ_i de X^* dans Y^* en une fonction ρ'_i de $X^*_{\{i\}}$ dans Y^*, pour obtenir un monôme

$$M' = \prod_{i=1}^{n} \rho'_i \quad \text{de } Z^* = (\bigcup_{i=1}^{n} X_{\{i\}})^* \text{ dans } Y^*$$

qui est fonctionnel. En effet, pour tout mot de Z^*, il existe une seule factorisation sur $X^*_{\{1\}}X^*_{\{2\}}...X^*_{\{n\}}$, donc une seule image par M ;

$$\begin{aligned} M(L) &= \{\rho_1(w_1)\rho_2(w_2)...\rho_n(w_n) \ / \ w_1w_2...w_n \in L\} \\ &= \{\rho'_1(w_{1\{1\}})\rho'_2(w_{2\{2\}})...\rho'_n(w_{n\{n\}}) \ / \ w_1w_2...w_n \in L\} \\ &= M'(L'') \\ &\in EDT0L. \end{aligned}$$

□

La famille des $EDT0L$-langages étant fermée par union, on en déduit la proposition suivante :

Proposition IV.C.4

$EDT0L$ est fermée par polynôme de fonctions rationnelles.

Corollaire IV.C.5

$EDT0L$ est fermée par facteur gauche, facteur droit, facteur, quotient rationnel.

Les transductions rationnelles polynômialement bornées sont des polynômes de transductions rationnelles fonctionnelles ou cycliques d'image finie. On peut donc s'interroger à juste titre sur la fermeture de $EDT0L$ par transduction rationnelle polynômialement bornée. Les transductions rationnelles polynômialement bornées étant les polynômes de transductions rationnelles fonctionnelles ou cycliques d'image finie, on obtiendrait cette propriété de clôture en répondant positivement à la question de M.P. Schützenberger : toute transduction rationnelle cyclique est-elle un polynôme de fonctions rationnelles ? Malheureusement, nous montrerons au chapitre suivant que la réponse est négative et le problème de savoir si $EDT0L$ est fermée par transduction rationnelle polynômialement bornée reste ouvert.

Il suffit de transformer chaque fonction μ_i de X dans Y^* en une fonction μ'_i de $X_{(i)}^*$ dans Y^* pour obtenir un morphisme

$$M' = \prod_{i=1}^{n} \mu'_i \text{ de } Z^* = (\bigcup_{i=1}^{n} X_{(i)})^* \text{ dans } Y^*$$

qui est fonctionnel. En effet, pour tout mot de Z^*, il existe une seule factorisation sur $X_{(1)}^* \dots X_{(n)}^*$, donc une seule image par M'.

$$M'(L) = \{\mu_1(w_1) \dots \mu_n(w_n) \,/\, w_1 \dots w_n \in L\}$$
$$= \{\mu'_1(w_{1(1)}) \dots \mu'_n(w_{n(n)}) \,/\, w_1 \dots w_n \in L\}$$
$$= M'(L')$$
$$\in EDT0L$$

□

La famille des EDT0L-langages étant fermée par union, on en déduit la proposition suivante

Proposition IV.C.4

EDT0L est fermée par polynôme de fonctions rationnelles.

Corollaire IV.C.5

EDT0L est fermée par factor gauche, facteur droit, facteur, sous-mot rationnels.

Les transductions rationnelles polynômialement bornées sont des polynômes de transductions rationnelles fonctionnelles ou cycliques d'image finie. On peut donc s'intéresser à la fermeture de EDT0L par transduction rationnelle polynômialement bornée. Les transductions rationnelles polynômialement bornées étant les polynômes de transductions rationnelles fonctionnelles ou cycliques d'image finie, on obtiendrait cette propriété de clôture en répondant positivement à la question de M.P. Schützenberger : toute transduction rationnelle cyclique est-elle un polynôme de fonctions rationnelles ? Malheureusement, nous montrerons au chapitre suivant que la réponse est négative et le problème de savoir si EDT0L est fermée par transduction rationnelle polynômialement bornée reste ouvert.

Chapitre V

POLYNÔMES DE FONCTIONS RATIONNELLES

ET TRANSDUCTIONS RATIONNELLES POLYNÔMIALEMENT BORNÉES

Le problème de la fermeture de la famille des $EDT0L$-langages par transduction rationnelle polynômialement bornée nous a amené à étudier plus profondément cet ensemble de transductions.

La caractérisation des transductions rationnelles polynômialement bornées, effectuée par M.P. Schützenberger [64], indique que l'ensemble de ces transductions est l'ensemble des polynômes (somme finie de produits finis) en des transductions rationnelles fonctionnelles ou cycliques d'image finie ; nous pourrions dire aussi que c'est la clôture par somme et produit de l'ensemble des transductions rationnelles linéairement bornées puisque les transductions fonctionnelles ou cycliques d'image finie sont des transductions linéairement bornées et qu'inversement une transduction rationnelle linéairement bornée est aussi un polynôme en des transductions rationnelles fonctionnelles ou cycliques d'image finie. Avant de montrer que toute transduction rationnelle polynômialement bornée peut être obtenue par composition de transductions rationnelles linéairement bornées, vérifions la propriété suivante :

Lemme V.1

L'ensemble des transductions rationnelles polynômialement bornées est clos par composition.

Preuve :

Soient τ_1, τ_2 deux transductions rationnelles polynômialement bornées. Pour tout u de X^*, le nombre d'images de u par τ_1, noté $||\tau_1(u)||$, est borné par $k_1(|u|+1)^{d_1}$. La transduction τ_1 étant d'image finie, pour tout v de $\tau_1(u)$, on a $|v| \leq l|u| + l'$; les entiers l et l' ne dépendent que τ_1. Nous pouvons majorer cette inégalité par

$$|v| + 1 \leq l|u| + l' + 1 \leq (l+1)(l'+1)(|u|+1).$$

Posons $m = (l+1)(l'+1)$. Pour tout v appartenant à $\tau_1(u)$, le nombre d'images de v par τ_2 peut être borné par

$$||\tau_2(v)|| \leq k_2(|v|+1)^{d_2} \leq k_2 m^{d_2}(|u|+1)^{d_2}.$$

Le nombre d'images de u par $\tau_2 \circ \tau_1$ est majoré par le produit du nombre d'images de u par τ_1 et du nombre maximum d'images de chacune d'elles par τ_2, soit :

$$\begin{aligned} ||\tau_2 \circ \tau_1(u)|| &\leq k_1(|u|+1)^{d_1} k_2 m^{d_2}(|u|+1)^{d_2} \\ &\leq k_1 k_2 m^{d_2}(|u|+1)^{d_1+d_2} \end{aligned}$$

donc, $\tau_2 \circ \tau_1$ est polynômialement bornée. □

Proposition V.2

L'ensemble des transductions rationnelles polynômialement bornées est la clôture par composition de l'ensemble des transductions rationnelles linéairement bornées.

Preuve :

Rappelons que toute transduction linéairement bornée est polynômialement bornée et que l'ensemble de ces dernières reste clos par composition.

Un polynôme de transductions rationnelles fonctionnelles ou cycliques d'image finie se décompose en une somme finie de monômes.

Un monôme $\tau_1 \times \tau_2 \times \ldots \times \tau_p$, avec $p > 1$, produit de transductions rationnelles fonctionnelles ou cycliques d'image finie de X^* dans Y^* peut se mettre sous la forme d'une composition de transductions rationnelles linéairement bornées :

$$(a) \qquad \tau_1 \times \tau_2 \times \ldots \times \tau_p = \tau'_p \circ \ldots \tau'_2 \circ \tau'_1 \circ \delta \circ \eta (\circ \eta)^{p-2}.$$

Pour cela on définit $X_{\{i\}} = \{x_{\{i\}} \ / \ x \in X\}$, pour $i \in [1,p]$. On notera $w_{\{i\}} = x_{1\{i\}} x_{2\{i\}} \ldots x_{n\{i\}}$ le mot $w = x_1 x_2 \ldots x_n$ écrit sur l'alphabet $X^*_{\{i\}}$. Soit $\sharp$ une lettre n'appartenant pas à X. On définit les transductions

$$\eta \text{ de } (X \cup \{\sharp\})^* \text{ dans } (X \cup \{\sharp\})^*$$

$$\eta(w) = \{w_1 \sharp w_2 \ / \ w_1, w_2 \in (X \cup \{\sharp\})^* \ \textit{et} \ w = w_1 w_2\}$$

$$\delta \text{ de } (X \cup \{\sharp\})^* \text{ dans } (\bigcup_{i=1}^{p} X_{\{i\}})^*$$

$$\delta(w) = w_{1\{1\}} w_{2\{2\}} \ldots w_{p\{p\}} \quad \text{si} \quad w = w_1 \sharp w_2 \sharp \ldots \sharp w_p$$
$$= \emptyset \quad \text{sinon}$$

$$\tau \text{ de } (\bigcup_{i=1}^{p} X_{\{i\}})^* \text{ dans } (Y \cup (\bigcup_{i=1}^{p} X_{\{i\}}))^*$$

$$\tau'_i(w) = u \tau_i(w_i) v \quad \text{si} \quad w = u w_{i\{i\}} v, \ u \in Y^*, \ w_{i\{i\}} \in X^*_{\{i\}}, \ v \in (\bigcup_{j \neq i} X_{\{j\}})^*$$
$$= \emptyset \quad \text{sinon}$$

On vérifie aisément que les transductions η, δ, τ'_1, ..., τ'_p sont linéairement bornées et que l'égalité (a) est correcte.

Considérons maintenant $T_1 + T_2 + \ldots + T_q$ une somme de monômes de $X^{*\prime}$ dans Y^*. En utilisant le même principe de recopie sur des alphabets différents, nous avons

$$(b) \qquad T_1 + T_2 + \ldots + T_q = T'_1 \circ \ldots \circ T'_q \circ \mu$$

avec

$$\mu \text{ de } X^* \text{ dans } \bigcup_{i=1}^{q} X_{\{i\}}^*$$

$$\mu(w) = \{w_{\{i\}} \ / \ 1 \leq i \leq q\}$$

$$T_i' \text{ de } Y^* \cup (\bigcup_{i=1}^{q} X_{\{i\}}^*) \text{ dans lui} - \text{même}$$

$$T_i'(w_{\{j\}}) = T_i(w) \text{ si } i = j$$

$$= w_{\{j\}} \text{ si } i \neq j$$

$$T_i'(w) = w \text{ si } w \in Y^*$$

En utilisant le résultat (a) dans l'égalité (b), nous obtenons une composition de transductions rationnelles linéairement bornées. □

Complétons les propriétés de clôture par composition avec la proposition suivante :

Proposition V.3

L'ensemble des polynômes de fonctions rationnelles est clos par composition.

Preuve :

Soient τ et τ' deux polynômes de fonctions rationnelles

$$\tau = \sum_{i=1}^{m} \rho_i \quad \text{et} \quad \tau' = \sum_{j=1}^{n} \rho_j'$$

où les ρ_i et ρ_j' sont des monômes. Nous avons

$$\tau \circ \tau' = \sum_{i=1}^{m} \sum_{j=1}^{n} \rho_i \circ \rho_j'.$$

Il suffit donc d'analyser la composition de deux monômes.

Soient

$$\rho' = \prod_{i=1}^{p} f_i \quad \text{et} \quad \rho = \prod_{j=1}^{q} g_j$$

deux monômes de fonctions rationnelles de X^* dans Z^* et de Z^* dans Y^* ; on supposera les alphabets disjoints. Posons $f_i = \pi_Z \circ \cap R_i \circ \pi_X^{-1}$ et $g_j = \pi_Y \circ \cap S_j \circ \pi_Z^{-1}$. La transduction ρ (resp. ρ') est caractérisée par le langage rationnel $R_1 R_2 ... R_p$ (resp. $S_1 S_2 ... S_q$). La composée de ces deux transductions est caractérisée par le langage rationnel $Q = (Y^* \sqcup\!\sqcup R_1 ... R_p) \cap (S_1 ... S_q \sqcup\!\sqcup X^*)$ et les deux morphismes π_X et π_Y. Il s'agit de mettre cette transduction sous forme de polynôme de fonctions. Pour mettre en évidence le produit dans le langage rationnel, introduisons des marques $\sharp$ (resp. $\natural$) pour séparer les langages rationnels R_i (resp. S_j). Nous obtenons un langage rationnel

$$Q' = ((Y \cup \{\natural\})^* \sqcup\!\sqcup R_1 \sharp R_2 \sharp ... \sharp R_p) \cap (S_1 \natural S_2 \natural ... \natural S_q \sqcup\!\sqcup (X \cup \{\sharp\})^*).$$

Nous pouvons représenter ce langage rationnel par une union finie de langages rationnels ayant la même projection sur $\{\sharp, \natural\}^*$; c'est-à-dire

$$Q' = \bigcup_{i=1}^{p+q-2} P_{\alpha_i} \text{ où } P_{\alpha_i} = \{w \in Q' \ / \ \pi_{\sharp,\natural}(w) = \alpha_i \in \sharp^{p-1} \sqcup\!\sqcup \natural^{q-1}\}.$$

Chacun de ces P_{α_i} étant lui-même une union finie de langages rationnels, nous pouvons récapituler par

$$Q' = \bigcup_{finie} T_{r,1}\alpha_{i,1}T_{r,2}\alpha_{i,2}...\alpha_{i,p+q-2}T_{r,p+q-1} = \bigcup_{finie} T_r$$

où chaque $T_{r,j}$ est un langage rationnel et où $\alpha_i = \alpha_{i,1}...\alpha_{i,p+q-2} \in \sharp^{p-1} \sqcup\!\sqcup \natural^{q-1}$. La transduction $\rho \circ \rho' = \pi_Y \circ \cap Q' \circ \pi_X^{-1}$ se décompose de nouveau en somme de transductions $\pi_Y \circ \cap T_r \circ \pi_X^{-1}$. Comme π_X est un morphisme alphabétique et que T_r est un produit de langages rationnels $T_{r,j}$, chaque transduction $\pi_Y \circ \cap T_r \circ \pi_X^{-1}$ est un produit de transductions $\pi_Y \circ \cap T_{r,j} \circ \pi_X^{-1}$. On remarquera que les $\sharp$ et $\natural$, inutiles dans la transduction, peuvent être éliminés : $\pi_Y \circ \cap T_r \circ \pi_X^{-1}$ est équivalent à $\pi_Y \circ \cap T_{r,1}...T_{r,p+q-1} \circ \pi_X^{-1}$.

Il reste à montrer le caractère fonctionnel de $\pi_Y \circ \cap T_{r,j} \circ \pi_X^{-1}$. Neuf cas sont à envisager selon l'environnement de $T_{r,j}$ dans le langage rationnel Q' :

- précédé de $\sharp$, précédé de $\natural$ ou $j = 1$

- suivi de $\sharp$, suivi de $\natural$ ou $j = p + q - 1$.

Pour chaque cas, l'analyse suit le même raisonnement ; on se reporte à la projection sur Z^*.

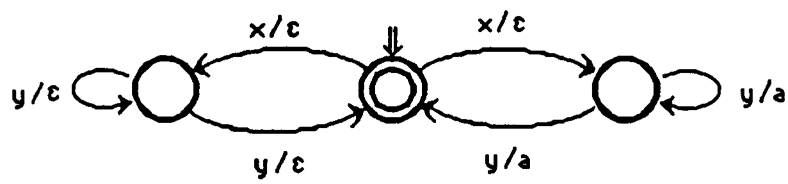
Examinons le cas où $T_{r,j}$ est précédé de $\sharp$ et suivi de $\natural$. Cela signifie que $\pi_{X \cup Z}(T_{r,j}) \subseteq FG(R_k)$ pour un certain k et que $\pi_{Z \cup Y}(T_{r,j}) \subseteq FD(S_{k'})$ pour un certain k'. Supposons qu'il existe $w \in X^*$ tel que $\|\pi_Y \circ \cap T_{r,j} \circ \pi_X^{-1}(w)\| \geq 2$. Il existe donc deux mots u_1, $u_2 \in T_{r,j}$ ayant même projection sur X^*. Alors, soit ces deux mots ont même projection sur Z^* et $\pi_Y \circ \cap S_{k'} \circ \pi_Z^{-1}$ n'est pas fonctionnelle, soit ces deux mots ont des projections différentes sur Z^* et $\pi_Z \circ \cap R_k \circ \pi_X^{-1}$ n'est pas fonctionnelle.

Les autres cas seraient traités de manière analogue.

Donc $\pi_Y \circ \cap T_{r,j} \circ \pi_X^{-1}$ est une fonction rationnelle et la tranduction $\rho \circ \rho'$ est polynôme de fonctions rationnelles. □

Nous poursuivons cette étude en montrant l'inclusion stricte de l'ensemble des polynômes en des transductions rationnelles fonctionnelles dans l'ensemble des transductions rationnelles polynômialement bornées. Nous vérifierons que la transduction cyclique d'image finie τ_a, décrite ci-dessous, ne peut être un polynôme de fonctions rationnelles.

$$\tau_a \ : \quad X^* = \{x, y\}^* \longrightarrow a^*$$



$$\tau_a = \pi_a \circ \cap R \circ \pi_X^{-1} \text{ avec } R = (x(ya)^+ + xy^+)^*$$

Etablissons d'abord quelques propriétés des fonctions rationnelles sur des alphabets à une lettre et des fonctions rationnelles ayant des domaines particuliers.

Le lemme suivant permet de caractériser les fonctions de y^* dans a^* (cf [27] exercice IX.8.1).

Lemme V.4

Soit $f = \pi_a \circ \cap R \circ \pi_y^{-1}$ une fonction rationnelle de y^* dans a^*, y et a étant deux lettres. Nous pouvons choisir R tel que celui-ci soit une union finie de langages rationnels de la forme uv^*, avec u, $v \in \{a, y\}^*$.

Preuve :

Tout langage rationnel R contient un langage rationnel borné R' tel que $\Psi(R) = \Psi(R')$, propriété établie dans [44]. Les fonctions $\pi_a \circ \cap R \circ \pi_y^{-1}$ et $\pi_a \circ \cap R' \circ \pi_y^{-1}$ sont équivalentes. Etant borné, le langage rationnel R' est une union finie de langages de la forme

$$w_1^{s_1}(w_1^{t_1})^* ... w_n^{s_n}(w_n^{t_n})^* = u_1 v_1^* u_2 ... u_m v_m^* u_{m+1} \quad \text{avec} \quad v_i \neq \varepsilon \quad \text{et} \quad m \geq 0.$$

Examinons le cas où $m \geq 2$. Comme f est une fonction, nous avons

$$|v_i|_y \neq 0 \quad \text{et} \quad \frac{|v_i|_a}{|v_i|_y} = \frac{|v_j|_a}{|v_j|_y} \text{ pour tout } i, j \in [1, m].$$

Posons $p = ppcm(|v_m|_y, |v_{m-1}|_y)$. Dans le langage rationnel R', nous pouvons remplacer

$$u_1 v_1^* u_2 \ldots u_{m-1} v_{m-1}^* u_m v_m^* u_{m+1} \quad \text{par} \quad \bigcup_{k=0}^{p-1} u_1 v_1^* u_2 \ldots u_{m-1} v_{m-1}^* u_m v_m^k u_{m+1} \ ;$$

un raisonnement par récurrence nous permet de remplacer $u_1 v_1^* u_2 \ldots u_{m-1} v_{m-1}^* u_m v_m^* u_{m+1}$ par une union finie de langages rationnels de la forme $u_1 v_1^* u_2'$. Donc il existe un langage

$$R'' = \bigcup_{i=1}^{r} u_i v_i^* u_i'$$

tel que $f = \pi_a \circ \cap R \circ \pi_y^{-1} = \pi_a \circ \cap R'' \circ \pi_y^{-1}$. Il est alors évident qu'en prenant

$$S = \bigcup_{i=1}^{r} u_i u_i' v_i^*$$

on a $f = \pi_a \circ \cap S \circ \pi_y^{-1}$. $\square$

Enonçons quelques propriétés qui nous permettrons de passer, sous certaines conditions, d'un produit de fonctions à une fonction et inversement.

Soient X et A deux alphabets disjoints.

Propriété V.5

Soient f_1 et f_2 deux fonctions de X^* dans A^*. Soit $x \in X$. Si $Dom(f_1) \subseteq X^*x$ et $Dom(f_2) \subseteq (X \setminus \{x\})^*$, le produit $f_1 \times f_2$ est une fonction.

Partant d'une relation rationnelle, il est possible de factoriser cette relation pour obtenir un polynôme de relations rationnelles. En particulier, nous avons les propriétés suivantes :

Propriété V.6

Soit f une fonction rationnelle de X^* dans A^* telle que $Dom(f) \subseteq (X \setminus \{x\})^* x X^*$. Alors, f est une somme finie de produits de fonctions rationnelles $h_i \times g_i$ avec $Dom(h_i) \subseteq (X \setminus \{x\})^*$ et $Dom(g_i) \subseteq xX^*$.

Propriété V.7

Soit f une fonction rationnelle de X^* dans A^* telle que $Dom(f) \subseteq X^* x (X \setminus \{x\})^*$. Alors, f est une somme finie de produits de fonctions rationnelles $g_i \times h_i$ avec $Dom(g_i) \subseteq X^*x$ et $Dom(h_i) \subseteq (X \setminus \{x\})*$.

Les propriétés précédentes concernaient des fonctions rationnelles particulières sans faire intervenir notre exemple τ_a. La suivante est liée à la façon dont le domaine de τ_a peut être factorisé.

Lemme V.8

Soit $f_1 \times ... \times f_p$ un monôme de fonctions rationnelles tel que, pour tout $w \in \{x,y\}^*$, $f_1 \times ... \times f_p(w) \subseteq \tau_a(w)$. Si $Dom(f_i) \subseteq y^*$ et $Dom(f_{i+1}) \subseteq y^*$, le produit $f_i \times f_{i+1}$ est une somme finie de fonctions rationnelles.

Preuve :

Le lemme V4 nous permet de considérer le produit $f_i \times f_{i+1}$ en tant que somme finie de produits $f_i' \times f_{i+1}'$ où $f_i' = \pi_a \circ \cap R_i \circ \pi_{\{x,y\}}^{-1}$ et $f_{i+1}' = \pi_a \circ \cap R_{i+1} \circ \pi_{\{x,y\}}^{-1}$ avec $R_i = u_i v_i^*$ et $R_{i+1} = u_{i+1} v_{i+1}^*$.

Considérons un produit $f_i \times f_{i+1}$ caractérisé par de tels langages rationnels.

Si $\pi_y(v_i) = \varepsilon$, alors, pour tout mot de $\{x,y\}^*$, il existe au plus une factorisation pouvant donner une image. Nous avons :

$$f_i \times f_{i+1}(w) = f_i(\pi_y(u_i)) f_{i+1}(w') \quad \text{si } w = \pi_y(u_i) w'$$
$$f_i \times f_{i+1}(w) = \emptyset \quad \text{sinon}$$

Même démarche si $\pi_y(v_{i+1}) = \varepsilon$.

Si $\pi_y(v_i) \neq \varepsilon$ et $\pi_y(v_{i+1}) \neq \varepsilon$, montrons que $\frac{|v_i|_a}{|v_i|_y} = \frac{|v_{i+1}|_a}{|v_{i+1}|_y}$. Supposons $\frac{|v_i|_a}{|v_i|_y} > \frac{|v_{i+1}|_a}{|v_{i+1}|_y}$. Posons $r = |v_i|_y$ et $s = |v_{i+1}|_y$. Posons $\bar{v}_i = v_i^s$ et $\bar{v}_{i+1} = v_{i+1}^r$. Nous avons $|\bar{v}_i|_y = |\bar{v}_{i+1}|_y$ et $|\bar{v}_i|_a > |\bar{v}_{i+1}|_a \geq 0$.

Les fonctions $f_1, f_2, ..., f_{i-1}, f_{i+2}, ..., f_p$ sont supposées non triviales, c'est-à-dire caractérisées par des langages rationnels non vides. Il existe donc des mots $w_1 \in R_1$, $w_2 \in R_2$,... Posons $w = w_1 ... w_{i-1}$ et $w' = w_{i+2} ... w_p$.

Il existe un entier k tel que $|\bar{v}_i^k|_a = k|\bar{v}_i|_a > |w u_i u_{i+1} w'|_y$ et un entier k' tel que $|\bar{v}_i^{k'}|_a - |\bar{v}_{i+1}^{k'}|_a = k'(|\bar{v}_i|_a - |\bar{v}_{i+1}|_a) > |w u_i u_{i+1} w'|_y$.

Les deux mots $w u_i \bar{v}_i^{k+k'} u_{i+1} w'$ et $w u_i \bar{v}_i^k u_{i+1} \bar{v}_{i+1}^{k'} w'$ appartiennent à $R_1 ... R_p$ et ont la même projection α sur $\{x,y\}^*$. Leurs projections respectives sur a^* appartiennent à l'image du mot α.

La définition des entiers k et k' nous permet de déduire que

$$\pi_a(w u_i \bar{v}_i^{k+k'} u_{i+1} w') = a^m \text{ avec } m > |w u_i u_{i+1} w'|_y$$
$$\pi_a(w u_i \bar{v}_i^k u_{i+1} \bar{v}_{i+1}^{k'} w') = a^{m'} \text{ avec } m' > |w u_i u_{i+1} w'|_y$$
$$\text{et } m - m' = k'(|\bar{v}_i|_a - |\bar{v}_{i+1}|_a) > |w u_i u_{i+1} w'|_y$$

La transduction τ_a transforme les lettres y, situées entre deux lettres x, soit en autant de a, soit en ε ; puisque $\pi_{\{x,y\}}(u_i \bar{v}_i u_{i+1} \bar{v}_{i+1}) \subseteq y^*$, la transduction τ_a effectue la même transformation sur tous les y apparaissant dans $\pi_{\{x,y\}}(u_i \bar{v}_i^k u_{i+1} \bar{v}_{i+1}^{k'})$.

On obtient alors la contradiction suivante :

L'inégalité $m' > |wu_iu_{i+1}w'|_y$ signifie que les y apparaissant dans $\pi_{\{x,y\}}(\bar{v}_i^k\bar{v}_{i+1}^{k'})$ sont modifiés en a par τ_a pour obtenir $a^{m'}$.

Par contre, l'inégalité $m - m' > |wu_iu_{i+1}w'|_y$ et le fait que a^m, $a^{m'} \in \tau_a(\alpha)$ signifient que les y apparaissant dans $\pi_{\{x,y\}}(\bar{v}_i^k\bar{v}_{i+1}^{k'})$ sont transformés en ε par τ_a pour obtenir $a^{m'}$.

Un raisonnement identique pour $\frac{|v_i|_a}{|v_i|_y} < \frac{|v_{i+1}|_a}{|v_{i+1}|_y}$ amène la même contradiction.

Sachant que $\frac{|v_i|_a}{|v_i|_y} = \frac{|v_{i+1}|_a}{|v_{i+1}|_y}$, posons $q = ppcm(|v_i|_y, |v_{i+1}|_y)$. Nous obtenons alors $f_i \times f_{i+1} = \pi_a \circ \cap R \circ \pi_{\{x,y\}}^{-1}$ avec

$$R = \bigcup_{j=0}^{q} u_iv_i^*u_{i+1}v_{i+1}^j$$

et donc, $f_i \times f_{i+1}$ est une fonction rationnelle. □

Lemme V.9

Soit τ'_a un polynôme de fonctions rationnelles tel que, pour tout $w \in \{x,y\}^*$, on ait $\tau'_a(w) \subseteq \tau_a(w)$. Il existe un polynôme de fonctions rationnelles équivalent dont chaque fonction possède un domaine inclus dans $(xy^+)^*$.

Preuve :

Nous allons opérer, sur les fonctions qui n'auraient pas un domaine inclus dans $(xy^+)^*$, des factorisations et des regroupements en utilisant les propriétés précédentes. L'application d'une propriété fournit un polynôme équivalent.

Considérons une telle fonction f_i. Si son domaine $Dom(f_i)$ est inclus dans y^*, on pose $h_i = f_i$; sinon la propriété V.6 nous permet de décomposer f_i en une union de $h_i \times g_i$ avec $Dom(h_i) \subseteq y^*$, $Dom(g_i) \subseteq (xy^*)^+$. Si $Dom(f_{i-1}) \subseteq y^*$, on pose $h_{i-1} = f_{i-1}$; sinon la propriété V.7 est appliquée sur f_{i-1}, devenant ainsi une union de $g_{i-1} \times h_{i-1}$ avec $Dom(h_{i-1}) \subseteq y^*$, $Dom(g_{i-1}) \subseteq (y^*x)^+$. Le lemme V.8 nous permet de regrouper chaque produit $h_{i-1} \times h_i$ en une somme de fonctions h'_{i-1}. Enfin, en utilisant la propriété V.5, nous transformons $g_{i-1} \times h'_{i-1}$ en g'_{i-1}. Nous avons donc remplacé un monôme par un polynôme dont toutes les i-èmes fonctions ont un domaine inclus dans $(xy^*)^*$. Si nous appliquons ce procédé sur toutes les fonctions, nous obtenons un polynôme dont chaque fonction a un domaine inclus dans $(xy^*)^*$. Comme le domaine de τ_a est $(xy^+)^*$, chaque fonction a un domaine inclus dans $(xy^+)^*$. □

Proposition V.10

L'ensemble des polynômes en des fonctions rationnelles est strictement inclus dans l'ensemble des transductions rationnelles polynômialement bornées.

Preuve :

La transduction τ_a est polynômialement bornée. Supposons que ce soit un polynôme de fonctions rationnelles. D'après le lemme V.9, nous pouvons choisir ce polynôme tel que chaque fonction possède un domaine inclus dans $(xy^+)^*$. Soit p le nombre maximum de fonctions utilisées dans un monôme et n le nombre de monômes. Le nombre de factorisations d'un mot $w = xy^{i_0}xy^{i_1}...xy^{i_r}$ pouvant donner des images différentes est borné par nC_r^{r+p-1}, ce qui est en contradiction avec le fait qu'il existe des mots w ayant 2^{r+1} images ; prenons, par exemple, $w = xyxy^2xy^4...xy^{2^r}$. □

Ainsi, la différence entre transductions rationnelles cycliques, polynômes de fonctions rationnelles et transductions rationnelles polynômialement bornées est établie.

Le contre-exemple τ_a est une transduction rationnelle étoilée décroissante. Donc il appartient à $\mathbf{H} \circ \mathbf{H}^{-1} \circ \mathbf{H}$ [45]. Examinons les compositions de morphismes et de morphismes inverses.

Un contre-exemple peut être trouvé dans $\mathbf{H} \circ \mathbf{H}^{-1}$. En effet, les morphismes sont des fonctions et l'ensemble des polynômes de fonctions rationnelles est clos par composition. Si toutes les transductions rationnelles cycliques de $\mathbf{H} \circ \mathbf{H}^{-1}$ étaient des polynômes de fonctions rationnelles, cette propriété serait alors vérifiée par les transductions rationnelles cycliques de $\mathbf{H} \circ \mathbf{H}^{-1} \circ \mathbf{H}$.

Proposition V.11

Il existe des transductions rationnelles cycliques d'image finie dans $\mathbf{H} \circ \mathbf{H}^{-1}$ *qui ne sont pas des polynômes de fonctions rationnelles.*

Remarquons qu'inversement toutes les transductions rationnelles cycliques d'image finie de $\mathbf{H}^{-1} \circ \mathbf{H}$ sont des fonctions.

Terminons cette étude en traitant le cas des transductions rationnelles cycliques appartenant à $\mathbf{H} \circ \mathbf{H}_\alpha^{-1} \circ \mathbf{H}$. Puisque nous ne considérons que des transductions d'image finie, si une transduction cyclique appartient à $\mathbf{H} \circ \mathbf{H}_\alpha^{-1} \circ \mathbf{H}$, nous pouvons trouver une transduction équivalente dans $\mathbf{H} \circ \mathbf{H}_{s\alpha}^{-1} \circ \mathbf{H}$.

L'ensemble $\mathbf{H} \circ \mathbf{H}_{s\alpha}^{-1}$ est celui des substitutions finies. La proposition suivante implique que toute transduction rationnelle cyclique d'image finie appartenant à $\mathbf{H} \circ \mathbf{H}_{\alpha}^{-1} \circ \mathbf{H}$ est un polynôme de fonctions rationnelles.

Proposition V.12

Toute substitution finie cyclique est un polynôme de fonctions rationnelles cycliques.

Preuve :

Soit σ une substitution finie cyclique de X^* dans le monoïde libre engendré par un mot m. L'image de chaque lettre de X est un sous-ensemble de m^*. Nous pouvons donc analyser cette substitution comme la composition d'une substitution de X^* dans a^* où $a \notin X$ et d'un morphisme (fonction) qui transforme la lettre a en un mot m.

Soit p_i le nombre d'images de la lettre x_i par la substitution σ. Nous pouvons noter $\sigma(x_i) = \{w_{ij} \in a^* \ / \ 1 \leq j \leq p_i\}$. Puisque tout mot w_{ij} appartient à a^*, l'ordre de la transformation des lettres x_i n'a pas d'importance. Seul compte le nombre de x_i transformés en un w_{ij} donné.

Soit n le cardinal de X. La substitution σ est équivalente à une composition de substitutions T_i sur une seule lettre x_i. Nous construisons donc $T_i = f_{i1} \times ... \times f_{ip_i}$ où les fonctions f_{ij} sont les morphismes de $(X \cup \{a\})^*$ dans $(X \cup \{a\})^*$ définis par

$$f_{ij}(x_i) = w_{ij}$$

$$f_{ij}(x_k) = x_k$$

$$f_{ij}(a) = a$$

$$\sigma = T_n \circ T_{n-1} \circ ... \circ T_1$$

Puisque l'ensemble des polynômes de fonctions rationnelles est clos par composition, la substitution σ appartient encore à cet ensemble. □

Dans cette dernière partie nous examinerons quelques propriété du produit de Hadamard sur les polynômes de fonctions rationnelles. Le produit de Hadamard est habituellement défini pour des séries formelles. Nous utiliserons la même notation pour des transductions. Ces propriétés nous permettrons de construire un exemple de transduction rationnelle cyclique d'image finie appartenant à $\mathbf{H} \circ \mathbf{H}^{-1}$ et qui n'est pas un polynôme de fonctions rationnelles.

Définition

Le produit de Hadamard de deux transductions τ_1 et τ_2 de X^* dans Y^*, noté $\tau_1 \odot \tau_2$, est défini par $\tau_1 \odot \tau_2(w) = \tau_1(w)\tau_2(w)$, pour tout mot w de X^*.

Généralement cette opération ne conserve pas la rationnalité des transductions. Il suffit pour s'en convaincre de prendre pour τ_1 l'identité sur $\{a\}$ et pour τ_2 la transduction rationnelle qui transforme la lettre a en b. On a alors, $\tau_1 \odot \tau_2(a^*) = \{a^n b^n \ / \ n \in \mathbb{N}\}$. C. Choffrut a établi certaines conditions pour que le produit de Hadamard de deux transductions rationnelles soit rationnel [16,17]. Pour les transductions rationnelles cycliques, nous avons le résultat suivant :

Proposition V.13

Le produit de Hadamard de deux transductions rationnelles cycliques est une transduction rationnelle cyclique.

Cette propriété est une conséquence directe de la proposition suivante, démontrée par C. Choffrut.

Proposition V.14 [16]

Soient τ_1 et τ_2 deux transductions rationnelles de X^* dans Y^* et B un langage rationnel inclus dans Y^* tel que $B^n \cap B^p \neq \emptyset$ entraîne $n = p$. Si, pour tout $w \in X^*$, il existe $I,\ J \in \mathbb{N}$ tels que l'on ait

$$\tau_1(w) = \bigcup_{i \in I} B^i \quad \text{et} \quad \tau_2(w) = \bigcup_{j \in J} B^j$$

alors la transduction $\tau_1 \odot \tau_2$ est rationnelle.

Nous pouvons aussi remarquer que la proposition V.13 se déduit immédiatement de résultats sur les séries formelles à coeffiçients dans un semi-anneau commutatif (cf [16,27]).

Montrons que la proposition V.13 reste vraie lorsqu'il s'agit de polynômes de fonctions rationnelles cycliques. Pour cela nous utiliserons le résultat de S. Eilenberg, connu sous le nom de "*Cross-Section Theorem*".

Théorème V.15 [27]

Soient h un morphisme de X^* dans Y^* et L un langage rationnel inclus dans X^*. Alors il existe un langage rationnel $L' \subseteq L$ tel que h est une bijection de L' dans $h(L)$.

Proposition V.16

Le produit de Hadamard de deux polynômes de fonctions rationnelles cycliques de X^* dans le monoïde libre engendré par une lettre a est un polynôme de fonctions rationnelles cycliques.

Preuve :

Soient
$$\tau_1 = \sum_{i=1}^{p_1}\prod_{j=1}^{q_1} f_{ij} \qquad \text{et} \qquad \tau_2 = \sum_{k=1}^{p_2}\prod_{l=1}^{q_2} g_{kl}$$
deux polynômes de fonctions rationnelles de X^* dans a^*. Nous supposerons que a n'appartient pas à X. Nous montrons que la transduction $\tau_1 \odot \tau_2$ peut être remplacée par une composition de polynômes de fonctions rationnelles.

Transformons le polynôme τ_1 en un polynôme τ_1' tel que les images d'un mot contiennent des informations sur le mot de départ et sur son image par τ_1. Chaque fonction f_{ij} peut être caractérisée par un langage rationnel tel que $f_{ij} = \pi_a \circ \cap R_{ij} \circ \pi_X^{-1}$. Définissons le morphisme θ qui marque les lettres a par $\theta(x) = x$, pour tout $x \in X$, et $\theta(a) = \bar{a}$. Du théorème V.15 nous déduisons l'existence d'un langage rationnel $R'_{ij} \subseteq \theta(R_{ij})$ tel que π_X est une bijection entre R'_{ij} . Donc $f'_{ij} = \cap R'_{ij} \circ \pi_X^{-1}$ est une fonction de X^* dans $(X \cup \{\bar{a}\})^*$. Cette fonction f'_{ij} posséde le même domaine que f_{ij} car $Dom(f'_{ij}) = \pi_X(\theta(R_{ij})) = \pi_X(R_{ij}) = Dom(f_{ij})$. De plus elle vérifie $|f'_{ij}(w)|_{\bar{a}} = |f_{ij}(w)|_a$ et $\pi_X(f'_{ij}(w)) = w$, pour tout $w \in X^*$. Posons
$$\tau_1' = \sum_{i=1}^{p_1}\prod_{j=1}^{q_1} f'_{ij}.$$

Nous devons maintenant effectuer la transformation due à τ_2 tout en préservant les lettres $\bar{a}$. Supposons que $g_{kl} = \pi_a \circ \cap S_{kl} \circ \pi_X^{-1}$. En utilisant une nouvelle fois le théorème V.15 nous déduisons l'existence d'un langage rationnel $S'_{kl} \subseteq S_{kl} \sqcup\!\sqcup\, \bar{a}^*$ tel que $\pi_{X\cup\bar{a}}$ est une bijection entre S'_{kl} et $\pi_{X\cup\bar{a}}(S_{kl} \sqcup\!\sqcup\, \bar{a}^*)$. Alors $g'_{kl} = \pi_{\{a,\bar{a}\}} \circ \cap S'_{kl} \circ \pi_{X\cup\bar{a}}^{-1}$ est une fonction de $(X \cup \{\bar{a}\})^*$ dans $\{a, \bar{a}\}^*$. La projection sur X du domaine de cette fonction est égale au domaine de g_{kl} car $\pi_X(Dom(g'_{kl})) = \pi_X(S'_{kl}) = \pi_X(S_{kl}) = Dom(g_{kl})$. De plus, pour tout $v \in (X \cup \{\bar{a}\})^*$, on a $|g'_{kl}(v)|_a = |g_{kl}(\pi_X(v))|_a$ et $|g'_{kl}(v)|_{\bar{a}} = |v|_{\bar{a}}$. Posons
$$\tau_2' = \sum_{k=1}^{p_2}\prod_{l=1}^{q_2} g'_{kl}.$$

A ce niveau, les mots, images par $\tau_2' \circ \tau_1'$, contiennent les informations sur l'image par τ_1, lettres $\bar{a}$, et par τ_2, lettres a. Le dernier polynôme sera le morphisme η défini par $\eta(a) = \eta(\bar{a}) = a$.

Les polynômes de fonctions rationnelles étant clos par composition, nous pouvons conclure que $\tau_1 \odot \tau_2 = \eta \circ \tau_2' \circ \tau_1'$ est un polynôme de fonctions rationnelles cycliques. □

Le problème suivant a été évoqué par S. Eilenberg dans [27] pour des séries formelles de même support :

Si $\sigma \odot \rho$ et ρ sont rationnelles, en est-il de même pour σ ?

Ce problème fut résolu négativement par J. Karhumäki [39]. Dans [16], C. Choffrut établit certaines conditions pour que la réponse soit affirmative. Dans le cas des transductions cycliques, nous avons :

Lemme V.17

Soient τ_1 et τ_2 deux transductions rationnelles cycliques de X^* dans a^* telles que, pour tout $u \in Dom(\tau_1) \cap Dom(\tau_2)$, pour tout $v_1 \in \tau_1(u)$ et $v_2 \in \tau_2(u)$ on ait $|v_2| \leq |v_1|$. La transduction τ, définie, pour tout u de $Dom(\tau_1) \cap Dom(\tau_2)$ par $\tau(u) = \{a^{i-j} \; / \; a^i \in \tau_1(u) \text{ et } a^j \in \tau_2(u)\}$ est rationnelle.

Preuve :

Posons $\tau_1 = \pi_a \circ \cap R_1 \circ \pi_X^{-1}$ et $\tau_2 = \pi_a \circ \cap R_2 \circ \pi_X^{-1}$. Nous définissons la transduction $\delta = \cap R_2 \circ \pi_X^{-1} \circ \pi_X \circ \cap R_1$ de $(X \cup \{a\})^*$ dans $(X \cup \{a\})^*$. A tout mot w_1 de R_1, cette transduction associe l'ensemble des mots de R_2 qui ont la même projection sur X^* que w_1. Si nous considérons une projection u appartenant au domaine, la différence de longueur entre $w_1 \in R_1$, tel que $\pi_X(w_1) = u$, et une de ses images par δ représente la longueur d'un mot de $\tau(u)$.

La transduction δ est décroissante. On déduit de la caractérisation des transductions décroissantes [47] l'existence de deux morphismes, g alphabétique et h strictement alphabétique, et d'un langage rationnel R tel que $\delta = g \circ \cap R \circ h^{-1}$.

Définissons alors g' par $g'(x) = a$ quand $g(x) = \varepsilon$ et $g'(x) = \varepsilon$ quand $g(x) \neq \varepsilon$. La transduction $g' \circ \cap R \circ h^{-1}$ associe, à tout mot w_1 de $Dom(\tau_1) \cap Dom(\tau_2)$, le mot a^m où $m = |w_1| - |\delta(w_1)|$.

On vérifie alors que $\tau = g' \circ \cap R \circ h^{-1} \circ \cap R_1 \circ \pi_X^{-1}$ convient. $\square$

Nous obtenons comme conséquence

Corollaire V.18

Soient $\tau_1 \odot \tau_2$ et τ_1 deux fonctions rationnelles cycliques de X^* dans a^*. La transduction τ_2 est une fonction rationnelle si $Dom(\tau_2) \subseteq Dom(\tau_1)$.

On remarque facilement que la condition sur le domaine est nécessaire. En effet considérons les transductions τ_1 et τ_2 définies par leurs graphes $\hat{\tau}_1 = \{(a^n b^n, a^n) \ / \ n \in \mathbf{N}\}$ et $\hat{\tau}_2 = \{(ab, a)\}$. Ces deux transductions sont fonctionnelles et τ_1 n'est pas rationnelle. Par contre $\tau_1 \odot \tau_2$ est rationnelle puisque son graphe est fini $\widehat{\tau_1 \odot \tau_2} = \{(ab, a^2)\}$.

Le problème du domaine ne se pose pas lorsque τ_1 est un morphisme.

Proposition V.19

Soit $h \odot \tau$ un polynôme de fonctions rationnelles cycliques de X^* dans le monoïde libre engendré par une lettre a. Si h est un morphisme, la transduction τ est un polynôme de fonctions rationnelles.

Preuve :

Posons

$$h \odot \tau = \sum_{i=1}^{p} \prod_{j=1}^{q} f_{ij}.$$

Constatons tout d'abord que, pour toute fonction f_{ij}, il existe une borne $d_{ij} \in \mathbf{N}$ telle que $|f_{ij}(u)| - |h(u)| \geq -d_{ij}$, pour tout $u \in Dom(f_{ij})$. Si cela n'était pas vérifié, il serait aisé de trouver des mots ayant des images de longueur négative.

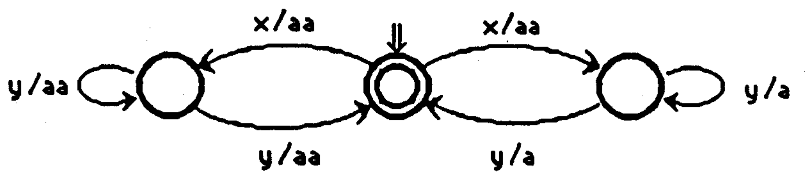
Soit D le plus grand des d_{ij}. Nous définissons alors les fonctions f'_{ij} par $f'_{ij}(u) = a^D f_{ij}(u)$, pour tout $u \in X^*$. Les fonctions f'_{ij} ont donc mêmes domaines que f_{ij}. Comme nous avons $|f'_{ij}(u)| \geq |h(u)|$, pour tout $u \in X^*$, nous pouvons considérer les fonctions g_{ij} telles que $f'_{ij} = h \odot g_{ij}$. D'aprés la propriété précédente, g_{ij} est une fonction rationnelle ; son domaine est celui de f_{ij}.

Posons

$$\tau' = \sum_{i=1}^{p} \prod_{j=1}^{q} g_{ij}.$$

Ce polynôme τ' vérifie l'équation : $\tau'(u) = a^{qD}\tau(u)$, pour tout $u \in X^*$. Définissons la fonction rationnelle η de a^* dans a^* par $\eta(v) = (a^{qD})^{-1}v$, pour tout $v \in a^*$. En tant que composée de polynômes de fonctions rationnelles, la transduction $\tau = \eta \circ \tau'$ est un polynôme de fonctions rationnelles. $\square$

Cette dernière proposition nous permet de construire d'autres exemples de transductions rationnelles cycliques d'image finie qui ne sont pas des polynômes de fonctions rationnelles. L'exemple τ'_a se construit à partir de l'exemple τ_a.



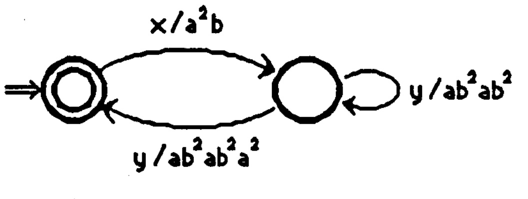
Si nous définissons le morphisme h de $\{x,y\}^*$ dans a^* par $h(x) = aa$ et $h(y) = a$. De façon évidente, $\tau'_a = h \odot \tau_a$ n'est pas un polynôme de fonctions rationnelles.

Terminons en donnant un exemple de transduction rationnelle cyclique d'image finie appartenant à $\mathbf{H}_{s\alpha}^{-1} \circ \mathbf{H}$ qui ne peut pas être un polynôme de fonctions rationnelles.

Définissons les morphismes g et h de $\{\delta_1, \delta_2, \delta_3, \alpha, \beta\}^*$ dans $\{a,b\}^*$ par

$h(\delta_1) = a^2b$	$g(\delta_1) = a$
$h(\delta_2) = a^2$	$g(\delta_2) = a$
$h(\delta_3) = ba^2$	$g(\delta_3) = a$
$h(\alpha) = ab^2ab^2$	$g(\alpha) = a$
$h(\beta) = bab$	$g(\beta) = a$

Si nous définissons la fonction rationnelle θ par le transducteur ci-dessous :



Nous obtenons $\tau'_a = g \circ h^{-1} \circ \theta$. Donc, si $g \circ h^{-1}$ était un polynôme de fonctions rationnelles, la transduction τ'_a en serait un aussi.

RÉFÉRENCES

[1] S. ARŠON, "Démonstration de l'existence de suites asymétriques infinies". Mat. Sb. 44 (1937), pp. 769-777.

[2] J.M. AUTEBERT, J. BEAUQUIER, L. BOASSON and M. LATTEUX, "Very small families of algebraic nonrational languages". in Formal Language Theory, Perspectives and Open Problems, R.V. Book (ed), Academic Press, 1980, pp. 89-107.

[3] J.M. AUTEBERT, J. BEAUQUIER, L. BOASSON and M. NIVAT, "Quelques problèmes en théorie des langages algébriques". RAIRO Informatique Théorique 13 (1979), pp. 363-379.

[4] J.M. AUTEBERT, P. FLAJOLET and J. GABARRÒ, "Prefixes of infinite words and ambiguous context-free languages". Information Processing Letters 25 (1987), pp.

[5] J.M. AUTEBERT and J. GABARRÒ, "Compléments des facteurs gauches de mots infinis et ambiguïté : quelques exemples". Report de Recerca RR85/15, Facultat d'Informàtica de Barcelona, 1985.

[6] J.M. AUTEBERT and J. GABARRÒ, "Iterated gsm's and co-CFL". Report de Recerca RR87/12, Facultat d'Informàtica de Barcelona, 1987.

[7] J. BEAUQUIER, "Deux familles de langages incomparables". Information and Control 43 (1979), pp. 101-122.

[8] J. BERSTEL, Transductions and Context-free Languages. Teubner, Stuttgart, 1979.

[9] J. BERSTEL, "Mots sans carré et morphismes itérés". Discrete Mathematics 29 (1980), pp. 235-244.

[10] J. BERSTEL, "Every iterated morphism yields a co-CFL". Information Processing Letters 22 (1986), pp. 7-9.

[11] J. BERSTEL, "Complémentaires de mots infinis". Séminaire du LITP, Paris, 1986.

[12] M. BLATTNER and M. LATTEUX, "Parikh-bounded languages". 8^{th} ICALP, Acre, 1981, Lecture Notes in Computer Science 115, pp. 316-323.

[13] L. BOASSON and M. NIVAT, "Adherences of languages".
Journal of Computer and System Sciences 20 (1980), pp. 285-309.

[14] R.V. BOOK, "On languages with a certain prefix property".
Mathematical Systems Theory 10 (1977), pp. 229-237.

[15] W. BUCHER, D. HAUSSLER and M.G. MAIN, "Applications of an infinite square free co-CFL".
12^{th} ICALP, Napflion, 1985, Lecture Notes in Computer Science 194.

[16] C. CHOFFRUT, "Contribution à l'étude de quelques familles remarquables de fonctions rationnelles".
Thèse de Doctorat d'état, 1978, Université de Paris VII.

[17] C. CHOFFRUT, "Sur les transductions reconnaissables".
RAIRO Informatique Théorique 12 (1978), pp. 203-212.

[18] A. COBHAM, "Uniform Tag Sequences".
Mathematical Systems Theory 6 (1972), pp. 164-192.

[19] K. ČULIK II, "On some families of languages related to developmental systems".
International Journal of Computer Mathematics 4 (1974), pp. 31-42.

[20] K. ČULIK II and I. FRIS, "The decidability of the equivalence problem for $D0L$-systems".
Information and Control 35 (1977), pp. 20-39.

[21] K. ČULIK II and T. HARJU, "The ω-sequence equivalence problem for $D0L$-systems is decidable"
Journal of the Association for Computing Machinery 31 (1984), pp. 282-298.

[22] K. ČULIK II and A. SALOMAA, "On infinite words obtained by iteraring morphisms".
Theoretical Computer Science 19 (1982), pp. 29-38.

[23] A. EHRENFEUCHT and G. ROZENBERG, "On inverse homomorphic images of deterministic $ET0L$-languages".
in Automata, Languages, Development, A. Lindenmayer and G. Rozenberg (eds), North-Holland Publ., Amsterdam, 1976, pp. 179-189.

[24] A. EHRENFEUCHT and G. ROZENBERG, "On some context-free languages that are not deterministic $ET0L$-languages".
RAIRO Informatique Théorique 11 (1977), pp. 273-291.

[25] A. EHRENFEUCHT and G. ROZENBERG, "Elementary homomorphisms and a solution of the $D0L$ sequence equivalence problem".
Theoretical Computer Science 7 (1978), pp. 169-183.

[26] A. EHRENFEUCHT, G. ROZENBERG and S. SKYUM, "A relation between $ET0L$ and $EDT0L$-languages".
Theoretical Computer Science 1 (1976), pp. 325-330.

[27] S. EILENBERG, Automata, Languages and Machines.
Vol. A, Academic Press, New York, 1974.

[28] C.C. ELGOT and G. MEZEI, "On relations defined by generalized finite automata".
IBM Journal of Research and Development 9 (1965), pp. 47-68.

[29] J. ENGELFRIET, "Top-down tree transducers with regular look-ahead".
Mathematical Systems Theory, 10 (1977), pp. 289-303.

[30] J. ENGELFRIET, G. ROZENBERG and G. SLUTZKI, "Tree transducers, L-systems and two-way machines".
Journal of Computer and System Sciences 20 (1980), pp. 150-202.

[31] D. FERMENT, "Principality results about some matrix languages families".
Lecture Notes in Computer Science 172, pp. 151-161.

[32] A. GRAZON, "Contribution à l'étude des petites familles de langages".
Thése de 3^e cycle, 1985, Université de Paris VII.

[33] A. GRAZON, "An infinite word language which is not co-CFL".
Information Processing Letters 24 (1987), pp. 81-85.

[34] T. HEAD, "Adherence equivalence is decidable for $D0L$-languages".
STACS 84, Paris, 1984, Lecture Notes in Computer Science 166, pp. 241-249.

[35] T. HEAD, "Adherences of $D0L$-languages".
Theoretical Computer Science 31 (1984), pp. 139-149.

[36] G.T. HERMAN and G. ROZENBERG, Developmental Systems and Languages. North-Holland Publ., Amsterdam, 1975.

[37] S. ISTAIL, "On irreductible languages and nonrational numbers". Bull. Soc. Math. Roumanie 21 (1977), pp. 301-308.

[38] A.K. JOSHI and T. YOKOMORI, "Semi-linearity, Parikh-boundedness and tree adjunct languages". Information Processing Letters 17 (1983), pp. 137-143.

[39] J. KARHUMÄKI, "Remarks on commutative N-rational series". Theoretical Computer Science 5 (1977), pp. 211-217.

[40] J. KARHUMÄKI and M. LINNA, "A note on morphic characterization of languages". Discrete Applied Math. 5 (1983), pp. 243-246.

[41] M. LATTEUX, "$EDT0L$-systèmes et opérateurs associés". Publication du Laboratoire de Calcul de l'Université de Lille I n 100 (1977).

[42] M. LATTEUX, "Substitutions dans les $EDT0L$-systèmes ultralinéaires". Information and Control 42 (1979), pp. 194-260.

[43] M. LATTEUX, "Sur les générateurs algébriques et linéaires". Acta Informatica 13 (1980), pp. 347-363.

[44] M. LATTEUX and J. LEGUY, "Une propriété de la famille GRE". FCT'79, Berlin, 1979, Akademie-Verlag, pp. 255-261.

[45] M. LATTEUX and J. LEGUY, "On the composition of morphisms and inverse morphisms". Lecture Notes in Computer Science 154, 1983, pp. 420-432.

[46] M. LATTEUX and P. TURAKAINEN, "A new normal form for composition of morphisms and inverse morphisms". à paraître dans Math. Systems Theory.

[47] J. LEGUY, "Transductions rationnelles décroissantes". RAIRO Informatique Théorique 5 (1981), pp.141-148.

[48] M. LINNA, "On periodic ω-sequences obtained by iterating morphisms". Ann. Univ. Turkuensis, Ser. A I 186 (1984), pp. 64-71.

[49] M.G. MAIN, "An infinite square free co-CFL". Information Processing Letters 20 (1985), pp. 105-107.

[50] E. MAKINEN, "On context-free and Szilard languages". BIT 24 (1984), pp.164-170.

[51] M. NIVAT, "Transductions des langages de Chomsky". Annales de l'Institut Fourier 18 (1968), pp. 339-456.

[52] J.J. PANSIOT, "Hiérarchie et fermeture de certaines classes de Tag-systèmes". Acta Informatica 20 (1983), pp. 179-196.

[53] J.J. PANSIOT, "Decidability of periodicity for infinite words". Informatique Théorique et Applications 20 (1986), pp. 43-46.

[54] G. PĂUN, "On the family of finite index matrix languages". Journal of Computer and System Sciences 18 (1979), pp. 267-280.

[55] G. PĂUN, "On the Parikh-boundedness of matrix-languages". Bulletin of EATCS 25 (1985), pp. 34-37.

[56] V. RAJLICH, "Absolutely parallel grammars and two-way finite state transducers". Journal of Computer and System Sciences 6 (1972), pp. 324-342.

[57] A. RESTIVO and C. REUTENAUER, "Some applications of a theorem of Shirshov to language theory". Information and Control 57 (1983), pp. 205-213.

[58] A. RESTIVO and C. REUTENAUER, "Cancellation, pumping and permutation in formal languages". 11^{th} ICALP, Antwerp, 1984, Lecture Notes in Computer Science 172, pp. 414-422.

[59] G. ROZENBERG and A. SALOMAA, The Mathematical Theory of L-Systems. Academic Press, New York, 1980.

[60] G. ROZENBERG and D. VERMEIR, "On $ET0L$-systems of finite index". Information and Control 38 (1978), pp. 103-133.

[61] G. ROZENBERG and D. VERMEIR, "On the effect of finite index restriction on several families of grammars".
Information and Control 39 (1978), pp. 284-302.

[62] B. ROZOY, "The Dyck language $D_1'^*$ is not generated by any matrix grammar of finite index".
Information and Control 74 (1987), pp. 64-89.

[63] P. SÉÉBOLD, "Ultimately periodic binary words and iterated morphisms".
Exposé aux journées LANFOR, 1987.

[64] M.P. SCHÜTZENBERGER, "Sur les relations rationnelles entre monoïdes libres".
Theoretical Computer Science 3 (1976), pp. 243-259.

[65] S. SKYUM, "Decomposition theorems for various kinds of languages parallel in nature".
SIAM Journal on Computing 5 (1976), pp. 284-296.

[66] A. THUE, "Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen".
Norske Videnskapsselsk Skrifter I, Math.-Natur. Kl. (1912), pp. 1-67.

[67] P. TURAKAINEN, "A homomorphic characterization of principal semi AFL's without using intersection with regular sets".
Information Sciences 27 (1982), pp. 141-149.

résumé

Nous poursuivons l'étude de certaines familles de langages de A. Lindenmayer.

La première partie est principalement consacrée aux D0L-langages. J. Berstel a montré que le centre d'un langage engendré par un morphisme prolongeable itéré est co-algébrique. Nous montrons que le résultat reste vrai quand le procédé itératif correspond à un D0L-système, à un gsm prolongeable ou à un Tag-système uniforme. Nous terminons cette partie en montrant que l'on peut décider si un mot infini ultimement périodique est la limite d'un D0L-langage.

Dans la seconde partie, nous étudions la famille des EDT0L-langages. Nous démontrons que les EDT0L-langages d'index fini sont Parikh-bornés. Nous apportons aussi une réponse à une conjecture de A. Ehrenfeucht et G. Rozenberg ; pour cela, nous montrons que tout langage appartenant au cône rationnel engendré par EDT0L peut être obtenu en appliquant, à un EDT0L-langage, un morphisme (non effaçant) inverse suivi d'un morphisme (alphabétique). Nous montrons ensuite que EDT0L est fermée par polynôme de fonctions rationnelles.

Les transductions rationnelles polynômialement bornées sont étudiées dans la troisième partie. Répondant à une question de M.P. Schützenberger, nous démontrons l'inclusion stricte de l'ensemble des polynômes de fonctions rationnelles dans l'ensemble des transductions rationnelles polynômialement bornées.

Mots clés :

- morphismes itérés
- D0L-langages
- EDT0L-langages
- centre de langages
- langages Parikh-bornés
- transductions rationnelles
- polynômes de transductions rationnelles
- transductions rationnelles à images finies
- transductions rationnelles cycliques