N° d'ordre : 411

THESES

présentées à

L'UNIVERSITE DES SCIENCES ET TECHNIQUES DE LILLE

pour obtenir le grade de

DOCTEUR ES SCIENCES MATHEMATIQUES

par

Michel LATTEUX

PREMIERE THESE : LANGAGES COMMUTATIFS

DEUXIEME THESE : Synchronisation de processus

Thèses soutenues le 31 mai 1978 devant la commission d'examen

MM.	P.	BACCHUS	Président
	J.	BERSTEL	Examinateur
	L.	BOASSON	Rapporteur
	V.	CORDONNIER	Examinateur
	G.	JACOB	Rapporteur
	M.	NIVAT	Rapporteur
	J.P.	VERJUS	Examinateur

UNIVERSITE DES SCIENCES
ET TECHNIQUES DE LILLE

01.02.1976

DOYENS HONORAIRES de l'Ancienne Faculté des Sciences

MM. R. DEFRETIN, H. LEFEBVRE, M. PARREAU.

PROFESSEURS HONORAIRES des Anciennes Facultés de Droit
et Sciences Economiques, des Sciences et des Lettres

M. ARNOULT, Mme BEAUJEU, MM. BROCHARD, CHAPPELON, CHAUDRON, CORDONNIER, CORSIN, DEHEUVELS, DEHORS, DION, FAUVEL, FLEURY, P. GERMAIN, HEIM DE BALSAC, HOCQUETTE, KAMPE DE FERIET, KOUGANOFF, LAMOTTE, LASSERRE, LELONG, Mme LELONG, MM. LHOMME, LIEBAERT, MARTINOT-LAGARDE, MAZET, MICHEL, NORMANT, PEREZ, ROIG, ROSEAU, ROUBINE, ROUELLE, SAVART, WATERLOT, WIEMAN, ZAMANSKI.

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

MM. R. DEFRETIN, M. PARREAU.

PRESIDENT DE L'UNIVERSITE
DES SCIENCES ET TECHNIQUES DE LILLE

M. J. LOMBARD.

PROFESSEURS TITULAIRES

M.	BACCHUS Pierre	Astronomie
M.	BEAUFILS Jean-Pierre	Chimie Physique
M.	BECART Maurice	Physique Atomique et Moléculaire
M.	BILLARD Jean	Physique du Solide
M.	BIAYS Pierre	Géographie
M.	BONNEMAN Pierre	Chimie Appliquée
M.	BONNOT Ernest	Biologie Végétale
M.	BONTE Antoine	Géologie Appliquée
M.	BOUGHON Pierre	Algèbre
M.	BOURIQUET Robert	Biologie Végétale
M.	CELET Paul	Géologie Générale
M.	CONSTANT Eugène	Electronique
M.	DECUYPER Marcel	Géométrie
M.	DELATTRE Charles	Géologie Générale
M.	DELHAYE Michel	Chimie Physique
M.	DERCOURT Michel	Géologie Générale
M.	DURCHON Maurice	Biologie Expérimentale
M.	FAURE Robert	Mécanique
M.	FOURET René	Physique du Solide
M.	GABILLARD Robert	Electronique
M.	GLACET Charles	Chimie Organique
M.	GONTIER Gérard	Mécanique
M.	GRUSON Laurent	Algèbre
M.	GUILLAUME Jean	Microbiologie
M.	HEUBEL Joseph	Chimie Minérale
M.	LABLACHE-COMBIER Alain	Chimie Organique
M.	LANSRAUX Guy	Physique Atomique et Moléculaire
M.	LAVEINE Jean-Pierre	Paléontologie
M.	LEBRUN André	Electronique
M.	LEHMANN Daniel	Géométrie

Mme	LENOBLE Jacqueline	Physique Atomique et Moléculaire
M.	LINDER Robert	Biologie et Physiologie Végétales
M.	LOMBARD Jacques	Sociologie
M.	LOUCHEUX Claude	Chimie Physique
M.	LUCQUIN Michel	Chimie Physique
M.	MAILLET Pierre	Sciences Economiques
M.	MONTARIOL Frédéric	Chimie Appliquée
M.	MONTREUIL Jean	Biochimie
M.	PARREAU Michel	Analyse
M.	POUZET Pierre	Analyse Numérique
M.	PROUVOST Jean	Minéralogie
M.	SALMER Georges	Electronique
M.	SCHILTZ René	Physique Atomique et Moléculaire
Mme	SCHWARTZ Marie-Hélène	Géométrie
M.	SEGUIER Guy	Electrotechnique
M.	TILLIEU Jacques	Physique Théorique
M.	TRIDOT Gabriel	Chimie Appliquée
M.	VIDAL Pierre	Automatique
M.	VIVIER Emile	Biologie Cellulaire
M.	WERTHEIMER Raymond	Physique Atomique et Moléculaire
M.	ZEYTOUNIAN Radyadour	Mécanique

PROFESSEURS SANS CHAIRE

M.	BELLET Jean	Physique Atomique et Moléculaire
M.	BODARD Marcel	Biologie Végétale
M.	BOILLET Pierre	Physique Atomique et Moléculaire
M.	BOILLY Bénoni	Biologie Animale
M.	BRIDOUX Michel	Chimie Physique
M.	CAPURON Alfred	Biologie Animale
M.	CORTOIS Jean	Physique Nucléaire et Corpusculaire
M.	DEBOURSE Jean-Pierre	Gestion des entreprises
M.	DEPREZ Gilbert	Physique Théorique
M.	DEVRAINNE Pierre	Chimie Minérale
M.	GOUDMAND Pierre	Chimie Physique
M.	GUILBAULT Pierre	Physiologie Animale
M.	LACOSTE Louis	Biologie Végétale
Mme	LEHMANN Josiane	Analyse
M.	LENTACKER Firmin	Géographie
M.	LOUAGE Francis	Electronique
Mle	MARQUET Simone	Probabilités
M.	MIGEON Michel	Chimie Physique
M.	MONTEL Marc	Physique du Solide
M.	PANET Marius	Electrotechnique
M.	RACZY Ladislas	Electronique
M.	ROUSSEAU Jean-Paul	Physiologie Animale
M.	SLIWA Henri	Chimie Organique

MAITRES DE CONFERENCES (et chargés d'Enseignement)

M.	ADAM Michel	Sciences Economiques
M.	ANTOINE Philippe	Analyse
M.	BART André	Biologie Animale
M.	BEGUIN Paul	Mécanique
M.	BKOUCHE Rudolphe	Algèbre
M.	BONNELLE Jean-Pierre	Chimie
M.	BONNEMAIN Jean-Louis	Biologie Végétale
M.	BOSCQ Denis	Probabilités
M.	BREZINSKI Claude	Analyse Numérique
M.	BRUYELLE Pierre	Géographie

.../...

M. CARREZ Christian	Informatique
M. CORDONNIER Vincent	Informatique
M. COQUERY Jean-Marie	Psycho-Physiologie
Mle DACHARRY Monique	Géographie
M. DEBENEST Jean	Sciences Economiques
M. DEBRABANT Pierre	Géologie Appliquée
M. DE PARIS Jean-Claude	Mathématiques
M. DHAINAUT André	Biologie Animale
M. DELAUNAY Jean-Claude	Sciences Economiques
M. DERIEUX Jean-Claude	Microbiologie
M. DOUKHAN Jean-Claude	Physique du Solide
M. DUBOIS Henri	Physique
M. DYMENT Arthur	Mécanique
M. ESCAIG Bertrand	Physique du Solide
Me EVRARD Micheline	Chimie Appliquée
M. FONTAINE Jacques-Marie	Electronique
M. FOURNET Bernard	Biochimie
M. FORELICH Daniel	Chimie Physique
M. GAMBLIN André	Géographie
M. GOBLOT Rémi	Algèbre
M. GOSSELIN Gabriel	Sociologie
M. GRANELLE Jean-Jacques	Sciences Economiques
M. GUILLAUME Henri	Sciences Economiques
M. HECTOR Joseph	Géométrie
M. JACOB Gérard	Informatique
M. JOURNEL Gérard	Physique Atomique et Moléculaire
Mle KOSMAN Yvette	Géométrie
M. KREMBEL Jean	Biochimie
M. LAURENT François	Automatique
Mle LEGRAND Denise	Algèbre
Mle LEGRAND Solange	Algèbre
M. LEROY Jean-Marie	Chimie Appliquée
M. LEROY Yves	Electronique
M. LHENAFF René	Géographie
M. LOCQUENEUX Robert	Physique Théorique
M. LOUCHET Pierre	Sciences de l'Education
M. MACKE Bruno	Physique
M. MAHIEU Jean-Marie	Physique Atomique et Moléculaire
Me N'GUYEN VAN CHI Régine	Géographie
M. MAIZIERES Christian	Automatique
M. MALAUSSENA Jean-Louis	Sciences Economiques
M. MESSELYN Jean	Physique Atomique et Moléculaire
M. MONTUELLE Bernard	Biologique Appliquée
M. NICOLE Jacques	Chimie Appliquée
M. PAQUET Jacques	Géologie Générale
M. PARSY Fernand	Mécanique
M. PECQUE Marcel	Chimie Physique
M. PERROT Pierre	Chimie Appliquée
M. PERTUZON Emile	Physiologie Animale
M. PONSOLLE Louis	Chimie Physique
M. POVY Lucien	Automatique
M. RICHARD Alain	Biologie
M. ROGALSKI Marc	Analyse
M. ROY Jean-Claude	Psycho-Physiologie
M. SIMON Michel	Sociologie
M. SOMME Jean	Géographie
Mle SPIK Geneviève	Biochimie
M. STANKIEWICZ François	Sciences Economiques
M. STEEN Jean-Pierre	Informatique

M.	THERY Pierre	Electronique
M.	TOULOTTE Jean-Marc	Automatique
M.	TREANTON Jean-René	Sociologie
M.	VANDORPE Bernard	Chimie Minérale
M.	VILLETTE Michel	Mécanique
M.	WALLART Francis	Chimie
M.	WERNIER Georges	Informatique
M.	WATERLOT Michel	Géologie Générale
Mme	ZINN-JUSTIN Nicole	Algèbre

Je tiens à exprimer ma reconnaissance à tous les membres du Jury,

à Monsieur le Professeur BACCHUS qui me fait l'honneur de présider le jury de cette thèse,

à Jean BERSTEL qui fut le premier à m'encourager et qui, depuis, n'a jamais cessé de s'intéresser à mon travail,

à Luc BOASSON dont la profonde connaissance des langages algébriques a bien facilité ma tâche et avec qui j'ai toujours pris un grand plaisir à travailler,

à Vincent CORDONNIER, Directeur du Laboratoire d'Informatique de Lille, qui a accepté de juger mon travail,

à Gérard JACOB dont la venue à Lille a été déterminante dans la conduite de ma thèse,

à Maurice NIVAT dont l'intérêt pour mes recherches et les conseils toujours judicieux m'ont incité à poursuivre mes efforts,

à Monsieur le Professeur VERJUS qui, en me proposant un sujet de seconde thèse, m'a fait découvrir un domaine de l'informatique particulièrement intéressant.

Il est clair, pour moi, que cette thèse n'aurait jamais abouti s'il n'y avait eu à Lille une véritable équipe d'informatique théorique et je tiens à remercier, de leur collaboration, mes amis André ARNOLD, Max DAUCHET et Jean-Philippe RETHORE. En particulier, l'apport des discussions quasi-quotidiennes que j'ai eues avec André ARNOLD est difficilement mesurable.

Je m'en voudrais d'oublier la remarquable qualité du travail effectué par Patricia CARON pour la frappe et par Madame DEBOCK pour l'impression. Je les en remercie très vivement.

A mes parents,

A ma femme.

TABLE DES MATIÈRES

INTRODUCTION 1 - 6

CHAPITRE I : PRELIMINAIRES 7 - 23

A - LANGAGES RATIONNELS ET LANGAGES ALGÉBRIQUES 7 - 10

B - TRANSDUCTIONS RATIONNELLES ET FAMILLES DE LANGAGES 11 - 15

C - PSL-LANGAGES, LANGAGES COMMUTATIFS ET LANGAGES BORNÉS 16 - 20

D - CIL-LANGAGES ET LANGAGES SANS INSERTION 21 - 23

CHAPITRE II : LANGAGES COMMUTATIFS 24 - 65

A - LANGAGES COMMUTATIFS ET LANGAGES BORNÉS 24 - 32

B - COMPARAISON DE C(COM) AVEC DIVERSES FAMILLES DE LANGAGES 33 - 41

C - EFFACEMENT ET LANGAGES COMMUTATIFS 42 - 53

D - MOTS INFINIS ET LANGAGES COMMUTATIFS 54 - 60

E - LANGAGES K-COMMUTATIFS 61 - 65

CHAPITRE III : PSL-LANGAGES COMMUTATIFS 66 - 100

A - LE CÔNE RATIONNEL C(C(RAT)) 66 - 70

B - LANGAGES EFFAÇABLES DANS L_Ψ 71 - 75

C - PSL-LANGAGES K-COMMUTATIFS 76 - 80

D - UN LANGAGE COMMUTATIF MINIMAL 81 - 86

E - LANGAGES SANS PRODUIT 87 - 93

F - SUR LES LANGAGES ALGÉBRIQUES COMMUTATIFS 94 - 100

*

* *

BIBLIOGRAPHIE 101 - 104

INTRODUCTION

La théorie des langages formels est née, vers 1959, de préoccupations linguistiques. On cherchait alors un modèle mathématique pour décrire les langues naturelles, en se basant sur l'idée qu'on pouvait caractériser les phrases bien écrites par leur processus de génération, les grammaires. Dans ce but, Chomsky définissait quatre types de grammaire, les grammaires à structure de phrase, qui génèrent exactement les langages récursivement énumérables, les grammaires à contexte lié qui engendrent une partie des langages récursifs, les grammaires "context-free" correspondant aux langages algébriques et les grammaires linéaires à droite qui donnent la famille, notée Rat, des langages rationnels, une famille fondamentale en théorie des langages, déjà connue et étudiée à cette époque.

Simultanément, cette étude a puisé d'autres motivations dans l'informatique et, en particulier, dans la conception de langages de programmation évolués, destinés aux utilisateurs d'ordinateurs. Le rôle primordial des langages algébriques s'en est trouvé souligné. Ainsi, le terme "Algol-like languages" pour désigner ces langages est particulièrement significatif du rapport existant entre la famille Alg et un langage de programmation tel que Algol 60. Les travaux effectués sur les langages algébriques ont, aussi, été à l'origine d'importants progrès dans des domaines tels que l'analyse syntaxique, la compilation, la théorie des langages de programmation et, plus récemment, la théorie des schémas de programmes. Enfin, la famille Alg a servi d'exemple pour définir la notion de famille abstraite de langages ou FAL, introduite par Ginsburg et Greibach dans [28]. Il est apparu, en effet, que de nombreuses familles de langages obtenues à partir de grammaires ou d'automates particuliers possédaient des propriétés de clôture comparables à celles de la famille Alg et que certains résultats sur ces familles ne dépendaient, en fait, que de ces propriétés. La notion de famille abstraite de langages introduit un concept général dont l'apport s'est révélé très fécond, non seulement pour l'unification des résultats déjà connus et la simplification de démonstrations antérieures, mais aussi pour la compréhension de certains phénomènes (plusieurs propriétés qui apparaissaient de nature différente deviennent des corollaires d'un seul résultat plus général). D'autre part, cette notion a été à l'origine de nombreuses questions qui ont donné un nouvel essor à l'étude des langages, ce dont témoignent les travaux concernant la théorie des FAL (voir, par exemple, Beauquier [4], Berstel [5], [6], Boasson [8] à [11], Ginsburg [25], Ginsburg et Greibach [29], Ginsburg et Spanier [33], [34], Goldstine [35], [36], [37], Greibach [39] à [45], Nivat [65], Rovan [73] et Ullian [79]).

Dans cette optique, les transductions rationnelles, opérations particulières sur les langages, introduites par Elgot et Mezei [21], jouent un rôle primordial. Nivat [64] en a fait une étude approfondie et en a donné une caractérisation en terme de bimorphismes alphabétiques qui les rend manipulables et que nous utiliserons tout au long de ce travail. Nous nous intéresserons principalement aux cônes rationnels (terminologie due à S. Eilenberg [19]) qui sont, justement, les familles de langages fermées par transductions rationnelles ou, ce qui est équivalent, fermées par homomorphismes, homomorphismes inverses et intersections avec les langages rationnels. Les résultats que nous obtiendrons sur les FAL (cônes rationnels clos par union, produit et étoile) ou sur les FAL closes par substitution s'appuiront dans tous les cas sur des résultats concernant les cônes rationnels.

Une autre famille de langages, tout à fait "orthogonale" aux familles de langages définies par Chomsky, la famille, notée Bor, des langages bornés (un langage L est borné s'il existe des mots $w_1,\ldots,w_k$ tels que L soit inclus dans $w_1^* \ldots w_k^*$), introduite par Ginsburg et Spanier [31], occupe une place très importante en théorie des langages. Une particularité des langages bornés est qu'on peut représenter sans ambiguité de tels langages par des parties de $\mathbb{N}^k$ et utiliser des propriétés de nature algébrique pour résoudre certains problèmes les concernant. Ces langages se sont, aussi, révélés une source d'exemples et de contre-exemples. Ainsi, pour montrer l'existence, pour tout entier positif k, de langages exprimables comme intersection de k+1 langages algébriques sans l'être de k langages algébriques, Liu et Weiner [62] ont utilisé des langages bornés. De même, les premiers exemples de hiérarchies infinies décroissantes de cônes rationnels ont été construites à partir de langages bornés (cf. [6], [7]). D'autre part, certains problèmes indécidables dans le cas général, deviennent décidables si l'on se restreint aux langages bornés. C'est le cas, en particulier, pour le problème de l'égalité de deux langages algébriques qui devient décidable si l'un des deux langages est borné [24]. Ceci explique l'intérêt de ces langages et le grand nombre de travaux concernant ces langages ainsi que les cônes rationnels et les FAL qu'ils engendrent (voir, entre autres. Berstel [6], Berstel et Boasson [7], Durieux [18], Ginsburg [24], Ginsburg et Spanier [31], [32], Goldstine [35], [36], [37], Ibarra [48], Rovan [73] et Sudborough [78]).

Les langages commutatifs (fermés par permutation) qui sont, dans cette thèse, au centre de nos préoccupations, ont la même particularité que les langages bornés : on peut représenter sans ambiguité un langage commutatif défini sur un alphabet T par une partie de $\mathbb{N}^k$ où k est le cardinal de T. Le lien entre ces deux familles de langages est renforcé par le fait que tout langage borné est l'image par une transduction rationnelle d'un langage commutatif. Nous préciserons ce rapport en établissant l'égalité entre C(Com), le cône rationnel engendré par les langages commutatifs et $C_\cap$(Bor), le plus petit cône rationnel clos par intersection et contenant la famille Bor. Notre travail sur les langages commutatifs va, en particulier, nous permettre d'étendre à C(Com) certaines propriétés déjà connues pour les langages de C(Bor). Un fait est à remarquer, c'est que, bien souvent, l'extension de ces résultats ne nécessite pas une démonstration plus compliquée et que, même, dans certains cas, cette démonstration se simplifie. Ceci peut s'expliquer de la manière suivante : contrairement à la famille Bor, la famille des langages commutatifs, Com, est close par homomorphisme inverse. Le théorème de Nivat conduit, alors, à une caractérisation particulièrement simple des langages de C(Com) : ce sont les images par un homomorphisme strictement alphabétique de l'intersection d'un langage commutatif et d'un langage rationnel. Nous montrons, ainsi, simplement, que *Lin*, la famille des langages algébriques linéaires n'est pas incluse dans F_σ(Com), résultat qui implique immédiatement : *Lin* $\subsetneq$ C(AB) [39] (AB désigne la famille des langages algébriques bornés), Alg $\subsetneq$ F(Bor) et Alg $\subsetneq$ F_σ(AB) [36], *Lin* $\subsetneq$ C(Bor) [18] et Alg $\subsetneq$ F_σ(Bor) [37].

Une importante propriété des langages algébriques a été démontrée par Parikh dans [69] : Soient T un alphabet, L un langage algébrique inclus dans T^* et Ψ_T la fonction de Parikh (i.e. l'homomorphisme canonique de T^* dans le monoïde commutatif libre engendré par T), alors $\Psi_T(L)$ est un ensemble semi-linéaire ou langage rationnel commutatif (cf. [20]). Il existe, maintenant, un certain nombre de résultats d'une grande utilité, concernant les ensembles semi-linéaires (cf. Eilenberg et Schützenberger [20], Ginsburg [24], Ito [49], Liu et Weiner [61]). Du point de vue des transductions rationnelles, le seul fait de savoir qu'un langage est un PSL-langage (son image par la fonction de Parikh est un ensemble semi-linéaire) n'apporte aucun renseignement, puisque tout langage est rationnellement équivalent à un PSL-langage. Par contre, l'information devient beaucoup plus intéressante quand le langage ne domine rationnellement que des PSL-langages. Ces langages ne sont pas nécessairement algébriques et il n'est pas sans intérêt d'étudier les cônes rationnels ne contenant que des PSL-langages, pour mettre en évidence que certaines propriétés des langages algébriques ne dépendaient que de la semi-linéarité. Ainsi, tout cône rationnel vérifiant effectivement cette propriété (pour tout $L \in \mathcal{L}$, on peut construire l'ensemble semi-linéaire $\Psi(L)$) ne contient que des langages récursifs et, ce qui est équivalent, pour tout langage L de $\mathcal{L}$, il est décidable de déterminer si $L = \emptyset$. De plus, tout langage borné de $\mathcal{L}$ est l'image par un homomorphisme alphabétique de l'intersection de deux langages algébriques bornés [53].

Une grande partie de notre travail sera consacré aux langages appartenant à $\mathcal{L}_\psi = C(c(\mathrm{Rat}))$, le cône rationnel engendré par les PSL-langages commutatifs. Ginsburg et Spanier ont montré [33] que ce cône rationnel ne contenait que des PSL-langages et ils conjecturaient que $\mathcal{L}_\psi$ était non principal. Nous démontrons, ici, cette conjecture qui a été à l'origine de notre travail sur les langages commutatifs. Ce résultat nous permet de répondre à une autre question de Ginsburg ([25] page 180) ; en effet, nous démontrons que $\mathcal{L}_\psi$ est égal à $C_\cap(C_1)$, le plus petit cône rationnel clos par intersection, contenant le langage $C_1 = \{a^n b^n \ / \ n \geq 0\}$ et nous obtenons le premier exemple de cône rationnel principal $\mathcal{L}$ tel que $C_\cap(\mathcal{L})$ soit non principal.

Remarquons, enfin, que notre travail contient quelques résultats intéressants en eux-mêmes qui ne concernent pas directement les langages commutatifs. C'est le cas de la proposition II.20 se rapportant aux langages linéaires et de la proposition III.56 où nous montrons que Rat est le seul cône rationnel clos par intersection contenu dans Alg. Ces deux propositions sont ensuite utilisées dans le cadre des langages commutatifs. Par contre, c'est parce que la technique de démonstration est la même que pour les langages commutatifs, que nous établissons (proposition II.40) que tout langage algébrique inclus dans $a^* b^*$ est effaçable.

Passons à une présentation un peu plus précise de notre travail que nous avons divisé en trois chapitres.

Le premier chapitre, préliminaire, est consacré principalement aux définitions et propriétés classiques de la théorie des langages. En particulier, nous rappelons un certain nombre de résultats concernant les transductions rationnelles (fidèles), leurs caractérisations en terme de bimorphismes, leurs propriétés de composition. La section C qui se rapporte aux PSL-langages, aux langages commutatifs et aux langages bornés se termine par la démonstration d'une caractérisation des langages bornés appartenant à un cône rationnel qui ne contient que des PSL-langages (proposition I.22). Dans la dernière section de ce chapitre nous établissons des propriétés générales concernant les FAL, les CIL-langages (langages ne contenant pas le produit de deux langages infinis) et les langages sans insertion.

Le deuxième chapitre est divisé en cinq sections. Dans la première, nous démontrons quelques propriétés concernant les cônes rationnels clos par intersection qui nous permettent de démontrer l'égalité entre $C(\mathrm{Com})$ et $C_n(\mathrm{Bor})$ (proposition II.7). Nous démontrons, ensuite, quelques lemmes de base dont nous ferons un large usage tout au long de ce travail.

Dans la section B, pour montrer que certains langages "assez bas" dans la hiérarchie des langages algébriques, tels que D'^*_1, $\mathrm{Sym} = \{w\,\bar{w}^R \,/\, w \in \{a, b\}^*\}$, $C^*_1 = \{a^n b^n \,/\, n \geq 0\}^*$, ne sont pas dominés rationnellement par des langages commutatifs, nous démontrons un théorème de factorisation (proposition II.17) pour les langages de $C(\mathrm{Com})$ qui met en évidence ce qui peut encore permuter dans ces langages. En montrant, pour le langage Sym, une propriété déjà connue pour les générateurs de tout cône rationnel fermé par substitution, nous établissons que la famille des langages algébriques linéaires n'est pas incluse dans $F_\sigma(\mathrm{Com})$, le plus petit cône rationnel clos par substitution, contenant les langages commutatifs (proposition II.25). Nous montrons, aussi, que l'intersection est nécessaire pour passer des langages bornés aux langages commutatifs. En effet, il existe (corollaire II.33) des PSL-langages commutatifs n'appartenant pas à $F_\sigma(\mathrm{Bor})$.

Dans la section C, nous donnons une condition nécessaire et suffisante (proposition II.42) pour qu'un langage commutatif L soit effaçable, c'est-à-dire pour que $C(L)$ soit égal à $C^f(L)$, la famille des langages que l'on peut obtenir à partir de L par une transduction rationnelle fidèle (transduction rationnelle n'effaçant qu'un nombre borné de lettres consécutives). Cette condition nous amène à étudier les cônes rationnels fidèles clos par quotient rationnel et nous répondons à une question de Ginsburg ([25] page 34) en montrant l'existence d'un langage (borné) L tel que $F^f(L) \subsetneq F(L)$ soit clos par quotient rationnel à droite et à gauche (proposition II.37). Cette condition, que nous utiliserons au chapitre suivant dans le cas des PSL-langages commutatifs, nous permet aussi de caractériser exactement les langages commutatifs effaçables inclus dans D^*_1 (proposition II.49).

Le résultat principal de la section D est que, pour tout langage de $C(\mathrm{Com})$, il existe des mots $w_1,\dots,w_k$ tels que L et $L \cap w^*_1 \dots w^*_k$ aient la même image par la fonction de Parikh (proposition II.54). D'une part, nous en déduisons que tout langage infini de $F_\sigma(\mathrm{Com})$ contient un langage borné infini et, en utilisant une caractérisation des langages d'un mot infini rationnels (proposition II.51), nous obtenons, alors, que tout langage d'un mot infini appartenant à $F_\sigma(\mathrm{Com})$ est rationnel. D'autre part, ce résultat va nous permettre de donner une condition nécessaire et suffisante pour que $F_\sigma(L)$, avec L inclus dans $C(\mathrm{Com})$, contienne un langage borné non rationnel (proposition II.59). Comme corollaire immédiat, nous retrouvons un résultat de Goldstine [36] concernant le langage $L_G = \{a^{j_o} c a^{j_1} \dots c a^{j_t} / t \geq 0, \exists i \neq j_i\}$.

Dans la section E, nous nous intéresserons aux langages de Com_k, c'est-à-dire aux langages commutatifs définis sur un alphabet de k lettres. En utilisant les résultats de la section précédente, nous montrons principalement que tout langage infini de $F_\sigma(Com_k)$ contient un langage borné infini de dimension k (proposition II.66). Nous en déduisons que $E_{k+1} = \{a_1^n \dots a_{k+1}^n / n \geq 0\}$ n'appartient pas à $F_\sigma(Com_k)$, ce qui permet de démontrer la conjecture de Ginsburg et Spanier concernant la non principalité de $F(c(Rat))$.

Le troisième chapitre concerne principalement les PSL-langages commutatifs et contient les résultats pour lesquels nous faisons intervenir la propriété de semi-linéarité. Dans la première section, nous étudions le cône rationnel $L_\psi = C(c(Rat))$, engendré par les PSL-langages commutatifs et nous établissons (proposition III.14) que L_ψ est égal à $C_\cap(C_1)$, le plus petit cône rationnel clos par intersection contenant le langage $C_1 = \{a^n b^n / n \geq 0\}$. Nous obtenons, ainsi, le premier exemple de cône rationnel principal, $L = C(C_1)$ tel que $C_\cap(L)$ soit non principal (cf. [25] page 180).

Dans la section B, en utilisant les résultats du chapitre précédent, nous démontrons que tout PSL-langage commutatif est effaçable (proposition III.15), ce qui n'est pas vrai pour les PSL-langages bornés. Nous en déduisons que $C_\cap(C_1)$ est égal à $C_\cap^f(C_1)$, le plus petit cône rationnel fidèle clos par intersection et contenant le langage C_1. Ceci nous permet d'améliorer un résultat de Baker et Book [3] concernant les automates non déterministes multi-compteur "one-reversal".

Dans la section C, nous montrons que tout PSL-langage k-commutatif appartient au cône rationnel engendré par E'_{k+1}, le plus petit langage commutatif contenant $(a_1 \dots a_{k+1})^*$ (proposition III.27). Nous établissons, ensuite, que tout PSL-langage commutatif est une intersection finie de langages commutatifs de $C(D_1^*)$ (proposition III.31). Nous retrouvons, donc, en particulier, un résultat énoncé par Perrot dans [71].

Dans la section D, en utilisant une caractérisation des langages commutatifs appartenant à $C(\bar{D}_1^*)$ (lemme III.35) où $\bar{D}_1^* = \{a, b\}^* \setminus D_1^*$, nous démontrons que $\bar{D}_1^*$ est un langage commutatif minimal, c'est-à-dire que tout langage commutatif non rationnel appartenant à $C(\bar{D}_1^*)$ est rationnellement équivalent à $\bar{D}_1^*$. Ce résultat nous amène à conjecturer, d'une part, que tout PSL-langage commutatif non rationnel domine rationnellement $\bar{D}_1^*$ et d'autre part, que tout langage quasi-rationnel commutatif est rationnel.

Le résultat principal de la section E concerne les langages commutatifs inclus dans D_1^*. Nous démontrons que tout langage commutatif L inclus dans D_1^* est sans produit, c'est-à-dire que $C(L)$ ne contient pas le produit de deux langages non rationnels définis sur des alphabets disjoints. Comme

D'^*_1 n'est pas sans produit (proposition III.49), nous mettons, ainsi, en évidence une nouvelle distinction entre les langages de Dyck et de semi-Dyck sur une lettre.

La dernière section est consacrée aux langages algébriques commutatifs. Nous allons, d'abord, établir l'existence de tels langages n'appartenant pas à $C(D^*_1)$. Pour obtenir ce résultat, nous établirons en premier lieu (proposition III.54) une propriété des prépaires itérantes très strictes (introduites par Boasson dans [8]) pour les langages de $C(D^*_1)$. L'autre résultat de base de cette section ne concerne les langages commutatifs que par l'un de ses corollaires : nous montrons que si L_1 et L_2 sont des langages non rationnels définis sur des alphabets disjoints, le langage shuf (L_1, L_2) n'est pas algébrique. Nous en déduisons que Rat est le seul cône rationnel clos par intersection inclus dans Alg et, à l'aide d'un résultat démontré au chapitre II, que tout langage algébrique commutatif est sans produit.

Chapitre I

PRELIMINAIRES

A - Langages rationnels et langages algébriques

B - Transductions rationnelles et familles de langages

C - PSL-Langages, langages commutatifs et langages bornés

D - CIL-Langages et langages sans insertion

A - LANGAGES RATIONNELS ET LANGAGES ALGÉBRIQUES

Un **alphabet** est un ensemble fini non vide, T, dont les éléments sont appelés **lettres**. Les éléments du monoïde libre T^* engendré par T sont des **mots** et les parties de T^* sont les **langages**. Pour tout mot w de T^* et toute lettre a de T, $\ell_a(w)$ désignera le nombre d'occurrences de la lettre a dans le mot w. La **longueur** de w, notée $\ell(w)$, est égale au nombre total d'occurrences de lettres dans w. Le **mot vide**, noté ε, qui est l'élément neutre du monoïde libre T^*, est le seul mot de longueur nulle.

Dans l'ensemble des parties de T^*, on définit les opérations suivantes :

- l'**union**, notée $\cup$, et l'**intersection**, notée $\cap$, qui sont des opérations ensemblistes.
- le **produit**, qui est induit par le produit (ou concaténation) dans T^*, est

défini par :

$$L_1 L_2 = \{w_1 w_2 \ / \ w_1 \in L_1, w_2 \in L_2\}$$

- l'**étoile**, qui est le passage au sous-monoïde engendré, est définie par :

$$L = \bigcup_{n\geq 0} L^n \text{ avec } L^o = \{\varepsilon\} \text{ et } L^{n+1} = LL^n = L^nL \text{ pour tout } n \geq 0.$$

De façon classique, pour tout $w \in T^*$, nous écrirons w à la place de $\{w\}$ et w^* à la place de $\{w\}^*$.

Nous pouvons, maintenant, définir la famille, notée Rat, des langages rationnels, qui est la famille fondamentale en théorie des langages :

Définition : *La famille des langages rationnels du monoïde libre T^* est la plus petite famille de langages de T^* qui contienne les langages finis et qui soit close par union, produit et étoile.*

Cette famille peut être, aussi caractérisée par les automates d'état fini :

Définition : *Un automate d'état fini est un 5-uple, $M = (Q, T, f, q_o, F)$ où*

- *Q est un ensemble fini non vide* (**états**),
- *T est l'alphabet d'entrée,*
- *f est une application de $Q \times T$ dans Q (fonction de transition),*
- *$q_o \in Q$ est l'*état initial,
- *F est une partie de Q (ensemble des* états terminaux).

On peut, de façon naturelle, étendre f en une application de $Q \times T^*$ dans Q par :

$$\forall\, q \in Q,\ f(q, \varepsilon) = q \text{ et } f(q, wa) = f(f(q, w), a),\ \forall\, w \in T^*,\ \forall\, a \in T.$$

Le langage **reconnu** par l'automate M, noté T(M), est égal à $\{w \in T^* \ / \ f(q_o, w) \in F\}$.

Un langage $L \subseteq T^*$ est dit **régulier** s'il existe un automate d'état fini M tel que $L = T(M)$.

Le théorème de Kleene montre le lien entre les langages rationnels et les langages réguliers :

Théorème de Kleene : Un langage L inclus dans le monoïde libre T^* est rationnel si et seulement si il est régulier.

Si, à chaque "transition" d'un automate d'état fini M, on associe un mot défini sur un alphabet Δ, on obtient une application de T^* dans Δ^* :

Définition : *Un* gsm *ou* **transducteur unilatère** *est un 6-uple* $(Q, T, \Delta, \delta, g, q_0)$ *où*

- *Q, T, δ, q_0 ont la même signification que dans la définition de l'automate d'état fini,*
- *Δ est l'alphabet de sortie,*
- *g est une application de $Q \times T$ dans Δ^* que l'on étend classiquement à $Q \times T^*$ en posant pour tout $q \in Q$, $g(q, \varepsilon) = \varepsilon$ et $g(q, wa) = g(q, w)\, g(\delta(q, w), a)$, $\forall\, w \in T^*$, $\forall\, a \in T$.*

La fonction $\hat{g}$ définie par $\hat{g}(w) = g(q_0, w)$ pour tout $w \in T^*$ est appelée **gsm-fonction** ou **application séquentielle**.

Définition : *Soient u et v deux mots sur l'alphabet T. Nous dirons que u est un* **facteur** *de v si et seulement si v s'écrit $v_1\, u\, v_2$ avec $v_1, v_2 \in T^*$. Si $v = u\, v_2$, nous dirons que v est un* **facteur gauche** *(ou* **préfixe***) de v et si $v = v_1\, u$, nous dirons que u est un* **facteur droit** *(ou suffixe) de v.*

Définition : *Soient L un langage inclus dans T^* et w un mot de T^*. Nous dirons que u est un* **facteur itérant** de w dans L *si et seulement si :*

- *$w = w_1\, u\, w_2$ avec $w_1, u, w_2 \in T^*$ et $u \neq \varepsilon$,*
- *$w_1\, u^*\, w_2$ est inclus dans L.*

Nous dirons, aussi, que u est un facteur itérant dans L, si et seulement si il existe un mot w tel que u soit un facteur itérant de w dans L.

Cette définition nous permet d'énoncer un résultat de base pour les langages rationnels :

Lemme de l'étoile : *Si R est un langage rationnel inclus dans T^*, il existe un entier n tel que tout mot w, de longueur supérieure à n, admette un facteur itérant dans R.*

Ce résultat a été précisé par Ogden [67] :

Lemme d'Ogden : *Si R est un langage rationnel inclus dans* T^*, *il existe un entier n tel que si l'on distingue au moins n occurrences d'un mot w de R, ce mot se factorise en* $w_1\ u\ w_2$ *où u est un facteur itérant de w dans R qui contient au moins une et au plus n occurrences distinguées et où* $w_1\ w_2$ *en contient au moins une.*

Nous utiliserons, par la suite, une caractérisation des langages rationnels définis sur un alphabet d'une seule lettre (cf. [24], [75]) :

Proposition I.1 : *Un langage* $R \subseteq a^*$ *est rationnel si et seulement si* $\{i\ /\ a^i \in R\}$ *est un ensemble finalement périodique, c'est-à-dire si et seulement si R est une union finie de langages de la forme* $a^i(a^j)^*$ *avec* $i,\ j \geq 0$.

Une autre famille de langages, la famille des langages algébriques, notée Alg, joue un rôle primordial aussi bien en théorie des langages, qu'en théorie de la programmation et en linguistique. Cette famille a fait l'objet d'études intensives (cf. par exemple [2], [4], [5], [8], [24], [41], [64]).

Définition : *Une* grammaire algébrique *est un quadruplet* $G = (N, T, P, S)$ *où*

- *T est l'alphabet* terminal,
- *N est l'alphabet* non-terminal, *disjoint de T,*
- *P, l'ensemble des* règles, *est un ensemble fini inclus dans* $N \times (T \cup N)^*$ *(on notera* $\{B \to w\}$ *à la place de* (B, w) *les éléments de P),*
- *S,* *l'*axiome, *appartient à N.*

On définit sur $V^* = (N \cup T)^*$ la relation binaire $\underset{G}{=>}$ par $w \underset{G}{=>} w'$ si et seulement si $w = w_1\ B\ w_2$, $w' = w'_1\ w'_2\ w'_3$ et $\{B \to w'_2\} \in P$. Et on notera $\overset{*}{\underset{G}{=>}}$ la clôture reflexive et transitive de la relation $\underset{G}{=>}$.

Le langage **engendré** par la grammaire G, noté L(G), est égal à $\{w \in T^*\ /\ S \overset{*}{\underset{G}{=>}} w\}$. Nous dirons, alors, que $L \subseteq T^*$ est un **langage algébrique** si et seulement si il existe une grammaire algébrique qui engendre L.

Exemples : Considérons D_1^* (resp. D'^*_1) le langage de Dyck (resp. de semi-Dyck) sur une lettre, c'est-à-dire la classe du mot vide ε dans la congruence engendrée par $b\ a = a\ b = \varepsilon$ (resp. $ab = \varepsilon$). Les langages D_1^* et D'^*_1 sont des langages algébriques engendrés respectivement par les grammaires algébriques $G = (N, T, P, S)$ et $G' = (N, T, P', S)$ avec $N = \{S\}$, $T = \{a, b\}$, $P' = \{S \to S\ S\} \cup \{S \to \varepsilon\} \cup \{S \to a\ S\ b\}$ et $P = P' \cup \{S \to b\ S\ a\}$.

Une grammaire algébrique $G = (N, T, P, S)$ est **linéaire** si et seulement si P est inclus dans $N \times T^*(N \cup \{\varepsilon\})\ T^*$, c'est-à-dire si toute partie droite de règle possède au plus une occurrence d'un non-terminal.

Un langage L est un langage algébrique linéaire si et seulement si il existe une grammaire linéaire qui engendre L. Nous dirons, alors, que L appartient à *Lin*, la famille des langages algébriques linéaires.

Le langage Sym = $\{w\, \bar{w}^R \,/\, w \in \{a, b\}^*\}$ où $\bar{w}^R$ est l'image miroire de $\bar{w}$ que l'on obtient en "marquant" toutes les occurrences de lettres dans w, est un langage algébrique linéaire. Considérons, en effet, la grammaire linéaire $G = (N, T, P, S)$ avec $N = \{S\}$, $T = \{a, b, \bar{a}, \bar{b}\}$ et $P = \{S \rightarrow a\ S\ \bar{a}\} \cup \{S \rightarrow b\ S\ \bar{b}\} \cup \{S \rightarrow \varepsilon\}$. Il est clair que Sym = L(G).

B - TRANSDUCTIONS RATIONNELLES ET FAMILLES DE LANGAGES

Une autre opération, plus puissante que les gsm-fonctions définies à la section précédente, peut être associée aux langages rationnels. Les transductions rationnelles introduites par Elgot et Mezei [21] se sont révélées un outil essentiel pour l'étude des familles agréables de langages après que Nivat en ait donné une caractérisation en terme de bimorphismes alphabétiques [64].

<u>Définition</u> : *Une* **transduction** *est une application d'un monoïde libre T^* dans $P(\Delta^*)$, l'ensemble des parties du monoïde libre Δ^*.*

A une telle application, notée τ, nous associons $\hat{\tau} \subseteq T^* \times \Delta^*$ définie par :

$$\hat{\tau} = \{(w, z) \ / \ w \in T^*, \ z \in \tau(w)\}$$

Si T^* et Δ^* sont deux monoïdes libres, $T^* \times \Delta^*$ est un monoïde dont le produit est défini par :

$$(y, z)\ (y', z') = (y\,y', z\,z'), \ \forall\ (y, z), (y', z') \in T^* \times \Delta^*$$

et dont l'élément neutre est égal à $(\varepsilon, \varepsilon)$.

On peut, alors, définir le produit et l'étoile de langages inclus dans $T^* \times \Delta^*$ qui sont des opérations induites du produit dans $T^* \times \Delta^*$. La famille des langages rationnels de $T^* \times \Delta^*$ est la plus petite famille de langages contenant les parties finies de $T^* \times \Delta^*$ et fermée par union, produit et étoile.

<u>Définition</u> : *Une transduction τ de T^* dans Δ^* est une* **transduction rationnelle** *si et seulement si $\hat{\tau}$ est un langage rationnel de $T^* \times \Delta^*$.*

Un homomorphisme h de T^* dans Δ^* est **alphabétique** (resp. **strictement alphabétique**, **continu** ou **non-effaçant**) si h(T) est inclus dans $\Delta \cup \{\varepsilon\}$ (resp. Δ, $\Delta^+ = \Delta\,\Delta^*$).

<u>Théorème de Nivat</u> : Une transduction τ de T^* dans Δ^* est rationnelle, si et seulement si il existe un alphabet Z, un langage rationnel $R \subseteq Z^*$ et deux homomorphismes (alphabétiques) h et g, respectivement de Z^* dans T^* et Δ^*, tels que :

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

<u>Définition</u> : *Une transduction τ est* **fidèle** *si et seulement si pour tout mot z, l'ensemble de mots y vérifiant $(y, z) \in \hat{\tau}$ est fini.*

Boasson et Nivat ont donné, pour les transductions rationnelles fidèles, une caractérisation que nous utiliserons par la suite à chaque fois que nous nous intéresserons aux cônes rationnels fidèles :

Proposition I.2 [14] *Une transduction* τ *de* T^* *dans* Δ^* *est rationnelle fidèle si et seulement si il existe un alphabet* Z, *un langage rationnel* $R \subseteq Z^*$ *et deux homomorphismes alphabétiques* h *et* g, *respectivement de* Z^* *dans* T^* *et* Δ^*, *tels que :*

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

ii) Il existe un entier k *tel que* $R \cap Z^* Z_0^k Z^*$ *soit vide avec* $Z_0 = \{z \in Z \ / \ g(z) = \varepsilon\}$.

(Nous dirons, alors, que g *est* k-limité *sur* R*).*

Définition : *Une transduction* τ *est* continue *si et seulement si* $(z, \varepsilon) \in \hat{\tau}$ *implique* $z = \varepsilon$.

Proposition I.3 [14] *Une transduction* τ *de* T^* *dans* Δ^* *est rationnelle fidèle et continue si et seulement si il existe un alphabet* Z, *un langage rationnel* $R \subseteq Z^*$ *et deux homomorphismes* h *de* Z^* *dans* T^*, g, *strictement alphabétique, de* Z^* *dans* Δ^* *tels que :*

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

Proposition I.4 [8] *Si* τ *est une transduction rationnelle fidèle de* T^* *dans* Δ^* *et* $L \subseteq T^*$, *un langage contenant le mot vide, il existe une transduction rationnelle fidèle et continue* τ' *telle que* $\tau'(L) = \tau(L)$.

Enfin, rappelons une propriété très importante de composition des transductions rationnelles :

Proposition I.5 [21] *Soient* τ_1 *et* τ_2 *deux transductions rationnelles (resp. rationnelles fidèles, rationnelles fidèles et continues), la première de* T^* *dans* Δ^* *et la seconde de* Δ^* *dans* X^*. *Alors la transduction* $\tau_3 = \tau_2 \circ \tau_1$ *de* T^* *dans* X^* *est rationnelle (resp. rationnelle fidèle, rationnelle fidèle et continue).*

Nous dirons que le langage L_2 est **rationnellement supérieur** ou **domine rationnellement** le langage L_1 s'il existe une transduction rationnelle τ, telle que $L_1 = \tau(L_2)$. Les langages L_1 et L_2 seront dits **rationnellement équivalent** si chacun domine rationnellement l'autre et **rationnellement incomparable** si aucun ne domine rationnellement l'autre. Le lecteur trouvera une étude plus complète sur les transductions rationnelles dans le livre de Berstel [5].

Définition : *Une* famille de langages *est un couple* $(\Sigma, \mathcal{L})$ *où* Σ *est un ensemble infini dénombrable de lettres et* $\mathcal{L}$ *est un ensemble de parties de* Σ^* *vérifiant :*

i) Pour tout $L \in \mathcal{L}$ *il existe une partie finie* $T \subseteq \Sigma$ *telle que* $L \subseteq T^*$

ii) $\mathcal{L}$ *contient un langage non vide.*

Par la suite, nous remplacerons systématiquement $(\Sigma, \mathcal{L})$ par $\mathcal{L}$ puisque, dans notre travail, cela n'entraine aucune ambiguïté. D'autre part, nous écrirons L à la place de $\{L\}$, la famille de langages réduite au seul élément L.

Définition : *Un* **cône rationnel** *est une famille de langages fermée par transduction rationnelle.*

Définition : *Un* **cône rationnel fidèle** *est une famille de langages fermée par transduction rationnelle fidèle.*

Remarquons que cette définition diffère légèrement de celle qu'en donne Boasson dans [8] qui se limite aux transductions rationnelles fidèles et continues. Nous évitons, ainsi, d'avoir à "distinguer" le langage L du langage $L \cup \{\varepsilon\}$. La proposition I.4 montre que, pour l'essentiel, ces deux notions sont les mêmes et permet d'établir que $\mathcal{L}$ est un cône rationnel fidèle si et seulement si $\mathcal{L}$ est fermé par homomorphisme continu, homomorphisme inverse, intersection avec un langage rationnel, union avec le langage $\{\varepsilon\}$.

Pour toute famille de langages $\mathcal{L}$, notons :

- $C(\mathcal{L})$ (resp. $C^f(\mathcal{L})$), le plus petit (au sens de l'inclusion) cône rationnel (resp. rationnel fidèle) contenant $\mathcal{L}$.
- $H(\mathcal{L})$ (resp. $H_\alpha(\mathcal{L})$, $H_c(\mathcal{L})$), la plus petite famille de langages contenant $\mathcal{L}$ et fermée par homomorphisme (resp. homomorphisme alphabétique, homomorphisme continu).
- $H^{-1}(\mathcal{L})$ (resp. $H_\alpha^{-1}(\mathcal{L})$), la plus petite famille de langages contenant $\mathcal{L}$ et fermée par homomorphisme inverse (resp. homomorphisme alphabétique inverse).

Si $\mathcal{L}_1$ et $\mathcal{L}_2$ sont des familles de langages, $\mathcal{L}_1 \wedge \mathcal{L}_2$ désignera la famille de langages $\{L_1 \cap L_2 \ / \ L_1 \in \mathcal{L}_1, L_2 \in \mathcal{L}_2\}$. Les propositions précédentes permettent, alors, d'écrire

- $C(\mathcal{L}) = H(H^{-1}(\mathcal{L}) \wedge \mathrm{Rat}) = H_\alpha(H_\alpha^{-1}(\mathcal{L}) \wedge \mathrm{Rat})$.
- $C^f(\mathcal{L}) = H_c(H^{-1}(\mathcal{L}') \wedge \mathrm{Rat})$ avec $\mathcal{L}' = \{L \cup \{\varepsilon\} \ / \ L \in \mathcal{L}\}$.

Les cônes rationnels fermés par union sont les "full semi-AFL" et nous noterons $C_\cup(\mathcal{L})$ (resp. $C_\cup^f(\mathcal{L})$) le plus petit cône rationnel (resp. rationnel fidèle) clos par union et contenant $\mathcal{L}$.

Soient L_1 et L_2 deux langages inclus dans T^*. Le **quotient à droite** (resp. **à gauche**) de L_1 par L_2, noté $L_1 \ / \ L_2$ (resp. $L_2 \setminus L_1$), est égal à $\{w \in T^* \ / \ \exists\, w_2 \in L_2$ tel que $w\, w_2 \in L_1\}$ (resp. $\{w \in T^* /$ $\exists\, w_2 \in L_2$ tel que $w_2\, w \in L_1\}$).

Une famille de langages $\mathcal{L}$ est close par **quotient rationnel à droite** (resp. à gauche) si et seulement si pour tout $L \in \mathcal{L}$ et tout $R \in \mathrm{Rat}$, $L \ / \ R$ (resp. $R \setminus L$) appartient à $\mathcal{L}$.

Tout cône rationnel est clos par quotient rationnel à droite et à gauche. Donc, pour tout langage $L \subseteq T^*$, $\mathrm{Init}(L) = L \ / \ T^*$ appartient à $C(L)$. Par contre, un cône rationnel fidèle n'est pas nécessairement clos par quotient rationnel à droite ou à gauche (cf. [25]).

Définition : *Un cône rationnel (resp. rationnel fidèle)* $\mathcal{L}$ *est* principal *si et seulement si il existe un langage* L *tel que* $\mathcal{L} = C(L)$ *(resp.* $\mathcal{L} = C^{f}(L)$*). Nous dirons, alors, que le langage* L *est un* générateur *(resp. générateur fidèle) de* $\mathcal{L}$.

Définition : *Un langage* L *est* effaçable *si et seulement si il est générateur fidèle de* C(L), *c'est-à-dire si* $C^{f}(L) = C(L)$.

Proposition I.6 [29] *Tout cône rationnel (fidèle) principal est fermé par union.*

Définition : *Soient* s *un homomorphisme de* T^* *dans* $P(\Delta^*)$ *et* $\mathcal{L}$ *une famille de langages. Nous dirons que* s *est une* $\mathcal{L}$-substitution *si et seulement si* $s(a) \in \mathcal{L}$ *pour tout* $a \in T$.

A cette opération sur les langages on fait correspondre un opérateur binaire sur les familles de langages, défini de la manière suivante :

Si $\mathcal{L}$ et $\mathcal{L}'$ sont des familles de langages, posons

$$\mathcal{L} \,\square\, \mathcal{L}' = \{s(L) \;/\; L \in \mathcal{L},\ s \text{ est une } \mathcal{L}'\text{-substitution}\}.$$

Cet opérateur est fondamental en théorie des langages et se trouve à l'origine de la théorie des opérateurs syntaxiques (cf. [13], [40], [42], [56], [60], [65], [72]).

Définition : *Une famille de langages* $\mathcal{L}$ *est* close par substitution *si et seulement si* $\mathcal{L} \,\square\, \mathcal{L} \subseteq \mathcal{L}$.

Si pour tout mot $w \in L$, il existe au plus une occurrence d'une lettre a telle que $s(a) \neq \{a\}$, nous dirons que la substitution s est une insertion sur L (cf. [42]). Plus précisément :

Définition : *Soit* s *une substitution définie sur* T^*. *Alors* s *est une* $\mathcal{L}$-intersection sur L *s'il existe* $T' \subseteq T$ *tel que :*

i) $L \subseteq (T \setminus T')^* \, (T' \cup \{\varepsilon\}) \, (T \setminus T')^*$,

ii) $s(a) = \{a\}, \ \forall\, a \in T \setminus T'$,

iii) $s(a) \in \mathcal{L} \quad \forall\, a \in T'$.

Définition : *Une famille de langage* $\mathcal{L}$ *est* close par insertion *si* $\{s(L) \;/\; L \in \mathcal{L},\ s\ \mathcal{L}\text{-insertion sur } L\}$ *est inclus dans* $\mathcal{L}$.

Définition : *Une famille de langages* $\mathcal{L}$ *constitue une famille agréable (fidèle) de langages, en abrégé FAL (fidèle) si et seulement si* $\mathcal{L}$ *est un cône rationnel (fidèle), fermé par union et étoile.*

Pour toute famille de langages $\mathcal{L}$, notons $F(\mathcal{L})$ (resp. $F^f(\mathcal{L})$), la plus petite FAL (resp. FAL fidèle) contenant $\mathcal{L}$. La plus petite FAL close par substitution contenant $\mathcal{L}$, notée $F_\sigma(\mathcal{L})$, est égale à $\bigcup_{i\geq 1} \mathcal{L}_i$ où $\mathcal{L}_i$ est défini par $\mathcal{L}_1 = C(\mathcal{L})$ et $\mathcal{L}_{j+1}=\mathcal{L}_j \,\square\, \mathcal{L}_1$ ($= \mathcal{L}_1 \,\square\, \mathcal{L}_j$), $\forall_j \geq 1$. Ainsi, $F_\sigma(\text{Lin})$ désignera la famille des langages quasi-rationnels (cf. [64]).

Proposition I.7 [28] *Pour toute famille de langages* $\mathcal{L}$, $F(\mathcal{L}) = \mathit{Rat} \,\square\, C(\mathcal{L})$ *et* $F^f(\mathcal{L}) = \mathit{Rat} \,\square\, C^f(\mathcal{L})$.

Proposition I.8 [29] *Si* L *est un langage inclus dans* T^* *et* c *une lettre n'appartenant pas à* T, $F(L) = C((Lc)^*)$.

Introduisons, maintenant, une opération sur les langages, l'opération de shuffle ou produit de Hurwitz que nous utiliserons par la suite à de nombreuses reprises. Si L_1 et L_2 sont des langages inclus respectivement dans T_1^* et T_2^*,

$$\text{shuf}\ (L_1, L_2) = \{x_1\, y_1 \ldots x_p\, y_p \ /\ x_i \in T_1^*,\ y_j \in T_2^*,\ x_1 \ldots x_p \in L_1,\ y_1 \ldots y_p \in L_2\}.$$

Il est clair que cette opération est commutative et associative, ce qui justifie l'écriture de shuf $(L_1,\ldots,L_n)$ pour n supérieur à deux. Pour les cônes rationnels (fidèles) le lien entre l'intersection et l'opération shuffle est mise en évidence par les deux propositions suivantes :

Proposition I.9 [25] *Un cône rationnel (fidèle) est clos par intersection si et seulement si il est clos pour l'opération shuffle.*

Proposition I.10 [25] *Soient les langages* $L_1 \subseteq T_1^*$, $L_2 \subseteq T_2^*$ *avec* $T_1 \cap T_2 = \emptyset$. *Alors,* $H(C(L_1) \wedge C(L_2)) = C(L)$ *et* $H_c(C^f(L_1) \wedge C^f(L_2)) = C^f(L)$ *avec* $L = \mathit{shuf}\ (L_1, L_2)$.

Pour toute famille de langages $\mathcal{L}$, notons $C_n(\mathcal{L})$ (resp. $C_n^f(\mathcal{L})$) le plus petit cône rationnel (resp. rationnel fidèle) clos par intersection et contenant $\mathcal{L}$. Désignons par $\wedge\,\mathcal{L}$, la famille $\bigcup_{i\geq 1} \mathcal{L}_i$ où $\mathcal{L}_1 = \mathcal{L}$ et $\forall\ j \geq 1$, $\mathcal{L}_{j+1} = \mathcal{L}_j \wedge \mathcal{L}_1$. Nous avons, alors :

Proposition I.11 [25] *Pour toute famille de langages* $\mathcal{L}$, $C_n(\mathcal{L})$ *est égal à* $H(\wedge\, C(\mathcal{L}))$ *et* $C_n^f(\mathcal{L})$ *est égal à* $H_c(\wedge\, C^f(\mathcal{L}))$.

Considérons maintenant un langage $L_1 \subseteq T_1^*$. Pour tout entier i supérieur à 1, construisons, par induction, un alphabet T_i et un langage $L_i \subseteq T_i^*$. Prenons Δ_i un alphabet disjoint de T_i qui soit en bijection avec T_1 par un homomorphisme h_i. Posons $T_{i+1}=T_i \cup \Delta_i$, $L_{i+1} = \text{shuf}(L_i, L'_i)$ où $L'_i = h_i(L_1)$ est une "recopie" de L_1 sur un alphabet disjoint de T_i. La famille $\mathcal{L} = \{L_i \ /\ i \geq 1\}$ vérifie, alors :

Proposition I.12 [25]

- $C_n(L_1) = C(\mathcal{L})$,
- $C_n^f(L_1) = C^f(\mathcal{L})$.

C - PSL-LANGAGES, LANGAGES COMMUTATIFS ET LANGAGES BORNÉS.

Notons $\mathbb{N}$, l'ensemble des entiers naturels et $\mathbb{Z}$, l'ensemble des entiers relatifs. Posons $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$ et pour tout $p \in \mathbb{N}_+$, $\mathbb{N}^p = \{(i_1,\dots,i_p) \ / \ i_j \in \mathbb{N},\ \forall\, j \in \{1,\dots,p\}\}$. Un ensemble $S \subseteq \mathbb{N}^p$ est un **ensemble linéaire** s'il existe $x_o \in \mathbb{N}^p$, $P = \{x_1,\dots,x_n\}$ ensemble fini inclus dans $\mathbb{N}^p$, qui est l'ensemble des périodes tels que $S = \{x_o + \sum_{i=1}^{n} \lambda_i x_i \ / \ \lambda_i \in \mathbb{N}\}$, ensemble que l'on notera $L(x_o\ ;\ P)$.

L'ensemble $L(x_o\ ;\ P)$ sera dit **linéaire stratifié** si P vérifie les deux conditions suivantes :

i) Chaque élément de P a au plus deux coordonnées non nulles.

ii) Il n'existe pas d'entiers i, j, k, m, tels que $1 \leq i < j < k < m \leq p$ et tels que $(u_1,\dots,u_p)$, $(u'_1,\dots,u'_p)$ soient dans P avec $u_i\, u'_j\, u_k\, u'_m \neq 0$.

<u>Définition</u> *Un ensemble $S \subseteq \mathbb{N}^p$ avec $p \in \mathbb{N}_+$ est un* **ensemble semi-linéaire (stratifié)** *si et seulement si S est une union finie d'ensembles linéaires (stratifiés) inclus dans $\mathbb{N}^p$.*

A tout langage inclus dans $T^* = \{a_1,\dots,a_p\}^*$, on peut faire correspondre une partie de $\mathbb{N}^p$ grace à la **fonction de Parikh** notée Ψ_T(ou Ψ s'il n'y a pas d'ambiguité) qui est définie sur T^* par :

$$\Psi_T(w) = (\ell_{a_1}(w),\dots,\ell_{a_p}(w)) \text{ pour tout } w \in T^*.$$

En fait, Ψ_T est l'homomorphisme canonique de T^* dans le monoïde commutatif libre engendré par T.

<u>Définition</u> : *Un langage $L \subseteq T^*$ est un* **PSL-langage** *si et seulement si $\Psi_T(L)$ est un ensembe semi-linéaire.*

Une importante propriété des langages algébriques est donnée par le théorème suivant :

<u>Théorème de Parikh</u> [69], [70] Tout langage algébrique est un PSL-langage.

Du point de vue des transductions rationnelles, le fait de savoir qu'un langage est un PSL-langage n'apporte guère d'indication. En effet :

<u>Proposition I.13</u> : *Tout langage est rationnellement équivalent à un PSL-langage.*

<u>Démonstration</u> : Soient $L \subseteq T^*$ avec $T = \{a_1,\dots,a_k\}$, et $T' = T \cup \{d\}$ où d est une lettre n'appartenant pas à T. Montrons que $L' = L\, d\, a_1^* \dots a_k^*$ est un PSL-langage. Posons $S = \Psi_T(L)$, S_1 l'ensemble des éléments minimaux de S et $L_1 = \Psi_T^{-1}(S)$. Il est facile de vérifier que $\Psi_{T'}(L') = \Psi_{T'}(L_1\, d\, a_1^* \dots a_k^*)$ et comme S_1 est fini (cf. [24]), L_1 est fini donc $L_1\, d\, a_1^* \dots a_k^*$ est un langage rationnel et le théorème de Parikh entraine que $\Psi_{T'}(L') = \Psi_{T'}(L_1\, d\, a_1^* \dots a_k^*)$ est un ensemble semi-linéaire . D'autre part L' est le produit de L par un langage rationnel donc $L' \in C(L)$. Enfin, $L = (L' \cap T^*d) \ / \ d \in C(L')$ et L est rationnellement équivalent à L'.□

Pour tout mot w, c(w) désignera l'ensemble des mots que l'on peut obtenir par permutation à partir de w. La **fermeture commutative** de L, notée c(L), est égale à $\{z \in c(w) \ / \ w \in L\}$ c'est-à-dire $\Psi^{-1} \circ \Psi(L)$.

Définition : *Un langage* L *est* **commutatif** *si et seulement si* $L = c(L) = \Psi^{-1} \circ \Psi(L)$.

Pour toute famille de langage $\mathcal{L}$, notons $c(\mathcal{L}) = \{c(L) \ / \ L \in \mathcal{L}\}$. Il est clair que c(Rat) = c(Alg) est égal à la famille des PSL-langages commutatifs (lemme III.1). On dira que la famille $\mathcal{L}$ est **commutativement close** si et seulement si $\mathcal{L}$ contient $c(\mathcal{L})$.

La famille de tous les langages commutatifs sera notée Com et pour tout $k \in \mathbb{N}_+$, Com_k désignera la famille des langages k-commutatifs, c'est-à-dire des langages commutatifs définis sur un alphabet d'au plus k lettres.

Définition : *Un langage* $L \subseteq T^*$ *est* **k-borné** (*ou* **borné**) *si et seulement si il existe des mots* $w_1, \ldots, w_k \in T^*$ *tels que* $L \subseteq w_1^* \ldots w_k^*$.

Si les mots $w_1, \ldots, w_k$ sont des lettres toutes distinctes, nous dirons que L est **strictement borné**.

La famille de tous les langages bornés est notée Bor et pour tout $k \in \mathbb{N}_+$, Bor_k désignera la famille des langages k-bornés. Il est clair que Com = c(Bor), par contre pour tout $k \in \mathbb{N}_+$, $\mathrm{Com}_k \subsetneq c(\mathrm{Bor}_k)$.

La proposition I.13 montre que tout langage borné est rationnellement équivalent à un PSL-langage borné. Cette propriété n'est plus vraie quand on remplace "borné" par "strictement borné". En effet le cône rationnel engendré par un PSL-langage L strictement borné est inclus dans $C(c(L)) \subseteq C(c(\mathrm{Rat}))$ qui est un PSL-cône, c'est-à-dire un cône rationnel ne contenant que des PSL-langages (corollaire III.4).

Nous allons, maintenant, donner une caractérisation des langages bornés appartenant à un PSL-cône rationnel que nous utiliserons au chapitre III pour décrire les langages bornés de C(c(Rat)). Remarquons d'abord, que tout PSL-langage 1-borné est rationnel. Par contre un PSL-langage 2-borné peut dominer rationnellement un langage qui n'est pas un PSL-langage. Prenons, en effet, $L = \{(ab)^i \ (ba)^j \ / \ i \geq 0, \ j \geq i^2\}$. Le cône rationnel C(L) contient le langage $\{a^i \ b^j \ / \ i \geq 0, \ j \geq i^2\}$, alors que L est un PSL-langage 2-borné (on peut même montrer que pour tout langage rationnel R, $L \cap R$ est un PSL-langage). Nous avons cependant :

Proposition I.14 [52] *Soit* L *un langage 2-borné. Si pour tout homomorphisme h,* $h^{-1}(L)$ *est un PSL-langage, le cône rationnel engendré par* L *ne contient que des PSL-langages.*

Si L est un langage k-borné avec $k \geq 3$, la proposition précédente n'est plus vraie. Prenons $L = \{a^{i^2} \ b \ a^j \ / \ i, j \geq 0\}$. Le langage $L' = \{a^{i^2} \ b^j \ / \ i, j \geq 0\}$ appartient à C(L), alors qu'on peut établir que pour tout homomorphisme h, $h^{-1}(L)$ est un PSL-langage. Pour obtenir un résultat comparable à la proposition précédente pour tous les langages bornés, il suffit de prendre un gsm à la place de l'homomorphisme :

Proposition I.15 [52] *Soit* L *un langage borné. Si, pour toute application séquentielle* g, *le langage* $g^{-1}(L) = \{w \,/\, g(w) \in L\}$ *est un PSL-langage, le cône rationnel engendré par* L *ne contient que des PSL-langages.*

Nous allons, maintenant, établir un lien entre ces langages et les langages algébriques bornés. Pour tout n-uple de mots $w = \langle w_1,\ldots,w_n\rangle$, f_w désignera la fonction définie sur $\mathbb{N}^n$ par $f_w(i_1,\ldots,i_n) = w_1^{i_1} \ldots w_n^{i_n}$.

Montrons d'abord que tout langage L inclus dans $w_1^* \ldots w_n^*$, tel que $f_w^{-1}(L) = \{(i_1,\ldots,i_n) \,/\, w_1^{i_1} \ldots w_n^{i_n} \in L\}$ est un semi-linéaire, est une intersection finie de langages algébriques. Dans le cas où $w_1,\ldots,w_n$ sont des lettres toutes distinctes, ce résultat découle immédiatement des deux propositions suivantes :

Proposition I.16 [31] *Soient* $L \subseteq w_1^* \ldots w_n^*$ *et* w *le n-uple* $\langle w_1,\ldots,w_n\rangle$. *Le langage* L *est algébrique si et seulement si* $f_w^{-1}(L)$ *est un semi-linéaire stratifié.*

Proposition I.17 [61] *Tout ensemble semi-linéaire est une intersection finie de semi-linéaires stratifiés.*

Dans le cas où $w_1,\ldots,w_n$ sont des mots quelconques que l'on peut toujours supposer différents du mot vide, nous utiliserons une autre propriété :

Lemme I.18 [24] *Soient un alphabet* $\Delta = \{a_1,\ldots,a_n\}$ *et un homomorphisme* h *défini sur* Δ *par* $h(a_i) = w_i$, $\forall\, i \in \{1,\ldots,n\}$. *Il existe un langage rationnel* $R \subseteq a_1^* \ldots a_n^*$ *qui est en bijection par* h *avec* $w_1^* \ldots w_n^*$.

Nous pouvons alors établir :

Lemme I.19 : *Soit* $L \subseteq w_1^* \ldots w_n^*$ *tel que* $f_w^{-1}(L)$ *est un semi-linéaire. Alors* L *est une intersection finie de langages algébriques.*

Démonstration : D'après la proposition I.17 il existe des semi-linéaires stratifiés $S_1,\ldots,S_p$, tels que $S = f_w^{-1}(L) = \bigcap_{i=1}^{p} S_i$. Considérons l'alphabet $\Delta = \{a_1,\ldots,a_n\}$. L'application f_a défini sur $\mathbb{N}^n$ par $f_a(i_1,\ldots,i_n) = a_1^{i_1} \ldots a_n^{i_n}$ établit une bijection entre $\mathbb{N}^n$ et $a_1^* \ldots a_n^*$. Donc $f_a(S) = f_a(\bigcap_{i=1}^{p} S_i) = \bigcap_{i=1}^{p} f_a(S_i)$ où, $\forall\, i \in \{1,\ldots,p\}$, $f_a(S_i)$ est un langage algébrique [24]. Considérons le langage rationnel R et l'homomorphisme h définis au lemme précédent. Posons, pour tout $i \in \{1,\ldots,p\}$, $L'_i = f_a(S_i) \cap R$. Comme $f_a(S) = h^{-1}(L) \cap a_1^* \ldots a_n^*$ et que $R \subseteq a_1^* \ldots a_n^*$, nous avons $h^{-1}(L) \cap R = f_a(S) \cap R = \bigcap_{i=1}^{p} L'_i$. L'inclusion de L dans $h(R) = w_1^* \ldots w_n^*$ implique $L = L \cap h(R) = h(h^{-1}(L) \cap R)$ et comme h est injectif sur R qui contient

L'_i, $\forall\, i \in \{1,\ldots,p\}$, nous en déduisons que $L = h(\bigcap_{i=1}^{p} L'_i) = \bigcap_{i=1}^{p} h(L'_i)$ est une intersection finie de langages algébriques.□

Pour $w = \langle w_1,\ldots,w_n\rangle$ et $S \subseteq \mathbb{N}$, on n'a pas toujours l'égalité entre S et $f_w^{-1}(f_w(S))$. Cependant, montrons :

<u>Lemme I.20</u> : *Pour tout semi-linéaire $S \subseteq \mathbb{N}^n$ et tout n-uple de mot $w = \langle w_1,\ldots,w_n\rangle$, $f_w^{-1} \circ f_w(S)$ est un semi-linéaire.*

<u>Démonstration</u> : Pour tout semi-linéaire $S \subseteq \mathbb{N}^n$, il existe un langage rationnel $R \subseteq \{a_1,\ldots,a_n\}^*$ tel que $\Psi(R) = S$. La clôture commutative de R, $c(R) = \Psi^{-1} \circ \Psi(R)$ appartient au plus petit cône rationnel contenant $c(\mathrm{Rat})$ qui ne contient que des PSL-langages [33], [54].

Considérons l'homomorphisme h défini par : $h(a_i) = w_i$, $\forall\, i \in \{1,\ldots,n\}$. Les langages $L = f_w(S) = h(c(R) \cap a_1^* \ldots a_n^*)$ et $L' = h^{-1}(L) \cap a_1^* \ldots a_n^*$ appartiennent à $C(c(\mathrm{Rat}))$. Donc $f_w^{-1} \circ f_w(S) = f_w^{-1}(L) = \Psi(L')$ est un semi-linéaire.□

Pour tout entier positif i, nous noterons :

- $AB_i = \{L_1 \cap \ldots \cap L_i \ /\ \forall\, s \in \{1,\ldots,i\},\ L_s$ est un langage algébrique borné$\}$.

Liu et Weiner ont démontré [62] que, pour tout entier $k \geq 2$, le langage $L_k = \{w\,w \ /\ w \in a_1^* \ldots a_k^*\}$ appartient à $AB_k \setminus AB_{k-1}$, d'où la hiérarchie infinie $AB_1 \subsetneq AB_2 \subsetneq \ldots \subsetneq AB_{k-1} \subsetneq AB_k \subsetneq \ldots$. Nous allons établir que $AB_1 = H(AB_1) \subsetneq H(AB_2) = H(AB_3) = \ldots$. Montrons d'abord :

<u>Lemme I.21</u> : *Soient S un semi-linéaire de $\mathbb{N}^n$ et $w = \langle w_1,\ldots,w_n\rangle$ un n-uple de mots. Alors $f_w(S) = \{w_1^{i_1} \ldots w_n^{i_n} \ /\ (i_1,\ldots,i_n) \in S\}$ appartient à $H(AB_2)$, la famille des images homomorphes de l'intersection de deux langages algébriques bornés.*

<u>Démonstration</u> : Comme AB_1 est clos par union, il est facile de vérifier que $H(AB_2) = H(AB_1 \wedge AB_1)$ est aussi clos par union, donc on peut supposer que S est un linéaire égal à $L(u_o\ ;\ \{u_1,\ldots,u_k\})$.

Posons $\Delta = \{a_1,\ldots a_n\}$, $Y = \{a_{1o}, \ldots a_{1k}, a_{2o}, \ldots a_{nk}\}$,

$$L_1 = \{a_{1o}\, a_{11}^{P_1} \ldots a_{1k}^{P_k}\, a_{2o}\, a_{2k}^{P_k} \ldots a_{21}^{P_1}\, a_{3o}\, a_{31}^{t_1} \ldots a_{3k}^{t_k}\, a_{4o}\, a_{4k}^{t_k} \ldots a_{41}^{t_1} \ldots \ /\ P_i, t_i, \ldots \in \mathbb{N}\},$$

$$\text{et}\quad L_2 = \{a_{1o}\, a_{11}^{s_1} \ldots a_{1k}^{s_k}\, a_{2o}\, a_{2k}^{P_k} \ldots a_{21}^{P_1}\, a_{3o}\, a_{31}^{P_1} \ldots a_{3k}^{P_k} \ldots \ /\ s_i, P_i, \ldots \in \mathbb{N}\}.$$

Les langages L_1 et L_2 sont algébriques, bornés et il est clair que $L_1 \cap L_2 = \{a_{1o}\, a_{11}^{P_1} \ldots a_{1k}^{P_k}\, a_{2o}\, a_{2k}^{P_k} \ldots a_{21}^{P_1}\, a_{3o}\, a_{31}^{P_1} \ldots a_{3k}^{P_k} \ldots \ /\ P_i \in \mathbb{N}\}$. Considérons, alors, l'homomorphisme g défini sur Y par : $\forall\, i \in \{1,\ldots,n\}$ $\forall\, j \in \{0,\ldots,k\}$, $g(a_{ij}) = w_i^{r_{ij}}$ où r_{ij} est égal à la $i^{\text{ème}}$ composante du vecteur u_j. Alors $f_w(S) = g(L_1 \cap L_2)$ appartient à $H(AB_2)$.□

Prenons maintenant L inclus dans $w_1^* \ldots w_n^*$. Si L est une intersection finie de langages algébriques $L_1,\ldots,L_p$, $f_w^{-1}(L) = \bigcap_{i=1}^{p} f_w^{-1}(L_i \cap w_1^* \ldots w_n^*)$. D'après la proposition I.16, $\forall\ i \in \{1,\ldots,p\}$, $f_w^{-1}(L_i \cap w_1^* \ldots w_n^*)$ est un semi-linéaire et comme la classe des semi-linéaires est close par intersection [24], $S = f_w^{-1}(L)$ est un semi-linéaire. Pour tout homomorphisme g, $g(L) = g \circ f_w(S)$ est égal à $f_y(S)$ avec $y = \langle g(w_1),\ldots,g(w_n)\rangle$. Considérons, maintenant $L' \subseteq w_1^* \ldots w_n^*$ et h l'homomorphisme défini sur $\Delta = \{a_1,\ldots,a_n\}$ par $h(a_i) = w_i$, $\forall\ i \in \{1,\ldots,n\}$. Si L' appartient à un cône rationnel qui ne contient que des PSL-langages, $L'' = h^{-1}(L') \cap a_1^* \ldots a_n^*$ est un PSL-langage et $S' = \Psi(L'')$ est un semi-linéaire, donc $L' = f_w(S')$ avec S' semi-linéaire. Enfin, comme AB_1 est clos par homomorphisme, il est facile de vérifier que tout langage de $H(AB_2) = H(AB_1 \wedge AB_1)$ est l'image homomorphe alphabétique de l'intersection de deux langages algébriques bornés. Des lemmes précédents, nous pouvons alors, déduire une caractérisation multiple des langages bornés d'un PSL-cône rationnel :

<u>Proposition I.22</u> : *Soient $w = \langle w_1,\ldots,w_n\rangle$ un n-uple de mots et un langage $L \subseteq w_1^* \ldots w_n^*$. Les propriétés suivantes sont équivalentes :*

- *i) $f_w^{-1}(L)$ est un semi-linéaire,*
- *ii) il existe un semi-linéaire S tel que $L = f_w(S)$,*
- *iii) L est une intersection finie de langages algébriques,*
- *iv) L est l'image homomorphe (alphabétique) de l'intersection de deux langages algébriques bornés,*
- *v) C(L) ne contient que des PSL-langages.*

D - CIL-LANGAGES ET LANGAGES SANS INSERTION

Les IRS-langages (ou langages sans facteur itérant) sont les langages qui ne contiennent aucun langage rationnel infini. Ces langages jouissent de propriétés très intéressantes qui se sont révélées, à plusieurs reprises, d'une grande utilité (cf. [14], [26], [43]). Nous allons définir des langages qui vérifient une propriété plus restrictive :

Définition : *Un langage* L *est un* CIL-langage *si et seulement si* $L_1 \ L_2 \subseteq L$ *implique* L_1 *ou* L_2 *fini.*

Remarquons que tout CIL-langage est un langage sans facteur itérant. En effet si L posséde un facteur itérant, il existe x, u, y tels que $x\ u^* \ y \subseteq L$ avec $u \neq \varepsilon$ et le langage L contient le produit $L_1 \ L_2$ où $L_1 = x\ u^*$ et $L_2 = u^*\ y$ sont des langages infinis. Par contre, la réciproque est fausse puisque le produit de deux IRS-langages définis sur des alphabets disjoints est un IRS-langage. D'autre part, il est clair que tout CIL-langage algébrique est un langage algébrique linéaire et nous conjecturons que tout IRS-langage algébrique linéaire est un CIL-langage.

Remarquons, aussi, que les "c-finite" langages, définis par Greibach ([42], page 50, exemple 3.1) sont des CIL-langages ainsi que les langages $C_1 = \{a^n\ b^n \ / \ n \geq 0\}$ et Sym.

Montrons, maintenant, pour les CIL-langages, une propriété déjà établie pour les langages qui engendrent un cône rationnel translatable [14] ou fermé par un opérateur syntactique [42].

Proposition I.23 : *Soient* $\mathcal{L}$ *un cône rationnel fermé par union et* L *un CIL-langage. Alors* $L \in F(\mathcal{L})$ *implique* $L \in \mathcal{L}$.

Démonstration : Comme $L \in F(\mathcal{L}) = \mathrm{Rat} \ \square \ \mathcal{L}$, il existe un langage rationnel $R \subseteq T^*$ et une $\mathcal{L}$-substitution s, définie sur T^*, tels que $L = s(R)$. La famille Rat étant close pour les substitutions finies, on peut supposer, sans nuire à la généralité de la démonstration, que pour tout $a \in T$, soit s(a) est un langage infini de $\mathcal{L}$, soit $s(a) = \{a\}$, ce qui implique que R est fini. En effet dans le cas contraire $L = s(R)$ contient un facteur itérant et ne peut être un CIL-langage. Comme $\mathcal{L}$ est clos pour l'union, il nous reste à montrer que $s(x) \in \mathcal{L}$ pour tout $x \in R$. Tout mot $x \in R$ peut se factoriser en $x_1\ bx_2$ avec $s(x_1) = x_1$, $s(x_2) = x_2$ et $b \in T \cup \{\varepsilon\}$. Donc $s(x) = x_1\ s(b)\ x_2 \in \mathcal{L}$.□

Proposition I.24 : *Tout cône rationnel* $\mathcal{L}$ *clos par intersection est clos par insertion.*

Démonstration : Soient $L \subseteq T^*$ un langage de $\mathcal{L}$ et s une $\mathcal{L}$-insertion sur L. Il existe $T' \subseteq T$ tel que $L \subseteq (T \setminus T')^* \ (T' \cup \{\varepsilon\}) \ (T \setminus T')^*$. Alors, $s(L) = L' \cup (\bigcup_{a \in T'} s(L_a))$ avec $L' = L \cap (T \setminus T')^* \in \mathcal{L}$ et $\forall\ a \in T'$, $L_a = L \cap (T \setminus T')^* \ \{a\} \ (T \setminus T')^*$. Comme $\mathcal{L}$ est clos par union, il nous reste à montrer que

$s(L)_a \in \mathcal{L}$. Pour cela, considérons l'alphabet Δ tel que $\forall\, a \in T'$, $s(a) \subseteq \Delta^*$ et posons $\bar{\Delta} = \{\bar{b} \,/\, b \in \Delta\}$ de telle façon que $T \cap \bar{\Delta} = \emptyset$. Soient h et g les homomorphismes définis respectivement sur Δ^* et $(T \cup \bar{\Delta})^*$ par : $h(b) = \bar{b}$, $\forall\, b \in \Delta$, $g(a) = \varepsilon$, $g(x) = x$, $\forall\, x \in T \setminus \{a\}$ et $g(\bar{y}) = y$, $\forall\, \bar{y} \in \bar{\Delta}$. Il est, alors, facile de vérifier que $s(L_a)$ est égal à $g(\text{Shuf}\,(L_a, h(s(a))) \cap (T \setminus T')^* \{a\}\, \bar{\Delta}^*\, (T \setminus T')^*)$, appartient à $\mathcal{L}$ qui est clos pour l'opération "Shuffle".□

Nous pouvons, maintenant, établir un résultat concernant les CIL-langages, que nous utiliserons au chapitre II.

<u>Proposition I.25</u> : *Soient $\mathcal{L}$ un cône rationnel fermé par intersection ou insertion et* L *un CIL-langage. Alors, si* L *appartient à $F_\sigma(\mathcal{L})$,* L *est aussi dans $\mathcal{L}$.*

<u>Démonstration</u> : Prenons L_2 un CIL-langage de $\mathcal{L} \,\square\, \mathcal{L}$. Il existe, donc, $L_1 \subseteq T^*$, $L_1 \in \mathcal{L}$ et une $\mathcal{L}$-substitution s, tels que $L_2 = s(L_1)$. Comme $\mathcal{L}$ est fermé par substitution finie, on peut supposer, sans nuire à la généralité de la démonstration, que $\forall\, a \in T$, soit $s(a) = \{a\}$, soit $s(a)$ est infini. Comme $L_2 = s(L_1)$ ne peut pas contenir le produit de langages infinis, s est une $\mathcal{L}$-insertion et d'après la proposition précédente $L_2 \in \mathcal{L}$. Donc tout CIL-langage de $\mathcal{L} \,\square\, \mathcal{L}$ appartient à $\mathcal{L}$.

Posons, maintenant, $\mathcal{L}_1 = \mathcal{L}$ et $\forall\, n \geq 1$, $\mathcal{L}_{n+1} = \mathcal{L}_n \,\square\, \mathcal{L}$. Montrons, par récurrence sur n, que tout CIL-langage de $\mathcal{L}_n$ appartient à $\mathcal{L}$. Nous venons de vérifier cette propriété pour $n = 2$. Prenons L un CIL-langage de $\mathcal{L}_{k+1}$. Comme $\mathcal{L}_k$ est un cône rationnel, donc clos par substitution finie, il existe $L' \subseteq T^*$, $L' \in \mathcal{L}_k$ et une $\mathcal{L}$-substitution s tels que $s(L') = L$ et $\forall\, a \in T$, $s(a) \neq \{\varepsilon\}$. Donc, L' est nécessairement un CIL-langage. D'après l'hypothèse de récurrence, L' est un CIL-langage de $\mathcal{L}$ et L, CIL-langage de $\mathcal{L} \,\square\, \mathcal{L}$ appartient à $\mathcal{L}$. La démonstration se termine en remarquant que $F_\sigma(\mathcal{L}) = \bigcup_{n\geq 1} \mathcal{L}_n$.□

<u>Définition</u> : L *est un* **langage sans insertion** *si* $\forall\, L_1, L_2$ *langages infinis, il existe* $x_1 \in L_1$ *tel que toute factorisation de* x_1 *en* $x'_1\, x''_1$ *vérifie* $x'_1\, L_2\, x''_1 \not\subseteq L$.

Il est clair que tout langage sans insertion est un CIL-langage. D'autre part, la notion de langage sans insertion n'a guère d'intérêt dans le cadre des langages algébriques, puisque tout langage algébrique sans insertion est fini. Cependant, il est facile de construire des langages sans insertion. Pour tout langage L, notons L [c], le langage $\{w\, c^n \,/\, w \in L,\ n = \ell(w)\}$ (cf. [26]).

<u>Lemme I.26</u> : *Pour tout langage* $L \subseteq T^*$ *avec* $c \notin T$, L [c] *est un CIL-langage. De plus* L [c] *est un langage sans insertion si et seulement si* L *est un CIL-langage.*

<u>Démonstration</u> : Supposons, en effet, qu'il existe deux langages infinis L_1 et L_2 tels que $L_1 L_2 \subseteq L\,[c]$. Si L_2 est inclus dans c^*, pour tout $x_1 \in L_1$, il existe $i > \ell(x_1)$ tel que $c^i \in L_2$, donc $x_1\, c^i \in L_1 L_2 \setminus L[c]$.

Dans le cas contraire, L_1 est inclus dans T^* et pour tout $x_2 \in L_2$, il existe $x_1 \in L_1$ avec $\ell(x_1) > \ell(x_2)$, donc $x_1 x_2 \in L_1 L_2 \setminus L[c]$, d'où la contradiction.

Supposons, maintenant, que $L[c]$ ne soit pas sans insertion. Il existe, alors, deux langages infinis L_1 et L_2 tels que pour tout $x_1 \in L_1$, il existe une factorisation $x_1 = x'_1 x''_1$ avec $x'_1 L_2 x''_1 \subseteq L[c]$. Il existe aussi des entiers n_1 et n_2 tels que $\forall x_1 \in L_1$, $\forall x_2 \in L_2$, $\ell_c(x_1) - \ell_T(x_1) = n_1$ et $\ell_c(x_2) - \ell_T(x_2) = n_2$ avec $n_1 + n_2 = 0$ ($\ell_T(x)$ désigne le nombre d'occurrences, dans x, de lettres appartenant à T). Comme L_2 est infini, $L_2 \cap T^+ c^+ \neq \emptyset$ et $x'_1 L_2 x''_1 \subseteq L[c]$ implique $x'_1 \in T^*$ et $x''_1 \in c^*$. Soit h_c, l'homomorphisme défini sur $T \cup \{c\}$ par $h_c(c) = \varepsilon$ et $h_c(a) = a$, $\forall a \in T$. Comme L_1 est infini et $\ell_c(x_1) - \ell_T(x_1) = n_1$, $h_c(L_1)$ est infini. De même $h_c(L_2)$ est infini et comme $x'_1 L_2 x''_1 \subseteq L[c]$ implique $h_c(x'_1 x''_1) = x'_1$, L contient $h_c(L_1) h_c(L_2)$, le produit de deux langages infinis.

Supposons, enfin, que L contienne le produit $L'_1 L'_2$ de deux langages infinis. Les langages $L_1 = L'_1[c]$ et $L_2 = L'_2[c]$ sont infinis et vérifient : tout mot x_1 de L_1 se factorise en $x'_1 c^n$ avec $n = \ell(x'_1)$ et $x'_1 L_2 c^n$ est inclus dans $L[c]$ qui n'est donc pas un langage sans insertion.□

Si L est un langage sans insertion, nous obtenons un résultat équivalent à la proposition I.25, mais, cette fois-ci, il n'est plus nécessaire de supposer que le cône rationnel $\mathcal{L}$ est clos par insertion. Pour tout cône rationnel clos par union $\mathcal{L}$, posons $I_1(\mathcal{L}) = \mathcal{L}$ et $I_{n+1}(\mathcal{L}) = \{s(L) \ / \ L \in I_n(\mathcal{L})$ et s est une $\mathcal{L}$-insertion sur L$\}$, $\forall n \in \mathbb{N}_+$. Alors le plus petit cône rationnel clos par insertion et contenant $\mathcal{L}$, noté $C_i(\mathcal{L})$, est égal à $\bigcup_{n\geq 1} I_n(\mathcal{L})$.

<u>Proposition I.27</u> : *Soit $\mathcal{L}$ un cône rationnel fermé par union. Tout langage sans insertion appartenant à $F_\sigma(\mathcal{L})$ appartient à $\mathcal{L}$.*

<u>Démonstration</u> : Montrons que pour tout $n \in \mathbb{N}_+$ et tout langage L sans insertion,

$L \in I_{n+1}(\mathcal{L})$ implique $L \in I_n(\mathcal{L})$. Prenons, en effet, $L \in I_{n+1}(\mathcal{L})$. Il existe un langage $L' \in I_n(\mathcal{L})$ et s une $\mathcal{L}$-insertion sur L' tels que $s(L') = L$. Donc il existe un alphabet T et $T' \subseteq T$ tels que $L' \subseteq (T \setminus T')^* (T' \cup \{\varepsilon\}) (T \setminus T')^*$, $s(a) = a$, $\forall a \in T \setminus T'$ et $s(a) \in \mathcal{L}$, $\forall a \in T'$. Alors $s(L') = L'_1 \cup (\bigcup_{a\in T'} s(L'_a))$ où $L'_1 = L' \cap (T \setminus T')^*$ et $\forall a \in T'$, $L'_a = L' \cap (T \setminus T')^* \{a\} (T \setminus T')^*$ sont des langages de $I_n(\mathcal{L})$ qui est un cône rationnel clos par union. Il nous reste, donc, à montrer que $\forall a \in T'$, $s(L'_a)$ appartient à $I_n(\mathcal{L})$. Comme L est sans insertion, pour tout $a \in T'$, $s(L'_a) \subseteq L$ est aussi sans insertion. Nécessairement, ou bien L'_a est fini, ou bien s(a) est fini. Dans les deux cas $s(L'_a)$ appartient à $I_n(\mathcal{L})$. Comme $I_1(\mathcal{L}) = \mathcal{L}$ et $C_i(\mathcal{L}) = \bigcup_{n\geq 1} I_n(\mathcal{L})$, on en déduit que tout langage sans insertion appartenant à $C_i(\mathcal{L})$, appartient à $\mathcal{L}$. Le résultat cherché découle, alors, immédiatement de la proposition I.25.□

En particulier, pour tout langage $L \subseteq T^*$ avec $d \notin T$, et tout entier k supérieur à deux, le langage $L' = \{(w\,d)^k \ / \ w \in L\}$ est un langage sans insertion, ce qui implique : si $\mathcal{L}$ est un cône rationnel clos par union, $L' \in F_\sigma(\mathcal{L})$ entraine $L' \in \mathcal{L}$ (cf. [45]).

CHAPITRE II

LANGAGES COMMUTATIFS

A - LANGAGES COMMUTATIFS ET LANGAGES BORNÉS

B - COMPARAISON DE C(COM) AVEC DIVERSES FAMILLES DE LANGAGES

C - EFFACEMENT ET LANGAGES COMMUTATIFS

D - MOTS INFINIS ET LANGAGES COMMUTATIFS

E - LANGAGES K-COMMUTATIFS

A - Langages commutatifs et langages bornés

Nous allons commencer par étudier les propriétés de clôture de $C(\mathrm{Com})$, le plus petit cône rationnel contenant tous les langages commutatifs, ainsi que les liens entre les langages bornés et les langages commutatifs. Une première remarque d'importance est que la famille Com est fermée par homomorphisme inverse. En utilisant la caractérisation des transductions rationnelles en terme de bimorphisme (cf. théorème de Nivat), nous obtenons l'égalité : $C(\mathrm{Com}) = H_\alpha(\mathrm{Com} \wedge \mathrm{Rat})$, c'est à dire que tout langage appartenant à $C(\mathrm{Com})$ est l'image homomorphe (alphabétique) de l'intersection d'un langage commutatif et d'un langage borné. Cette égalité explique pourquoi certaines propriétés qui s'étendent de $C(\mathrm{Bor})$, le plus petit cône rationnel contenant tous les langages bornés, à $C(\mathrm{Com})$, admettent, alors parfois, des démonstrations plus simples.

La famille Com est aussi fermée par intersection. Etant donné que toute famille de langages fermée par intersection et homomorphisme inverse engendre un cône rationnel clos par intersection (lemme II.2), nous en déduisons que $C(\mathrm{Com})$ est clos par intersection. Nous donnons ensuite une condition nécessaire et suffisante pour qu'un cône rationnel, L, clos par intersection soit commutativement clos (i.e. $L \in L$ implique $c(L) = \Psi^{-1} \circ \Psi(L) \in L$). Cette condition, l'appartenance à L du langage $C_1 = \{a^n b^n \ / \ n \geq 0\}$ est vérifiée pour $L = C_\cap(\mathrm{Bor})$, le plus petit cône rationnel clos par intersection contenant les langages bornés. Nous obtenons, ainsi, l'un des résultats de base de cette section, l'égalité entre $C(\mathrm{Com})$ et $C_\cap(\mathrm{Bor})$.

L'autre résultat de base, le lemme II.10, est plus technique mais se trouve directement à l'origine de deux corollaires concernant tout cône rationnel L engendré à partir de langages commutatifs. Le premier, que nous utiliserons à de nombreuses reprises tout au long de cette thèse, affirme que si L est langage strictement borné, $L \in L$ implique $c(L) \in L$. Le deuxième montre, pour L, l'équivalence entre le fait d'être clos par produit et celui d'être clos par intersection.

Lemme II.1 : *Pour toute famille de langages L close par homomorphisme inverse (alphabétique) et tout cône rationnel L', $C(L) \wedge L'$ est inclus dans $H(L' \wedge L)$.*

Démonstration : Comme L est clos par homomorphisme inverse alphabétique, le théorème de Nivat, énoncé au chapitre I, implique $C(L) = H_\alpha(L \wedge \mathrm{Rat})$ et pour tout $L \in C(L)$, il existe $L_1 \in L$, $R \in \mathrm{Rat}$ et un homomorphisme alphabétique h tels que $L = h(L_1 \cap R)$. Alors pour tout langage $L_2 \in L'$, $L \cap L_2 = h(L_1 \cap R) \cap L_2 = h(h^{-1}(L_2) \cap L_1 \cap R)$ avec $h^{-1}(L_2) \cap R \in L'$, donc $h^{-1}(L_2) \cap L_1 \cap R \in L' \wedge L$ et $L \cap L_2 \in H(L' \wedge L)$.□

Remarquons que si L n'est pas clos par homomorphisme inverse alphabétique, le lemme précédent n'est plus vérifié. Prenons, en effet, pour L, la famille des langages algébriques inclus dans $a^* b^*$. Alors $C(L)$ est égal à $C(C_1)$ où $C_1 = \{a^n b^n \ / \ n \geq 0\}$ [6] et contient strictement Rat. La proposition III.56 implique alors que $C(L) \wedge C(L)$ n'est pas inclus dans Alg et donc $C(L) \subsetneq C(L) \wedge C(L)$. Par contre, il est clair que $C(L) \wedge L = L \wedge L$ et comme $L \wedge L$ est égal à L [24], nous obtenons $C(C(L) \wedge L) = C(L) \subsetneq C(L) \wedge C(L)$.

De même, $\mathcal{L} \wedge \mathcal{L}in = \mathcal{L} \wedge \mathcal{L} = \mathcal{L}$ où $\mathcal{L}in$ désigne la famille des langages algébriques linéaires. Nous avons $F(\mathcal{L}) = F^f(\mathcal{L}) \subseteq \mathrm{Alg}$ [41] , alors que $F(\mathcal{L}) \wedge \mathcal{L}in \supseteq C(\mathcal{L}) \wedge \mathcal{L}in \supseteq C(\mathcal{L}) \wedge C(\mathcal{L})$ qui n'est pas inclus dans Alg. Donc $F(\mathcal{L})$ et $F^f(\mathcal{L})$ ne sont pas clos par intersection avec les langages algébriques linéaires ce qui répond à une question de S. Ginsburg ([25] page 55).

Si nous considérons, maintenant, une famille close par intersection et homomorphisme inverse, nous obtenons :

<u>Lemme II.2</u> : *Soit $\mathcal{L}$ une famille de langages close par intersection et homomorphisme inverse (alphabétique). Alors, $C(\mathcal{L})$ est clos par intersection, produit et union.*

<u>Démonstration</u> : Comme $C(\mathcal{L}) = \mathcal{H}_\alpha(\mathcal{L} \wedge \mathrm{Rat})$, pour tout $L \in C(\mathcal{L})$, il existe $L_1 \in \mathcal{L}$, $R \in \mathrm{Rat}$ et un homomorphisme alphabétique h tels que $L = h(L_1 \cap R)$. Alors, pour tout $L_2 \in \mathcal{L}$, $L \cap L_2 = h(L_1 \cap R) \cap L_2 = h(L_1 \cap h^{-1}(L_2) \cap R)$ et comme $h^{-1}(L_2) \in \mathcal{L}$, $L_1 \cap h^{-1}(L_2) \in \mathcal{L} \wedge \mathcal{L} = \mathcal{L}$ et $L \cap L_2$ appartient à $C(L)$. Donc, $C(\mathcal{L}) \wedge \mathcal{L}$ est inclus dans $C(\mathcal{L})$ et le lemme précédent entraine $C(\mathcal{L}) \wedge C(\mathcal{L}) \subseteq C(C(\mathcal{L}) \wedge \mathcal{L}) \subseteq C(\mathcal{L})$. Il est clair, d'autre part, que $C(\mathcal{L}) \subseteq C(\mathcal{L}) \wedge C(\mathcal{L})$ et $C(\mathcal{L}) = C(\mathcal{L}) \wedge C(\mathcal{L})$ est un cône rationnel clos par intersection.

D'après la proposition I.9 , $C(\mathcal{L})$ est clos pour l'opération shuffle. Prenons alors, dans $C(\mathcal{L})$, $L_1, L_2 \subseteq T^*$ et considérons $T' = \{a' \mid a \in T\}$ avec $T \cap T' = \emptyset$ ainsi que l'homomorphisme g défini sur T par $g(a) = a'$, $\forall a \in T$. Le langage $L' = \mathrm{shuf}\,(L_1, g(L_2)) \in C(\mathcal{L})$ et $L_1 L_2 = g^{-1}(L' \cap T^* T'^*)$ appartient à $C(\mathcal{L})$ qui est donc clos par produit. Le langage $L'' = (L_1 \cup \{\varepsilon\})\,(g(L_2) \cup \{\varepsilon\})$ appartient à $C(\mathcal{L})$ ainsi que $L_1 \cup L_2$ qui est égal à $g^{-1}(L'' \cap (T^* \cup T'^*))$ et $C(\mathcal{L})$ est clos par union.□

Nous allons pouvoir appliquer ce lemme à la famille de tous les langages commutatifs. Nous avons, en effet :

<u>Lemme II.3</u> : *La famille Com est close par union, intersection, homomorphisme alphabétique et homomorphisme inverse.*

<u>Démonstration</u> : Soient L_1 et L_2 deux langages commutatifs. Alors, $L_1 \cap L_2 = c(L_1) \cap c(L_2) = \Psi^{-1} \circ \Psi(L_1) \cap \Psi^{-1} \circ \Psi(L_2) = \Psi^{-1}(\Psi(L_1) \cap \Psi(L_2)) = \Psi^{-1} \circ \Psi \circ \Psi^{-1}(\Psi(L_1) \cap \Psi(L_2)) = \Psi^{-1} \circ \Psi(c(L_1) \cap c(L_2)) = c(L_1 \cap L_2)$ est un langage commutatif. On montre de la même façon que $L_1 \cup L_2 = c(L_1) \cup c(L_2) = c(L_1 \cup L_2)$ est un langage commutatif.

Prenons, maintenant, un homomorphisme g. Il est clair que pour tout langage L', $g \circ c(L')$ est inclus dans $c \circ g(L')$. En particulier, en prenant $L' = g^{-1}(L)$, on obtient $g \circ c \circ g^{-1}(L) \subseteq c \circ g \circ g^{-1}(L) = c(L)$ ce qui implique $c \circ g^{-1}(L) \subseteq g^{-1} \circ c(L)$. Si L est un langage commutatif, $L = c(L)$ et $g^{-1}(L) \subseteq c \circ g^{-1}(L) \subseteq g^{-1} \circ c(L)$ qui est donc un langage commutatif.

Enfin, si g est un homomorphisme alphabétique, il est clair que g et c commutent, c'est-à-dire que pour tout langage L, $c \circ g(L)$ est égal à $g \circ c(L)$ et si L est commutatif $c \circ g(L) = g \circ c(L) = g(L)$ qui est donc un langage commutatif.□

Des deux lemmes précédents, nous pouvons déduire immédiatement :

<u>Proposition II.4</u> : *La famille C(Com) est close par intersection, union et produit.*

Considérons un langage borné L. Il existe, alors, des mots $w_1,\ldots,w_n$ tels que L soit inclus dans $w_1^*\ldots w_n^*$. Soit g l'homomorphisme défini sur $T = \{a_1,\ldots,a_n\}$ par $g(a_i) = w_i$, $\forall\, i \in \{1,\ldots,n\}$. Le langage $L' = g^{-1}(L) \cap a_1^*\ldots a_n^*$ est un langage strictement borné qui vérifie $L' = c(L') \cap a_1^*\ldots a_n^*$ et $L = g(L')$. Le langage $c(L')$ est commutatif, donc L' et g(L') appartiennent à C(Com). Nous en déduisons que la famille Bor est incluse dans C(Com) et $C_\cap(\mathrm{Bor})$, le plus petit cône rationnel clos par intersection, contenant tous les langages bornés, est inclus dans $C_\cap(\mathrm{Com}) = C(\mathrm{Com})$. Pour montrer l'inclusion inverse, nous utiliserons le résultat suivant :

<u>Proposition II.5</u> : *Soient $\mathcal{L}$ un cône rationnel et $\mathcal{L}_1$, le plus petit cône rationnel contenant le langage $C_1 = \{a^n b^n \mid n \geq 0\}$. Alors, $\mathcal{L} \wedge \mathcal{L}_1 \subseteq \mathcal{L}$ implique $\mathcal{L}$ commutativement clos.*

<u>Démonstration</u> : Prenons $T = \{a_1,\ldots,a_n\}$, $\bar{T} = \{\bar{a}_1,\ldots,\bar{a}_n\}$ et $L \subseteq T^*$ un langage de $\mathcal{L}$. Pour tout $i \in \{1,\ldots,n\}$, définissons l'homomorphisme h_i de $(T \cup \bar{T})^*$ dans $\{a,b\}^*$ par : $h_i(a_i) = a$, $h_i(\bar{a}_i) = b$ et $h_i(a_j) = h_i(\bar{a}_j) = \varepsilon$ pour $j \neq i$. Posons $L' = L\bar{T}^* \cap (\bigcap_{i=1}^{n} h_i^{-1}(C_1))$ et montrons que $c(L) = g(L')$ où g est l'homomorphisme défini sur $T \cup \bar{T}$ par $g(a_i) = \varepsilon$ et $g(\bar{a}_i) = a_i$, $\forall\, i \in \{1,\ldots,n\}$.

Soit $x \in L'$. On peut factoriser x en yz avec $y \in L$ et $z \in \bar{T}^*$. Pour tout $i \in \{1,\ldots,n\}$, $yz \in h_i^{-1}(C_1)$ donc $\ell_{a_i}(y) = \ell_{\bar{a}_i}(z)$. Ceci implique que $g(x) = g(z) \in c(y) \subseteq c(L)$. Réciproquement, si $x \in c(L)$, il existe $y \in L$ et $\bar{x} \in \bar{T}^*$ tels que $x \in c(y)$ et $g(\bar{x}) = x$. Pour tout $i \in \{1,\ldots,n\}$, $h_i(y\bar{x}) \in C_1$ et comme $y\bar{x} \in L\bar{T}^*$, $y\bar{x} \in L'$. Donc $x = g(y\bar{x})$ appartient à g(L'). Comme $\mathcal{L} \wedge C(C_1)$ est inclus dans $\mathcal{L}$, $g(L') \in \mathcal{L}$ qui est commutativement clos. □

Comme tout cône rationnel commutativement clos contient C_1, qui est égal à $c((ab)^*) \cap a^*b^*$, la proposition précédente implique :

<u>Corollaire II.6</u> : *Soit $\mathcal{L}$ un cône rationnel fermé par intersection. Alors $\mathcal{L}$ est commutativement clos si et seulement si $C_1 \in \mathcal{L}$.*

Remarquons qu'il existe des cônes rationnels différents de Rat, fermés par intersection sans être commutativement clos. Considérons, en effet, D'^*_1, le langage de semi-Dyck sur une lettre et posons $L = \mathrm{Init}(D'^*_1)$. Il est montré dans [1] que $F_\cap(L)$ la plus petite FAL close par intersection et contenant L, est incluse dans la famille des langages récursifs. Comme $F_\cap(C_1)$ est égal à la famille de tous les langages récursivement énumérables [46], $C_1 \notin F_\cap(L)$ et à fortiori $C_1 \notin C_\cap(L)$ qui n'est donc pas commutativement clos.

De même, il existe des cônes rationnels commutativement clos sans être fermé par intersection. Prenons, par exemple, $\mathcal{L}_S$, la plus grande FAL ne contenant que des PSL-langages (cf.[33]). Comme C(c(Rat)) ne contient que des PSL-langages (corollaire III.4), C(c(Rat)) est inclus dans $\mathcal{L}_S$ et $c(\mathcal{L}_S) = c(\text{Rat}) \subseteq \mathcal{L}_S$ qui est donc un cône rationnel commutativement clos. Et, comme $C_1 \in \mathcal{L}_S$, $F(\mathcal{L}_S) = \mathcal{L}_S \subsetneq F_\cap(\mathcal{L}_S)$ qui contient tous les langages récursivement énumérables. Donc $\mathcal{L}_S$ est un cône rationnel commutativement clos, qui n'est pas fermé par intersection.

Considérons, maintenant, la famille $C_\cap$(Bor). Elle contient le langage borné C_1, donc $C_\cap$(Bor) est commutativement clos et contient Com = c(Bor) ce qui entraine C(Com) inclus dans $C_\cap$(Bor). Nous pouvons, alors, énoncer le résultat principal de cette section :

<u>Proposition II.7</u> : *La famille C(Com) est égale à $C_\cap$(Bor), le plus petit cône rationnel clos par intersection et contenant tous les langages bornés.*

Etudions; maintenant, les rapports entre les homomorphismes alphabétiques, l'opération de shuffle et l'opération de clôture commutative.

<u>Lemme II.8</u> : *Soient $L_1,\ldots,L_n$ des langages commutatifs. Alors, $c(L_1,\ldots,L_n) = \text{shuf}(L_1,\ldots,L_n)$.*

<u>Démonstration</u> : En vertu de l'associativité du produit et du shuffle, il suffit de montrer le lemme pour n = 2. Il existe des homomorphismes alphabétiques h, g_1, g_2 tels que $L = \text{shuf}(L_1, L_2) = h(g_1^{-1}(L_1) \cap g_2^{-1}(L_2))$ [25]. D'après le lemme II.3, L est un langage commutatif et comme $L_1 L_2 \subseteq L \subseteq c(L_1 L_2)$, nous obtenons $L = c(L_1 L_2)$.□

<u>Lemme II.9</u> : *Soient h un homomorphisme alphabétique défini sur un alphabet T, $L_1,\ldots,L_n$ des langages inclus dans T^* et $L = \text{shuf}(L_1,\ldots,L_n)$. Alors h(L) est égal à $\text{shuf}(h(L_1),\ldots,h(L_n))$ et si les langages $h(L_1),\ldots,h(L_n)$ sont commutatifs, h(L) est aussi commutatif.*

<u>Démonstration</u> : Plaçons nous d'abord dans le cas n = 2. Prenons $y \in h(L)$. Il existe $x_1 \in L_1$ et $x_2 \in L_2$ tels que $y = h(x)$ avec $x \in \text{shuf}(x_1, x_2)$ ce qui implique $h(x) \in \text{shuf}(h(x_1), h(x_2)) \subseteq \text{shuf}(h(L_1), h(L_2))$. Réciproquement si $y \in \text{shuf}(h(L_1), h(L_2))$ il existe $x_1 \in L_1$ et $x_2 \in L_2$ tels que $y \in \text{shuf}(h(x_1), h(x_2))$. Comme h est alphabétique, $\forall x \in T^*$, $h(x) = z' z''$ implique qu'il existe une factorisation de x en $x' x''$ avec $h(x') = z'$ et $h(x'') = z''$. Nous pouvons en déduire que $\text{shuf}(h(x_1), h(x_2)) \subseteq h \circ \text{shuf}(x_1, x_2)$. Donc $y = h(x)$ avec $x \in \text{shuf}(x_1, x_2) \subseteq \text{shuf}(L_1, L_2)$ et $y \in L'$.

Enfin, si $h(L_1)$ et $h(L_2)$ sont des langages commutatifs, d'après le lemme II.8, $L' = \mathrm{shuf}(h(L_1), h(L_2))$ est un langage commutatif.

En utilisant l'associativité de l'opération shuffle, il est alors facile d'obtenir, par induction, le résultat dans le cas général.□

Nous pouvons, maintenant, énoncer un lemme dont découlent la plupart des résultats ultérieurs de cette section :

Lemme II.10 : *Soient $T_1,\ldots,T_n$ des alphabets disjoints deux à deux, L' un langage inclus dans $T_1^*\ldots T_n^*$ et L'' égal à $\{y \in \mathrm{shuf}(x_1,\ldots,x_n) \ / \ x_1\ldots x_n \in L'$ et $\forall\, i \in \{1,\ldots,n\},\ x_i \in T_i^*\}$. Alors, pour tout langage commutatif L, on a :*

i) $L' \in C(L)$ ssi $L'' \in C(L)$

ii) $L' \in C^f(L)$ ssi $L'' \in C^f(L)$.

Démonstration : Comme $L' = L'' \cap T_1^*\ldots T_n^*$, $L'' \in C(L)$ (resp. $L'' \in C^f(L)$) implique $L' \in C(L)$ (resp. $L' \in C^f(L)$).

Prenons, maintenant, $L' \in C(L)$. Il existe un alphabet Z, un langage rationnel R et deux homomorphismes alphabétiques h et g tels que $L' = g(h^{-1}(L) \cap R)$. Pour tout $i \in \{1,\ldots,n\}$, Posons $Z_i = \{z \in Z \ / \ g(z) \in T_i \cup \{\varepsilon\}\}$. Comme L' est inclus dans $T_1^*\ldots T_n^*$, L' est égal à $g(h^{-1}(L) \cap R')$ où $R' = R \cap Z_1^*\ldots Z_n^*$. Alors il existe des langages rationnels $R_{i,j}$ tels que $R' = \bigcup_{i=1}^{p} R_{i,1}\ldots R_{i,n}$ où $\forall\, i \in \{1,\ldots,p\}$, $\forall\, j \in \{1,\ldots,n\}$, $R_{i,j}$ est inclus dans Z_j^*. Considérons le langage rationnel $R'' = \bigcup_{i=1}^{p} \mathrm{shuf}(R_{i,1},\ldots,R_{i,n})$ et montrons que $L'' = g(h^{-1}(L) \cap R'')$.

Prenons $y \in L''$. Il existe $x_1,\ldots,x_n$ tels que $y \in \mathrm{shuf}(x_1,\ldots,x_n)$, $x = x_1\ldots x_n \in L'$ et $\forall\, j \in \{1,\ldots,n\}$, $x_j \in T_j^*$. Il existe alors, $z \in Z$ tel que $z \in h^{-1}(L)$, $z \in R'$ et $g(z) = x$. On peut trouver $i \in \{1,\ldots,p\}$ et une factorisation de z en $z_1\ldots z_n$ tels que $\forall\, j \in \{1,\ldots,n\}$, $z_j \in R_{i,j}$ et donc $g(z_j) = x_j$. D'après le lemme précédent, on a $\mathrm{shuf}(x_1,\ldots,x_n) = \mathrm{shuf}(g(z_1),\ldots,g(z_n)) = g \circ \mathrm{shuf}(z_1,\ldots,z_n)$ et comme $\mathrm{shuf}(z_1,\ldots,z_n) \subseteq \mathrm{shuf}(R_{i,1},\ldots,R_{i,n}) \subseteq R''$ et que $\mathrm{shuf}(z_1,\ldots,z_n) \subseteq c(z_1 \ldots z_n) \subseteq h^{-1}(L)$ qui est commutatif, nous avons $y \in g(h^{-1}(L) \cap R'')$ ce qui implique $L'' \subseteq g(h^{-1}(L) \cap R'')$. Réciproquement si $y \in g(h^{-1}(L) \cap R'')$, il existe $z \in Z^*$ tel que $y = g(z)$, $z \in h^{-1}(L)$ et $z \in R''$ donc $z \in \mathrm{shuf}(R_{i,1},\ldots,R_{i,n})$ pour un certain $i \in \{1,\ldots,p\}$ et $z \in \mathrm{shuf}(z_1,\ldots,z_n)$ avec $\forall\, j \in \{1,\ldots,n\}$, $z_j \in R_{i,j}$. Le mot $z' = z_1\ldots z_n$ appartient à R' et $c(z') = c(z) \subseteq h^{-1}(L)$ donc $g(z') = g(z_1)\ldots g(z_n)$ appartient à L' avec $\forall\, j \in \{1,\ldots,n\}$, $g(z_j) \in T_j^*$ ce qui entraine d'après le lemme précédent que $y \in g \circ \mathrm{shuf}(z_1,\ldots,z_n) = \mathrm{shuf}(g(z_1),\ldots,g(z_n)) \subseteq L''$ donc $L'' = g(h^{-1}(L) \cap R'') \in C(L)$.

Maintenant, si $L' \in C^f(L)$, on peut reprendre les mêmes notations en ajoutant le fait que g est k-limité sur R pour un certain $k \in \mathbb{N}_+$. Il est clair que g est encore k-limité sur les $R_{i,j}$ et il est facile de vérifier que g est k'-limité sur R'' avec $k' = kn$. En effet si w est un facteur d'un mot du langage shuf $(R_{i,1},\dots,R_{i,n})$, il existe $w_1,\dots,w_n$ tels que $w \in$ shuf $(w_1,\dots,w_n)$ avec $\forall\, j \in \{1,\dots,n\}$, w_j est un facteur d'un mot de $R_{i,j}$. Comme $\ell(w) = \ell(w_1)+\dots+\ell(w_n)$, si $g(w) = \varepsilon$ avec $\ell(w) > kn$, il existe $j \in \{1,\dots,n\}$ tel que $g(w_j) = \varepsilon$ avec $\ell(w_j) > k$ ce qui implique que g n'est pas k-limité sur $R_{i,j}$ d'où la contradiction.□

Le langage L" construit au lemme précédent est inclus dans c(L'). D'autre par, si $x_1,\dots,x_n$ sont des mots définis chacun sur un alphabet d'une seule lettre, $\mathrm{shuf}(x_1,\dots,x_n)$ est égal à $c(x_1\dots x_n)$. Donc, dans le cas où T_i ne contient qu'une lettre, L" est égal à c(L') et nous obtenons un résultat auquel nous ferons de nombreuses références par la suite :

Corollaire II.11 : *Pour tout langage commutatif L et tout langage L' strictement borné on a :*

i) $L' \in C(L)$ *ssi* $c(L') \in C(L)$

ii) $L' \in C^f(L)$ *ssi* $c(L') \in C^f(L)$.

En particulier en prenant pour L un langage rationnel commutatif, nous retrouvons un résultat bien connu :

Un langage L, strictement borné est rationnel si et seulement si c(L) est rationnel.

Remarquons que ce résultat n'est plus valable pour un langage L borné quelconque. En effet, $L = (ab)^*$ est rationnel, alors que $c(L) = D_1^*$ est non rationnel.

Considérons, maintenant, deux langages L et L' strictement bornés tels que $L' \in C(c(L))$. Ceci n'implique pas nécessairement $L' \in C(L)$. En effet, prenons $L = C_1$ et $L' = \{a_1^{n_1}a_2^{n_2}a_3^{n_3} \;/\; n_1, n_2, n_3 \geq 0,\ n_2 = n_1 + n_3\}$. Il est clair que $L' \in C(c(L)) = C(D_1^*)$. Par contre L' n'est pas un langage linéaire et n'appartient pas à C(L). Nous pouvons cependant montrer que :

Proposition II.12 : *Soit L un langage commutatif. Pour tout langage borné L' appartenant à C(L) (resp.* $C^f(L)$*), il existe un langage rationnel borné B tel que* $L' \in C(L \cap B)$ *(resp.* $C^f(L \cap B)$*).*

Démonstration : Il existe des mots $w_1,\dots,w_n$ tels que $L' \subseteq w_1^*\dots w_n^*$. Considérons l'alphabet $T = \{a_1,\dots,a_n\}$, l'homomorphisme ϕ défini sur T par $\phi(a_i) = w_i$, $\forall\, i \in \{1,\dots,n\}$ et posons $L'' = \phi^{-1}(L') \cap a_1^*\dots a_n^*$. Alors $L'' \in C^f(L')$ et $L' = \phi(L'') \in C^f(L'')$ donc $C^f(L') = C^f(L'')$. Si $L'' \in C(L)$, il existe un alphabet Z et des homomorphismes alphabétiques h, g tels que $L'' = g(h^{-1}(L) \cap R)$. Posons $Z' = Z \cap g^{-1}(\varepsilon)$ et $\forall\, i \in \{1,\dots,n\}$, $Z_i = \{z \in Z \;/\; g(z) \in a_i^*\}$. Alors $L'' = g(h^{-1}(L) \cap R')$ où $R' = R \cap Z_1^*\dots Z_n^*$ peut s'écrire sous la forme $\bigcup_{i=1}^{p} R_{i,1}\dots R_{i,n}$ avec $\forall\, j \in \{1,\dots,n\}$, $R_{i,j}$ langage rationnel inclus dans Z_j^*. Pour tout $i \in \{1,\dots,p\}$ et $j \in \{1,\dots,n\}$, il existe un langage rationnel borné $R'_{i,j} \subseteq Z_j$ tel que $c(R'_{i,j}) = c(R_{i,j})$. Considérons le langage rationnel borné $R'' = \bigcup_{i=1}^{p} R'_{i,1}\dots R'_{i,n}$ et montrons que L" est égal à $g(h^{-1}(L \cap B) \cap R'')$ ou $B = h(R'')$ est un langage rationnel borné. Prenons $x \in L''$. Il existe, alors, $y = y_1\dots y_n \in h^{-1}(L)$ et

$i \in \{1,\dots,p\}$ tels que $g(y) = x$ et $\forall\, j \in \{1,\dots,n\}$, $y_j \in R_{i,j}$. Pour tout $j \in \{1,\dots,n\}$, il existe $y'_j \in R'_{i,j}$ tel que $c(y'_j) = c(y_j)$ donc $g(y'_j) = g(y_j)$. Alors $y' = y'_1 \dots y'_n \in R'' \subseteq h^{-1}(B)$ et comme $h^{-1}(L)$ est commutatif et que $c(y') = c(y)$, $y' \in h^{-1}(L)$. Donc $y' \in h^{-1}(L \cap B) \cap R''$, $x = g(y) = g(y') \in g(h^{-1}(L \cap B) \cap R'')$ et $L'' \subseteq g(h^{-1}(L \cap B) \cap R'')$. L'inclusion inverse se montre de la même façon et l'on obtient $L'' = g(h^{-1}(L \cap B) \cap R'') \in C(L \cap B)$ et $L' \in C(L'') \subseteq C(L \cap B)$.

Maintenant, si L'' appartient à $C^f(L)$, on peut garder les mêmes notations en ajoutant l'hypothèse que g est limité sur R. Supposons que h ne soit pas limité sur R''. Alors, il existe $i \in \{1,\dots,p\}$, $j \in \{1,\dots,n\}$ tels que g n'est pas limité sur $R'_{i,j}$, donc $R'_{i,j}$ possède un facteur itérant appartenant à Z'^* et il existe $u, v \in Z^*$, $x \in Z'\, Z'^*$ tels que $u\, x^*\, v \subseteq R'_{i,j}$. Soient m le nombre d'états d'un automate d'état fini reconnaissant $R_{i,j}$, t la longueur du mot $u\,v$ et posons $k = m(t+1)$.

Comme $c(R_{i,j}) = c(R'_{i,j})$, il existe dans $R_{i,j}$ un mot $w \in c(u\, x^k\, v)$. Comme $\ell(w) \geq k$, il existe une factorisation de w en $u'\, w'\, v'$ avec $w' \in Z'^m\, Z'^*$, donc $R_{i,j}$ possède un facteur itérant appartenant à $Z'\, Z'^*$ et g n'est pas limité sur $R_{i,j}$, donc à fortiori g n'est pas limité sur R, d'où la contradiction. On a, alors, $L'' \in C^f(L \cap B)$ et $L' \in C^f(L'') \subseteq C^f(L \cap B)$.□

Cette propriété n'est plus nécessairement vérifiée quand le langage L n'est pas commutatif. Considérons, en effet, le langage :

$$L = \{c\, a\, c\, a^2 \dots c\, a^n \;/\; n \geq 0\}$$

Il est facile de voir que L domine rationnellement le langage borné $C_1 = \{a^n b^n \;/\; n \geq 0\}$. Donc L et $L' = \mathrm{Init}(L)$ qui sont rationnellement équivalent ($L = (L' \cap \{a,c\}^*\, c) / c$), sont des langages non rationnels. Et comme L' est le langage d'un mot infini, la proposition II.51 entraine que pour tout langage rationnel borné B, $L \cap B \subseteq L' \cap B$ est un langage fini, donc $C(L \cap B) = \mathrm{Rat}$ ne contient pas C_1 qui appartient pourtant à $C(L)$.

Le corollaire II.11 va nous permettre, maintenant, de transposer certains résultats des langages bornés aux langages commutatifs :

Corollaire II.13 : *Pour toute famille finie $\mathcal{L}$ de langages m-commutatifs, il existe un langage m-commutatif L tel que $C(L) = C_\cup(\mathcal{L})$ et $C^f(L) = C^f_\cup(\mathcal{L})$.*

Démonstration : On peut supposer que $\mathcal{L} = \{L_1,\dots,L_p\}$ où $\forall\, i \in \{1,\dots,p\}$, L_i est un langage commutatif défini sur l'alphabet $T = \{a_1,\dots,a_m\}$. Considérons la famille $\mathcal{L}' = \{L'_1,\dots,L'_p\}$ où $\forall\, i \in \{1,\dots,p\}$, $L'_i = L_i \cap a_1^* \dots a_m^*$. Il existe un langage $L' \in a_1^* \dots a_m^*$ tel que $C(L') = C_\cup(\mathcal{L}')$ et $C^f(L') = C^f_\cup(\mathcal{L}')$ [25] [35]. Posons $L = c(L')$. D'après le corollaire II.11, comme $L'_i \in C^f(L') \subseteq C^f(L)$, $L_i = c(L') \in C^f(L)$, $\forall\, i \in \{1,\dots,p\}$ et $\mathcal{L} \subseteq C^f(L) \subseteq C(L)$ qui sont des cônes rationnels clos par union, donc $C^f_\cup(\mathcal{L}) \subseteq C^f(L)$ et

$C_U(\mathcal{L}) \subseteq C(L)$. Il reste à montrer que $L \in C_U^f(\mathcal{L})$. Comme $L' \in C_U^f(\mathcal{L}')$, il existe $L''_1,\ldots,L''_p$ inclus dans $a_1^*\ldots a_m^*$ tels que $L' = \bigcup_{i=1}^{p} L''_i$ et $\forall\ i \in \{1,\ldots,p\}$, $L''_i \in C^f(L'_i)$, ce qui implique $c(L''_i) \in C^f(L_i)$. Donc $L = c(L') = \bigcup_{i=1}^{p} c(L''_i) \in C_U^f(\mathcal{L})$.□

Considérons, maintenant, une famille $\mathcal{L}$ de langages commutatifs et un langage L strictement borné, appartenant à $C_U(\mathcal{L})$. Il existe, alors, une famille finie $\mathcal{L}' \subseteq \mathcal{L}$ telle que $L \in C_U(\mathcal{L}')$ et d'après le corollaire précédent un langage commutatif L' tel que $C_U(\mathcal{L}') = C(L')$. Le corollaire II.11 entraine, alors, $c(L) \in C(L') = C_U(\mathcal{L}') \subseteq C_U(\mathcal{L})$. Si $L \in C_U^f(\mathcal{L})$ on raisonne de la même façon et l'on obtient :

Proposition II.14 : *Pour tout langage L strictement borné et pour toute famille $\mathcal{L}$ de langages commutatifs on a :*

$$\begin{aligned} &i)\quad L \in C_U(\mathcal{L}) \quad ssi \quad c(L) \in C_U(\mathcal{L}),\\ &ii)\quad L \in C_U^{f}(\mathcal{L}) \quad ssi \quad c(L) \in C_U^{f}(\mathcal{L}). \end{aligned}$$

Nous allons maintenant déduire du lemme II.10, un résultat que nous utiliserons au chapitre suivant pour montrer une propriété des langages algébriques commutatifs :

Proposition II.15 : *Soient $T_1,\ldots,T_n$ des alphabets disjoints deux à deux, $L_1,\ldots,L_n$ des langages définis respectivement sur $T_1,\ldots,T_n$, $L = L_1 \ldots L_n$ et $L' = shuf(L_1,\ldots,L_n)$. Pour toute famille $\mathcal{L}$ de langages commutatifs on a :*

$$\begin{aligned} &i)\quad L \in C_U(\mathcal{L}) \quad ssi \quad L' \in C_U(\mathcal{L}),\\ &ii)\quad L \in C_U^{f}(\mathcal{L}) \quad ssi \quad L' \in C_U^{f}(\mathcal{L}). \end{aligned}$$

Démonstration : Si $L \in C_U(\mathcal{L})$, il existe une famille finie $\mathcal{L}' \subseteq \mathcal{L}$ telle que $L \in C_U(\mathcal{L}')$ et un langage commutatif L'_1 tel que $C(L'_1) = C_U(\mathcal{L}')$. D'après le lemme II.10, $L' = shuf(L_1,\ldots,L_n) = \{y \in shuf(x_1,\ldots,x_n)\ /\ x_1\ldots x_n \in L$ et $\forall\ i \in \{1,\ldots,n\},\ x_i \in T_i^*\}$ appartient à $C(L'_1)$ donc à $C_U(\mathcal{L})$.

Réciproquement si $L' \in C_U(\mathcal{L})$, $L = L' \cap T_1^*\ldots T_n^*$ appartient encore à $C_U(\mathcal{L})$.

On raisonnerait de la même façon pour démontrer ii).□

Nous pouvons en déduire :

Proposition II.16 : *Soient $\mathcal{L}$ une famille de langages commutatifs et $\mathcal{L}' = C(\mathcal{L})$ (resp. $C_U(\mathcal{L})$, $C^{f}(\mathcal{L})$, $C_U^{f}(\mathcal{L})$). Alors $\mathcal{L}'$ est clos par produit ssi $\mathcal{L}'$ est clos par intersection.*

Démonstration : Tout cône rationnel (fidèle) clos par intersection est clos par produit, donc, si $\mathcal{L}'$ est clos par intersection, il est clos par produit. Réciproquement, si $\mathcal{L}'$ est clos par produit, il est clos par union. Prenons donc $\mathcal{L}' = \mathcal{C}_{\cup}(\mathcal{L})$ (la démonstration est similaire pour $\mathcal{L}' = \mathcal{C}^{f}_{\cup}(\mathcal{L})$). $\mathcal{L}'$ est clos par intersection ssi il est clos pour l'opération shuffle. Prenons deux langages L_1, $L_2 \in \mathcal{L}'$, définis sur un alphabet T. Considérons un alphabet Δ disjoint de T, et un homomorphisme h strictement alphabétique, défini sur T qui établit une bijection entre T et Δ. Le langage $L_1\ h(L_2)$ appartient à $\mathcal{C}_{\cup}(\mathcal{L})$ et d'après la proposition précédente, $L = \mathrm{shuf}(L_1, h(L_2))$ appartient à $\mathcal{C}_{\cup}(\mathcal{L})$ ainsi que $\mathrm{shuf}\ (L_1, L_2) = g(L)$ où g est l'homomorphisme défini sur $T \cup \Delta$ par $g(a) = a$ si $a \in T$, $h^{-1}(a)$ sinon.□

B - COMPARAISON DE C(Com) AVEC DIVERSES FAMILLES DE LANGAGES

Dans cette section, nous allons d'abord montrer que certains langages "assez bas" dans la hièrarchie des langages algébriques n'appartiennent pas à C(Com). Pour cela, nous établissons un théorème de factorisation qui met en évidence ce qui peut encore commuter dans les langages appartenant à C(Com). Nous en déduisons très aisément que les langages $C_1^* = \{a^n b^n \ / \ n \geq 0\}^*$, D'^*_1, le langage de semi-Dyck sur une lettre et Sym = $\{w \tilde{w}^R \ / \ w \in \{a,b\}^*\}$ n'appartiennent pas à C(Com). Alors, comme Sym est un CIL-langage et que C(Com) est clos par intersection, la proposition I.25 implique Sym $\notin F_\sigma$(Com). En fait, ce résultat peut être aussi obtenu en montrant, pour le langage Sym (proposition II.20), une propriété vérifiée par tout générateur d'un cône rationnel clos par un opérateur syntactique : Pour toute famille de langages L telle que Sym $\in F_\sigma(L)$, il existe L $\in$ L tel que Sym $\in$ C(L).

Nous allons, ensuite, montrer l'inclusion stricte de C(Bor) dans C(Com). Plus précisément, nous établissons l'existence de PSL-langages commutatifs n'appartenant pas à F_σ(Com). Pour obtenir ce résultat, nous démontrons (lemme II.31) que pour tout cône rationnel clos par union, L, $D_1^* \in F(L)$ si et seulement si il existe L $\in$ L tel que $D'^*_1 \subseteq L \subseteq D_1^*$. Ceci va nous permettre, en utilisant un théorème de Durieux, d'établir que D_1^* n'appartient pas à F(Bor) et d'en déduire, à l'aide d'une propriété des langages sans insertion (proposition I.27), que c(Rat) n'est pas inclus dans F_σ(Bor)

Prenons, maintenant, un langage L appartenant à C(Com). Montrons qu'il existe un entier k tel que si $x_1,\ldots,x_k$ sont des facteurs disjoints de longueur supérieure à k d'un mot de L, on peut faire permuter un facteur non vide d'un x_i et un facteur d'un $x_j (i \neq j)$ sans sortir du langage : Plus précisément : nous obtenons :

<u>Proposition II.17</u> : *Soit* $L \subseteq T^*$ *un langage appartenant à* C(Com). *Il existe* $k \in \mathbb{N}_+$ *tel que, pour tout mot* $w \in L$ *et toute factorisation de* w *en* $w_0 \, x_1 \, w_1 \ldots x_{k+1} \, w_{k+1}$ *vérifiant* $\ell(x_i) \geq k$, $\forall \, i \in \{1,\ldots,k+1\}$, *il existe* $s < t \in \{1,\ldots,k+1\}$ *tels que :*

i) $x_s = y'_s \, y_s \, y''_s$, $x_t = y'_t \, y_t \, y''_t$ *avec* $y_s, y_t \neq \varepsilon$,

ii) L *contient* $\{z \, u \, z' \, v \, z'' \ / \ u, v \in \{\varepsilon, y_s, y_t, y_s y_t, y_t y_s\}, \ u v \in \{y_s y_t, y_t y_s\}\}$

où $z = w_0 \ldots w_{s-1} \, y'_s$, $z' = y''_s \, w_s \ldots y'_t$ *et* $z'' = y''_t \, w_{t+1} \ldots w_{k+1}$.

<u>Démonstration</u> : Comme C(Com) = H_α(Com $\wedge$ Rat), il existe un langage commutatif L', un langage rationnel R et un homomorphisme alphabétique g tels que L = g(L' $\cap$ R). Soit M = (Q, Δ, f, q_0, F) un automate d'état fini à k états qui reconnait R. Si $w = w_0 \, x_1 \ldots x_{k+1} \, w_{k+1}$ appartient à L, il existe $w' = w'_0 \, x'_1 \ldots x'_{k+1} \, w'_{k+1} \in L' \cap R$ tels que $g(w'_i) = w_i$, $\forall \, i \in \{0,\ldots,k+1\}$ et $g(x'_j) = x_j$, $\forall \, j \in \{1,\ldots,k+1\}$. Pour tout $i \in \{1,\ldots,k+1\}$, comme $\ell(x_i) \geq k$, x_i peut se factoriser en $x_{i,1} \ldots x_{i,k}$ avec $x_{i,j} \in T^+$ et donc $x'_i = x'_{i,1} \ldots x'_{i,k}$ avec $g(x'_{i,j}) = x_{i,j}$. Pour tout $i, j \in \{1,\ldots,k+1\}$, posons $z'_{i,j} = w'_0 \ldots w'_{i-1} \, x'_{i,1} \ldots x'_{i,j-1}$ et $q_{i,j} = f(q_0, z'_{i,j})$. Comme l'automate M n'a que k états, pour tout $i \in \{1,\ldots,k+1\}$, il existe $j'_i < j''_i \in \{1,\ldots,k+1\}$ tels que $q_{i,j'_i} = q_{i,j''_i}$. Nous en déduisons qu'il existe $s < t \in \{1,\ldots,k+1\}$ tels que

$q_{s,j'_s} = q_{s,j''_s} = q_{t,j'_t} = q_{t,j''_t} = q$. Pour $p \in \{s, t\}$, posons $\bar{y}'_p = x'_{p,1} \dots x'_{p,j'_p - 1}$, $\bar{y}_p = x'_{p,j'_p} \dots x'_{p,j''_p - 1}$ et $\bar{y}''_p = x'_{p,j''_p} \dots x'_{p,k}$ et considérons l'ensemble $A = \{\bar{z}\ u\ \bar{z}'\ v\ \bar{z}'' \ /\ u, v \in \{\varepsilon, \bar{y}_s, \bar{y}_t, \bar{y}_s \bar{y}_t, \bar{y}_t \bar{y}_s\},\ u v \in \{\bar{y}_s \bar{y}_t, \bar{y}_t \bar{y}_s\}\}$ où $\bar{z} = w'_o \dots w'_{s-1} \bar{y}'_s$, $\bar{z}' = \bar{y}''_s w'_s \dots \bar{y}'_t$ et $\bar{z}'' = \bar{y}''_t w'_{t+1} \dots w'_{k+1}$. Comme $f(q_o, \bar{z}) = q = f(q, \bar{z}')$, $f(q, \bar{z}'') \in F$ et $\forall\, u \in \{\bar{y}_s, \bar{y}_t\}$, $f(q, u) = q$, A est inclus dans R. D'autre part, si $u v \in \{\bar{y}_s \bar{y}_t, \bar{y}_t \bar{y}_s\}$, $\bar{z}\ u\ \bar{z}'\ v\ \bar{z}'' \in c(w') \subseteq L'$ qui est commutatif. Donc A est inclus dans $L' \cap R$ et $g(A)$ est inclus dans L. Il est, alors, clair que la proposition est vérifiée en prenant $z = g(\bar{z})$, $z' = g(\bar{z}')$, $z'' = g(\bar{z}'')$ et pour $p \in \{s, t\}$, $y'_p = g(\bar{y}'_p)$, $y_p = g(\bar{y}_p)$, $y''_p = g(\bar{y}''_p)$.□

Cette proposition va nous permettre de montrer très facilement que certains langages "assez bas" dans la hiérarchie des langages algébriques n'appartiennent pas à C(Com).

<u>Corollaire II.18</u> : *Les langages $C_1^* = \{a^n b^n \ /\ n \geq 0\}^*$, D'^*_1 et $Sym = \{w\, \tilde{w}^R \ /\ w \in \{a,b\}^*\}$ n'appartiennent pas à C(Com).*

<u>Démonstration</u> : Prenons $L = C_1^*$ ou D'^*_1 et supposons que L appartienne à C(Com). Soit k, l'entier de la proposition précédente. Posons $w_o = \varepsilon$ et $\forall\, i \in \{1,\dots,k+1\}$ $w_i = b^k$, $x_i = a^k$. Le mot $w = w_o\, x_1 \dots x_{k+1}\, w_{k+1} = (a^k b^k)^{k+1}$ appartient à L et d'après la proposition précédente, il existe $s \in \{1,\dots,k+1\}$ tel que $(a^k b^k)^{s-1}\, a^{k'} b^k$ soit facteur gauche d'un mot de L avec $k' < k$, d'où la contradiction.

Supposons, maintenant, que Sym appartienne à C(Com). Pour tout $i \in \{1,\dots,k+1\}$, posons $x_i = b^k$ et $w_{i-1} = a^k$. Prenons $x = w_o\, x_1 \dots w_k\, x_{k+1}$, $w_{k+1} = \bar{x}^R$. Alors, $w = w_o\, x_1 \dots x_{k+1} w_{k+1} = x\, \bar{x}^R \in Sym$. Si nous avons pris k suffisamment grand, la proposition précédente implique qu'il existe $x' \neq x$ tel que $x'\, \bar{x}^R \in Sym$, d'où la contradiction.□

Le raisonnement utilisé pour montrer que Sym n'appartient pas à C(Com) permet aussi d'établir que le langage (non algébrique) $\{w\, \bar{w} \ /\ w \in \{a,b\}^*\} \notin C(Com)$ qui n'est donc pas clos par duplication.

Comme C(Bor) est inclus dans C(Com), le corollaire précédent nous permet de retrouver un théorème de Durieux qui étendait et précisait des résultats de Greibach [39] et Golstine [36].

<u>Corollaire II. 19</u> [18] *La famille des langages algébriques linéaires n'est pas incluse dans le cône rationnel engendré par les langages bornés.*

En particulier, aucun générateur du cône rationnel des langages algébriques n'est commutatif. Ce fait est établi par Beauquier dans [4] d'une toute autre manière en étudiant la structure des générateurs algébriques. Steyaert obtient ce résultat [77] en examinant le taux de croissance de l'index rationnel des langages algébriques commutatifs (cf. [12]).

Le langage C_1 est un langage borné, donc $C_1 \in \mathcal{C}(\mathrm{Com})$ et $C_1^* \in \mathcal{F}(\mathrm{Com})$. Par contre, on peut montrer que Sym n'appartient pas à $\mathcal{F}_\sigma(\mathrm{Com})$. En effet, Sym est un CIL-langage qui n'appartient pas à $\mathcal{C}(\mathrm{Com})$ qui est clos par intersection, ce qui entraine, d'après la proposition I.25, Sym $\notin \mathcal{F}_\sigma(\mathrm{Com})$. Ce résultat peut être obtenu d'une autre façon en montrant pour la famille Lin une propriété vérifiée par tout cône rationnel principal clos par substitution [40]. Nous avons, en effet :

Proposition II.20 : *Pour toute famille de langage $\mathcal{L}$ telle que $\mathcal{L}in \subseteq \mathcal{F}_\sigma(\mathcal{L})$, il existe un langage $L \in \mathcal{L}$ vérifiant $\mathcal{L}in \subseteq \mathcal{C}(L)$.*

La démonstration de cette proposition repose sur les deux lemmes suivants. Dans [42], Greibach montre que si L_1 et L_2 sont des langages algébriques linéaires vérifiant $L_1 \cup L_2 = L = \{w\, c\, w^R \;/\; w \in \{a,b\}^*\}$, alors L appartient à $\mathcal{C}(L_1) \cup \mathcal{C}(L_2)$. En fait, l'hypothèse de linéarité des langages L_1 et L_2 n'est pas utile. Plus généralement, on peut montrer :

Lemme II.21 : *Soient L_1 et L_2 deux langages tels que $\mathcal{C}(L_1 \cup L_2)$ contienne $\mathcal{L}in$ la famille des langages algébriques linéaires. Alors, soit $\mathcal{L}in \subseteq \mathcal{C}(L_1)$, soit $\mathcal{L}in \subseteq \mathcal{C}(L_2)$.*

Démonstration : Si $\mathcal{L}in$ est inclus dans $\mathcal{C}(L_1 \cup L_2)$, le langage $\mathrm{Sym}' = \{w\, \bar{w}^R \;/\; w \in \{a,b,c\}^*\}$ appartient à $\mathcal{C}(L_1 \cup L_2)$ et d'après le théorème de Nivat, il existe un langage R et deux homomorphismes h et g tels que $\mathrm{Sym}' = h(g^{-1}(L_1 \cup L_2) \cap R)$. Donc $\mathrm{Sym}' = L'_1 \cup L'_2$ avec $L'_1 = h(g^{-1}(L_1) \cap R) \in \mathcal{C}(L_1)$ et $L'_2 = h(g^{-1}(L_2) \cap R) \in \mathcal{C}(L_2)$. Posons $T = \{a,b,c\}$ et pour $i \in \{1,2\}$, $A_i = \{w \in T^* \;/\; w\, \bar{w}^R \in L'_i\}$. Distinguons deux cas :

i) Si $\mathrm{Init}\,(A_1) = T^*$, alors, $\forall\, x \in \{a,b\}^*$, il existe $y \in T^*$ tel que $x\, c\, y\, \bar{y}^R\, \bar{c}\, \bar{x}^R \in L'_1$. Comme on peut effaçer par une transduction rationnelle dans tout mot de L'_1 le facteur qui débute par la première occurrence de c et se termine par la dernière occurrence de $\bar{c}$, il est clair que $\mathrm{Sym} \in \mathcal{C}(L'_1)$ et donc $\mathcal{L}in = \mathcal{C}(\mathrm{Sym}) \subseteq \mathcal{C}(L'_1) \subseteq \mathcal{C}(L_1)$.

ii) Si $\mathrm{Init}\,(A_1) \subsetneq T^*$, il existe $x_1 \in T^*$ tel que $x_1\, T^* \cap A_1 = \emptyset$, ce qui entraine $x_1\, T^* \subseteq A_2$. Alors, $\forall\, w \in T^*$, $x_1\, w\, \bar{w}^R\, \bar{x}_1^R \in L'_2$ et il et il est clair que $\mathrm{Sym}' \in \mathcal{C}(L_2)$, d'où le résultat.□

Remarquons que, d'après un résultat de Ginsburg et Spanier [34], le lemme précédent n'est plus vérifié si on remplace la famille $\mathcal{L}in$ par $\mathcal{C}(C_1)$, le cône rationnel engendré par $C_1 = \{a^n b^n \;/\; n \geq 0\}$.

Lemme II.22 : *Soient $\mathcal{L}_1$ et $\mathcal{L}_2$ deux cônes rationnels clos par union tels que $\mathcal{L}_1 \,\square\, \mathcal{L}_2$ contienne la famille $\mathcal{L}in$. Alors, soit $\mathcal{L}in \subseteq \mathcal{L}_1$, soit $\mathcal{L}in \subseteq \mathcal{L}_2$.*

Démonstration : Si $\mathcal{L}in$ est inclus dans $\mathcal{L}_1 \,\square\, \mathcal{L}_2$, il existe $L \in \mathcal{L}_1$, $L \subseteq Z^*$, et une $\mathcal{L}_2$-substitution s tels que $\mathrm{Sym}' = s(L)$. Posons $Z' = \{x \in Z \;/\; s(x) \text{ est infini}\}$ et $Z'' = Z \setminus Z'$. Comme $\mathcal{L}_1$ et $\mathcal{L}_2$ sont des cônes rationnels, on peut supposer que $L \cap Z^* x Z^* \neq \emptyset$, $\forall\, x \in Z$ et $s(x) = \{x\}$, $\forall\, x \in Z''$. D'autre part, comme

Sym' est un CIL-langage et que $s(x) \neq \emptyset$, $\forall x \in Z$, L est inclus dans $Z''^* \cup Z''^* Z' Z''^*$. Pour tout $x \in Z'$ notons L_x, le langage $L \cap Z''^* x Z''^*$.

Prenons $x \in Z'$. Comme $L_x \neq \emptyset$, il existe u, v tels que $u x v \in L_x$ et comme s(x) est infini, $s(x) \cap T^+ \bar{T}^+ \neq \emptyset$ et $u \in T^*$, $v \in \bar{T}^*$ avec $T = \{a,b,c\}$ et $\bar{T} = \{\bar{a},\bar{b},\bar{c}\}$. Etant donné que $s(u x v) = u\, s(x)\, v$ est inclus dans Sym', l'une des deux hypothèses suivantes est nécessairement vérifiée :

i) Il existe $z \in T^*$ tel que $u = u'z$ avec $v = \bar{u}'^R$ et donc $s(x) \subseteq T^* \bar{T}^* \bar{z}^R$. Alors définissons la substitution s' et l'homomorphisme h en x par $h(x) = x\, \bar{z}^R$ et $s'(x) = s(x)/\, \bar{z}^R = \{y \,/\, y\, \bar{z}^R \in s(x)\}$. Comme $u'\, z\, s'(x)\, \bar{z}^R\, \bar{u}'^R \subseteq$ Sym', s'(x) est inclus dans Sym'.

ii) Il existe $z \in T^*$ tel que $v = \bar{z}^R\, \bar{u}^R$ et donc $s(x) \subseteq z\, T^* \bar{T}^*$. Posons, alors, $h(x) = z\, x$ et $s'(x) = \{y \,/\, z\, y \in s(x)\}$. Il est clair que $s'(x) \subseteq$ Sym' et $s'(x) \in C(s(x)) \subseteq \mathcal{L}_2$.

Achevons de définir h et s' sur Z en posant $h(x) = x$ et $s'(x) = \{x\}$, $\forall x \in Z''$. Alors $L' = h(L) \in \mathcal{L}_1$ et $s'(L') = s(L) =$ Sym'. Prenons $d \notin T \cup \bar{T}$, posons $L'_1 = \{u\, d\, v \,/\, u \in T^*, v \in \bar{T}^*, \exists\, x \in Z'$ tel que $L' \cap \{u v, u x v\} \neq \emptyset\}$, $L'_2 = \{\varepsilon\} \cup (\bigcup_{x \in Z'} s'(x))$ et considérons la substitution s'' définie sur $T \cup \bar{T} \cup \{d\}$ par $s''(d) = L'_2$ et $s''(x) = \{x\}$, $\forall x \in T \cup \bar{T}$. Par construction, si $u\, d\, v \in L'_1$, $v = \bar{u}^R$, $L'_2 \subseteq$ Sym' et Sym' $= s'(L') \subseteq s''(L'_1) \subseteq$ Sym'. D'autre part, comme $\mathcal{L}_1$ et $\mathcal{L}_2$ sont clos par union, $L'_1 \in \mathcal{L}_1$ et $L'_2 \in \mathcal{L}_2$. L'égalité entre Sym' et $s''(L'_1)$ implique que $A_1 A_2 = T^*$ avec $A_1 = \{w \in T^* \,/\, w\, d\, \bar{w}^R \in L'_1\}$ et $A_2 = \{w \in T^* \,/\, w\, \bar{w}^R \in L'_2\}$. Nous pouvons alors raisonner comme au lemme précédent. Si $\mathrm{Init}(A_1)$ est égal à T^*, $\mathcal{L}in \subseteq C(L'_1) \subseteq \mathcal{L}_1$. Sinon, il existe $z \in T^*$ tel que $z\, T^* \cap A_1 = \emptyset$. Alors, $\forall\, w \in T^*$, il existe une factorisation de z en $z'\, z''$ avec $z' \in A_1$, $z''\, w \in A_2$. Donc pour tout mot $w\, \bar{w}^R \in$ Sym', il existe un facteur droit z'' de z tel que $z''\, w\, \bar{w}^R\, \bar{z}''^R \in L'_2$ ce qui entraine Sym' $\in C(L'_2) \subseteq \mathcal{L}_2$ et $\mathcal{L}in \subseteq \mathcal{L}_2$.□

Considérons, maintenant, une famille de langages $\mathcal{L}$ telle que $\mathcal{L}in \subseteq F_\sigma(\mathcal{L})$, posons $\mathcal{L}'_1 = C_\cup(\mathcal{L}) = \mathcal{L}'$ et définissons pour tout $i \in \mathbb{N}_+$, $\mathcal{L}'_i$ par la relation $\mathcal{L}'_{i+1} = \mathcal{L}'_i \,\square\, \mathcal{L}'$. Pour tout $i \in \mathbb{N}_+$, $\mathcal{L}'_i$ est un cône rationnel clos par union et comme Sym $\in F_\sigma(\mathcal{L}) = \bigcup_{i \geq 1} \mathcal{L}'_i$, il existe i tel que Sym $\in \mathcal{L}'_i$ et $\mathcal{L}in \subseteq \mathcal{L}'_i$. Le lemme précédent entraine, alors, par induction $\mathcal{L}in \subseteq \mathcal{L}'$.

Alors Sym $\in \mathcal{L}' = C_\cup(\mathcal{L})$ et il existe des langages $L_1,\ldots,L_n \in \mathcal{L}$ tels que Sym $\in C_\cup(\{L_1,\ldots,L_n\})$, donc $\mathcal{L}in \subseteq C(\bar{L}_1 \cup \ldots \cup \bar{L}_n)$ où pour tout $i \in \{1,\ldots,n\}$, $\bar{L}_i$ est une recopie de L_i sur un alphabet disjoint des autres. Alors, le lemme II.21 permet de montrer par induction qu'il existe $i \in \{1,\ldots,n\}$ tel que $\mathcal{L}in \subseteq C(\bar{L}_i) = C(L_i)$ avec $L_i \in \mathcal{L}$ ce qui achève la démonstration de la proposition II.20.□

Cette proposition peut s'écrire autrement :

<u>Corollaire II.23</u> : *La famille* $\mathrm{Ext}(\mathcal{L}in) = \{L \,/\, \mathcal{L}in \not\subseteq C(L)\}$ *est une FAL close par substitution.*

Si nous nous restreignons aux langages appartenant à $\mathcal{L}in$ nous obtenons une propriété qui est aussi vérifiée par tout cône rationnel principal clos par produit :

Corollaire II.24 : *La famille* $Ng(\mathcal{L}in) = \mathcal{L}in \cap Ext(\mathcal{L}in)$ *est un cône rationnel clos par union.*

Du corollaire II.18 et de la proposition II.20 découle immédiatement :

Proposition II.25 : *La famille des langages algébriques linéaires n'est pas incluse dans la plus petite FAL contenant les langages commutatifs et close par substitution.*

La proposition II.17 n'est plus vérifiée si on remplace $C(Com)$ par $F(Com)$. En effet, le langage C_1^* appartient à $F(Com)$ sans vérifier les propriétés de cette proposition. Nous allons, cependant, pouvoir montrer par un raisonnement analogue que $D'^*_1 \notin F(Com)$. Pour obtenir ce résultat, établissons, d'abord, que pour tout cône rationnel $\mathcal{L}$ clos par union, $D'^*_1 \in F(\mathcal{L})$ implique l'existence dans $\mathcal{L}$ d'un langage "assez proche" de D'^*_1. Précisons cela en définissant pour tout m, n $\in \mathbb{N}$, les langages $L_{m,n}$.

Pour $w \in \{a,b\}^*$, posons $d(w) = \ell_a(w) - \ell_b(w)$ où $\ell_a(w)$ désigne le nombre d'occurrences de la lettre a dans w, et $\bar{d}(w) = \inf \{d(x) \ / \ \exists\, y$ tel que $w = xy\}$.

Pour m, n $\in \mathbb{N}$, notons : $L_{m,n} = \{w \in \{a,b\}^* \ / \ d(w) = m-n,\ \bar{d}(w) \geq -n\}$.

Les langages $L_{m,n}$ sont rationnellement équivalents à $D'^*_1 = L_{o,o}$, le semi-Dyck sur une lettre.

Lemme II.26 : *Soit* $\mathcal{L}$ *un cône rationnel fermé par union. Si* D'^*_1 *appartient à* $F(\mathcal{L})$, *il existe* $k \in \mathbb{N}$ *et* $L \in \mathcal{L}$ *tels que* $D'^*_1 \subseteq L \subseteq L_{k,k}$.

Démonstration : Il existe un rationnel $R \subseteq \{a_1,\dots,a_p\}^*$, des langages non vides $L_1,\dots,L_p$ de $\mathcal{L}$ et une substitution s définie par $s(a_i) = L_i$, $\forall\, i \in \{1,\dots,p\}$ tels que $D'^*_1 = s(R)$. Prenons un mot u x v de R, $z \in s(u)$, $z' \in s(v)$ et posons $n = d(z)$, $m = -d(z')$. Pour tout $y \in s(x)$, $z\,y\,z' \in D'^*_1$, donc $\bar{d}(zy) \geq \bar{d}(zyz') \geq 0$ et $d(zyz') = 0$. Nous en déduisons que $n = d(z)$ et $m = d(zy)$ qui sont supérieurs à $\bar{d}(zy)$, appartiennent à $\mathbb{N}$. De plus, $\bar{d}(zy) \geq 0$ implique $\bar{d}(y) \geq -d(z) = -n$ et comme $d(y) = -d(zz') = m-n$, $s(x)$ est inclus dans $L_{m,n}$. On peut supposer, sans nuire à la généralité, que chaque a_i occurre au moins dans un mot de R, donc on en déduit que pour tout $i \in \{1,\dots,p\}$, il existe m_i, $n_i \in \mathbb{N}$ tels que $L_i = s(a_i) \subseteq L_{m_i,n_i}$.

Posons $\mu = \sup \{m_i \ / \ i \in \{1,\dots,p\}\}$ et $\gamma = \sup \{n_i \ / \ i \in \{1,\dots,p\}\}$. Comme $u\,x\,v \in R$ qui est rationnel, il existe $t \in \mathbb{N}$, u', u" tels que $\ell(u')$, $\ell(u'') \leq t$ et u' x v, u" v $\in$ R. Il existe m', n' $\in \mathbb{N}$ tels que

$s(u') \subseteq L_{m',n'}$. Tout mot $y' \in s(u')$ est un facteur gauche d'un mot de D'^*_1 donc $\bar{d}(y) \geq 0$ et l'on peut prendre $n' = 0$. En outre $m' = d(y') \leq \sum_{i=1}^{\ell(u')} \mu \leq \sum_{j=1}^{t} \mu = t\mu$. De même, $s(u")$ est inclus dans $L_{m",o}$ avec $m" \leq t\mu$. Comme $s(u")\, s(v)$ est inclus dans D'^*_1, nous en déduisons $s(v) \subseteq L_{o,m"}$ et $s(u')\, s(x)\, s(v) \subseteq D'^*_1$ implique $s(x) \subseteq L_{m",m'}$. Nous en déduisons la propriété :

(*) Il existe $k' = \mu t + 1$ tel que $u\, x\, v \in R$ implique $s(x) \subseteq L_{m,n}$ avec $0 \leq m,n < k'$.

Posons $B = \{(q, q', i) \,/\, 0 \leq q, q' \leq k', i \in \{1,\dots,p\}$ et $q - q' = m_i - n_i\}$ et pour $\rho = (q, q', i) \in B$, $A_\rho = \{w \,/\, a^q\, w b^{q'} \in L_i\}$.

Nous allons montrer que $D'^*_1 \subseteq L \subseteq L_{k,k}$ avec $k = k' + \gamma$ et $L = \bigcup_{\rho \in B} A_\rho$.

Prenons $w \in D'^*_1$, $a^{k'}\, wb^{k'} \in D'^*_1$ et d'après la propriété (*), il existe pas de mots x_1, x_2 tels que $x_1\, x_2 \in R$ et $a^{k'}\, w_1 \in s(x_1)$ avec w_1 facteur gauche de w. Et comme $a^{k'}\, wb^{k'} \in s(R)$, il existe $q, q' \leq k'$, $i \in \{1,\dots,p\}$ tels que $a^q\, wb^{q'} \in s(a_i) = L_i \subseteq L_{m_i,n_i}$. $d(w) = 0$, donc $q - q' = m_i - n_i$, $\rho = (q, q', i) \in B$ et $w \in A_\rho \subseteq L$.

Prenons, maintenant, $w' \in A_\rho$ avec $\rho = (q, q', i) \in B$. Donc $a^q\, w' b^{q'} \in L_i \subseteq L_{m_i,n_i}$ et comme $q - q' = m_i - n_i$, $d(w') = 0$. De plus $\bar{d}(a^q\, w') \geq -n_i$ implique $\bar{d}(w') \geq -n_i - q \geq -k$, donc $w' \in L_{k,k}$.

Il nous reste à montrer que $L \in \mathcal{L}$, ce qui vient du fait que pour $\rho = (q, q', i) \in B$, $A_\rho \in C(L_i) \subseteq \mathcal{L}$. Et $L = \bigcup_{\rho \in B} A_\rho \in \mathcal{L}$ car B est fini et $\mathcal{L}$ est clos par union.□

Lemme II.27 : *Soit* L *un langage de* $C(Com)$. *Si* D'^*_1 *est inclus dans* L, *alors, pour tout* $m \in \mathbb{N}$, $L \setminus L_{m,n} \neq \emptyset$.

Démonstration : Si $L \in C(Com)$, il existe un langage commutatif L' et un langage rationnel R inclus dans T^* ainsi qu'un homomorphisme alphabétique h de T^* dans $\{a,b\}^*$ tels que $L = h(L' \cap R)$. Soit $M = (Q, T, f, p_o, F)$ un automate d'état fini qui reconnait R. Prenons $k = m + 1$ et n supérieur au nombre d'états de M. Alors $w = (a^{nk}\, b^{nk})^n \in D'^*_1 \subseteq L$, donc il existe $z = x_1\, y_1 \dots x_n\, y_n \in L' \cap R$ tel que $h(x_i) = a^{nk}$ et $h(y_i) = b^{nk}$ pour tout $i \in \{1,\dots,n\}$. Posons pour tout $j \in \{0,\dots,n-1\}$, $z_{j+1} = x_1\, y_1 \dots x_j\, y_j$ et $p_j = f(p_o, z_{j+1})$. Comme $h(x_j) = a^{nk}$, x_j peut se factoriser en $u_j\, x'_j\, v_j$ avec $p'_j = f(p_j, u_j) = f(p'_j, x'_j)$ et $h(x'_j) \in a^k\, a^*$. De plus, il existe $t < \ell$ tels que $p'_t = p'_\ell$, ce qui entraine que $z' = z_t\, u_t\, v_t\, y_t \dots u_\ell\, x'_\ell\, x'_t\, v_\ell\, y_\ell \dots x_n\, y_n \in R$. Comme L' est commutatif, z' appartient aussi à L' et $h(z') \in L$. Par contre $y = h(z')$ n'appartient pas à $L_{m,m}$ car $\bar{d}(y) = d(h(z_t\, u_t\, v_t\, y_t)) \leq -k < -m$.□

Nous pouvons en déduire immédiatement :

Proposition II.28 : *Le langage de semi-Dyck sur une lettre,* D'^*_1 *n'appartient pas à* $F(Com)$, *la plus petite FAL contenant tous les langages commutatifs.*

Comme D_1^*, le langage de Dyck sur une lettre est un langage commutatif, nous retrouvons un résultat dû à Boasson :

Corollaire II.29 : [9] *Le langage de semi-Dyck sur une lettre,* D'^*_1 *n'appartient pas à* $F(D_1^*)$.

Comparons, maintenant, les familles engendrées à partir de langages bornés et de langages commutatifs. Dans la première section de ce chapitre, il était montré l'égalité entre C(Com) et $C_\cap$(Bor). Montrons que l'intersection est nécessaire, c'est-à-dire que C(Bor) $\subsetneq$ C(Com).

Plus précisément, nous allons établir que c(Rat) n'est pas inclus dans F_σ(Bor), la plus petite FAL contenant les langages bornés et close par substitution. Pour obtenir ce résultat, nous utiliserons la proposition suivante qui est due à J.L. Durieux et qui concerne l'image par une transduction rationnelle des mots définis sur un alphabet d'une seule lettre :

Proposition II.30 : [18] *Soient s une transduction rationnelle de* X^* *vers* Y^*, *x une lettre de X, u et v deux mots de* YY^*. *Alors, il existe deux entiers positifs k, t ne dépendant que de s et des longueurs de u et v et vérifiant :*

Si $\alpha u^p \beta v^q \gamma \in s(x^n)$ *et si* $p, q \geq k$, *alors il existe* $i, j \in \mathbb{N}$ *avec* $i + j = t$ *tels que :*

1) $\alpha u^{p-i} \beta v^{q+j} \gamma \in s(x^n)$ *et* $\alpha u^{p+i} \beta v^{q-j} \gamma \in s(x^n)$

2) *si* $i = 0$ *alors* $j = t$ *et* $\alpha u^p \beta v^{q-j} (v^j)^* \gamma \subseteq s(x^n)$

3) *si* $j = 0$ *alors* $i = t$ *et* $\alpha u^{p-i} (u^i)^* \beta v^q \gamma \subseteq s(x^n)$.

Cette proposition va nous servir à montrer que D_1^* n'appartient pas à $F(\mathcal{B})$. Nous utiliserons, aussi, un autre résultat similaire au lemme II.26 :

Lemme II.31 : *Soit* $\mathcal{L}$ *un cône rationnel clos par union. Alors* D_1^* *appartient à* $F(\mathcal{L})$ *si et seulement si il existe dans* $\mathcal{L}$ *un langage L tel que* $D'^*_1 \subseteq L \subseteq D_1^*$.

Démonstration : Considérons l'homomorphisme h défini sur $\{a,b\}^*$ par h(a) = b et h(b) = a. si $D'^*_1 \subseteq L \subseteq D_1^*$, nous avons $(D'^*_1\, h(D'^*_1))^* \subseteq (L\, h(L))^* \subseteq (D_1^*\, h(D_1^*))^*$ et comme, clairement, $(D_1^*\, h(D_1^*))^*$ et $D'^*_1\, h(D'^*_1))^*$ sont égaux à D_1^*, $(L\, h(L))^* = D_1^*$. Donc, $L \in \mathcal{L}$ implique $D_1^* = (L\, h(L))^* \in F(\mathcal{L})$.

Réciproquement si $D_1^* \in F(\mathcal{L})$, il existe un langage rationnel $R \subseteq T^*$ et une $\mathcal{L}$-substitution s tels que $s(R) = D_1^*$. On peut supposer que $\forall\, d \in T$, $s(d) \neq \emptyset$ et que la lettre d occurre au moins dans un mot de R. Pour tout $i \in \mathbb{Z}$, notons $L_i = \{w \in \{a,b\}^* \ /\ \ell_a(w) - \ell_b(w) = i\}$. Il est clair que pour tout $d \in T$, il existe $i \in \mathbb{Z}$ tel que $s(d) \subseteq L_i$. De plus, si x est un facteur d'un mot de R, il existe x', x" de longueur inférieure au nombre d'états de l'automate qui reconnait R, tels que x' x x" $\in$ R. Donc il existe un entier positif k vérifiant :

(*) Si $u\,x\,v \in R$, $s(x) \subseteq L_j$ avec $-k \leq j \leq k$.

Notons, alors, $B = \{(q, t, d) \,/\, 0 \leq q,t \leq 2k+1,\ d \in T$ et $s(d) \subseteq L_i$ avec $i = q - t\}$ et pour $p = (q, t, d) \in B$, posons $A_p = \{w \,/\, a^q\, w\, b^t \in s(d)\}$. Montrons maintenant, que $L = \bigcup_{p\in B} A_p$ vérifie les propriétés du lemme.

Prenons d'abord, $w \in A_p$ avec $p = (q, t, d) \in B$. Alors $a^q\, w\, b^t \in s(d)$ donc $q + \ell_a(w) - \ell_b(w) - t = q - t$ ce qui implique $\ell_a(w) - \ell_b(w) = 0$ et $w \in D_1^*$.

Prenons maintenant, $w \in D'^*_1$. Alors $y = a^{2k+1}\, w\, b^{2k+1}$ appartient à D_1^* et il existe $x \in R$ tel que $y = s(x)$. Il existe une factorisation de x en $x'\, d\, x''$ avec $d \in T$ telle que $a^i \in s(x')$, $a^j\, y' \in s(d)$ avec $j = 2k + 1 - i$, $y'' \in s(x'')$ et $y'\, y'' = w\, b^{2k+1}$. D'après la propriété (*), i est inférieur à $k + 1$, donc $j = 2k + 1 - i$ est supérieur à k et comme $w \in D'^*_1$, pour tout facteur gauche w' de w, on a $\ell_a(w') - \ell_b(w') \geq 0$ et $\ell_a(a^j\, w') - \ell_b(a^j\, w') \geq j > k$ ce qui implique $a^j\, w' \notin s(d)$. Nous en déduisons que $y' = w\, b^m$, donc $a^j\, w\, b^m \in s(d) \subseteq L_i$ avec $i = j - m$. Par construction, nous avons $0 \leq j, m \leq 2k + 1$, donc $p = (j, m, d) \in B$ et $w \in A_p \subseteq L$.

Enfin, il est clair que pour tout $p = (q, t, d) \in B$, A_p appartient à $C(s(d)) \subseteq \mathcal{L}$ et comme B est fini, $L = \bigcup_{p\in B} A_p$ appartient à $\mathcal{L}$ qui est clos par union.□

Nous pouvons maintenant établir :

Proposition II.32 : *Le langage D_1^* n'appartient pas à $F(\mathrm{Bor})$, la plus petite FAL contenant les langages bornés.*

Démonstration : D'après le lemme précédent, l'appartenance de D_1^* à $F(\mathrm{Bor})$ implique l'existence d'un langage $L \in C(\mathrm{Bor})$ tel que $D'^*_1 \subseteq L \subseteq D_1^*$. Il existe alors, un alphabet $T = \{a_1,\ldots,a_n\}$, un langage $L' \subseteq a_1^* \ldots a_n^*$ et une transduction rationnelle s telle que $\hat{s} = \{(g(w), h(w)) \,/\, w \in R\}$ et $L = s(L')$. Considérons un automate d'état fini $M = (Q, Y, f, q_o, F)$ qui reconnait R. Pour tout $q, q' \in Q$, posons $R_{q,q'} = \{w \in Y^* \,/\, f(q,w) = q'\}$ et $\hat{s}_{q,q'} = \{(g(w), h(w)) \,/\, w \in R_{q,q'}\}$. A chaque transduction rationnelle $s_{q,q'}$ et pour $\ell(u) = \ell(v) = 1$, la proposition II.30, fait correspondre un entier positif k. Prenons un entier p supérieur à chacun des k ainsi définis. Alors $z = (a^p b^p)^{n+1} \in D'^*_1$ et il existe $y \in L'$ tel que $z \in s(y)$. Il est clair qu'il existe une factorisation de y en $y_1\, y_2\, y_3$ et une factorisation de z en $z_1\, z_2\, z_3$ de telle façon que $y_2 \in a_j^*$, $z_1 \in s_{q_o,q}(y_1)$ $z_2 = z'_2\, a^p\, b^p\, z''_2 \in s_{q,q'}(y_2)$, $z_3 \in s_{q',q''}(y_3)$ avec $q,q' \in Q$ et $q'' \in F$. La proposition III.30 implique alors qu'il existe deux entiers $i_1, i_2 \geq 0$ tels que $i_1 + i_2 > 0$ et $\bar{z}_2 = z'_2\, a^{p+i_1}\, b^{p-i_2}\, z''_2 \in s_{q,q'}(y_2)$ ce qui entraine $z_1\, \bar{z}_2\, z_3 \in L \subseteq D_1^*$, d'où la contradiction puisque $\ell_a(z_1\bar{z}_2z_3) - \ell_b(z_1\bar{z}_2z_3) = i_1 + i_2 > 0$.□

Considérons, maintenant, le langage $L = c((abcd)^*)$. Supposons que L appartienne à $F_\sigma(Bor)$. Alors $L \cap \{a,b\}^* c^* d^* = \{x\, c^n d^n \;/\; x \in D_1^*,\ n = \ell(x)\}$ est un langage sans insertion qui appartient à $F_\sigma(Bor)$, donc à $C(Bor)$ (cf. proposition I.27), d'où la contradiction d'après la proposition précédente et nous pouvons énoncer :

Corollaire II.33 : *La famille $c(Rat)$ n'est pas incluse dans la plus petite FAL close par substitution et contenant les langages bornés.*

C - Effacement et langages commutatifs

Les transductions rationnelles peuvent faire correspondre à un langage récursif, un langage non récursif. Ainsi tout langage récursivement énumérable est l'image par une transduction rationnelle du langage shuf(Sym, $\overline{\text{Sym}}$) où $\overline{\text{Sym}}$ est une recopie de Sym = $\{w\,\bar{w}^R \ / \ w \in \{a,b\}^*\}$ sur un alphabet disjoint [3]. Par contre, pour les transductions rationnelles fidèles qui ne peuvent effacer qu'un nombre borné de lettres consécutives, l'image d'un langage récursif est un langage récursif. Ces deux notions sont, donc, assez différentes et il est intéressant d'examiner les langages effaçables, c'est-à-dire les langages L qui vérifient l'égalité entre $C(L)$ et $C^f(L)$, le plus petit cône rationnel fidèle contenant L. C'est ce que nous faisons dans cette section où nous donnons une condition nécessaire et suffisante pour qu'un langage commutatif soit effaçable (proposition II.42). Cette condition fait intervenir la notion de quotient rationnel (quand nous ne précisons pas, il s'agira du quotient rationnel à droite). Nous sommes donc amenés à étudier les cônes rationnels fidèles qui sont clos par quotient rationnel et nous exhibons un langage $L \subseteq a^*b^*c^*$, non effaçable, qui engendre, cependant, un cône rationnel fidèle clos par quotient rationnel à droite et à gauche. Nous apportons, ainsi, une réponse à une question de Ginsburg ([25] page 34). Par contre, nous démontrons qu'un langage $L \subseteq a^*b^*$ est effaçable si et seulement si $C^f(L)$ est clos par quotient rationnel à droite et à gauche et nous en déduisons que tout langage algébrique inclus dans a^*b^* est effaçable. Nous montrons, ensuite, le résultat principal de cette section :

Un langage commutatif L est effaçable si et seulement si pour tout langage rationnel R, L/R appartient à $C^f(L)$.

Nous utiliserons cette propriété au chapitre suivant pour montrer que tout PSL-langage commutatif est effaçable. D'autre part, elle permet de retrouver immédiatement un résultat de Ullian [79] : tout langage $L \subseteq a^*$ est effaçable. Enfin, nous en déduirons une caractérisation des langages commutatifs effaçables inclus dans D_1^* (proposition II.49).

Pour toute famille de langage $\mathcal{L}$, nous noterons

$$\mathcal{L}/\text{Rat} = \{L' \ / \ L' = L/R \text{ avec } L \in \mathcal{L} \text{ et } R \in \text{Rat}\}.$$

Nous dirons que $\mathcal{L}$ est clos par quotient rationnel si $\mathcal{L}$ contient $\mathcal{L}/\text{Rat}$ ce qui implique, en fait, $\mathcal{L} = \mathcal{L}/\text{Rat}$.

Tout cône rationnel est clos par quotient rationnel (cf. [25]). Par contre, ce n'est pas toujours vrai pour un cône rationnel fidèle. Nous allons donner, maintenant, une condition pour que le plus petit cône rationnel fidèle clos par union contenant une famille de langage $\mathcal{L}$ soit clos par quotient rationnel :

Lemme II.34 : *Pour toute famille de langages L, $C^f_U(L)$ est clos par quotient rationnel si et seulement si L/Rat est inclus dans $C^f_U(L)$.*

Démonstration : Comme $L/\mathrm{Rat} \subseteq C^f_U(L)$ / Rat, $C^f_U(L)$ clos par quotient rationnel implique $L/\mathrm{Rat} \subseteq C^f_U(L)$.

Réciproquement, prenons $L' \in L$, $L'' \in C^f(L')$, $R' \in$ Rat et montrons que L''/R' appartient à $C^f_U(L)$. Comme $L'' \in C^f(L')$, il existe un alphabet Δ, un langage rationnel $R \subseteq \Delta^*$ et deux homomorphismes alphabétiques g et h tels que $L'' = g(B)$ avec $B = h^{-1}(L') \cap R$ et g limité sur R. L'homomorphisme g étant alphabétique $g(B) / R' = g(B / g^{-1}(R'))$. Posons $R'' = g^{-1}(R')$ et considérons l'automate d'état fini $M = (Q, \Delta, f, q_o, F)$ qui reconnait R. Pour tout $q \in Q$, notons $R_q = \{w \in \Delta^* \ / \ f(q_o, w) = q\}$ et $R_{\bar{q}} = \{w \in \Delta^* \ / \ f(q,w) \in F\}$. Montrons que B/R'' est égal à $\bigcup_{q\in H} (h^{-1}(L_q) \cap R_q)$ où $H = \{q \in Q \ / \ R_{\bar{q}} \neq \emptyset\}$ et $L_q = L' \ / \ h(R'' \cap R_{\bar{q}})$.

Prenons $x \in B/R''$. Il existe, alors, $y \in R''$ et $q \in Q$ tels que $xy \in B = h^{-1}(L') \cap R$ et $x \in R_q$. Comme $xy \in R$ avec $x \in R_q$, $y \in R_{\bar{q}}$ et $q \in H$. De plus, $h(x)\ h(y) = h(xy) \in L'$ avec $y \in R'' \cap R_{\bar{q}}$ ce qui implique $h(x) \in L' \ / \ h(R'' \cap R_{\bar{q}}) = L_q$ et donc $x \in h^{-1}(L_q) \cap R_q$ avec $q \in H$.

Inversement, si $x \in h^{-1}(L_q) \cap R_q$ avec $q \in H$, $h(x) \in L_q = L' \ / \ h(R'' \cap R_{\bar{q}})$ ce qui entraine qu'il existe $y \in R'' \cap R_{\bar{q}}$ tel que $h(x)\ h(y) = h(xy) \in L'$. Comme $x \in R_q$ et $y \in R_{\bar{q}}$, $xy \in R$ et $xy \in h^{-1}(L') \cap R = B$ avec $y \in R''$ donc $x \in B/R''$.

Nous avons alors, $L'' \ / \ R' = g(B/R'') = \bigcup_{q\in H} g(h^{-1}(L_q) \cap R_q)$. Par hypothèse, $L_q = L'/h(R'' \cap R_{\bar{q}}) \in L$/Rat, qui est inclus dans $C^f_U(L)$. D'autre part, pour tout $q \in H$, l'homomorphisme g est limité sur R_q. En effet, dans le cas contraire, comme $q \in H$, il existe $y \in R_{\bar{q}}$ et g serait non limité sur $R_q\, y \subseteq R$, d'où la contradiction. Donc L''/R' appartient à $C^f_U(L)$, pour tout langage $L'' \in C^f(L)$. Enfin, si $L \in C^f_U(L)$, il existe $L_1,\dots,L_n \in C^f(L)$ tels que $L = \bigcup_{i=1}^{n} L_i$ et pour tout rationnel R', $L/R' = \bigcup_{i=1}^{n} (L_i \ / \ R')$ appartient à $C^f_U(L)$.□

Lemme II.35 : *Soit L un cône rationnel fidèle clos par union. Si L est clos par quotient rationnel, $F^f(L)$ = Rat □ L est aussi clos par quotient rationnel.*

Démonstration : Si $L \in F^f(L)$, il existe un langage rationnel $R_1 \subseteq T_1^*$ et une L-substitution σ telle que $L = \sigma(R_1)$. Considérons un langage rationnel $R_2 \subseteq T_2^*$ et montrons que $\sigma(R_1) \ / \ R_2 \in F^f(L)$. Pour $i = 1, 2$, soit $M_i = (Q_i, T_i, f_i, q_{i,o}, F_i)$ un automate d'état fini qui reconnait R_i. Pour $i = 1, 2$ et $q \in Q_i$, notons $R_{i,q} = \{w \in T_i^* \ / \ f_i(q_{i,o}, w) = q\}$ et $R_{i,\bar{q}} = \{w \in T_i^* \ / \ f_i(q, w) \in F_i\}$. Considérons alors, l'ensemble H inclus dans $Q_1 \times T_1 \times Q_2$ défini de la manière suivante :

$H = \{(q, a, q') \ / \ \sigma(R_{1,\bar{q}_a}) \cap R_{2,q'} \neq \emptyset$ avec $q_a = f_1(q, a)\}$ et montrons l'égalité

$$\sigma(R_1) \ / \ R_2 = V \cup \Big(\bigcup_{(q,a,q')\in H} \sigma(R_1, q)\ (\sigma(a) \ / R_{2,q'})\Big)$$ où $V = \{\varepsilon\}$ si $\varepsilon \in \sigma(R_1) \ / \ R_2$, $\emptyset$ sinon.

Prenons $x \in \sigma(R_1) / R_2$ avec $x \neq \varepsilon$. Il existe alors $y \in R_2$ tel que $xy \in \sigma(R_1)$. Comme $x \neq \varepsilon$, x et y se factorisent respectivement en $x_1\, x_2$ et $y_1\, y_2$ avec $x_1 \in \sigma(u)$, $x_2\, y_1 \in \sigma(a)$, $y_2 \in \sigma(v)$, $u, v \in T_1^*$, $a \in T_1$ et $u\, a\, v \in R_1$. Posons $q = f_1(q_{1,o}, u)$, $q_a = f_1(q, a)$ et $q' = f_2(q_{2,o}, y_1)$. Comme $f_1(q_{1,o}, u\, a) = q_a$ et $u\, a\, v \in R_1$, $v \in R_{1,\bar{q}_a}$ et $y_2 \in \sigma(R_{1,\bar{q}_a})$. De même, comme $y_1\, y_2 = y \in R_2$ avec $q' = f_2(q_{2,o}, y_1)$, $y_2 \in R_{2,\bar{q}'}$, ce qui implique $(q, a, q') \in H$. En outre, $x_1 \in \sigma(u) \subseteq \sigma(R_{1,q})$ et $x_2\, y_1 \in \sigma(a)$ avec $y_1 \in R_{2,q'}$, donc $x_2 \in \sigma(a) / R_{2,q'}$ et $x = x_1\, x_2$ appartient à $\sigma(R_{1,q})\, (\sigma(a) / R_{2,q'})$.

Réciproquement, si $x \in \sigma(R_{1,q})\, (\sigma(a) / R_{2,q'})$ avec $(q, a, q') \in H$, x se factorise en $x_1\, x_2$ avec $x_1 \in \sigma(R_{1,q})$ et $x_2 \in \sigma(a) / R_{2,q'}$. Donc, il existe $u \in R_{1,q}$, $y_1 \in R_{2,q'}$ tels que $x_1 \in \sigma(u)$, $x_2\, y_1 \in \sigma(a)$. Comme $(q, a, q') \in H$, il existe $y_2 \in \sigma(R_{1,\bar{q}_a}) \cap R_{2,q'}$ avec $q_a = f_1(q, a)$. Ceci implique qu'il existe $v \in R_{1,\bar{q}_a}$ tel que $y_2 \in \sigma(v)$ et $u\, a\, v \in R_1$, donc $x_1\, x_2\, y_1\, y_2 \in \sigma(R_1)$ avec $y_1\, y_2 \in R_2$ et $x = x_1\, x_2 \in \sigma(R_1) / R_2$.

La démonstration s'achève en remarquant que $\sigma(R_{1,q})$ appartient à $F^f(L)$ puisque $R_{1,q}$ est un langage rationnel et que $\sigma(a) / R_{2,q'} \in L/\mathrm{Rat} = L$ par hypothèse. Comme H est fini, $\sigma(R_1) / R_2 \in F^f(L)$ qui est donc clos par quotient rationnel.□

En utilisant ces deux résultats on peut alors énoncer le lemme suivant :

Lemme II.36 : *Soit une famille de langage $L = \{L_1, L_2, L_3\}$ avec $L_1 = \{a^n\, b^p\, c^n \;/\; p \geq n \geq o\}$, $L_2 = \{a^n\, b^p\, c^q \;/\; p \geq n \geq q \geq o\}$ et $L_3 = \{a^n\, b^p\, c^q \;/\; p \geq q \geq n \geq o\}$. La plus petite FAL fidèle contenant L, $F^f(L)$, est close par quotient rationnel.*

Démonstration : D'après les lemmes II.34 et II.35, il nous suffit de montrer que L/Rat est inclus dans $C^f_\cup(L)$ et donc d'établir que $L_i / R \in C^f_\cup(L)$ pour tout $i \in \{1, 2, 3\}$ et tout langage rationnel R. Comme $L / (R_1 \cup R_2) = L / R_1 \cup L/R_2$ et que tout langage rationnel est une union finie de langages rationnels dont l'image par la fonction de Parikh est un ensemble linéaire [55], on peut supposer que $\Psi(R)$ est un ensemble linéaire. Enfin, tout facteur d'un mot appartenant à L_i, est dans $a^*\, b^*\, c^*$, donc $L_i / R = L_i / R'$ où $R' = R \cap a^*\, b^*\, c^*$ est rationnel et on peut supposer $R \subseteq a^*\, b^*\, c^*$.

Considérons d'abord le cas où R est inclus dans c^*. Si R est fini, nous avons $L_i / R \in C^f(L_i) \subseteq C^f_\cup(L)$, $\forall\, i \in \{1, 2, 3\}$. Si R est infini, comme $\Psi(R)$ est un ensemble linéaire, $\exists\, j \in \mathbb{N}$ et $k \in \mathbb{N}_+$ tels que $R = c^j (c^k)^*$. Soit h l'homomorphisme défini sur $T = \{a, b, c\}$ par $h(x) = x^k$, $\forall\, x \in T$. Pour tout $i \in \{0,\dots,k-1\}$, $L'_i = \mathrm{shuf}\,(h(L_2), a^i\, b^i\, b^*\, c^i) \cap a^*\, b^*\, c^*$ appartient à $C^f(L_2)$, donc $L''_1 = L_1 / (c^k)^* = \bigcup_{i=o}^{k-1} L'_i \in C^f(L_2) \subseteq C^f_\cup(L)$ et $L_1 / R = L''_1 / c^j \in C^f(L''_1) \subseteq C^f_\cup(L)$. Comme $L_2 / (c^k)^*$ est égal à L_2, nous avons aussi $L_2 / R \in C^f(L_2) \subseteq C^f_\cup(L)$. Enfin, montrons que $L''_3 = L_3 / (c^k)^*$ appartient à $C(L_1 \cup L_2)$. Pour tout $i \in \{0,\dots,k-1\}$, soient Y_i le langage rationnel $\{a^n\, b^p\, c^q \;/\; p-q = \lambda k+i$ avec $\lambda \in \mathbb{Z}\}$ et $Z_i = \{a^n\, b^p\, c^q \;/\; p \geq q,\ p \geq n+i\} \cap Y_i$. Vérifions que $L''_3 = \bigcup_{i=o}^{k-1} Z_i$.

Prenons $w = a^n b^p c^q \in L_3 / (c^k)^*$. Il existe $j \in \mathbb{N}$ et $i \in \{0,\dots,k-1\}$, $w c^{kj} \in L_3 \cap Y_i$, donc $p \geq q + jk \geq n$ et $p - q = \lambda k + i$ avec $\lambda \in \mathbb{Z}$ et $i \in \{0,\dots,k-1\}$ ce qui implique $p-q-jk = (\lambda-j)k+i \geq 0$ donc $\lambda \geq j$, $p \geq q+jk+i \geq n+i$ et $w \in Z_i$.

Réciproquement, si $w = a^n b^p c^q \in Z_i$, nous avons $p \geq q$, $p \geq n+i$ et $p-q = \lambda k+i$ donc $\lambda \in \mathbb{N}$, $q+\lambda k = p-i \geq n$, ce qui implique $a^n b^p c^q c^{\lambda k} \in L_3$ et $w \in L_3 / (c^k)^*$.

D'autre part, il est facile de vérifier que pour tout $i \in \{0,\dots,k-1\}$, $Z_i = \text{shuf}(L_2 \cup L_3, \{b^i c^j / 0 \leq j \leq i\}) \cap Y_i$, donc $Z_i \in C^f(L_2 \cup L_3)$, $L''_3 \in C^f(L_2 \cup L_3) \subseteq C^f(\mathcal{L})$ et $L_3 / R = L''_3 / c^j \in C^f(L''_3) \subseteq C^f(\mathcal{L})$.

Considérons maintenant un langage rationnel R inclus dans $\{a, b\}^+ c^*$. Comme R peut s'écrire comme l'union finie de produit de la forme $R_1 R_2$ où R_1 et R_2 sont des langages rationnels inclus respectivement dans $\{a, b\}^+$ et c^*, il nous suffit de montrer que $L_i / R_1 R_2 \in C^f(\mathcal{L})$. Or, comme tout mot de R_1 contient au moins une lettre de $\{a, b\}$, $L_i / R_1 R_2$ est égal à A_i / R_1 avec $A_i = (L_i / R_2) \cap a^* b^*$. D'après ce que nous avons montré ci-dessus, $A_i \in C^f(\mathcal{L})$ et comme A_i est un PSL-langage inclus dans $a^* b^*$, A_i est un langage algébrique inclus dans $a^* b^*$ et il est facile de voir que A_i est effaçable donc $A_i / R_1 \in C^f(A_i) \subseteq C^f(\mathcal{L})$.□

Nous sommes, maintenant, en mesure de répondre à une question de Ginsburg ([25] page 34) en établissant l'existence d'une FAL fidèle close par quotient rationnel mais pas par homomorphisme effaçant. Nous avons, en effet :

<u>Proposition II.37</u> : *Il existe un langage* $L \subseteq a^* b^* c^*$ *tel que* $F^f(L)$ *soit clos par quotient rationnel mais pas par homomorphisme effaçant, c'est-à-dire* $F^f(L) \subsetneq F(L)$.

<u>Démonstration</u> : Considérons la famille $\mathcal{L} = \{L_1, L_2, L_3\}$. D'après Goldstine [37], il existe un langage $L \subseteq a^* b^* c^*$ tel que $F^f(L) = F^f(\mathcal{L})$. Il nous reste donc à montrer que $F^f(\mathcal{L})$ est inclus strictement dans $F(\mathcal{L})$. En utilisant un raisonnement analogue à celui qui est utilisé dans [26], on peut vérifier que tout langage infini appartenant à $C^f_{\cup}(\mathcal{L})$ possède un facteur itérant. Prenons $L \in C^f(L_i)$ avec $i \in \{1, 2, 3\}$. Il existe, alors, deux homomorphismes alphabétique h,g et un langage rationnel R tels que $L = g(h^{-1}(L_i) \cap R)$ et g est limité sur R. Soit k le nombre d'états d'un automate d'état fini qui reconnait R. Si pour tout $w \in L_i$ de longueur supérieure à 3k, $h^{-1}(w) \cap R = \emptyset$, $L = g(h^{-1}(L'_i) \cap R)$ où $L'_i = \{w \in L_i / \ell(w) \leq 3k\}$ est un langage fini, donc L est un langage rationnel infini et possède un facteur itérant. Dans le cas contraire il existe $w = a^n b^p c^q \in L_i$ tel que $h^{-1}(w) \cap R \neq \emptyset$ et $n+p+q > 3k$. On peut alors trouver dans R, un mot $y = y_1 y_2 y_3$ avec $h(y_1) = a^n$, $h(y_2) = b^p$, $h(y_3) = c^q$. Par définition des langages L_i, $n+p+q > 3k$ implique $p > k$ et il existe une factorisation de y_2 en $y'_2 u y''_2$ avec $u \neq \varepsilon$ telle que $y_1 y'_2 u u^* y''_2 y_3 \subseteq R$. Comme $h(u) \in b^*$, $h(y_1 y'_2 u u^* y''_2 y_3) \subseteq L_i$. Enfin, g étant limité sur R, $g(u) \neq \varepsilon$ et $g(y_1 y'_2 u) (g(u))^* g(y''_2 y_3) \subseteq L$ qui possède un facteur itérant.

Tout langage $L' \in C_U^f(L)$ peut s'écrire $L' = L'_1 \cup L'_2 \cup L'_3$ avec $L'_i \in C^f(L_i)$ pour $i \in \{1, 2, 3\}$ et si L' est infini, l'un des trois langages est infini et possède un facteur itérant donc L', aussi.

Le langage C_1 est un CIL-langage. Donc C_1 n'appartient pas à $C_U^f(L)$ et d'après la proposition I.23, $C_1 \notin F^f(L) = \mathrm{Rat} \,\square\, C_U^f(L)$. Par contre, il est clair que $C_1 \in C(L_1) \subseteq F(L)$ et $F^f(L) \subsetneq F(L)$.□

Etant donné que L est clos par miroir, $F^f(L)$ est aussi fermé par quotient rationnel à gauche et vérifie donc : $\forall\, L \in F^f(L)$, $\mathrm{Init}(L) \in F^f(L)$ et $\mathrm{Fac}(L) = \{w \,/\, \exists\, u, v$ tels que $u w v \in L\} \in F^f(L)$.

D'autre part, on peut remarquer que la proposition précédente n'est plus valable pour les langages bornés sur moins de trois lettres. En effet, tout langage $L \subseteq a^*$ est effaçable (cf. corollaire II.43) et pour les langages inclus dans $a^* b^*$, nous avons :

<u>Lemme II.38</u> : *Pour tout langage $L \subseteq a^* b^*$ et tout $i, j \in \mathbb{N}_+$, le langage $L' = (a^i)^* \setminus (L / (b^j)^*) = \{w \in a^* b^* \,/\, \exists\, s, t \in \mathbb{N}$ tels que $a^{is} w b^{jt} \in L\}$ est un langage rationnel.*

<u>Démonstration</u> : Posons $A = \{0,\dots i-1\} \times \{0,\dots,j-1\}$ et pour tout $(p, q) \in A$, notons $R_{p,q} = a^p(a^i)^* b^q(b^j)^*$ et $L_{p,q} = L \cap R_{p,q}$. Comme L' est égal à $\bigcup_{(p,q)\in A} (R_{p,q} \cap (a^* \setminus (L_{p,q}/b^*)))$, on peut se ramener au cas où $i = j = 1$ et il reste à montrer que $L' = a^* \setminus (L / b^*)$ est rationnel pour tout langage $L \subseteq a^* b^*$. Si pour tout $k \in \mathbb{N}$, il existe $s, t > k$ tels que $a^s b^t \in L$, alors $L' = a^* b^*$ est rationnel. Sinon L' est égal à $L_1 \cup L_2$ où $L_1 = \{a^s b^t \in L \,/\, s \leq k\}$ et $L_2 = \{a^s b^t \in L \,/\, t \leq k\}$. Alors, $L_1 = \bigcup_{0\leq s\leq k} a^s L'_s$ avec $L'_s = \{b^j \,/\, a^s b^j \in L_1\}$ et $a^* \setminus (a^s L'_s / b^*)$ est égal à $\bigcup_{0\leq n\leq s} (a^n(L'_s / b^*))$. Si L'_s est infini, $L'_s / b^* = b^*$, donc L'_s / b^* est toujours rationnel ainsi que L_1. Par symétrie L_2 est rationnel et $L' = L_1 \cup L_2$ est aussi rationnel.□

<u>Proposition II.39</u> : *Un langage L inclus dans $a^* b^*$ est effaçable si et seulement si $C^f(L)$ est clos par quotient rationnel à droite et à gauche.*

<u>Démonstration</u> : Si L est effaçable, $C^f(L) = C(L)$ qui est clos par quotient rationnel à droite et à gauche.

Réciproquement, prenons $L' \in C(L)$. Il existe, alors, un alphabet Δ, un langage rationnel $R \subseteq \Delta^*$ et deux homomorphismes alphabétiques h et g tels que $L' = g(h^{-1}(L) \cap R)$. Notons $A_1 = \{x \in \Delta \,/\, h(x) \in a^*\}$ et $A_2 = \{x \in \Delta / h(x) \in b^*\}$. Comme L est inclus dans $a^* b^*$, $L' = g(h^{-1}(L) \cap R')$ avec $R' = R \cap A_1^* A_2^*$. Le langage R' est un langage rationnel inclus dans $A_1^* A_2^*$, donc R' est une union finie de produit, $R' = \bigcup_{i=1}^{n} R'_{i,1} R'_{i,2}$ où pour tout $i \in \{1,\dots,n\}$, $R'_{i,1}$ et $R'_{i,2}$ sont des langages rationnels inclus respectivement dans A_1^* et A_2^*. Alors $L' = \bigcup_{i=1}^{n} L'_i$ avec $L'_i = g(h^{-1}(L) \cap R'_{i,1} R'_{i,2})$ et pour montrer que $L' \in C^f(L)$, il suffit de montrer que $L'_i \in C^f(L)$ pour tout $i \in \{1,\dots,n\}$. On peut, donc, supposer sans nuire à la généralité de la démonstration que $R = R_1 R_2$ où R_1 et R_2 sont des langages rationnels inclus respectivement dans A_1^* et A_2^*. Pour $i \in \{1, 2\}$, soit $M_i = (Q_i, A_i, f_i, q_i, F_i)$ un automate d'état fini à k_i états qui reconnait R_i. Pour $i \in \{1, 2\}$ et tout $w \in A_i^*$, notons $Q_i(w) = \{q \in Q_i \ / \ \exists\ w'$ tel que $w = w'\ w''$ avec $f_i(q_i, w') = q\}$ et pour tout $X_i \subseteq Q_i$, posons $R_{X_i} = \{w \in R_i \ / \ Q_i(w) = X_i\}$, $R'_{X_i} = R_{X_i} \setminus A_i^*\ Z^{k_i}\ A_i^*$ où Z désigne $\{y \in \Delta \ / \ g(y) = \varepsilon\}$. Pour $i \in \{1, 2\}$, il est clair que $R_i = \bigcup_{X_i \subseteq Q_i} R_{X_i}$, pour tout $q \in Q_i$, $Z_{i,q} = \{w \in (Z \cap A_i)^* \ / \ f_i(q, w) = q_i\}$ et pour tout $X_i \subseteq Q_i$, posons $Z_{X_i} = (\bigcup_{q \in X_i} Z_{i,q})^*$. Montrons, maintenant que pour $X_1 \subseteq Q_1$ et $X_2 \subseteq Q_2$, le langage $L_{X_1,X_2} = g(h^{-1}(L) \cap R_{X_1}\ R_{X_2})$ est égal à $g(h^{-1}(L) \cap Z_{X_1}\ R'_{X_1}\ R'_{X_2}\ Z_{X_2})$.

Raisonnons comme pour la démonstration du lemme II.41. Prenons $w \in L_{X_1,X_2}$. Alors $w = g(w_1 w_2)$ avec $w_1 \in R_{X_1}$, $w_2 \in R_{X_2}$ et $h(w_1\ w_2) \in L$. Pour $i \in \{1, 2\}$, il existe $w'_i \in R'_{X_i}$ et $z_i \in Z_{X_i}$ tels que $w_i \in \text{shuf}(w'_i, z_i)$ et comme $g(z_i) = \varepsilon$, on obtient $g(w'_i\ z_i) = g(z_i\ w'_i) = g(w_i)$. Enfin comme $h(w_1) \subseteq a^*$ et $h(w_2) \subseteq b^*$, $h(z_1\ w'_1) = h(w_1)$, $h(w'_2\ z_2) = h(w_2)$ et $w = g(w_1\ w_2) = g(z_1\ w'_1\ w'_2\ z_2) \in g(h^{-1}(L) \cap Z_{X_1}\ R'_{X_1}\ R'_{X_2}\ Z_{X_2})$.

Réciproquement, prenons $w = g(z_1\ w'_1\ w'_2\ z_2)$ avec $\forall\ i \in \{1, 2\}$, $z_i \in Z_{X_i}$, $w'_i \in R'_{X_i} \subseteq R_{X_i}$ donc $Q_i(w'_i) = X_i$. Alors, pour tout $i \in \{1, 2\}$, il existe $z'_i \in c(z_i)$ et $w_i \in R_{X_i}$ tels que $w_i \in \text{shuf}\ (w'_i, z'_i)$ et nous obtenons $w = g(w'_1\ w'_2) = g(w_1\ w_2)$ avec $h(w_1\ w_2) = h(z_1\ w'_1\ w'_2\ z_2) \in L$, donc $w \in g(h^{-1}(L) \cap R_{X_1}\ R_{X_2})$.

Comme $g(Z_{X_1}) = g(Z_{X_2}) = \{\varepsilon\}$, nous avons $L_{X_1,X_2} = g(h^{-1}(L'') \cap R'_{X_1}\ R'_{X_2})$ avec $L'' = h(Z_{X_1}) \setminus (L \ / \ h(Z_{X_2}))$. Etant donné que g est limité par construction sur R'_{X_1} et R'_{X_2}, g est limité sur $R'_{X_1}\ R'_{X_2}$ et $L_{X_1,X_2} \in \mathcal{C}^f(L'')$. Enfin, comme $L' = \{w \in L_{X_1,X_2} \ / \ X_1 \subseteq Q_1$ et $X_2 \subseteq Q_2\}$, il nous reste à montrer que $L'' \in \mathcal{C}^f(L)$. Le langage $h(Z_{X_1})$ est un langage rationnel inclus dans a^*, c'est donc une union finie de langages de la forme $a^s(a^t)^*$ et on peut se ramener au cas où $h(Z_{X_1}) = a^s(a^t)^*$ et où $h(Z_{X_2}) = b^i(b^j)^*$. Si le produit tj est non nul, d'après le lemme précédent, L" est un langage rationnel qui appartient donc à $\mathcal{C}^f(L)$. Sinon, supposons $t = 0$. Alors, par hypothèse, $L''_1 = L \ / \ (b^j)^*$ appartient à $\mathcal{C}^f(L)$ et $L'' = a^s \setminus (L''_1 \ / \ b^i) \in \mathcal{C}^f(L)$. Le cas où $j = 0$ se traite symétriquement, d'où le résultat.□

Si, de plus, le langage $L \subseteq a^*\ b^*$ est un PSL-langage, nous pouvons déduire de la proposition précédente :

<u>Proposition II.40</u> : *Tout langage algébrique inclus dans $a^*\ b^*$ est effaçable.*

Démonstration : Soit L un langage algébrique inclus dans $a^* b^*$. Alors, $S = \Psi(L)$ peut s'écrire $S = \bigcup_{i=1}^{n} S_i$ où les $S_i = L(c_i ; P_i)$ sont des ensembles linéaires propres disjoints deux à deux [20], [49]. Pour tout $i \in \{1,\dots,n\}$, P_i contient au plus deux éléments. Posons $I = \{i \in \{1,\dots,n\} / P_i \cap \mathbb{N}_+ \times \mathbb{N}_+ \neq \emptyset\}$ et $I' = \{1,\dots,n\} \setminus I$. Le langage $R' = \Psi^{-1}(\bigcup_{i\in I'} S_i) \cap a^* b^*$ est rationnel et nous avons $L = L' \cup R'$ et $L' = L \cap \bar{R}'$ avec $L' = \Psi^{-1}(\bigcup_{i\in I} S_i) \cap a^* b^*$. Montrons que L' / b^* appartient à $C^f(L')$. Pour cela, effectuons la même construction que Berstel dans [6]. Posons $P' = \{u \in \mathbb{N}_+ / \exists\, v \in \mathbb{N},\ i \in I \text{ tels que } (u, v) \in P_i\}$ et soit p le plus petit commun multiple des éléments de P'. Prenons $i \in I$ et posons $P'_i = P_i$ si P_i a deux éléments, $P_i \cup \{(0,0)\}$ sinon. Donc $P'_i = \{x, y\}$ avec $x = (x', x'') \in \mathbb{N}_+ \times \mathbb{N}_+$ et $y = (y', y'')$. Posons $s = p / x'$, $t = p / y'$ si $y' \neq 0$, 1 sinon, $x_1 = sx$, $y_1 = ty$, $P''_i = \{x_1, y_1\}$ et $C = \{c_i + \lambda x + \mu y / 0 \le \lambda < s \text{ et } 0 \le \mu < t\}$. Il est clair que S_i est égal à $\bigcup_{u\in C} L(u ; P''_i)$. On peut alors, supposer, sans nuire à la généralité de la démonstration, que le langage L' vérifiait : Il existe $p \in \mathbb{N}_+$ tel que pour tout $x = (x', x'') \in P = \bigcup_{i\in I} P_i$, soit $x' = p$, soit $x' = 0$. Posons $m = \sup \{u \in \mathbb{N}_+ / \exists\, i \in I,\ v \in \mathbb{N} \text{ tels que } (u,v) \in P_i \cup \{c_i\}\}$ et montrons que $L' / (b^*) = \bigcup_{j=o}^{m} ((a^p)^* L' / b^j)$.

Prenons $j \in \{0,\dots,m\}$ et $w \in (a^p)^* L' / b^j$. Alors, il existe $r \in \mathbb{N}$, $w_1 \in a^* b^*$ tels que $w = (a^p)^r w_1$ et $w' = w_1 b^j \in L'$. Il existe, donc, $i \in I$ tel que $z = \Psi(w') \in L(c_i ; P_i)$ et $(p, y'') \in P_i$. Alors $z' = z + r(p, y'') \in L(c_i ; P_i)$ ce qui entraine $(a^p)^r w_1 b^{j+ry''} \in L'$ et $w \in L' / b^*$.

Réciproquement si $w \in L' / b^*$, il existe $r \in \mathbb{N}$ tel que $w b^r \in L'$ et $i \in I$ tel que $z = \Psi(w b^r) \in L(c_i ; P_i)$. Il existe $\lambda, \mu \in \mathbb{N}$ tels que $z = c_i + \lambda x + \mu y$ où $\{x , y\} = P'_i$, $x = (p, x'')$, $y = (y', y'')$ et $c_i = (c'_i, c''_i)$. Par définition de m, il est clair que l'on peut trouver $\lambda' \in \{0,\dots,\lambda\}$ et $\mu' \in \{0,\dots,\mu\}$ tels que $c''_i + \lambda' x'' + \mu' y'' = \ell_b(w) + j$ avec $0 \le j \le m$. Posons, alors $z' = c_i + \lambda' x + \mu' y$, $w' = \Psi^{-1}(z') \cap a^* b^*$ et $k = \lambda - \lambda'$ si $y' = 0$, $\lambda - \lambda' + \mu - \mu'$ sinon. Comme $w' \in L'$ et que $w = (a^p)^k w'/b^j$ avec $0 \le j \le m$, on obtient bien $w \in \bigcup_{j=o}^{m} ((a^p)^* L' / b^j)$.

Etant donné que $(a^p)^* L' \in C^f(L')$, l'égalité que nous venons d'établir implique $L' / b^* \in C^f(L')$. Nous en déduisons que $L / b^* = L' / b^* \cup R' / b^*$ appartient $C^f(L)$ et comme $L' = L \cap \bar{R}' \in C^f(L)$, pour tout langage algébrique L inclus dans $a^* b^*$, $L / b^* \in C^f(L)$.

Prenons, maintenant, un langage rationnel $R \subseteq a^* b^*$ et montrons que $L / R \in C^f(L)$. Le langage R est égal à $R_1 \cup R_2$ avec $R_1 = R \cap a^+ b^*$ et $R_2 = R \cap b^*$, donc $L/R = L/R_1 \cup L/R_2$. Comme L/R_1 est un PSL-langage inclus dans a^*, L/R_1 est un langage rationnel et appartient à $C^f(L)$. Le langage R_2 est une union finie de langages de la forme $b^i(b^j)^*$, il suffit donc de montrer que $L'' = L /(b^j)^* \in C^f(L)$. Si $j = 0$, $L'' = L \in C^f(L)$. Sinon $L'' = \bigcup_{s=o}^{j-1} ((L_s / b^*) \cap a^* b^s(b^j)^*)$ avec $L_s = L \cap a^* b^s(b^j)^*$ et $L_s / b^* \in C^f(L_s) \subseteq C^f(L)$, donc $L'' \in C^f(L)$ et tout langage algébrique inclus dans $a^* b^*$ est clos par quotient rationnel.

Par symétrie, nous obtenons le même résultat pour le quotient rationnel à gauche et le résultat se déduit de la proposition II.39.□

En particulier, cette proposition implique un résultat démontré par Greibach dans [41], [44] : le langage $C_1 = \{a^n b^n \ / \ n \geq 0\}$ est effaçable.

Montrons, maintenant, que la proposition II.39 est aussi vérifiée pour touts les langages commutatifs. Mais, en raison de la commutativité, il est clair qu'il n'est pas utile de considérer le quotient rationnel à gauche. Etablissons d'abord :

Lemme II.41 : *Soit L un langage commutatif inclus dans T^*. Pour tout $L' \in C(L)$, il existe des langages rationnels $R_1, \ldots, R_p \subseteq T^*$ tels que $L' \in C_U^f(\mathcal{L})$ avec $\mathcal{L} = \{L \ / \ R_1, \ldots, L \ / \ R_p\}$.*

Démonstration : Il existe un langage rationnel $R \subseteq Y^*$ et deux homomorphismes g et h tels que $L' = g(h^{-1}(L) \cap R)$. Soient $M = (Q, T, f, q_o, F)$ un automate d'état fini qui reconnait R et k le nombre d'états de cet automate. Pour tout $x \in T^*$, notons $Q(x) = \{q \in Q \ / \ \exists \ x'$ tel que $x = x' \ x''$ et $f(q_o, x') = q\}$ et pour tout $X \subseteq Q$ posons $R_X = \{x \in R \ / \ Q(x) = X\}$, $R'_X = R_X \setminus T^* Z^k T^*$, où Z désigne $\{y \in Y \ / \ g(y) = \varepsilon\}$. Il est clair que R'_X est un langage rationnel et que $R = \bigcup_{X \subseteq Q} R_X$. Pour tout $q \in Q$, notons $Z_q = \{x \in Z^* \ / \ f(q, x) = q\}$ et pour tout $X \subseteq Q$, $Z_X = (\bigcup_{q \in X} Z_q)^*$ et $L_X = g(h^{-1}(L) \cap R_X)$. Montrons que $\forall \ X \subseteq Q$, $L_X = g(h^{-1}(L) \cap R'_X \ Z_X)$.

Prenons $y \in L_X$. Alors $y = g(x)$ avec $x \in h^{-1}(L) \cap R_X$. Comme l'automate M n'a que k états, on peut trouver une factorisation de x en $x'_1 \ z_1 \ x'_2 \ \ldots \ x'_n \ z_n$ telle que $x' = x'_1 \ldots x'_n \in R'_X$ et $\forall \ i \in \{1, \ldots, n\}$, $\exists \ q_i \in X$ tel que $z_i \in Z_{q_i}$ ce qui implique $z = z_1 \ldots z_n \in Z_X$. Donc $x' \ z \in R'_X \ Z_X$ et comme $x' \ z \in c(x)$ et que $h^{-1}(L)$ est commutatif, $x' \ z \in h^{-1}(L)$. Enfin, comme $g(z) = g(z_1) \ldots g(z_n) = \varepsilon$, nous obtenons $y = g(x) = g(x' \ z) \in g(h^{-1}(L) \cap R'_X \ Z_X)$.

Réciproquement, si $y \in g(h^{-1}(L) \cap R'_X \ Z_X)$, il existe un mot $x' \ z$ tel que $y = g(x' \ z)$, $x' \in R'_X$, $z \in Z_X$ et $h(x' \ z) \in L$. Comme $x' \in R_X$, x' se factorise en $x'_1 \ x'_2 \ \ldots \ x'_t$ avec $\forall \ i \in \{1, \ldots, t\}$, $x'_i \in Y$ et $f(q_o, x'_1 \ldots x'_i) = q_i$ donc $X = \{q_o, q_1, \ldots, q_t\}$. D'autre part, comme $z \in Z_X$, il existe $z_o, \ldots, z_t$ tels que $\forall \ i \in \{0, \ldots, t\}$, $z_i \in Z_{q_i}$ et $c(z) = c(z_o \ldots z_t)$. Alors $x = z_o \ x'_1 \ z_1 \ \ldots \ x'_t \ z_t$ et comme $x \in c(x' \ z_o \ldots z_t) = c(x' \ z)$ et que $h^{-1}(L)$ est commutatif, $x \in h^{-1}(L)$ et $y = g(x' \ z) = g(x') = g(x) \ \in L_X$.

Etant donné que $g(Z_X) = \{\varepsilon\}$, $L_X = g(h^{-1}(L) \cap R'_X \ Z_X)$ est égal à $g((h^{-1}(L) \ / \ Z_X) \cap R'_X)$ et comme $h^{-1}(L) \ / \ Z_X = h^{-1}(L \ / \ h(Z_X))$, $L_X = g(h^{-1}(L \ / \ h(Z_X)) \cap R'_X)$. Par définition de R'_X, l'homomorphisme g est k-limité sur R'_X, donc L_X appartient à $C^f(L \ / \ h(Z_X))$ où $h(Z_X)$ est un langage rationnel inclus dans T^*. Enfin, $L = g(h^{-1}(L) \cap R) = g(h^{-1}(L) \cap (\bigcup_{X \subseteq Q} R_X)) = \bigcup_{X \subseteq Q} L_X \in C_U^f(\mathcal{L})$ où $\mathcal{L} = \{L \ / \ R_1, \ldots, L \ / \ R_p\}$ avec $\{R_1, \ldots, R_p\} = \{h(Z_X) \ / \ X \subseteq Q\}$.□

Nous pouvons en déduire immédiatement le résultat de base de cette section :

Proposition II.42 : *Un langage commutatif L est effaçable (i.e. $C^{f}(L) = C(L)$) si et seulement si pour tout langage rationnel R, L / R appartient à $C^{f}(L)$.*

Cette proposition se généralise aisément :

Proposition II.43 : *Pour toute famille $\mathcal{L}$ de langages commutatifs, $C_{\cup}(\mathcal{L}) = C^{f}_{\cup}(\mathcal{L})$ si et seulement si $C^{f}_{\cup}(\mathcal{L})$ est clos par quotient rationnel.*

Démonstration : Si $L \in C_{\cup}(\mathcal{L})$, il existe une famille finie $\mathcal{L}' \subseteq \mathcal{L}$ telle que $L \in C_{\cup}(\mathcal{L}')$ et d'après le corollaire II.13, il existe un langage commutatif L' vérifiant $C^{f}(L') = C^{f}_{\cup}(\mathcal{L}')$ et $C(L') = C_{\cup}(\mathcal{L}')$. Donc $L \in C(L')$ avec $L' \in C^{f}_{\cup}(\mathcal{L}') \subseteq C^{f}_{\cup}(\mathcal{L})$. Comme $C^{f}_{\cup}(\mathcal{L})$ est clos par quotient rationnel, le lemme II.41 implique, alors, $L \in C^{f}_{\cup}(\mathcal{L})$ qui est donc égal à $C_{\cup}(\mathcal{L})$.□

Comme, pour tout langage commutatif L, L / R est encore commutatif, du corollaire II.13 et du lemme II.41, nous pouvons déduire aussi :

Corollaire II.44 : *Soit L un langage m-commutatif. Pour tout langage $L' \in C(L)$, il existe un langage L'' m-commutatif tel que $L'' \in C(L)$ et $L' \in C^{f}(L'')$.*

Corollaire II.45 : *Pour toute famille $\mathcal{L}$ de langages (m-)commutatifs, il existe une famille $\mathcal{L}'$ de langages (m-)commutatifs vérifiant $C^{f}(\mathcal{L}') = C_{\cup}(\mathcal{L}')$. En particulier, $C(Com) = C^{f}(Com)$.*

Démonstration : Il suffit de prendre $\mathcal{L}' = C_{\cup}(\mathcal{L}) \cap COM_m$ si $\mathcal{L} \subseteq COM_m$ et $\mathcal{L}' = C_{\cup}(\mathcal{L}) \cap COM$ sinon.□

Considérons le cas d'un langage $L \subseteq a^*$ où a est une lettre. D'après le lemme II.38, pour tout $i \in \mathbb{N}_+$, $L / (a^i)^* = (a^i)^* \setminus (L / b^*)$ est un langage rationnel. Il est, alors, clair que pour tout langage rationnel $R \subseteq a^*$, $L / R \in C^f(L)$ et la proposition II.42 permet de retrouver un résultat démontré en premier lieu par Ullian :

Corollaire II.46 [25] [79] : *Tout langage L inclus dans a^* est effaçable.*

Donnons, maintenant, une condition suffisante pour que L/R appartienne à $C^f(L)$ avec L langage commutatif et R langage rationnel. Ce résultat servira non seulement pour caractériser les langages commutatifs effaçables inclus dans D_1^*, mais aussi, au chapitre suivant, pour établir que tout PSL-langage commutatif est effaçable.

Lemme II.47 : *Soient L un langage commutatif, R un langage rationnel inclus dans T^* et $L' = L/R$. S'il existe $k \in \mathbb{N}$ vérifiant : $\forall x \in L'$, $\exists y \in R$ tel que $xy \in L$ et $\ell(y) \leq k(\ell(x) + 1)$, alors L' appartient à $C^f(L)$.*

Démonstration : On peut supposer, sans nuire à la généralité de la démonstration, que $\Psi(R)$ est un ensemble linéaire. Donc il existe des mots $w_o, w_1, \ldots, w_n \in T^*$ tels que $\Psi(R) = \Psi(w_o\, w_1^* \ldots w_n^*)$. Comme L est commutatif $L / R = L_1 / w_o$ où $L_1 = L / R'$ avec $R' = \{w_1, \ldots, w_n\}^*$. Soient $\bar{T} = \{\bar{a} \;/\; a \in T\}$ avec $T \cap \bar{T} = \emptyset$, $\Delta = T \cup \bar{T}$ et les homomorphismes h et g définis sur Δ par $h(\bar{a}) = h(a) = a$, $g(\bar{a}) = a$, $g(a) = \varepsilon$, $\forall a \in T$. Considérons le langage rationnel $R'' = \text{shuf}\,(\bar{T}^*, R') \setminus \Delta^* T^{k+1} \Delta^*$ et montrons que L_1 est égal à $g(h^{-1}(L) \cap R'')$.

Prenons $x = x_1 \ldots x_p \in L_1$ avec $\forall i \in \{1, \ldots, p\}$ $x_i \in T$. L'hypothèse du lemme implique, alors, qu'il existe $y \in R'$ tel que $xy' \in L$ avec $\ell(y) \leq k(\ell(x) + 1) = k(p + 1)$. Donc il existe une factorisation de y en $y_1 \ldots y_{p+1}$ avec $\forall i \in \{1, \ldots, p+1\}$, $\ell(y_i) \leq k$. Nous en déduisons que $z = y_1\, \bar{x}_1\, y_2 \ldots \bar{x}_p\, y_{p+1}$ appartient à R'' et comme $h^{-1}(L)$ est commutatif, $z \in h^{-1}(L)$ et $x = g(z) \in g(h^{-1}(L) \cap R'')$.

Il est clair, réciproquement, que $g(h^{-1}(L) \cap R'')$ est inclus dans L / R' et comme, par définition de R'', g est $k+1$-limité sur R'', $L_1 = L / R' = g(h^{-1}(L) \cap R'')$ appartient à $C^f(L)$. Enfin, $L / R = L_1 / w_o \in C^f(L_1) \subseteq C^f(L)$.□

Remarquons que la condition du lemme précédent n'est pas une condition nécessaire. En effet, prenons $L = \{a^{n!} \;/\; n \geq 0\}$ et $R = a^*$. Alors, $L / R = a^*$ appartient à $C^f(L)$ bien qu'il n'existe pas $k \in \mathbb{N}$ vérifiant : $\forall n \in \mathbb{N}$, $(n+1)! - (n!+1) \leq k(n!+2)$.

Nous allons maintenant, utiliser la proposition II.42 et le lemme II.47 pour montrer que certains langages commutatifs sont effaçables. Plus précisément, nous obtenons une condition nécessaire et suffisante qui ne dépend que de la longueur des mots de L, pour qu'un langage commutatif $L \subseteq D_1^*$ soit effaçable. Montrons d'abord :

Lemme II.48 : *Soient L un langage commutatif inclus dans D_1^* et t un entier positif. Alors $L / (a^t\, b^t)^*$ appartient à $C(D_1^*)$.*

Démonstration : Pour tout $i \in \{0, \ldots, t-1\}$, R_i désignera le langage rationnel égal à $\{x \in \{a, b\}^* \;/\; \ell_a(x) \equiv i (\text{mod. } t)\}$ et posons $I = \{i \in \{0, \ldots, t-1\} \;/\; L \cap R_i$ est infini$\}$. Il est clair que pour tout $i \in I$, $(L \cap R_i) / (a^t b^t)^* = D_1^* \cap R_i$ et que pour $i \in \bar{I} = \{0, \ldots, t-1\} \setminus I$, $(L \cap R_i) / (a^t b^t)^*$ est fini. Donc $L / (a^t b^t)^* = (\bigcup_{i \in \{0, \ldots, t-1\}} (L \cap R_i)) / (a^t b^t)^* = \bigcup_{i \in \{0, \ldots, t-1\}} ((L \cap R_i) / (a^t b^t)^*) = F \cup (\bigcup_{i \in I} (D_1^* \cap R_i))$ où F est fini. Donc $L / (a^t b^t)^* \in C(D_1^*)$.□

Proposition II.49 : *Soit L un langage commutatif infini inclus dans D_1^*. Alors, les trois propriétés suivantes sont équivalentes :*

i) *L est effaçable,*

ii) *D_1^* appartient à $C^f(L)$,*

iii) *$\exists\ k \in \mathbb{N}$ vérifiant $\forall\ x \in L$, $\exists\ y \in L$ tel que $\ell(x) < \ell(y) \leq k(\ell(x)+1)$.*

Démonstration : Montrons d'abord que iii) implique ii). Prenons $z \in D_1^* = L\ /\ (ab)^*$ et soit t le plus petit entier positif ou nul tel que $y = z(ab)^t \in L$. Comme L est commutatif, l'hypothèse entraine, alors, $\ell(y) \leq k(\ell(z) + 1)$ et à fortiori, $\ell((ab)^t) \leq k(\ell(z)+1)$. Donc, d'après le lemme II.47, $D_1^* = L\ /\ (ab)^* \in C^f(L)$.

Montrons, maintenant, que ii) implique i). D'après la proposition II.42, il suffit de montrer que pour tout langage rationnel $R \subseteq \{a,b\}^*$, $L\ /\ R \in C^f(L)$. On peut supposer que $\Psi(R)$ est un ensemble linéaire, donc il existe $x_o, x_1,\ldots,x_s \in \{a, b\}^*$ tels que $L\ /\ R = L\ /\ x_o\ x_1^* \ldots x_s^* = L_1\ /\ x_o$ avec $L_1 = L\ /\ x_1^* \ldots x_s^*$. Comme $L \in C^f(L_1)$ il suffit de montrer que $L_1 \in C^f(L)$. Distinguons plusieurs cas :

a) Il existe $t \in \mathbb{N}_+$ tel que $(a^tb^t) \in c(x_1^* \ldots x_s^*)$. Comme L est commutatif, $L_1 = L_2\ /\ x_1^* \ldots x_s^*$ avec $L_2 = L_1\ /\ (a^tb^t)^*$. Le lemme précédent implique, alors, $L_2 \in C(D_1^*)$et comme D_1^* et L_2 sont des PSL-langages commutatifs, nous obtenons, en utilisant la proposition III.15 $L_1 \in C(L_2) = C^f(L_2) \subseteq C^f(D_1^*)$ qui est inclus, par hypothèse dans $C^f(L)$.

b) Pour tout $i \in \{1,\ldots,s\}$, $\ell_a(x_i) > \ell_b(x_i)$. Donc, il existe $k \in \mathbb{N}$, qui vérifie : $\forall\ i \in \{1,\ldots,s\}$, $k\ \ell_a(x_i) \geq (k+1)\ \ell_b(x_i)$ ce qui entraine que pour tout y appartenant à $x_1^* \ldots x_s^*$, $k\ \ell_a(y) \geq (k+1)\ \ell_b(y)$. Si $x \in L_1$, il existe $y \in x_1^* \ldots x_s^*$ tel que $xy \in L \subseteq D_1^*$. Donc $\ell_a(xy) = \ell_b(xy) = n$, $k\ \ell_a(y) \geq (k+1)\ \ell_b(y)$ et comme $\ell_a(y) \leq n$, nous obtenons $\ell_b(y) \leq \frac{kn}{k+1}$, $\ell_b(x) = n - \ell_b(y) \geq n - \frac{kn}{k+1} = \frac{n}{k+1}$ et $\ell(y) = \ell_a(y) + \ell_b(y) \leq \frac{n(2k+1)}{k+1} \leq (2k+1)\ \ell_b(x) \leq (2k+1)\ (\ell(x)+1)$. Le lemme II.47 entraine alors, $L_1 \in C^f(L)$.

c) Pour tout $i \in \{1,\ldots,s\}$, $\ell_b(x_i) > \ell_a(x_i)$. On utilise alors, le même raisonnement qu'au b).

Montrons, enfin, que i) implique iii). Si L est effaçable, $C_1 = D_1^* \cap a^*\ b^* = (L\ /\ (ab)^*) \cap a^*\ b^* \in C^f(L)$. Donc, il existe un alphabet Y, un langage rationnel $R' \subseteq Y^*$ et deux homomorphismes alphabétiques h et g de Y^* dans $\{a, b\}^*$ tels que $C_1 = g(h^{-1}(L) \cap R')$ et g est k-limité sur R'. On peut supposer que, pour tout $y \in Y$, soit $g(y) \neq \varepsilon$, soit $h(y) \neq \varepsilon$. Soit k' le nombre d'états d'un automate d'état fini qui reconnait R'. Si $u\ x\ v \in R'$ avec $h(x) = \varepsilon$, $h(u\ v) \in L$ et $\ell(x) \geq k$, x se factorise en $x_1\ x_2\ x_3$ avec $x_2 \neq \varepsilon$, $u\ x_1\ x_2^*\ x_3\ v \subseteq R'$, et donc $g(x_2) \neq \varepsilon$, $h(u\ x_1\ x_2^*\ x_3\ v) = h(u\ v) \in L$ ce qui implique $g(u\ x_1)\ (g(x_2))^*\ g(x_3\ v) \subseteq C_1$ qui ne contient pas de facteur itérant d'où la contradiction. Donc, si $u\ x\ v \in h^{-1}(L) \cap R'$ avec $x \in Z^*$ $(Z = \{y \in Y\ /\ h(y) = \varepsilon\})$, $\ell(x) \leq k'$. Alors, il est facile de vérifier que $C_1 = g(h^{-1}(L) \cap R'')$ où $R''=R' \setminus Y^*\ Z^{k'}\ Y^*$. Prenons $w \in L$ avec $\ell(w) = 2\ n$ et $t = k'(n+1)$. Comme $a^tb^t \in C_1$, il existe $w' \in R''$ tel que $h(w') \in L$ et $g(w') = a^t b^t$. Comme h est k'-limité sur R'', $\ell(w') \leq k'(\ell(h(w'))+1)$ et donc $\ell(h(w')) \geq \frac{\ell(w')}{k'} - 1$ ce qui implique, puisque g est alphabétique, $\ell(h(w')) \geq \frac{2t}{k'} - 1 > \ell(w)$. D'autre part, h est alphabétique et g est k-limité sur R', donc aussi sur $R'' \subseteq R'$ ce qui implique $\ell(h(w')) \leq \ell(w') \leq k(\ell(g(w'))+1) = k(2t+1) = k(2k'(n+1)+1) \leq k''(2n+1)$ avec $k'' = 2kk' + 1$. Donc $h(w') \in L$ et vérifie $\ell(w) < \ell(h(w')) \leq k''(\ell(w) + 1)$.$\square$

D - Mots infinis et langages commutatifs

Rappelons la définition d'un mot infini donné par Nivat dans [66].

Un **mot infini** sur un alphabet X, est une application de $\mathbb{N}_+$ dans X et on écrit $u \in X^{\omega}$.

Le **langage du mot infini** u, noté Pr(u), est égal à $\{\varepsilon\} \cup \{u(1) \dots u(k) \ / \ k \in \mathbb{N}_+\}$.

Tout langage d'un mot infini, L, possède la propriété de préfixe, c'est-à-dire que $\forall$ x, y $\in$ L, soit x est un préfixe (facteur gauche) de y, soit y est un préfixe de x. Dans [15], Book montre que certaines familles de langages ne contiennent pas de langages non rationnels possédant la propriété de préfixe. En fait, dés que la famille considérée est close par intersection avec les langages rationnels et par homomorphisme (strictement alphabétique), il suffit d'examiner les langages définis sur un alphabet d'une seule lettre (cf. [58]).

Comme pour tout $L \subseteq a^*$, Init (L) = $\{x \ / \ \exists\ y,\ xy \in L\}$ est un langage rationnel, on ne peut, en aucun cas, utiliser la même méthode pour montrer qu'une famille de langages $\mathcal{L}$ vérifie la propriété (*) : Tout langage d'un mot infini appartenant à $\mathcal{L}$ est rationnel. La famille C(Com) contient tous les langages inclus dans a^*, donc il existe dans C(Com) des langages non rationnels ayant la propriété de préfixe. Par contre, nous allons montrer que F_σ(Com) vérifie la propriété (*). Pour ce faire, nous allons, d'abord, montrer pour les langages d'un mot infini l'équivalence entre le fait d'être rationnel, d'être un PSL-langage, de contenir un langage borné infini. Nous allons ensuite établir que tout langage rationnel contient un langage rationnel borné qui possède la même image par la fonction de Parikh. Nous en déduisons, pour les langages de C(Com) une propriété que nous utiliserons à plusieurs reprises par la suite : Si $L \in$ C(Com), il existe des mots $w_1,\dots,w_k$ tels que $\Psi(L) = \Psi(L \cap w_1^* \dots w_k^*)$. Une première conséquence est que tout langage infini appartenant à F_σ(Com) contient un langage borné infini, ce qui nous permet de montrer que la famille des DOL-langages n'est pas incluse dans F_σ(Com). Nous en déduisons, ensuite, une propriété des transductions rationnelles d'image bornée (proposition II.58) et une condition nécessaire et suffisante pour que $F_\sigma(\mathcal{L})$, avec $\mathcal{L} \subseteq C$(Com), contienne un langage borné non rationnel. Cette condition permet de retrouver très simplement un résultat de Goldstine [36] concernant le langage $L_G = \{a^{j_0} c \dots c\, a^{j_t} \ / \ t \geq 0, \exists\ i \neq j_i\}$.

D'après la définition donnée au début de cette section, il est clair que pour tout langage infini L, si L a la propriété de préfixe, Init(L) est le langage d'un mot infini. Montrons, d'abord, que cette condition peut être remplacée par l'égalité des mots de L de même longueur.

Proposition II.50 : *Soit* L *un langage tel que* $L = \mathrm{Init}(L)$. *Alors* L *vérifie la propriété de préfixe si et seulement si* L *ne possède pas deux mots différents de même longueur.*

Démonstration : Si L vérifie la propriété de préfixe, et si w, w' sont des mots de L de même longueur, soit w est un préfixe de w', soit w' est un préfixe de w. Dans les deux cas, $w = w'$.

Réciproquement, prenons $w, w' \in L$ avec $\ell(w) \leq \ell(w')$. Alors $w' = w'_1 w'_2$ avec $\ell(w'_1) = \ell(w)$ ce qui implique $w'_1 = w$ puisque $w'_1 \in \mathrm{Init}(L) = L$. Donc w est un préfixe de w' et L vérifie la propriété de préfixe.□

Nous dirons qu'un mot infini u est **finalement périodique** s'il existe $r \in \mathbb{N}$ et $q \in \mathbb{N}_+$ tels que pour tout i supérieur à r, $u(i+q) = u(i)$.

Un ensemble $A \subseteq \mathbb{N}$ sera dit rationnel, si le langage $\{a^i \mid i \in A\}$ est un langage rationnel.

Caractérisons, maintenant, les langages d'un mot infini qui sont rationnels :

Proposition II.51 : *Soit* L *le langage d'un mot infini* $u \in X^\omega$. *Les propriétés suivantes sont équivalentes :*

i) $\exists\, x, y \in X^+$ *tels que* $L = \mathrm{Init}(xy^*)$,

ii) L *est un langage rationnel,*

iii) $\forall\, b \in X$, $u^{-1}(b)$ *est un ensemble rationnel,*

iv) L *contient un langage borné infini,*

v) *Le langage* L *est PSL-langage.*

Démonstration : L'équivalence entre i), ii) et iii) a été établie dans [58]. D'autre part, il est clair que ii) implique iv) et v).

Montrons que iv) entre i). Par hypothèse, il existe $w_1, \ldots, w_p \in X^+$ tels que $L \cap w_1^* \ldots w_p^*$ soit infini. Donc, il existe $i \in \{1, \ldots, p\}$ tel que $\forall\, k \in \mathbb{N}$, $L \cap w_1^* \ldots w_i^k\ w_i^* \ldots w_p^*$ soit non vide. Prenons le plus petit i vérifiant cette propriété. Alors, $F = \{z \in w_1^* \ldots w_{i-1}^* \mid \exists\, z' \in w_i^* \ldots w_p^*,\ zz' \in L\}$ est fini et non vide. Il existe donc, dans F, un mot x tel que $\forall\, k \in \mathbb{N}$, $L \cap x\, w_i^k\, w_i^* \ldots w_p^*$ soit non vide. Comme $L = \mathrm{Init}(L)$ et que $w_i \neq \varepsilon$, on en déduit que $L \cap x\, w_i^*$ est infini. Alors pour tout $z \in L$ et tout $z' \in x\, w_i^*$, il existe $y \in L \cap x\, w_i^*$ qui possède z et z' comme facteur gauche ce qui entraine $L \subseteq \mathrm{Init}(x\, w_i^*) \subseteq \mathrm{Init}(L)$ et donc $L = \mathrm{Init}(x\, w_i^*)$.

Supposons, maintenant, que le langage L soit un PSL-langage. Alors $\Psi(L) = S = \bigcup_{i=1}^{n} L(c_i\, ; P_i)$.

Posons $X = \{a_1,\ldots,a_k\}$ et pour tout $z = (z_1,\ldots,z_k) \in \mathbb{N}^k$, $|z| = \sum_{i=1}^{k} z_i$. Montrons que si $z = (z_1,\ldots,z_k)$ et $z' = (z'_1,\ldots,z'_k)$ sont deux périodes de S, $\forall\, t \in \{1,\ldots,k\}$, $z_t\,|z'| = z'_t\,|z|$. En effet, il existe, dans S, $y = (y_1,\ldots,y_k)$ et $y' = (y'_1,\ldots,y'_k)$ tels que $L(y\ ;\ \{z\})$ et $L(y'\ ;\ \{z'\})$ soient inclus dans S. Supposons qu'il existe $t \in \{1,\ldots,k\}$ tel que $z_t\,|z'| > z'_t\,|z|$. Prenons, alors, $s \in \mathbb{N}$ et posons $x = (x_1,\ldots,x_k) = y + s\,|z'|\,z$ et $x' = (x'_1,\ldots,x'_k) = y' + (s + |y|)\,|z|\,z'$. Comme $z_t\,|z'| \geq z'_t\,|z| + 1$, nous avons $x_t - x'_t \geq s + y_t - y'_t - |y|\,|z|\,z'_t$ et comme $|z|\,|z'| \geq 1$, $|x'| \geq |x|$. Si s a été choisi suffisamment grand, $x_t - x'_t > 0$. Il existe $w, w' \in L$ tels que $\Psi(w) = x$ et $\Psi(w') = x'$ avec $\ell(w') = |x'| \geq |x| = \ell(w)$ et comme L a la propriété de préfixe, $\Psi(w') \geq \Psi(w)$ ce qui contredit le fait que $x_t - x'_t > 0$. Cette propriété implique l'existence de $y \in \mathbb{N}^k$ tel que, $\forall\, x \in P = \bigcup_{i=1}^{n} P_i$, $\exists\, s \in \mathbb{N}$ vérifiant $y = sx$ et on peut trouver deux ensembles finis $F, F' \subseteq \mathbb{N}^k$ tels que $S = F \cup \{z + \lambda y \ /\ z \in F', \lambda \in \mathbb{N}\}$. Comme F est fini et $\ell(L) = \{|z| \ /\ z \in S\} = \mathbb{N}$, $\forall\, i \in \{0,\ldots,|y|\}$, il existe $z_i \in F'$ tel que $|z_i| \equiv i \pmod{|y|}$. Prenons $j \in \{0,\ldots,|y| - 1\}$. Il existe, alors, $\alpha, \beta, r \in \mathbb{N}$ tels que $r = |z_{j+1}| + \alpha|y| = |z_j| + \beta|y| + 1$. Prenons $\lambda \in \mathbb{N}$ et considérons w, le mot de L de longueur $r + \lambda|y|$. Le mot w peut se factoriser en $w'\,bw''\,c$ avec $b, c \in X$, $\ell(w'b)=r$ et $\ell(w''c)=\lambda|y|$. Comme ℓ est une fonction injective sur L et que tout facteur gauche de w appartient à L, on obtient $\Psi(bw'') = \Psi(w'bw'') - \Psi(w') = \lambda y$ et $\Psi(w''c) = \Psi(w) - \Psi(w'b) = \lambda y$, ce qui implique $\Psi(b) = \Psi(c)$ et donc $b = c$ puisque b et $c \in X$. Donc, $\forall\, \lambda \in \mathbb{N}$, $u(r + \lambda|y|) = u(r)$. Comme j a été choisi de façon quelconque dans $\{0,\ldots,|y|-1\}$, on en déduit que u est finalement périodique et L est un langage rationnel [58].□

Le fait que tout langage d'un mot infini qui est semi-linéaire est aussi rationnel est à opposer au résultat de Book [15] qui montre l'existence d'un langage semi-linéaire, non rationnel ayant la propriété de préfixe. En fait, l'existence d'un tel langage se ramène à l'existence de mots infinis qui ne sont pas finalement périodiques. En effet, la proposition suivante montre le lien très étroit qui existe entre la famille $\mathcal{L}$ des langages d'un mot infini et la famille $\mathcal{L}'$ des PSL-langages ayant la propriété de préfixe :

Proposition II.52 : *Pour tout langage* L *de* $\mathcal{L}$ *(resp.* $\mathcal{L}'$*) il existe un langage appartenant à* $\mathcal{L}'$ *(resp.* $\mathcal{L}$*), rationnellement équivalent à* L.

Démonstration : Soit $L \subseteq X^*$, un langage de $\mathcal{L}$ avec $X = \{a_1,\ldots,a_k\}$. Considérons l'homomorphisme h défini sur X par : $\forall\, i \in \{1,\ldots,k\}$, $h(a_i) = a_i \ldots a_k\, a_1 \ldots a_{i-1}$ et posons $L' = h(L)$. Il est clair que $\Psi(L') = \Psi((a_1\ldots a_k)^*)$ et que $h^{-1}(L') = L$. Et comme la propriété de préfixe est conservée par homomorphisme, $L' \in \mathcal{L}'$.

Réciproquement, soit $L \subseteq X^*$ un langage de $\mathcal{L}'$ avec $X = \{a_1,\ldots,a_k\}$. Si L est fini, il suffit de prendre $L' = a^*$. Si L est infini, posons $L' = \mathrm{Init}(L)$. Le langage L' appartient à $\mathcal{L}$ et à $C(L)$. D'autre part, comme L est un PSL-langage, $B = \{w \in X^* \ /\ \ell(w) \in \ell(L)\}$ est un langage rationnel et $L = L' \cap B \in C(L')$.□

Nous allons, maintenant, montrer que si L est le langage d'un mot infini appartenant à C(Com), le plus petit cône rationnel contenant les langages commutatifs, alors L est un langage rationnel. Pour obtenir ce résultat, établissons, d'abord, que tout langage infini $L \in C(Com)$ contient un langage borné infini. Nous utiliserons une propriété concernant les langages rationnels qui nous servira aussi à la section suivante :

Lemme II.53 : *Tout langage rationnel R contient un langage rationnel borné R' tel que* $\Psi(R) = \Psi(R')$.

Démonstration : Comme la propriété est vraie pour les langages finis et quelle est conservée par union et par produit, il nous suffit, d'après le théorème de Kleene, de montrer que si R est un langage rationnel vérifiant cette propriété R^* la vérifie aussi. Comme $R' \subseteq R$ implique $R'^* \subseteq R^*$ et $\Psi(R') = \Psi(R)$ implique $\Psi(R'^*) = \Psi(R^*)$, on peut supposer que R est un langage rationnel borné. Supposons, d'abord, R inclus dans $a_1^* \ldots a_k^*$ où les a_i sont des symboles distincts. Alors R peut s'écrire, $R = \bigcup_{i=1}^{n} (R_{i,1} \ldots R_{i,k})$ où pour tout $j \in \{1,\ldots,k\}$ et tout $i \in \{1,\ldots,n\}$, $R_{i,j}$ est un langage rationnel inclus dans a_j^*. Comme tout langage rationnel inclus dans b^* est une union finie de langages de la forme $b^s(b^t)^*$ avec $s, t \in \mathbb{N}$, on peut supposer que chacun des $R_{i,j}$ est de cette forme. D'autre part, R^* contient $A_1^* \ldots A_n^*$ et $\Psi(R^*) = \Psi(A_1^* \ldots A_n^*)$où pour tout $i \in \{1,\ldots,n\}$, $A_i = R_{i,1} \ldots R_{i,k}$. Il suffit, alors, de montrer que pour tout $i \in \{1,\ldots,n\}$, il existe un langage rationnel borné B_i tel que $B_i \subseteq A_i^*$ et $\Psi(B_i) = \Psi(A_i^*)$. Prenons $i \in \{1,\ldots,n\}$. Le langage A_i est égal à $a_1^{s_1}(a_1^{t_1})^* \ldots a_k^{s_k}(a_k^{t_k})^*$. Les mots ε et $w = a_1^{s_1} \ldots a_k^{s_k}$ appartiennent à A_i^* et donc $B_i = \{\varepsilon\} \cup A_i w^*$ est un langage rationnel borné inclus dans A_i^*. Pour tout $y \in A_i^{p+1}$ avec $p \in \mathbb{N}$, il existe $y' \in A_i$ tel que $\Psi(y) = \Psi(y' w^p)$ donc $\Psi(B_i) = \Psi(A_i^*)$.

Supposons, maintenant, R inclus dans $w_1^* \ldots w_k^*$ et considérons l'homomorphisme h défini sur $\{a_1,\ldots,a_k\}$ par : $h(a_i) = w_i$, $\forall\, i \in \{1,\ldots,k\}$. Le langage $R_1 = h^{-1}(R) \cap a_1^* \ldots a_k^*$ est un langage rationnel inclus dans $a_1^* \ldots a_k^*$, donc il existe un langage rationnel borné $R'_1 \subseteq R_1^*$ tel que $\Psi(R'_1) = \Psi(R_1^*)$. Le langage $R' = h(R'_1)$ est un langage rationnel borné inclus dans $h(R_1^*) = R^*$ et vérifie $\Psi(R') = \Psi(R^*)$.□

Nous allons pouvoir en déduire le résultat de base de cette section qui est l'analogue du lemme précédent, pour les langages appartenant à C(Com) :

Proposition II.54 : *Pour tout langage* $L \in C(Com)$, *il existe des mots* $w_1^*,\ldots,w_k^*$ *tels que* $\Psi(L \cap w_1^* \ldots w_k^*) = \Psi(L)$.

Démonstration : Si $L \in C(Com)$, il existe un langage commutatif L', un langage rationnel R et deux homomorphismes h et g tels que $L = g(h^{-1}(L') \cap R)$ (cf. [64]). D'après le lemme précédent, il existe un langage rationnel $R' \subseteq x_1^* \ldots x_k^* \cap R$ tel que $\Psi(R') = \Psi(R)$ et comme $h^{-1}(L')$ est un langage commutatif, $\Psi(h^{-1}(L') \cap R') = \Psi(h^{-1}(L')) \cap \Psi(R') = \Psi(h^{-1}(L')) \cap \Psi(R) = \Psi(h^{-1}(L') \cap R)$ (cf. [54]). Alors, $\Psi(L) = \Psi(L'')$ avec $L'' = g(h^{-1}(L') \cap R') \subseteq w_1^* \ldots w_k^*$ où $\forall\, i \in \{1,\ldots,k\}$, $w_i = g(x_i)$. Donc $L'' \subseteq L \cap w_1^* \ldots w_k^* \subseteq L$, ce qui implique $\Psi(L'') \subseteq \Psi(L \cap w_1^* \ldots w_k^*) \subseteq \Psi(L) = \Psi(L'')$ et $\Psi(L \cap w_1^* \ldots w_k^*) = \Psi(L)$.□

Cette proposition nous permet, par exemple, de montrer que le langage $L = \{(a^n b)^p \ / \ n, p \geq 1\}$ n'appartient pas à $C(\mathrm{Com})$. En effet, il est clair que L n'est pas un PSL-langage. Par contre, nous allons montrer que pour tout $w_1,\dots,w_k \in \{a, b\}^*$, $L' = L \cap w_1^* \dots w_k^*$ est un PSL-langage. Posons $A = \{z \in \{a,b\}^* \ / \ \ell_b(z) > k\}$ et $\bar{A} = \{a,b\}^* \setminus A$. Si z appartient à $L' \cap A$, z se factorise en $z' z_i z''$ avec $z_i \in w_i^*$ et $\ell_b(z_i) > 1$. Alors, z_i possède un facteur $ba^t b$ avec $t \leq \ell_a(w_i)$. Nous en déduisons que $L' \cap A$ est égal à $L'' \cap A \cap w_1^* \dots w_k^*$ où $L'' = \{(a^n b)^p \ / \ p \geq 1, 1 \leq n \leq q\}$ avec $q = \sup \{\ell_a(w_i) \ / \ i \in \{1,\dots,k\}\}$. Comme $L'' = \bigcup_{i=1}^{q} (a^i b)^*$, le langage $L' \cap A$ est rationnel. D'autre part $L' \cap \bar{A}$ est égal à $L'_1 \cap w_1^* \dots w_k^*$ où $L'_1 = \{(a^n b)^p \ / \ n \geq 1, 1 \leq p \leq k\}$. Le langage L'_1 est rationnellement équivalent à $E_k = \{a_1^n \dots a_k^n \ / \ n \geq 0\}$ et appartient donc à $\mathcal{L}_\psi$, le plus petit cône rationnel clos par intersection et contenant $\{a^n b^n \ / \ n \geq 0\}$ (cf. [ ]). Et, comme $\mathcal{L}_\psi$ ne contient que des PSL-langages, $L' = (L'_1 \cap w_1^* \dots w_k^*) \cup (L' \cap A)$ est un PSL-langage.

Etant donné que $C(\mathrm{Com})$ est égal au plus petit cône rationnel contenant les langages bornés et clos par intersection, la proposition précédente reste à fortiori vraie si on remplace, dans l'énoncé, la famille Com par la famille Bor des langages bornés. Cette proposition implique, en particulier, que tout langage infini appartenant à $C(\mathrm{Com})$ contient un langage borné infini. Il est clair que si $\mathcal{L}_1$ et $\mathcal{L}_2$ sont deux cônes rationnels vérifiant cette propriété, $\mathcal{L}_1 \square \mathcal{L}_2 = \{s(L) \ / \ L \in \mathcal{L}_1$ et s est une $\mathcal{L}_2$-substitution$\}$ vérifie encore cette propriété et nous obtenons, par induction :

<u>Corollaire II.55</u> : *Si L est un langage infini appartenant à $F_\sigma(\mathrm{Com})$, le plus petit cône rationnel clos par substitution et contenant les langages commutatifs, alors L contient un langage borné infini.*

Des propositions II.51, II.52 et du corollaire précédent, nous pouvons déduire :

<u>Corollaire II.56</u> : *Soit L un langage de $F_\sigma(\mathrm{Com})$. Si L est le langage d'un mot infini, ou si L est un PSL-langage ayant la propriété de préfixe, alors L est un langage rationnel.*

Considérons, maintenant, la famille des DOL-langages, c'est-à-dire des langages de la forme $\{h^i(w) \ / \ i \in \mathbb{N}\}$ où $w \in X^*$, h est un homomorphisme de X^* dans X^* et h^i est défini par induction au moyen des relations $h^0(x) = x$, $\forall x \in X^*$ et $\forall j \in \mathbb{N}$, $h^{j+1} = h \circ h^j$ (cf. [47]). Tout langage défini sur un alphabet d'une lettre est commutatif, donc appartient à $F_\sigma(\mathrm{Com})$. Par contre, le corollaire II.56 va nous permettre de montrer qu'il existe un DOL-langage défini sur un alphabet de deux lettres, n'appartenant pas à $F_\sigma(\mathrm{Com})$. Prenons, en effet, $X = \{a, b\}$, $w = a$ et $L = \{h^i(w) \ / \ i \in \mathbb{N}\}$ où h est défini sur X par $h(a) = ab$, $h(b) = ba$. Si L appartient à $F_\sigma(\mathrm{Com})$, le langage $L_1 = \mathrm{Init}(L) \cap (X^2)^*$ appartient encore à $F_\sigma(\mathrm{Com})$. Or, le langage L_1 est un PSL-langage ayant la propriété de préfixe et non rationnel [15]. Le corollaire précédent entraine une contradiction et nous obtenons :

<u>Corollaire II.57</u> : *La famille des DOL-langages n'est pas incluse dans $F_\sigma(\mathrm{Com})$.*

La proposition II.54 va, aussi, nous permettre d'établir une propriété qui concerne les transductions rationnelles qui font passer d'un langage de $C(\mathrm{Com})$ à un langage borné :

Proposition II.58 : *Soient* L' *un langage borné et* L *un langage appartenant à* $C(\mathrm{Com})$. *Si* $L' \in C(L)$, *il existe un langage rationnel borné R et deux homomorphismes h et g tels que* $L' = g(h^{-1}(L) \cap R')$.

Démonstration : Il existe $w_1,\ldots,w_n$ tels que L' soit inclus dans $w_1^* \ldots w_n^*$. Posons $L'' = f^{-1}(L') \cap a_1^* \ldots a_n^*$ où f est l'homomorphisme défini sur $T = \{a_1,\ldots,a_n\}$ par $f(a_i) = w_i$, $\forall\, i \in \{1,\ldots,n\}$. Comme $L'' \in C(L)$, il existe un langage rationnel R et deux homomorphismes h et g' tels que $L'' = g'(h^{-1}(L) \cap R)$. Le langage $h^{-1}(L) \cap R$ appartient à $C(\mathrm{Com})$, donc il existe des mots $y_1,\ldots,y_k$ tels que $\Psi(h^{-1}(L) \cap R) = \Psi(h^{-1}(L) \cap R \cap y_1^* \ldots y_k^*)$ (proposition II.54). Posons $R' = R \cap y_1^* \ldots y_k^*$ et considérons le langage $L_1 = g(h^{-1}(L) \cap R')$. Comme $\Psi(L_1) = \Psi(L'')$ et que $L_1 \subseteq L'' \subseteq a_1^* \ldots a_n^*$, il est clair que $L'' = L_1 = g(h^{-1}(L) \cap R')$ et $L' = f(L'') = g(h^{-1}(L) \cap R')$ avec $g = f \circ g'$.□

Remarquons que la proposition précédente est, en particulier, vérifiée si L est un langage borné. D'autre part, nous aurions pu, par un raisonnement classique, montrer que les homomorphismes h et g pouvaient encore être pris alphabétiques. Enfin, comme $h^{-1}(L) \cap R' = h^{-1}(L \cap h(R')) \cap R'$ et que $h(R')$ est un langage rationnel borné, nous étendons la proposition II.12 à tous les langages de $C(\mathrm{Com})$.

Considérons, maintenant, une famille de langages incluse dans $C(\mathrm{Com})$ et donnons une condition nécessaire et suffisante pour que $F_\sigma(\mathcal{L})$ contienne un langage borné non rationnel :

Proposition II.59 : *Soit* $\mathcal{L}$ *une famille de langages incluse dans* $C(\mathrm{Com})$. *Alors, si* $F_\sigma(\mathcal{L})$ *contient un langage borné non rationnel, il existe* $L \in \mathcal{L}$ *et des mots* $w_1,\ldots,w_k$ *tels que* $L \cap w_1^* \ldots w_k^*$ *soit un langage non rationnel.*

Démonstration : Si $\mathcal{L}_1$ et $\mathcal{L}_2$ sont des cônes rationnels clos par union ne contenant aucun langage borné non rationnel, il en est de même pour $\mathcal{L}_1 \square \mathcal{L}_2$. On en déduit, par induction, que, si $F_\sigma(\mathcal{L})$ contient un langage borné non rationnel, $C_\cup(\mathcal{L})$ aussi. Ce qui implique, clairement, l'existence d'un langage $L \in \mathcal{L}$ tel que $C(L)$ contienne un langage borné non rationnel. Il existe, alors, d'après la proposition précédente, un homomorphisme h et un langage rationnel borné R' tel que $h^{-1}(L) \cap R'$ soit non rationnel. Comme $h^{-1}(L) \cap R' = h^{-1}(L \cap h(R')) \cap R'$, $L \cap h(R')$ est non rationnel. Le langage $h(R')$ est borné donc il existe $w_1,\ldots,w_k$ tels que $h(R') \subseteq w_1^* \ldots w_k^*$ ce qui entraine $L \cap h(R') = (L \cap w_1^* \ldots w_k^*) \cap h(R')$ et donc $L \cap w_1^* \ldots w_k^*$ est non rationnel avec $L \in \mathcal{L}$.□

Considérons, maintenant, le langage de Goldstine (cf. [36]) :

$$L_G = \{a^{j_0} c\, a^{j_1} c \ldots c\, a^{j_t} \ /\ t \geq 0,\ \exists\, i \neq j_i\}.$$

Il est clair que $L_G \in C(Com)$. Par ailleurs, $\bar{L}_G = \{a, c\}^* \setminus L_G = \{c\ a\ c\ a^2\ c \dots c\ a^k \ /\ k \geq 0\}$ a la propriété de préfixe et ne contient aucun langage borné infini (cf. section A de ce chapitre). Donc pour tout mot $w_1,\dots,w_k$, $\bar{L}_G \cap w_1^* \dots w_k^*$ est fini et $L_G \cap w_1^* \dots w_k^*$ est un langage rationnel. La proposition précédente permet, alors, de retrouver immédiatement une propriété du langage L_G, démontrée par Goldstine dans [36].

Corollaire II.60 [36] *La famille $F_\sigma(L_G)$ ne contient aucun langage borné non rationnel.*

E - LANGAGES K-COMMUTATIFS

Nous avons montré à la section précédente que tout langage infini de $F_\sigma(\text{Com})$ contient un langage borné infini (corollaire II.55). Nous allons, maintenant, préciser ce résultat en établissant que tout langage infini de $F_\sigma(\text{Com}_k)$ contient un langage borné infini de "dimension k". Nous obtenons, ainsi, un résultat analogue à celui de Goldstine dans [37], mais formulé de façon différente.

Comme le langage $E_{k+1} = \{a_1^n \ldots a_{k+1}^n / n \geq 0\}$ ne contient aucun langage borné infini de dimension k et appartient à $F(c(\text{Rat}))$, nous pouvons en déduire que la FAL engendrée par c(Rat) est non principale. Nous apportons, ainsi, pour une conjecture de Ginsburg et Spanier [33], une autre démonstration que celle que nous avions donnée dans [54]. Enfin, nous terminons cette section en montrant que pour tout $k \geq 2$, il existe $L \in \text{Com}_k$ tel que tout langage borné non rationnel appartenant à $F_\sigma(L)$ soit au moins de dimension k.

Définition : *Un langage* L *est un langage* **borné de dimension k** *s'il existe des mots* $x_0, y_1, \ldots, y_k, x_k$ *tels que* L *soit inclus dans* $x_0\, y_1^* \ldots y_k^*\, x_k$.

Les langages **minces** définis par Sakarovitch [74] sont, dans le cas du monoïde libre, des unions finies disjointes de langages rationnels bornés de dimension 1.

Pour tout langage rationnel R, $u\, x^*\, v \setminus R$ est un langage mince disjoint de R, donc toute union finie de langages rationnels bornés de dimension 1 est un langage mince. Le lemme II.53, nous permet, alors, de montrer :

Lemme II.61 : *Pour tout langage rationnel* $R \subseteq T^*$ *et tout homomorphisme* h *de* T^* *dans* a^*, *il existe un langage mince* $R' \subseteq R$ *tel que* $h(R') = h(R)$.

Démonstration : Il existe un langage rationnel borné R" tel que $\Psi(R'') = \Psi(R)$ et donc $h(R'') = h(R)$. Comme R" est une union finie de langages de la forme $x_0\, y_1^* \ldots y_t^*\, x_t$, on peut se ramener au cas où $R = x_0\, y_1^* \ldots y_t^*\, x_t$ avec $h(y_i) \neq \varepsilon$, $\forall\, i \in \{1,\ldots,t\}$. Si t est inférieur à 2, R est un langage mince. Sinon, posons $s = \ell(h(y_1)) - 1$ et $R_1 = \bigcup_{i=0}^{s} x_0\, y_1^* \ldots y_{t-1}^*\, x_{t-1}\, y_t^i\, x_t$. Il est, alors, clair que $h(R_1) = h(R)$ et nous obtenons le résultat par induction.□

Pour tout langage L, posons $\ell(L) = \{\ell(w) \,/\, w \in L\}$. Si nous prenons, dans l'énoncé du lemme précédent, l'homomorphisme h défini par $h(x) = a$, $\forall\, x \in T$, nous obtenons un résultat de même nature que le lemme II.53 :

Corollaire II.62 : *Tout langage rationnel* R *contient un langage mince* R' *tel que* $\ell(R) = \ell(R')$.

Ce résultat ne peut pas s'étendre de façon analogue à la proposition II.54, même pour les langages appartenant à $C(Com_1)$. C'est-à-dire qu'il existe $L \in C(Com_1)$ tel que pour tout langage mince $R, \ell(L) \neq \ell(L \cap R)$. Considérons, en effet $L' = \{a^{2^n} \ / \ n \geq 0\}$, l'homomorphisme h défini sur $X = \{a_1, a_2\}$ par $h(a_1) = a$, $h(a_2) = \varepsilon$ et posons $L = h^{-1}(L') \cap w_1^* w_2^*$ avec $w_1 = a_1 a_2$, $w_2 = a_1 a_2 a_2$. S'il existe un langage mince R tel que $\ell(L) = \ell(L \cap R)$, pour tout $n \in \mathbb{N}_+$, $r = 2^n + 3.2^{n-1} \in \ell(L) = \ell(L \cap R)$, donc il existe $i, j \in \mathbb{N}$ tels que $w_1^i w_2^j \in R$ avec $2i + 3j = r$. Il est, alors, facile de vérifier que $i = j = 2^{n-1}$. Pour tout $n \in \mathbb{N}_+$, $w_1^p w_2^p \in R$ avec $p = 2^{n-1}$, ce qui implique que R n'est pas un langage mince.

Cependant nous pouvons énoncer une version affaiblie du lemme II.61, pour les langages appartenant à $C(Com_1)$, que nous utiliserons pour la démonstration du lemme suivant :

<u>Lemme II.63</u> : *Pour tout langage infini $L \in C(Com_1)$, il existe un langage mince R tel que $L \cap R$ soit infini.*

<u>Démonstration</u> : Prenons $L' \subseteq a^*$ et $L \in C^f(L') = C(L')$. Il existe un langage rationnel R' et deux homomorphismes alphabétiques h et g tels que $L = g(h^{-1}(L') \cap R')$ avec g limité sur R'.

Considérons d'abord le langage $L'' = h^{-1}(L') \cap R'$. Si $L' \cap h(R')$ est infini, prenons R'_1 un langage mince inclus dans R' tel que $h(R') = h(R'_1)$ (cf. lemme II.61). Alors $L' \cap h(R'_1) = h(h^{-1}(L') \cap R'_1)$ est infini ce qui entraine $h^{-1}(L') \cap R'_1$ infini. Si $L' \cap h(R') = h(L'')$ est fini, comme L" est infini, L" contient un facteur itérant et il existe un langage mince $R'_1 = u x^* v$ tel que $L'' \cap R'_1 = R'_1$ soit un langage infini.

Le langage $R = g(R'_1)$ est un langage mince et $L \cap R = g(L'') \cap g(R'_1)$ contient $g(L'' \cap R'_1)$ qui est infini puisque g est limité sur R' et donc sur R'_1.□

Pour tout $k \in \mathbb{N}_+$, $\mathcal{B}_k$ désignera la famille des langages qui sont finis ou qui contiennent un langage borné infini de dimension k. Démontrons, d'abord, un lemme de réduction qui indique que pour établir l'inclusion $\mathcal{F}_\sigma^f(\mathcal{L}) \subseteq \mathcal{B}_k$, il suffit de considérer les langages $L \cap R$ avec $L \in \mathcal{L}$ et $R \in Rat$:

<u>Lemme II.64</u> : *Pour tout $k \in \mathbb{N}_+$ et toute famille de langages $\mathcal{L}$, $\mathcal{L} \wedge Rat \subseteq \mathcal{B}_k$ implique $\mathcal{F}_\sigma^f(\mathcal{L}) \subseteq \mathcal{B}_k$.*

<u>Démonstration</u> : Comme $\mathcal{B}_k$ est clos par substitution non effaçante, il suffit de montrer que $C^f(\mathcal{L})$ est inclus dans $\mathcal{B}_k$. Prenons $L \in \mathcal{L}$, $R \in Rat$ et deux homomorphismes h et g avec g limité sur R et montrons que $L' = g(h^{-1}(L) \cap R)$ appartient à $\mathcal{B}_k$. Comme g est limité sur R, pour tout langage $L'' \subseteq h^{-1}(L) \cap R$, $g(L'')$ est infini ssi L" est infini. Il nous reste, donc, à montrer que $h^{-1}(L) \cap R \in \mathcal{B}_k$. Supposons $L_1 = h^{-1}(L) \cap R$ infini. Alors, si $h(L_1) = L \cap h(R)$ est fini, L_1 possède un facteur itérant et appartient à $\mathcal{B}_1 \subseteq \mathcal{B}_k$. Si $L \cap h(R)$ est infini, par hypothèse, il existe un langage infini borné de dimension k inclus dans $L \cap h(R)$, donc il existe $R' = x_o y_1^* \ldots y_k^* x_k \subseteq h(R)$ tel que $L \cap R'$ soit infini. Soit $M = (Q, \Delta, f, q_o, F)$ l'automate

d'état fini qui reconnait R. Pour q, q' $\in$ Q, posons $R_{q,q'} = \{w \in \Delta^* \;/\; f(q,w) = q'\}$. Il existe $q_1, q'_1, \ldots, q_k, q'_k \in Q$, $w_o, \ldots, w_k \in \Delta^*$ tels que $f(q_o, w_o) = q_1$, $f(q'_k, w_k) \in F$, $f(q'_i, w_i) = q_{i+1}$, $\forall\, i \in \{1,\ldots,k-1\}$, $h(w_i) = x_i$, $\forall\, i \in \{0,\ldots,k\}$ et $h^{-1}(L) \cap w_o R_1 \ldots R_k w_k$ soit infini avec $\forall\, j \in \{1,\ldots,k\}$, $R_j = h^{-1}(y_j^*) \cap R_{q_j,q'_j}$ ce qui entraine $R'_1 = w_o R_1 \ldots R_k w_k \subseteq R$. Posons $L_1 = \{y_1^{i_1} \;/\; \exists\, i_2,\ldots,i_k \in \mathbb{N}_+$ tels que $x_o y_1^{i_1} \ldots y_k^{i_k} \in L \cap h(R'_1)\}$ et $L'_1 = h^{-1}(L_1) \cap R_1$. Si L'_1 est fini, il existe $z_1 \in L'_1$ tel que $h^{-1}(L) \cap R'_2$ soit infini avec $R'_2 = w_o z w_1 R_2 \ldots R_k w_k$. Si L'_1 est infini, comme $L'_1 \in C(\mathrm{Com}_1)$, le lemme II.63 entraine qu'il existe $u x^* v \subseteq R_1$ tel que $h^{-1}(L_1) \cap u x^* v$ soit infini. Nous en déduisons donc que $h^{-1}(L) \cap R'_2$ est infini avec $R'_2 = w_o u x^* v w_1 R_2 \ldots R_k w_k$.

Pour R_2, on recommence le même raisonnement, en considérant $L_2 = \{y_2^{i_2} \;/\; \exists\, i_1, i_3 \ldots i_k$ tels que $x_o y_1^{i_1} \ldots y_k^{i_k} x_k \in L \cap h(R'_2)\}$ et $L'_2 = h^{-1}(L_2) \cap R_2$. En itérant ce procédé, on construit un langage rationnel R'_k borné de dimension k, inclus dans R et tel que $h^{-1}(L) \cap R'_k$ soit infini, donc $h^{-1}(L) \cap R \cap R'_k = h^{-1}(L) \cap R'_k$ est un langage infini borné de dimension k inclus dans $h^{-1}(L) \cap R$ qui appartient, alors, à $\mathcal{B}_k$.$\square$

Prenons un langage k-borné L. Pour tout $R \in$ Rat, $L \cap R$ est k-borné et donc $L \cap R$ est un langage borné de dimension k. Comme $C(\mathrm{Bor}_k) = C^f(\mathrm{Bor}_k)$ [25], nous déduisons du lemme précédent un résultat qui découle aussi de ceux de Goldstine [37] :

Corollaire II.65 : *Pour tout $k \in \mathbb{N}_+$ et pour tout langage infini $L \in F_\sigma(\mathrm{Bor}_k)$, il existe un langage infini borné de dimension k inclus dans L.*

Montrons, maintenant, que le corollaire précédent reste valable si on remplace Bor_k par Com_k.

Proposition II.66 : *Soit $k \in \mathbb{N}_+$. Tout langage infini $L \in F_\sigma(\mathrm{Com}_k)$ contient un langage infini borné de dimension k.*

Démonstration : Comme $C(\mathrm{Com}_k) = C^f(\mathrm{Com}_k)$ (corollaire II.45), il nous suffit, d'après le lemme II.64, de montrer que pour tout langage commutatif $L \subseteq T^*$ avec $T = \{a_1,\ldots,a_k\}$ et tout langage rationnel $R \subseteq T^*$, $L \cap R \in \mathcal{B}_k$. D'après la proposition II.54, il existe $w_1,\ldots,w_n \in T^*$ tels que $\Psi(L \cap R) = \Psi(L \cap R \cap w_1^* \ldots w_n^*)$ et donc, si $L \cap R$ est infini, il existe $R' = x_o y_1^* \ldots y_t^* x_t \subseteq R$ tel que $L \cap R'$ soit infini avec $y_i \neq \varepsilon$ $\forall\, i \in \{1,\ldots,t\}$. Montrons que si t est supérieur à k, il existe dans R', un langage R", rationnel borné de dimension t-1, tel que L $\cap$ R" soit infini. En effet, si t est supérieur à k, les vecteurs $\Psi(y_1),\ldots\Psi(y_t)$ appartenant à $\mathbb{N}^k$ sont linéairement dépendants. Il existe, donc, deux ensembles non vides, disjoints $I, J \subseteq \{1,\ldots,t\}$ tels que $\sum_{i\in I} \alpha_i \Psi(y_i) = \sum_{i\in J} \alpha_i \Psi(y_i)$ avec $\forall\, i \in I \cup J$, $\alpha_i \in \mathbb{N}_+$. Posons $A = \{(\lambda_1,\ldots,\lambda_t) \in \mathbb{N}^t \;/\; x_o y_1^{\lambda_1} \ldots y_t^{\lambda_t} x_t \in L\}$. Il existe alors $s \in I \cup J$ et un sous-ensemble infini A' de A tels que $\lambda_s / \alpha_s \leq \lambda_i / \alpha_i$, $\forall\, i \in I \cup J$. Supposons, par exemple que $s \in I$. Alors, si $(\lambda_1,\ldots,\lambda_t) \in A'$, λ_s peut s'écrire $q\,\alpha_s + r_s$ avec $0 \leq r_s < \alpha_s$ et $\forall\, i \in I$, $\lambda_i = q\,\alpha_i + r_i$ avec $r_i \in \mathbb{N}$. Donc $\sum_{i\in I} \lambda_i \Psi(y_i) =$

$q \sum_{i\in I} \alpha_i \Psi(y_i) + \sum_{i\in I} r_i \Psi(y_i) = q \sum_{i\in J} \alpha_i \Psi(y_i) + \sum_{i\in I} r_i \Psi(y_i)$ avec $r_s < \alpha_s$ et il existe $\mu_1,\ldots,\mu_t \in \mathbb{N}$ tels que $\Psi(x_o\, y_1^{\mu_1} \ldots y_t^{\mu_t}\, x_t) = \Psi(x_o\, y_1^{\lambda_1} \ldots y_t^{\lambda_t}\, x_t)$ avec $\mu_s = r_s < \alpha_s$. Nous en déduisons qu'il existe $r < \alpha_s$ tel que $R'' = x_o\, y_1^* \ldots y_{s-1}^*\, x_{s-1}\, y_s^r\, x_s\, y_{s+1}^* \ldots y_t^*\, x_t$ vérifie $L \cap R''$ infini avec R'' borné de dimension t-1 inclus dans R'. Ceci entraine, par induction, l'existence d'un langage rationnel borné de dimension k, $R_1 \subseteq R$ tel que $L \cap R_1 = L \cap R \cap R_1$ soit infini.□

Pour tout entier $k \geq 2$, le langage k-borné $E_k = \{a_1^n \ldots a_k^n \ / \ n \geq 0\}$ n'appartient pas à $\mathcal{B}_{k-1}$ donc $E_k \in C(\mathrm{Bor}_k) \setminus F_\sigma(\mathrm{Com}_{k-1})$ et nous en déduisons :

<u>Corollaire II.67</u> [37] *Pour tout $k \in \mathbb{N}_+$, $C(\mathrm{Bor}_k) \subsetneq C(\mathrm{Bor}_{k+1})$, $F(\mathrm{Bor}_k) \subsetneq F(\mathrm{Bor}_{k+1})$ et $F_\sigma(\mathrm{Bor}_k) \subsetneq F_\sigma(\mathrm{Bor}_{k+1})$.*

<u>Corollaire II.68</u> : *Pour tout $k \in \mathbb{N}_+$, $C(\mathrm{Com}_k) \subsetneq C(\mathrm{Com}_{k+1})$, $F(\mathrm{Com}_k) \subsetneq F(\mathrm{Com}_{k+1})$ et $F_\sigma(\mathrm{Com}_k) \subsetneq F_\sigma(\mathrm{Com}_{k+1})$.*

En particulier, $F(\mathrm{Com}) = \bigcup_{k\geq 1} F(\mathrm{Com}_k)$ est une FAL non principale. Enfin, comme E_k est égal à $c(a_1 \ldots a_k)^*) \cap a_1^* \ldots a_k^*$, E_k appartient à $F(c(\mathrm{Rat}))$ où $c(\mathrm{Rat}) = \{c(R) \ / \ R \in \mathrm{Rat}\}$. Ceci apporte, pour une conjecture de Ginsburg et Spanier [33], une autre démonstration que celle que nous avions donnée dans [54] :

<u>Corollaire II.69</u> : *La famille $F(c(\mathrm{Rat}))$ est une FAL non principale.*

Terminons cette section en considérant les langages bornés appartenant à $F_\sigma(L)$ où L est un langage commutatif. Si le langage L est non rationnel, $F_\sigma(L)$ contient nécessairement un langage borné non rationnel. Nous allons, cependant, pouvoir montrer, pour les langages commutatifs, un résultat du même type que le corollaire II.60 :

<u>Proposition II.70</u> : *Pour tout entier $k \geq 2$, il existe $L \in \mathrm{Com}_k$ tel que $F_\sigma(L)$ ne contient aucun langage non rationnel borné de dimension inférieure à k.*

<u>Démonstration</u> : Considérons le langage $\bar{L} = c(\{a_1^n\, a_2^{n^2} \ldots a_k^{n^k} \ / \ n \geq 0\})$ et posons $L = \{a_1,\ldots,a_k\}^* \setminus \bar{L}$. Montrons, d'abord, que pour $y_o, y_1,\ldots,y_s \in T^*$ avec $T = \{a_1,\ldots,a_k\}$ et $s < k$, $\bar{L} \cap y_o\, y_1^* \ldots y_s^*$ est fini. Si $\bar{L} \cap y_o\, y_1^* \ldots y_s^*$ est infini, comme le rapport $\sum_{i=1}^{k-1} n^i / n^k$ tend vers zéro quand n tend vers l'infini, il existe $j \in \{1,\ldots,s\}$ tel que $y_j \in a_k^*$. Si $k = 2$, $j = s = 1$ et nous avons tout de suite une contradiction. Si $k > 2$, raisonnons par récurrence et considérons l'homomorphisme h défini sur T par $h(x) = \varepsilon$ si $x = a_k$, x sinon. Alors $h(\bar{L} \cap y_o\, y_1^* \ldots y_s^*)$ est inclus dans $h(\bar{L}) \cap h(y_o\, y_1^* \ldots y_s^*)$ qui est fini d'après l'hypothèse de récurrence que l'on peut appliquer puisque $h(y_j) = \varepsilon$, d'où la contradiction car $h(\bar{L} \cap y_o\, y_1^* \ldots y_s^*)$ fini implique $\bar{L} \cap y_o\, y_1^* \ldots y_s^*$ fini.

Alors, pour tout langage rationnel R borné de dimension inférieure à k, $\bar{L} \cap R$ est fini et $L \cap R$ est un langage rationnel. Si $C(L)$ contient un langage non rationnel borné de dimension $s < k$, il existe deux homomorphismes alphabétiques f et g et un langage rationnel R' tels que $g(f^{-1}(L) \cap R')$ soit un langage non rationnel inclus dans $b_1^* \dots b_s^*$. Comme $f^{-1}(L)$ est un langage commutatif, on en déduit, par un raisonnement classique, l'existence d'un langage rationnel $R'' = R_1 \dots R_s$ tel que $L_1 = g(f^{-1}(L) \cap R'')$ soit un langage non rationnel inclus dans $b_1^* \dots b_s^*$ avec $\forall\, i \in \{1,\dots,s\}$, $g(R_i) \subseteq b_i^*$ et $R_i = x_{i,o}\, x_{i,1}^* \dots x_{i,t_i}^*$.

Distinguons deux cas :

a) Pour tout $i \in \{1,\dots,1\}$ et tout $j, j' \in \{1,\dots,t_i\}$, $\Psi(f(x_{i,j}))\times\ell(g(x_{i,j'})) = \Psi(f(x_{i,j'}))$ $\times\, \ell(g(x_{i,j}))$. Alors, il est facile de voir que chaque R_i peut être remplacé par une union finiede langages rationnels bornés de dimension 1. Donc, $g(f^{-1}(L) \cap R'') = g(f^{-1}(L) \cap R''_1)$ où R''_1 est un langage rationnel borné de dimension s. Comme s est inférieur à k, $g(f^{-1}(L) \cap R''_1) = g(f^{-1}(L \cap f(R''_1)) \cap R''_1)$ est rationnel puisque $L \cap f(R''_1)$ est rationnel, d'où la contradiction.

b) Il existe $i \in \{1,\dots,s\}$, $j, j' \in \{1,\dots,t_i\}$ (prenons, pour simplifier $i = j = 1$ et $j' = 2$) tels que $\Psi(f(x_{1,1}))\ \ell(g(x_{1,2}))$ soit différent de $\Psi(f(x_{1,2}))\ \ell(g(x_{1,1}))$. Il est, alors, clair qu'on peut trouver $m \in \mathbb{N}$ tel que si $z_1 z' \in f^{-1}(\bar{L})$ avec $z_1 \in R_1$ et $g(z_1) \in b_1^m b_1^*$, il existe $z_2 \in R_1$ tel que $g(z_2) = g(z_1)$ et $z_2 z' \notin f^{-1}(\bar{L})$. Nous pouvons en déduire que $L'_1 = L_1 \cap b_1^m b_1^* \dots b_s^*$ est égal à $g(R'') \cap b_1^m b_1^* \dots b_s^*$ qui est un langage rationnel et L_1 est égal à $L'_1 \cup A$ où $A = \bigcup_{0 \le i < m} (L_1 \cap b_1^i b_2^* \dots b_s^*)$ est une union finie de langages bornés de dimension $s - 1$ appartenant à $C(L)$. Si $s = 1$, A est fini et L_1 est un langage rationnel, d'où la contradiction. Si s est supérieur à 1, on obtient la même contradiction, en raisonnant par induction. Donc tout langage borné de dimension inférieure à k, appartenant à $C(L)$ est rationnel.

Si L_1 et L_2 sont des cônes rationnels vérifiant cette propriété, il est clair que $L_1 \square L_2$ la vérifie aussi ce qui entraine, par induction, qu'il en est de même pour $F_\sigma(L)$.$\square$

Pour tout entier k supérieur à 1, posons, comme dans [7], $S_{\neq}^{(k)} = \{a_1^{n_1} \dots a_k^{n_k} \mid n_j \in \mathbb{N},$ $\exists\, j\ 0 < j < k$ tel que $n_j \neq n_{j+1}\}$. En reprenant exactement le même raisonnement que celui que nous avons utilisé pour la démonstration de la proposition précédente, nous pouvons montrer que pour tout $k \in \mathbb{N}$, $F_\sigma(S_{\neq}^{(k+1)})$ ne contient aucun langage k-borné non rationnel. Nous retrouvons, ainsi, un résultat de [7] : Pour tout entier $k \geq 2$, $S_{\neq}^{(k)} \notin F(S_{\neq}^{(k+1)})$. Par contre, si nous considérons les fermetures commutatives de ces langages, le résultat est différent. Nous verrons, en effet, au chapitre suivant, que pour tout entier $k \geq 2$, les langages $c(S_{\neq}^{(k)})$ sont rationnellement équivalent à $\bar{D}_1^* = \{a,b\}^* \setminus D_1^*$.

CHAPITRE III

PSL-LANGAGES COMMUTATIFS

A - LE CÔNE RATIONNEL $C(C(RAT)$

B - LANGAGES EFFAÇABLES DANS L_Ψ

C - PSL-LANGAGES K-COMMUTATIFS

D - UN LANGAGE COMMUTATIF MINIMAL

E - LANGAGES SANS PRODUIT

F - SUR LES LANGAGES ALGÉBRIQUES COMMUTATIFS

A - LE CÔNE RATIONNEL C(c(RAT))

Dans cette section, nous allons étudier L_Ψ, le plus petit cône rationnel contenant la famille c(Rat) qui est égale à la famille des PSL-langages commutatifs (lemme III.1). Pour ce faire, nous montrons d'abord, que la famille c(Rat) vérifie les mêmes propriétés de clôture que la famille Com. Nous en déduisons que L_Ψ = C(c(Rat)) est fermé par intersection et ne contient que des PSL-langages. Ceci nous permet d'établir que L_Ψ est le plus petit cône rationnel commutativement clos et de caractériser les langages bornés appartenant à L_Ψ, en utilisant un résultat général, montré au chapitre I, concernant les langages bornés d'un cône rationnel ne contenant que des PSL-langages. Nous démontrons, ensuite, que tout langage de L_Ψ peut être obtenu par transductions rationnelles et intersections à partir du langage $D_1^* = c((ab)^*)$. Comme D_1^* est égal à shuf (C_1, C_1^R) avec $C_1 = D_1^* \cap a^* b^*$ et C_1^R, l'image miroir de C_1, $= D_1^* \cap b^* a^*$ (lemme III.11), nous obtenons : $L_\Psi = C_\cap(D_1^*) = C_\cap(C_1)$ où pour tout langage L, $C_\cap(L)$ désigne le plus petit cône rationnel clos par intersection et contenant L. Comme nous avons montré au chapitre précédent que L_Ψ = C(c(Rat)) était un cône rationnel non principal, nous prouvons l'existence d'un langage L(D_1^* ou C_1) tel que $C_\cap(L) = H(\wedge\ C(L))$ soit un cône rationnel non principal ce qui répond à une question de S. Ginsburg ([25], page 180).

Lemme III.1 : *Un langage* L *appartient à* c(Rat) *si et seulement si* L *est un PSL-langage commutatif.*

Démonstration : Si L = c(R) avec R $\in$ Rat, L est un langage commutatif et $\Psi(L) = \Psi(R)$ est un semi-linéaire.

Réciproquement, soit L un PSL-langage commutatif. Il existe un langage rationnel R tel que $\Psi(L) = \Psi(R)$ [33]. Alors $L = c(L) = \Psi^{-1} \circ \Psi(L) = \Psi^{-1} \circ \Psi(R) = c(R) \in c(Rat)$.□

Si L est un langage commutatif, $L = \Psi^{-1} \circ \Psi(L)$ et pour tout langage L', $\Psi(L \cap L') = \Psi(\Psi^{-1} \circ \Psi(L) \cap L') = \Psi(L) \cap \Psi(L')$. Si, de plus, L et L' sont des PSL-langages, $\Psi(L) \cap \Psi(L')$, intersection de deux semi-linéaires, est un semi-linéaire et nous obtenons :

Lemme III.2 : *Si* L *est un langage commutatif, pour tout langage* L', $\Psi(L) \cap \Psi(L') = \Psi(L \cap L')$. *De plus, si* L *et* L' *sont des PSL-langages,* $L \cap L'$ *est un PSL-langage.*

Si h est un homomorphisme alphabétique, h et c commutent ainsi que h^{-1} et c. Pour tout langage L = c(R) avec R $\in$ Rat, nous avons $h(c(R)) = c(h(R)) \in c(Rat)$ et $h^{-1}(c(R)) = c(h^{-1}(R)) \in c(Rat)$, donc :

Lemme III.3 : *La famille* c(Rat) *est close par homomorphisme (inverse) alphabétique.*

Du théorème de Nivat, nous pouvons, alors, déduire que $C(c(Rat)) = H_\alpha(c(Rat) \wedge Rat)$. D'après le lemme III.2, c(Rat) $\wedge$ Rat ne contient que des PSL-langages et comme il est clair que la famille de tous les PSL-langages est fermée par substitution, nous retrouvons :

<u>Corollaire III.4</u> [33] *La FAL engendrée par c(Rat) ne contient que des PSL-langages.*

Tout cône rationnel L contient la famille Rat, donc, si L est commutativement clos, L contient c(Rat). Réciproquement, d'après le corollaire précédent, pour tout langage $L \in L_\Psi$, c(L) est un PSL-langage commutatif donc appartient à $c(Rat) \subseteq L_\Psi$ qui est donc commutativement clos. Nous avons, ainsi, une nouvelle caractérisation de L_Ψ.

<u>Corollaire III.5</u> : *La famille $L_\Psi = C(c(Rat))$ est le plus petit cône rationnel commutativement clos.*

Etablissons, maintenant, pour c(Rat), les mêmes propriétés de clôture que pour la famille Com :

<u>Lemme III.6</u> : *La famille c(Rat) est close par union, intersection, homomorphisme alphabétique et homomorphisme inverse.*

<u>Démonstration</u> : D'après les lemmes II.3, III.2 et III.3, c(Rat) est clos par union, intersection et homomorphisme alphabétique. Si $L \in c(Rat)$, pour tout homomorphisme g, $g^{-1}(L) \in$ Com (lemme II.3) et comme $g^{-1}(L) \in L_\Psi$, le corollaire III.4 implique que $g^{-1}(L)$ est un PSL-langage, qui appartient à c(Rat), d'après le lemme III.1.□

Nous pouvons en déduire aisément un résultat établi par Ginsburg d'une toute autre façon :

<u>Corollaire III.7</u> [24] *La classe des ensembles semi-linéaires est fermée pour les applications linéaires inverses.*

<u>Démonstration</u> : Soient θ une application linéaire de $\mathbb{N}^p$ dans $\mathbb{N}^q$, $T = \{a_1,\dots,a_p\}$ et $\Delta = \{b_1,\dots,b_q\}$. Il existe un homomorphisme h de T^* dans Δ^* qui vérifie $\Psi_\Delta \circ h = \theta \circ \Psi_T$. Prenons un semi-linéaire S inclus dans $\mathbb{N}^q$. Le langage $L' = \Psi_\Delta^{-1}(S)$ est un PSL-langage commutatif donc appartient à c(Rat). D'après le lemme III.6, $h^{-1}(L')$ appartient à c(Rat) et $\theta^{-1}(S) = \Psi_T(h^{-1}(L'))$ est un semi-linéaire.□

Des lemmes II.2 et III.6, nous déduisons immédiatement :

<u>Proposition III.8</u> : *La famille L_Ψ est close par intersection, union et produit.*

Intéressons nous, maintenant, aux langages bornés de L_Ψ. Considérons un langage algébrique $L \subseteq w_1^* \dots w_n^*$, un alphabet $T = \{a_1,\dots,a_n\}$ et l'homomorphisme h défini sur T par : $h(a_i) = w_i \ \forall\, i \in \{1,\dots,n\}$.

Le langage $L' = h^{-1}(L) \cap a_1^* \dots a_n^*$ est algébrique et $\Psi(L')$ est un semi-linéaire. Donc $c(L') \in c(Rat)$, $L' = c(L') \cap a_1^* \dots a_n^* \in \mathcal{L}_\Psi$ ainsi que $L = h(L')$. Nous en déduisons que $\mathcal{L}_\Psi$ contient toute intersection finie de langages algébriques bornés et la proposition I.22 permet de caractériser les langages bornés de $\mathcal{L}_\Psi$.

Proposition III.9 : *Pour tout langage borné* L, *les trois propriétés suivantes sont équivalentes :*

i) $L \in \mathcal{L}_\Psi$,

ii) L *est une intersection finie de langages algébriques,*

iii) L *est l'image homomorphe (alphabétique) de l'intersection de deux langages algébriques bornés.*

Tout langage récursivement énumérable est l'image homomorphe de l'intersection de deux langages linéaires [3]. Montrons qu'on ne peut pas remplacer l'un de ces deux langages par un langage de $\mathcal{L}_\Psi$. Considérons, en effet, $\mathcal{L}_S$, la plus grande FAL ne contenant que des PSL-langages (cf. [33]).

Proposition III.10 : *La famille* $\mathcal{L}_S$ *est fermée par intersection avec les langages de* $\mathcal{L}_\Psi$.

Démonstration : Comme $\mathcal{L}_S$ est un cône rationnel et que $c(Rat)$ est clos par homomorphisme inverse, le lemme II.1 implique $\mathcal{L}_S \wedge \mathcal{L}_\Psi = \mathcal{L}_S \wedge \mathcal{C}(c(Rat)) \subseteq \mathcal{H}(\mathcal{L}_S \wedge c(Rat))$. D'après le lemme III.2, $\mathcal{L}_S \wedge c(Rat)$ ne contient que des PSL-langages. Donc $\mathcal{C}(\mathcal{L}_S \wedge \mathcal{L}_\Psi) = \mathcal{H}(\mathcal{L}_S \wedge c(Rat))$ ne contient que des PSL-langages ainsi que $\mathcal{F}(\mathcal{L}_S \wedge \mathcal{L}_\Psi)$ qui est, donc, par définition de $\mathcal{L}_S$, inclus dans $\mathcal{L}_S$.□

En particulier, comme $Alg \subseteq \mathcal{L}_S$, $\mathcal{F}(Alg \wedge \mathcal{L}_\Psi)$, qui est inclus dans $\mathcal{L}_S$, ne contient pas tous les langages récursivement énumérables.

Montrons, maintenant, que tout langage de $\mathcal{L}_\Psi$ peut être obtenu à partir d'un seul langage au moyen des opérations de transduction rationnelle et d'intersection.

Rappelons que C_1 désigne le langage $\{a^n b^n \mid n \geq 0\}$ et que D_1^*, le langage de Dyck sur une lettre est égal à $c(C_1)$.

Lemme III.11 : $\mathcal{C}_\cap(C_1) = \mathcal{C}_\cap(D_1^*) \subseteq \mathcal{L}_\Psi$.

Démonstration : Comme $C_1 = D_1^* \cap a^* b^* \in \mathcal{C}_\cap(D_1^*)$ et que $D_1^* = c((ab)^*) \in c(Rat) \subseteq \mathcal{L}_\Psi$, nous obtenons $\mathcal{C}(C_1) \subseteq \mathcal{C}_\cap(D_1^*) \subseteq \mathcal{L}_\Psi$. Il nous reste donc à montrer que $D_1^* \in \mathcal{C}_\cap(C_1)$. Pour cela, établissons l'égalité :

$$D_1^* = Shuf(C_1, C_1^R) \text{ où } C_1^R = \{b^n a^n \mid n \geq 0\} \in \mathcal{C}(C_1).$$

Il est clair que $L = \text{Shuf}(C_1, C_1^R)$ est inclus dans D_1^*. Réciproquement, prenons $x \in D_1^*$. Ce mot eut se factoriser en $x'\, x''$ avec $\ell(x') = \ell(x'') = n$. Soit p le nombre d'occurrences de a dans x'. Nous avons, alors, $\ell_a(x') = \ell_b(x'') = p$ et $\ell_a(x'') = \ell_b(x') = n - p$ et on peut vérifier facilement que $x \in \text{Shuf}\,(a^p b^p, b^{n-p} a^{n-p}) \subseteq L$. Comme tout cône rationnel clos par intersection est clos pour l'opération "shuffle" [25], nous obtenons bien $D_1^* = L \in C_\cap(C_1)$.□

Pour montrer que L_Ψ est inclus dans $C_\cap(D_1^*)$, il nous suffit d'établir $c(\text{Rat}) \subseteq C_\cap(D_1^*)$. Nous avons besoin de deux nouveaux lemmes :

<u>Lemme III.12</u> : *Pour tout mot w de T^*, $c(w^*)$ appartient à $C_\cap(D_1^*)$.*

<u>Démonstration</u> : Posons $w = w_1 \ldots w_p$ avec $w_i \in T$, $\forall\, i \in \{1,\ldots,p\}$. Si $p \leq 1$, $c(w^*) = w^* \in \text{Rat} \subseteq C_\cap(D_1^*)$. Dans le cas contraire, prenons p symboles distincts $b_1,\ldots,b_p$, posons $B = \{b_1,\ldots,b_p\}$ et $y = b_1 \ldots b_p$. Pour $j \in \{2,\ldots,p\}$, définissons l'homomorphisme h_j sur B par $h_j(b_1) = a$, $h_j(b_j) = b$, $h_j(x) = \varepsilon$ pour $x \in B \setminus \{b_1, b_j\}$ et montrons que $c(y^*) = L$ avec $L = \bigcap_{j=2}^{p} h_j^{-1}\,(D_1^*)$.

D'après le lemme III.6, L est un langage commutatif et il nous suffit de vérifier que $\Psi(L) = \Psi(c(y^*)) = \Psi(y^*)$.

Pour tout $z \in L$ et tout $j \in \{2,\ldots,p\}$, $z \in h_j^{-1}(D_1^*)$, donc $\ell_{b_1}(z) = \ell_{b_j}(z)$ et $\Psi(z) \in \Psi(y^*)$. Réciproquement, $\forall\, k \in \mathbb{N}$ et $\forall\, j \in \{2,\ldots,p\}$, $b_1^k \ldots b_p^k \in h_j^{-1}(D_1^*)$, donc $c(y^*) = L \in C_\cap(D_1^*)$.

De plus, comme l'homomorphisme g défini, sur B, par $h(b_j) = w_j$, $\forall\, j \in \{1,\ldots,p\}$ est alphabétique $c(w^*) = c \circ g(y^*) = g \circ c(y^*) \in C_\cap(D_1^*)$.□

<u>Lemme III.13</u> : *La famille $c(\text{Rat})$ est incluse dans $C_\cap(D_1^*)$.*

<u>Démonstration</u> : Soit R un rationnel inclus dans T^*.

Comme $C_\cap(D_1^*)$ est clos par union, on peut supposer, sans nuire à la généralité de la démonstration, que $\Psi(R)$ est un ensemble linéaire

$$S = L(u_o\ ;\ \{u_1,\ldots,u_k\}).$$

Pour $i \in \{0,\ldots,k\}$, il existe $w_i \in T^*$ tel que $\Psi(w_i) = u_i$, donc $c(R) = c(w_o w_1^* \ldots w_k^*)$. $c(w_o)$ est fini, donc appartient à $C_\cap(D_1^*)$ et les lemmes II.8 et III.12 impliquent $c(R) = c(w_o w_1^* \ldots w_k^*) \in C_\cap(D_1^*)$ qui est clos pour l'opération "shuffle".□

Introduisons, maintenant, pour tout $n \geq 1$, le langage "origin-crossing" de dimension n (cf. [22]) en posant $O_1 = c((a_1\bar{a}_1)^*)$ et pour tout $i \geq 2$, $O_i = \mathrm{shuf}(O_{i-1}, c((a_i\bar{a}_i)^*))$.

Nous obtenons, alors, le résultat principal de cette section.

Proposition III.14 : $L_\psi = C_\cap(C_1) = C_\cap(D_1^*) = \bigcup_{n\geq 1} C(O_n)$.

Démonstration : Les deux premières égalités proviennent des lemmes III.11 et III.13. D'autre part, d'après le théorème 5.5.1 de [25], nous avons pour tout $n \geq 1$,

$$H(C(O_n) \wedge C(D_1^*)) = C(\mathrm{shuf}(O_n, D_1^*)) = C(O_{n+1}).$$

Posons

$$L'_n = H(L_1 \wedge \ldots \wedge L_n) \text{ où } L_1 = \ldots = L_n = C(D_1^*).$$

Il est facile de vérifier par induction que, $\forall\, n \geq 1$, $L'_n = C(O_n)$ et comme $C_\cap(D_1^*) = \bigcup_{n\geq 1} L'_n$ (proposition I.11), nous obtenons le résultat cherché.□

D'après le corollaire II.69, le cône rationnel L_ψ n'est pas principal. Alors $L = C(D_1^*)$ est un cône rationnel principal tel que $H(\wedge L) = C_\cap(L)$ ne soit pas un cône rationnel principal. Nous apportons, ainsi, une réponse à une question posée par S. Ginsburg ([25], page 180).

B - LANGAGES EFFAÇABLES DANS L_Ψ

Nous avons donné au chapitre II, une condition suffisante pour que L/R appartienne à $C^f(L)$ dans le cas où L est un langage commutatif et R un langage rationnel (cf. lemme II.47). Cette condition, jointe à la proposition II.43, va nous permettre de montrer le résultat principal de cette section : Tout PSL-langage commutatif L est effaçable, c'est-à-dire vérifie $C(L) = C^f(L)$. Nous en déduisons que pour toute famille L de PSL-langages commutatifs, $C_\cap(L) = C_\cap^f(L)$ et nous obtenons, en particulier : $L_\Psi = C_\cap^f(D_1^*) = C_\cap^f(C_1)$ avec $C_1 = \{a^n b^n \ / \ n \geq 0\}$. Ce résultat va nous permettre, d'une part, de préciser un théorème de Baker et Book [3] concernant les automates non déterministes multi-compteur où chaque pile ne peut faire qu'un passage de l'empilage au dépilage (cf. corollaire III.18), d'autre part, à l'aide d'un théorème de Book, Nivat et Paterson [17], de montrer que tout langage de L_Ψ est l'image par un homomorphisme strictement alphabétique de l'intersection de trois langages algébriques linéaires. Enfin, nous terminons cette section en démontrant que tout langage infini appartenant à $C_\cap(\bar{D}_1^*)$, avec $\bar{D}_1^* = \{a,b\}^* \setminus D_1^*$, est effaçable et possède un facteur itérant.

<u>Proposition III.15</u> : *Tout PSL-langage commutatif L est effaçable, c'est-à-dire* $C^f(L) = C(L)$.

<u>Démonstration</u> : Considérons d'abord un langage commutatif $L \subseteq T^*$ tel que $\Psi(L)$ soit un ensemble linéaire et un mot $y \in T^+$. Le langage $L' = L/y^*$ est un PSL-langage commutatif et nous allons établir que $L' \in C^f(L)$. Il existe alors, $u_o \in T^*$, $u_1,\ldots,u_q \in T^+$ tels que $L = c(u_o\, u_1^* \ldots u_q^*)$. Le langage L' est égal à $L'_1 \cup \ldots \cup L'_n$ où $\forall\, i \in \{1,\ldots,n\}$, $L'_i = c(L'_i)$ et $\Psi(L'_i)$ est un ensemble linéaire. Considérons le langage $L'_1 = c(x_o\, x_1^* \ldots x_s^*)$ avec $x_o \in T^*$ et $\forall\, j \in \{1,\ldots,s\}$, $x_i \in T^+$. Comme $x_o\, x_1^*$ est inclus dans L/y^*, pour tout $i \in \mathbb{N}$, il existe $r_i \in \mathbb{N}$ tel que $x_o\, x_1^i\, y^{r_i} \in c(u_o\, u_1^* \ldots u_q^*)$, donc il existe $\lambda_{i,1},\ldots,\lambda_{i,q} \in \mathbb{N}$ tels que $x_o\, x_1^i\, y^{r_i} \in c(u_o\, u_1^{\lambda_{i,1}} \ldots u_q^{\lambda_{i,q}})$. Comme tout sous-ensemble infini de $\mathbb{N}^{q+1}$ possède deux éléments comparables, il existe $j, j' \in \mathbb{N}$ tels que $j < j'$ et $z_j \leq z_{j'}$, où $\forall\, i \in \mathbb{N}$, z_i désigne le vecteur $(r_i, \lambda_{i,1},\ldots,\lambda_{i,q})$. Donc il existe un entier positif $\lambda_1 = j' - j$ et $\mu_1 = r_{j'} - r_j \in \mathbb{N}$ tels que $x_1^{\lambda_1}\, y^{\mu_1} \in c(u_1^* \ldots u_q^*)$. En raisonnant de la même façon pour les mots $x_2,\ldots,x_s$, on montre que, $\forall\, t \in \{1,\ldots,s\}$, il existe $\lambda_t \in \mathbb{N}_+$ et $\mu_t \in \mathbb{N}$ tels que $x_t^{\lambda_t}\, y^{\mu_t} \in c(u_1^* \ldots u_q^*)$. L'ensemble $I = \{(i_1,\ldots,i_s) \ / \ \forall\, t \in \{1,\ldots,s\},\ 0 \leq i_t < \lambda_t\}$ est fini, et comme L'_1 est inclus dans L/y^*, il existe $k'_1 \in \mathbb{N}$ vérifiant : $\forall\, (i_1,\ldots,i_s) \in I$, $\exists\, p \leq k'_1$ tel que $x_o\, x_1^{i_1} \ldots x_s^{i_s}\, y^p \in L$.

Montrons la propriété :

$$(*)\ \exists\, k \in \mathbb{N},\ \forall\, x \in L'_1,\ \exists\, w \in y^* \text{ tel que } x\, w \in L \text{ et } \ell(w) \leq k(\ell(x) + 1).$$

Posons, en effet, $k = \ell(y) \sup\{k'_1, \mu_1,\ldots,\mu_s\}$ et prenons $x \in L'_1$. Il existe, alors, $j_1,\ldots,j_s \in \mathbb{N}$ tels que $x \in c(x_o\, x_1^{j_1} \ldots x_s^{j_s})$. Pour tout $t \in \{1,\ldots,s\}$, j_t peut s'écrire $\lambda_t\, j'_t + j''_t$

avec $0 \le j''_t < \lambda_t$ donc $(j''_1,\dots,j''_s) \in I$ et il existe un entier $p \le k'_1$ tel que $x_o\, x_1^{j''_1} \dots x_s^{j''_s}\, y^p \in L$. De plus, $\forall\, t \in \{1,\dots,s\}$, $x_t^{j'_t\lambda_t}\, y^{j'_t\mu_t}$ appartient à $c(u_1^* \dots u_q^*)$, donc $x\, y^\alpha \in L$ avec $\alpha = p + \sum_{t=1}^{s} j'_t\, \mu_t$. Alors $\ell(y^\alpha) \le \ell(y)\, k'_1 + k \sum_{t=1}^{s} j'_t \le \ell(y)\, k'_1 + k\, \ell(x) \le k(\ell(x) + 1)$.

On peut raisonner de la même façon pour $L'_2,\dots,L'_n$, donc la propriété (*) reste vraie si on remplace L'_1 par $L' = L/y^* = L'_1 \cup \dots \cup L'_n$. Le lemme II.47 implique alors $L' \in C^f(L)$.

Si L est un PSL-langage commutatif quelconque, $L = \bigcup_{i=1}^{m} L_i$ où $\forall\, i \in \{1,\dots,m\}$, L_i est un PSL-langage commutatif et $\Psi(L_i)$ est un ensemble linéaire. Donc $L/y^* = (\bigcup_{i=1}^{m} L_i) / y^* = \bigcup_{i=1}^{m} (L_i / y^*)$ et pour tout $i \in \{1,\dots,m\}$, il existe $\alpha_i \in \mathbb{N}$ vérifiant : $\forall\, x \in L_i / y^*$, $\exists\, w \in y^*$ tel que $x\, w \in L_i$ et $\ell(w) \le \alpha_i (\ell(x) + 1)$. Il est alors, facile de montrer que $\alpha = \sup\, \{\alpha_i \,/\, i \in \{1,\dots,m\}\}$ vérifie : $\forall\, x \in L / y^* = \bigcup_{i=1}^{m} (L_i / y^*)$, $\exists\, w \in y^*$ tel que $x\, w \in L = \bigcup_{i=1}^{m} L_i$ et $\ell(w) \le \alpha(\ell(x) + 1)$. Le lemme II.47 entraine de nouveau $L/y^* \in C^f(L)$.

Prenons enfin, un langage rationnel $R \subseteq T^*$. Comme $L/(R_1 \cup R_2) = L/R_1 \cup L/R_2$ et que tout langage rationnel est une union finie de langages rationnels dont l'image par la fonction de Parikh est un ensemble linéaire [55], nous pouvons supposer $c(R) = c(y_o\, y_1^* \dots y_u^*)$ avec $y_o \in T^*$, $y_1,\dots,y_u \in T^+$. Donc $L/R = L/(y_o\, y_1^* \dots y_u^*)$ et comme $L/(R_1\, R_2) = (L/R_2) / R_1$, nous en déduisons, par induction, le résultat énoncé.□

Nous pouvons déduire de cette proposition un certain nombre de résultats intéressants :

<u>Corollaire III.16</u> : *Pour toute famille $\mathcal{L}$ de PSL-langages commutatifs $C_\cap(\mathcal{L}) = C_\cap^f(\mathcal{L})$, le plus petit cône rationnel fidèle clos par intersection et contenant $\mathcal{L}$.*

<u>Démonstration</u> : Remarquons, d'abord que $\mathcal{L} \subseteq \mathcal{L}_\Psi$, le plus petit cône rationnel rationnel commutativement clos, implique $C_\cap(\mathcal{L}) \subseteq C_\cap(\mathcal{L}_\Psi) = \mathcal{L}_\Psi$, donc $C_\cap(\mathcal{L})$ ne contient que des PSL-langages. Tout langage $L \in C_\cap(\mathcal{L})$ s'écrit $h(L_1 \cap \dots \cap L_n)$ où h est un homomorphisme et $\forall\, i \in \{1,\dots,n\}$, $L_i \in C_\cup^f(\mathcal{L})$. D'après le corollaire II.13, pour tout $i \in \{1,\dots,n\}$, il existe un langage commutatif $L'_i \in C_\cup^f(\mathcal{L})$ tel que $L_i \in C^f(L'_i)$. On peut supposer que les langages L'_i sont définis sur des alphabets disjoints. Alors $L' = \text{shuf}\,(L'_1,\dots,L'_n)$ est un PSL-langage commutatif appartenant à $C_\cap^f(\mathcal{L})$ et tel que $L = h(L_1 \cap \dots \cap L_n) \in C(L')$ (cf. [25]). La proposition précédente implique, alors, $C(L') = C^f(L')$ et $L \in C^f(L') \subseteq C_\cap^f(\mathcal{L})$.□

Prenons, en particulier, $\mathcal{L} = \{D_1^*\}$ avec D_1^*, le langage de Dyck sur une lettre, égal à $c((ab)^*)$. Comme $D_1^* = \text{shuf}\,(C_1, C_1^R)$ avec $C_1 = \{a^n\, b^n \,/\, n \ge 0\}$ et $C_1^R = \{b^n\, a^n \,/\, n \ge 0\}$, $D_1^* \in C_\cap^f(C_1)$ et $C_\cap(D_1^*) = C_\cap^f(D_1^*) = C_\cap^f(C_1)$.

Corollaire III.17 : *Le plus petit cône rationnel commutativement clos, L_ψ, est égal au plus petit cône rationnel fidèle clos par intersection et contenant le langage $C_1 = \{a^n b^n \mid n \geq 0\}$.*

Le cône rationnel $C(C_1)$ est la famille des langages reconnus par pile vide, par un automate non déterministe à compteur (l'alphabet de pile ne contient qu'un élément) qui ne peut plus empiler après avoir dépilé. (cf. [39]). Le langage C_1 est effaçable ([41], [44] et proposition II.40), c'est-à-dire que $C(C_1) = C^f(C_1)$ ou encore que tout langage $L \in C(C_1)$ peut être reconnu par un automate du type précédent qui travaille en temps réel (à chaque mouvement de l'automate, on lit une lettre sur le ruban d'entrée). La famille $C_\cap(C_1)$ est égal à l'ensemble des langages reconnus par piles vides, par un automate non déterministe multi-compteur où, pour chaque pile, on ne peut plus empiler après avoir dépilé [3]. De même, en utilisant les techniques de [30], on peut montrer que $L \in C^f_\cap(C_1)$ ssi il est reconnu par un automate de type précédent qui travaille en temps réel. L'égalité entre $C_\cap(C_1)$ et $C^f_\cap(C_1)$, qui découle du corollaire III.17, nous permet d'énoncer un résultat qui améliore un théorème de Baker et Book ([3] théorème 5) :

Corollaire III.18 : *Pour tout automate non déterministe multi-compteur où, pour chaque pile, on ne peut plus empiler après avoir dépilé, il existe un automate de même type qui accepte le même langage et qui travaille en temps réel.*

D'autre part, comme $C_1 \in Lin$ et que $Lin \subsetneq L_\psi$, nous en déduisons :

Corollaire III.19 : *La famille L_ψ est incluse strictement dans $C^f_\cap(Lin)$.*

En utilisant, alors, les résultats de Book, Nivat et Paterson [17], concernant la famille $C^f_\cap(Lin)$, nous obtenons :

Corollaire III.20 : *Tout langage de L_ψ est l'image par un homomorphisme strictement alphabétique de l'intersection de trois langages algébriques linéaires.*

Le langage L considéré lors de la démonstration de la proposition II.37 appartient clairement à L_ψ. Donc il existe des langages non effaçables dans $C_\cap(D_1^*)$. Par contre, nous allons montrer que tout langage de $C_\cap(\bar{D}_1^*)$ est effaçable avec $\bar{D}_1^* = \{a,b\}^* \setminus D_1^*$.

Pour tout $n \in \mathbb{N}_+$, considérons l'alphabet $\Delta_n = \{a_1, \bar{a}_1, \ldots, a_n, \bar{a}_n\}$ et définissons le langage $\bar{O}_n$ en posant $\bar{O}_1 = h_1(\bar{D}_1^*)$ et $\forall\ i \geq 2$, $\bar{O}_i = \text{shuf}\ (\bar{O}_{i-1}, h_i(\bar{D}_1^*))$ où h_i est l'homomorphisme défini sur $\{a, b\}$ par $h_i(a) = a_i$ et $h_i(b) = \bar{a}_i$. Alors, d'après la proposition I.12 , $C_\cap(\bar{D}_1^*) = \bigcup_{i \geq 1} C(\bar{O}_i)$. Montrons, d'abord, une propriété du langage $\bar{O}_n$:

Lemme III.21 : *Soient $n \in \mathbb{N}_+$, u, x, v des mots appartenant à Δ^*_n. Alors $u\, x^*\, v \cap \bar{D}_n$ est non vide ssi il existe $i \leq n$ tel que $u\, x^i\, v \in \bar{D}_n$.*

Démonstration : Pour tout $w \in \Delta^*_n$ et tout $i \in \{1,\dots,n\}$, notons $d_i(w) = \ell_{a_i}(w) - \ell_{\bar{a}_i}(w)$. Supposons que $\forall\, j \leq n$, $u\, x^j\, v \notin \bar{D}_n$. Alors, $\forall\, j \in \{0,\dots,n\}$, il existe $i_j \in \{1,\dots,n\}$ tel que $d_{i_j}(u\, x^j\, v) = 0$, donc il existe $j \neq j' \in \{0,\dots,n\}$ tel que $i_j = i_{j'} = t$. Nous en déduisons que $d_t(u\, x^j\, v) = d_t(u\, x^{j'}\, v)$ ce qui entraine $d_t(x) = 0$ et $d_t(u\, v) = 0$ puisque $j \neq j'$ et $d_t(u\, x^p\, v) = 0$, $\forall\, p \in \mathbb{N}$, donc $u\, x^*\, v \cap \bar{D}_n = \emptyset$.□

Lemme III.22 : *Soient $L \subseteq T^*$ un langage appartenant à $C_\cap(\bar{D}_1^*)$ et $\Delta \subseteq T$. Il existe un entier k vérifiant : Si $u\, x\, v \in L$ avec $x \in \Delta^*$, $\exists\, y \in \Delta^*$ tel que $u\, y\, v \in L$ et $\ell(y) < k$.*

Démonstration : Si $L \in C_\cap(\bar{D}_1^*)$, il existe $n \in \mathbb{N}_+$, un alphabet Y, un langage rationnel $R \subseteq Y^*$ et deux homomorphismes alphabétiques h et g tels que $L = g(h^{-1}(\bar{D}_n) \cap R)$. Posons $Z = \{y \in Y \;/\; g(y) \in \Delta \cup \{\varepsilon\}\}$ et considérons un automate d'état fini $M = (Q, Y, f, q_o, F)$ qui reconnait R. Si $u\, x\, v \in L$ avec $x \in \Delta^*$, il existe $u', v' \in Y^*$, $x' \in Z^*$ tels que $h(u'\, x'\, v') \in \bar{D}_n$, $u'\, x'\, v' \in R$, $g(u') = u$, $g(x') = x$ et $g(v') = v$. Notons $q = f(q_o, u')$, $q' = f(q, x')$ et $R' = \{w \in Z^* / f(q, w) = q'\}$. Le langage R' est un langage rationnel qui contient x', donc il existe $x_o, x_1,\dots,x_t \in Z^*$ tels que $\Psi(x') \in \Psi(x_o\, x_1^* \dots x_t^*) \subseteq \Psi(R')$. Il existe, donc, $i_1,\dots,i_t \in \mathbb{N}$ tels que $\Psi(x') = \Psi(x_o\, x_1^{i_1} \dots x_t^{i_t})$ et comme $\bar{D}_n$ est commutatif, il contient $h(u')\, v_o\, v_1^{i_1} \dots v_t^{i_t}\, h(v')$ où $\forall\, j \in \{0,\dots,t\}$, $v_j = h(x_j)$. D'après le lemme précédent, il existe $j_1,\dots,j_t$ tels que $h(u')\, v_o\, v_1^{j_1} \dots v_t^{j_t}\, h(v') \in \bar{D}_n$ avec $\forall\, s \in \{1,\dots,t\}$, $j_s \leq n$. Alors, il existe $y' \in R'$ tel que $\Psi(y') = \Psi(x_o\, x_1^{j_1} \dots x_t^{j_t})$ et, comme $h^{-1}(\bar{D}_n)$ est commutatif, $u'\, y'\, v' \in h^{-1}(\bar{D}_n) \cap R$. Donc $y = g(y')$ vérifie : $y \in \Delta^*$, $u\, y\, v \in L$ et, puisque g est alphabétique, $\ell(y) \leq \ell(y') \leq n\ell(x_o\, x_1 \dots x_t)$. Etant donné que $x_o, x_1,\dots,x_t$ ne dépendent que de M, Δ, q et q', il est clair qu'on peut trouver pour $\ell(y)$ une borne k, ne dépendant que de L et Δ.□

Proposition III.23 : *Tout langage infini appartenant à $C_\cap(\bar{D}_1^*)$ est effaçable et possède un facteur itérant.*

Démonstration : Prenons $L \in C(\bar{D}_n)$ et $L' \in C(L)$. Alors $L' = g(h^{-1}(L) \cap R)$ où R est un langage rationnel défini sur l'alphabet Y et h, g sont des homomorphismes alphabétiques. Posons $L'' = h^{-1}(L) \cap R$ et $\Delta = \{y \in Y \;/\; g(y) = \varepsilon\}$. Comme $L'' \in C_\cap(\bar{D}_1^*)$, le lemme précédent entraine qu'il existe $k \in \mathbb{N}$ tel que $L' = g(L'') = g(L'' \cap (Y^* \setminus Y^*\, \Delta^k\, Y^*)) = g(h^{-1}(L) \cap R')$ avec $R' = R \cap (Y^* \setminus Y^*\, \Delta^k\, Y^*)$. Donc g est k-limité sur R', $L' \in C^f(L)$ et L est effaçable.

Soit $H = g_1(g_2^{-1}(\bar{D}_n) \cap K)$ où K est un rationnel défini sur l'alphabet Z et g_1, g_2 sont des homomorphismes. Si H est infini, il existe $u, x, v \in Z^*$ tels que $g_2(u\, x\, v) \in \bar{D}_n$, $u\, x^*\, v \subseteq K$ et $g_1(x) \neq \varepsilon$. Pour tout $i \in \{1,\dots,n\}$ et tout $w \in \Delta^*_n$, $D_i(w)$ désignera la valeur absolue de $\ell_{a_i}(w) - \ell_{\bar{a}_i}(w)$. Alors, posons $t = \sup \{D_i(g_2(u\ v)) \ /\ i \in \{1,\dots,n\}\}$ et montrons que $g_1(u\ x^{t+1})\ (g_1(x))^*\ g_1(v)$ est inclus dans H. En effet, si $D_i(g_2(x)) = 0$, comme $g_2(u\ x\ v) \in \bar{O}_n$, $D_i(g_2(u\ v)) > 0$ ce qui implique $D_i(g_2(u\ x^j\ v)) > 0$, $\forall\ j \in \mathbb{N}$. par contre, si $D_i(g_2(x)) > 0$, $\forall\ j > t$, $D_i(g_2(x^j)) > t \geq D_i(g_2(u\ v))$ et $D_i(g_2(u\ x^j\ v)) > 0$. Donc, $\forall\ j > t$, $u\ x^j\ v \in g_2^{-1}(\bar{O}_n) \cap K$ et $g_1(u)\ g_1(x)^j\ g_2(v) \in H$ qui contient un facteur itérant puisque $g_1(x) \neq \varepsilon$.□

Nous en déduisons immédiatement :

<u>Corollaire III.24</u> : *La famille $C_\cap(\bar{D}_1^*)$ est incluse strictement dans L_ψ.*

Tout langage algébrique non rationnel inclus dans $a^*\ b^*$ est rationnellement supérieur à l'un des trois langages $L_{\neq} = \{a^n b^p \ /\ n \neq p\}$, $L_> = \{a^n b^p \ /\ n > p\}$ et $L_< = \{a^n b^p \ /\ n < p\}$ qui sont deux à deux rationnellement incomparables (cf. Berstel et Boasson [7]). Si nous considérons les clôtures commutatives de ces langages, $\bar{D}_1^* = c(L_{\neq})$, $L_1 = c(L_>)$ et $L_2 = c(L_<)$, nous obtenons $C(\bar{D}_1^*) \subseteq C(L_1) = C(L_2)$ et comme, clairement, $D_1^* \in C_\cap(L_1)$, l'inclusion précédente est stricte et nous avons $C_\cap(\bar{D}_1^*) \subsetneq C_\cap(L_1) = C_\cap(L_2) = L_\psi$. Par contre $C_\cap(L_>) \subsetneq L_\psi$ puisque $F_\cap(L_>)$ ne contient que des langages récursifs [1] alors que $F_\cap(D_1^*)$ contient tous les langages récursivement énumérables.

C - PSL-LANGAGES K-COMMUTATIFS

Pour tout $k \in \mathbb{N}_+$, nous désignerons par Rat_k la famille des langages rationnels définis sur un alphabet de k lettres. Alors, $c(Rat_k)$ est la famille des PSL-langages k-commutatifs. Le cône rationnel $L_\Psi = C(c(Rat)) = \bigcup_{k\geq 1} C(c(Rat_k))$ est non principal. On peut se poser la question de la principalité des cônes $C(c(Rat_k))$. Pour $k = 2$, les travaux de Berstel [6], Berstel et Boasson [7] concernant les langages algébriques inclus ans $a^* b^*$, permettent de montrer que $C(c(Rat_2))$ n'est autre que le cône rationnel engendré par le langage D_1^*. Pour k supérieur à 2, nous pouvons seulement montrer que $C(c(Rat_k))$ est inclus dans le cône rationnel engendré par E'_{k+1} qui est égal à la fermeture commutative de $(a_1 \ldots a_{k+1})^*$. Nous démontrons, ensuite, que si $L = c(w_o\, w_1^* \ldots w_n^*)$ est un PSL-langage k-commutatifs tel que les vecteurs $\Psi(w_1),\ldots,\Psi(w_n)$ sont linéairement indépendants, il existe k langages commutatifs $L_1,\ldots,L_k \in C(D_1^*)$ vérifiant $L = \bigcap_{i=1}^{k} L_i$ Nous en déduisons, alors, le résultat principal de cette section : tout PSL-langage commutatif est une intersection finie de langages commutatifs de $C(D_1^*)$. Nous retrouvons, ainsi en particulier, un résultat énoncé par Perot dans [71] : tout PSL-langage commutatif est une intersection finie de langages commutatifs algébriques. Ce résultat impliquant clairement une propriété démontrée par Liu et Weiner dans [61] : Tout ensemble semi-linéaire est une intersection finie d'ensembles semi-linéaires stratifiés.

Proposition III.25 : *Pour tout PSL-langage commutatif non rationnel* $L \subseteq \{a, b\}^*$, *on a* $C(\bar{D}_1^*) \subseteq C(L) \subseteq C(D_1^*)$ *et donc* $C(c(Rat_2)) = C(D_1^*)$.

Démonstration : Le langage $L' = L \cap a^* b^*$ est un langage algébrique non rationnel. D'après [6], $L' \in C(C_1)$ donc $L' \in C(D_1^*)$ et le corollaire II.11 implique $L = c(L') \in C(D_1^*)$ ce qui entraine $C(L) \subseteq C(D_1^*)$.

Comme L' est non rationnel, C(L') contient l'un des trois langages $L_{\neq}$, $L_>$, $L_<$ définis à la fin de la section précédente [7]. D'après le corollaire II.11, $C(L) = C(c(L'))$ contient l'un des trois langages $\bar{D}_1^* = c(L_{\neq})$, $L_1 = c(L_>)$, $L_2 = c(L_<)$. Comme $\bar{D}_1^* = L_1 \cup L_2 \in C(L_1) = C(L_2)$, nous avons $\bar{D}_1^* \in C(L)$ et $C(D_1^*) \subseteq C(L)$.□

Pour tout $k \in \mathbb{N}_+$, posons $E'_k = c((a_1 \ldots a_k)^*)$. Alors, $C(E'_2) = C(D_1^*) = C(c(Rat_2))$. Pour k supérieur à 2, nous obtenons un résultat moins précis. Montrons d'abord :

Lemme III.26 : *Soit w un mot défini sur l'alphabet* $\Delta = \{b_1,\ldots,b_k\}$. *Alors* $c(w^*)$ *appartient à* $C(E'_k)$.

Démonstration : Pour tout $i \in \{1,\ldots,k\}$, posons $t_i = \ell_{b_i}(w)$ et considérons l'homomorphisme h défini sur l'alphabet $\{a_1,\ldots,a_k\}$ par : $\forall\, i \in \{1,\ldots,k\}$, $h(a_i) = b_i^{t_i}$. Le langage $L = h(E'_k) \cap b_1^* \ldots b_k^* \in C(E'_k)$, est inclus dans $b_1^* \ldots b_k^*$ et il est clair que $\Psi(L) = \Psi(w^*)$. On en déduit alors, d'après le corollaire II.11 que $c(w^*) = c(L) \in C(E'_k)$.□

Si on remplace w^* par un langage rationnel défini sur le même alphabet on ne sait pas si le lemme précédent reste valable pour $k > 2$. On obtient cependant :

Proposition III.27 : *Soit R un langage rationnel défini sur l'alphabet* $\Delta = \{b_1,\ldots,b_k\}$. *Alors* $c(R)$ *appartient à* $C(E'_{k+1})$.

Démonstration : Comme $C(E'_{k+1})$ est clos par union, on peut supposer que $\Psi(R)$ est un ensemble linéaire. Donc, il existe $w_o, w_1,\ldots,w_p \in \Delta^*$ tels que $c(R) = c(w_o\, w_1^* \ldots w_p^*)$. Posons $w = w_1 \ldots w_p\, b_{k+1}$ et pour $i \in \{1,\ldots,p\}$, $w'_i = w_1 \ldots w_{i-1}\, w_{i+1} \ldots w_p\, b_{k+1}$. Considérons les homomorphismes h et g définis sur $Y = \{a_1,\ldots,a_k, b_1,\ldots,b_{k+1}\}$ par : $\forall\, i \in \{1,\ldots,k\}$, $h(a_i) = h(b_i) = b_i$, $g(a_i) = b_i$, $g(b_i) = \varepsilon$ et $h(b_{k+1}) = b_{k+1}$, $g(b_{k+1}) = \varepsilon$. Montrons que le langage $L = g(h^{-1}(c(w^*) \cap R))$ où $R = a_1^* \ldots a_k^*\, w'^*_1 \ldots w'^*_p$ vérifie $c(L) = c(w_1^* \ldots w_p^*)$.

Prenons $y \in c(w_1^* \ldots w_p^*)$. Il existe $i_1,\ldots,i_p \in \mathbb{N}$ tels que $y \in c(w_1^{i_1} \ldots w_p^{i_p})$. Posons $i = \sum_{j=1}^{p} i_j$. Comme $c(w_i) = c(w_1^i \ldots w_p^i\, b_k^i)$ et que $c(w'^{i_1}_1 \ldots w'^{i_p}_p) = c(w_1^{i-i_1} \ldots w_p^{i-i_p}\, b_k^i)$, il est facile de vérifier que $h^{-1}(c(w^i)) \cap a_1^* \ldots a_k^*\, w'^{i_1}_1 \ldots w'^{i_p}_p$ est égal à $a_1^{s_1} \ldots a_k^{s_k}\, w'^{i_1}_1 \ldots w'^{i_p}_p$ où $\forall\, j \in \{1,\ldots,k\}$, $s_j = \ell_{b_j}(w_1^{i_1} \ldots w_p^{i_p}) = \ell_{b_j}(y)$. Donc $y \in c(b_1^{s_1} \ldots b_k^{s_k})$ avec $b_1^{s_1} \ldots b_k^{s_k} = g(a_1^{s_1} \ldots a_k^{s_k}\, w'^{i_1}_1 \ldots w'^{i_p}_p) \in L$.

Réciproquement si $y \in c(L)$, il existe $s_1,\ldots,s_k, i_1,\ldots,i_p, t \in \mathbb{N}$ tels que $y \in c(b_1^{s_1} \ldots b_k^{s_k})$ et $z = h(a_1^{s_1} \ldots a_k^{s_k} w'^{i_1}_1 \ldots \quad \ldots w'^{i_p}_p) \in c(w^t)$. En particulier, $\ell_{b_{k+1}}(z) = \ell_{b_{k+1}}(w^t)$ ce qui entraine $t = \sum_{j=1}^{p} i_j$. Il est, alors, clair que $y \in c(w_1^{i_1} \ldots w_p^{i_p})$.

Comme L est inclus dans $b_1^* \ldots b_k^*$, le corollaire II.11 implique $c(L) \in C(c(w^*))$ et comme w est défini sur l'alphabet $\{b_1,\ldots,b_{k+1}\}$ le lemme précédent entraine que $c(w^*) \in C(E'_{k+1})$ donc $c(L) \in C(E'_{k+1})$. La démonstration se termine en remarquant que $c(R) = \text{shuf}\,(c(w_o), c(L)) \in C(c(L))$.□

Nous en déduisons que L_Ψ peut s'écrire $\bigcup_{k\geq 1} C(E'_k)$. Il est possible de comparer les langages E'_k avec les langages O_k définis à la section A de ce chapitre. Nous avons en effet :

Proposition III.28 : *Pour tout* $k \in \mathbb{N}_+$, $C(c(Rat_k)) \subsetneq C(E'_{k+1}) \subseteq C(O_k) \subseteq C(E'_{2k})$.

Démonstration : D'après la proposition précédente $C(c(Rat_k))$ est inclus dans $C(E'_{k+1})$ et comme $E'_{k+1} \cap a_1^* \ldots a_{k+1}^*$ ne contient aucun langage infini borné de dimension k, la proposition II.63 entraine $E'_{k+1} \notin C(c(Rat_k))$ et donc $C(c(Rat_k)) \subsetneq C(E'_{k+1})$.

Montrons, maintenant, que $E'_{k+1} \in C(O_k)$. Comme O_k est un langage commutatif, il suffit, d'après le corollaire II.11, de montrer que $E_{k+1} = E'_{k+1} \cap a_1^* \dots a_{k+1}^*$ appartient à $C(O_k)$. Raisonnons par récurrence sur k. Si $k = 1$, il est clair que $E_2 \in C(O_1)$. Le langage E_{k+2} est égal à $E_{k+1}\, a_{k+2}^* \cap a_1^* \dots a_k^*$ $\{a_{k+1}^n\, a_{k+2}^n \ /\ n \geq 0\}$. L'hypothèse de récurrence implique $E_{k+1}\, a_{k+2}^* \in C(O_k)$, donc $E_{k+2} \in C(O_k) \wedge C(D_1^*)$ qui est inclus, par construction, dans $C(O_{k+1})$.

Si $k = 1$, on a, $C(O_1) = C(E'_2)$. Supposons que $C(O_k)$ soit inclus dans $C(E'_{2k})$. Alors, O_{k+1}, qui est rationnellement équivalent à shuf (O_k, D_1^*), appartient à $H(C(O_k) \wedge C(D_1^*)) \subseteq H(C(E'_{2k}) \wedge C(D_1^*)) = C(L)$ avec $L = \mathrm{shuf}(E'_{2k}, D_1^*)$. Or, il est clair que $C(L) = C(E'_{2k+2} \ /\ R) \subseteq C(E'_{2k+2})$ où R est le langage rationnel $(a_1 \dots a_{2k})^* \ (a_{2k+1}\, a_{2k+2})^*$.□

Nous avons montré dans la section A de ce chapitre, que $c(\mathrm{Rat}) \subseteq C_\cap(D_1^*)$, c'est-à-dire que pour tout $R \in \mathrm{Rat}$, il existe des langages $L_1,\dots,L_p \in C(D_1^*)$ et un homomorphisme h tels que $c(R) = h(L_1 \cap \dots \cap L_p)$. Nous allons, maintenant, préciser ce résultat, en montrant que l'homomorphisme h peut se réduire à l'identité et que les langages $L_1,\dots,L_p$ peuvent être choisis commutatifs. Pour obtenir ce résultat, nous allons utiliser un raisonnement analogue à celui de Liu et Weiner dans [61] mais rendu plus simple par l'emploi d'un lemme de Ginsburg [24], concernant les ensembles linéaires propres qui sont les ensembles linéaires de la forme $L(u_o \ ;\ P)$ où P est un ensemble de vecteurs linéairement indépendants.

<u>Lemme III.29 [24]</u> : *Tout ensemble semi-linéaire est une union finie d'ensembles linéaires propres.*

En fait, ce résultat a été, depuis, précisé par Eilenberg et Schützenberger [20] d'une part et Ito [49] d'autre part, qui ont montré que l'union pouvait être prise disjointe.

Considérons, d'abord, les langages appartenant à $c(\mathrm{Rat}_n)$ et dont l'image par la fonction de Parikh est un ensemble linéaire propre :

<u>Lemme III.30</u> : *Soient un alphabet $T = \{a_1,\dots,a_n\}$ et un langage rationnel $R \subseteq T^*$. Si $S = \Psi(R)$ est un linéaire propre, il existe n langages commutatifs $L_1,\dots,L_n$ appartenant à $C(D_1^*)$ tels que $c(R) = \bigcap_{i=1}^{n} L_i$.*

<u>Démonstration</u> : Si S est fini, $c(R)$ est fini et appartient à $C(D_1^*)$. Il suffit alors de prendre $L_1 = \dots = L_n = c(R)$.

Si S est infini, il existe p vecteurs linéairement indépendants $(1 \leq p \leq n)$, $x_1,\dots,x_p \in \mathbb{N}^n$ et $x_o \in \mathbb{N}^n$ tels que $S = L(x_o \ ;\ \{x_1,\dots,x_p\})$. Posons $S' = L(0 \ ;\ \{x_1,\dots,x_p\})$ et considérons la matrice A à p lignes et n colonnes où pour tout $i \in \{1,\dots,p\}$, la $i^{\text{ème}}$ ligne de A, notée A_i est égale à x_i. Comme les lignes de A sont linéairement indépendantes, A possède une sous-matrice de rang p et on peut supposer sans nuire à la généralité de la démonstration que la matrice carrée A' obtenue à partir de A en ne gardant que les p premières colonnes est de rang p et donc régulière. Notons A", la matrice formée des n-p dernières colonnes de A. Nous avons, alors, $S' = \{\lambda A \ /\ \lambda \in \mathbb{N}^p\} = \{(y_1,\dots,y_n) \in \mathbb{N}^n \ /\ (y_1,\dots,y_p)\, A'^{-1} \in \mathbb{N}^p$ et

$(y_{p+1},\dots,y_n) = (y_1,\dots,y_p)\ A'^{-1}\ A''\}$. Pour $i \in \{1,\dots,p\}$, notons $S_i = \{(y_1,\dots,y_n) \in \mathbb{N}^n\ /\ (y_1,\dots,y_p)\ (A'^{-1})^i \in \mathbb{N}\}$ ($(A'^{-1})^i$ désigne la $i^{\text{ème}}$ colonne de la matrice A'^{-1}). Pour $i \in \{p+1,\dots,n\}$, notons $S_i = \{(y_1,\dots,y_n) \in \mathbb{N}^n\ /\ (y_1,\dots,y_p)\ (A'^{-1}\ A'')^{i-p} = y_i\}$. Comme A est une matrice à éléments dans $\mathbb{N}$ et que A' est inversible, les éléments des matrices A'^{-1} et $A'^{-1}\ A''$ sont des nombres rationnels. Alors, il est facile de vérifier que pour tout $i \in \{1,\dots,n\}$, $L'_i = \Psi^{-1}(S_i)$ est un langage commutatif appartenant à $C(D_1^*)$. D'autre part, il est clair que $S' = \bigcap_{i=1}^{n} S_i$, donc $L' = \Psi^{-1}(S') = \Psi^{-1}(\bigcap_{i=1}^{n} S_i) = \bigcap_{i=1}^{n} \Psi^{-1}(S_i) = L'_1 \cap \dots \cap L'_n$. Le langage $c(R)$ est égal à shuf $(\Psi^{-1}x_o), L') =$ shuf $(\Psi^{-1}(x_o), \bigcap_{i=1}^{n} L'_i)$ et comme les langages L'_i sont commutatifs, L est égal à $\bigcap_{i=1}^{n} L_i$ où $\forall\ i \in \{1,\dots,n\}, L_i =$ shuf $(\Psi^{-1}(x_o), L'_i)$. Etant donné que $\Psi^{-1}(x_o)$ est un langage commutatif, pour tout $i \in \{1,\dots,n\}$, L_i est un langage commutatif appartenant à $C(L'_i) \subseteq C(D_1^*)$.□

Pour tout langage rationnel R, $\Psi(c(R)) = \Psi(R)$ est un ensemble semi-linéaire. Donc, d'après le lemme III.29, $\Psi(R)$ est une union finie de linéaires propres. Alors, en utilisant la distributivité de l'union par rapport à l'intersection et le fait que $C(D_1^*)$ soit clos par union, nous pouvons déduire du lemme précédent :

Proposition III.31 : *Tout PSL-langage commutatif est une intersection finie de langages commutatifs appartenant à $C(D_1^*)$.*

Comme $C(D_1^*)$ est inclus dans Alg, cette proposition permet, en particulier de retrouver un résultat énoncé par Perrot dans [71].

Considérons, maintenant, un PSL-langage inclus dans $a_1^* \dots a_n^*$ où les a_i sont des lettres. Comme L est égal à $c(L) \cap a_1^* \dots a_n^*$, la proposition précédente implique :

Corollaire III.32 : *Tout PSL-langage strictement borné est une intersection finie de langages bornés appartenant à $C(D_1^*)$.*

Par contre si L est un PSL-langage inclus dans $w_1^* \dots w_n^*$ où les w_i sont des mots quelconques, L n'appartient pas nécessairement à $\mathcal{L}_\Psi = C_\cap(D_1^*)$ puisque $C(L)$ peut contenir des langages qui ne sont pas des PSL-langages (cf. [52]). Si on ajoute l'hypothèse que $L \in \mathcal{L}_\Psi$, l'égalité $L = c(L) \cap w_1^* \dots w_n^*$ n'est pas toujours vérifiée puisque l'homomorphisme h défini par $h(a_i) = w_i$, $\forall\ i \in \{1,\dots,n\}$, peut ne pas être injectif. Cependant, comme il existe un rationnel $R \subseteq \{a_1,\dots,a_n\}^*$ qui est en bijection par h avec $w_1^* \dots w_n^*$ [24], en utilisant le même raisonnement que dans le chapitre I (cf. lemme I.19), on obtient :

Corollaire III.33 : *Un langage borné L appartient à $\mathcal{L}_\Psi$ ssi L est une intersection finie de langages bornés appartenant à $C(D_1^*)$.*

Enfin, nous avons montré au chapitre I que tout langage borné appartenant à $\mathcal{L}_\psi$ (ou à un PSL-cône rationnel) était l'image par un homomorphisme alphabétique de l'intersection de deux langages algébriques bornés. Montrons que ces deux langages ne peuvent pas toujours être choisis dans $C(D_1^*)$. Plus précisément, $\forall$ k $\in \mathbb{N}_+$, il existe des langages bornés appartenant à $\mathcal{L}_\psi$ qui ne sont pas images homomorphes de l'intersection de k langages bornés de $C(D_1^*)$. Pour tout n $\in \mathbb{N}_+$, posons :

$BC_n = \{h(L_1 \cap \ldots \cap L_n)$ / h est un homomorphisme et $\forall$ i $\in \{1,\ldots,n\}$ L_i est un langage borné de $C(D_1^*)\}$.

Montrons que pour tout entier positif n, le langage $L_{n+1} = O_{n+1} \cap a_1^* \bar{a}_1^* \ldots a_{n+1}^* \bar{a}_{n+1}^*$ n'appartient pas à BC_n. Comme $C_n(D_1^*) = \bigcup_{n\geq 1} C(O_n)$ est non principal, le langage O_{n+1} n'appartient pas à $C(O_n)$ ce qui implique, d'après la proposition II.14, $L_{n+1} \notin C(O_n)$. D'autre part, $C(O_n) = \mathcal{H}(\mathcal{L}_1 \wedge \ldots \wedge \mathcal{L}_n)$ avec $\mathcal{L}_1 = \ldots = \mathcal{L}_n = C(D_1^*)$, donc BC_n est inclus dans $C(O_n)$ et $L_{n+1} \in BC_{n+1} \setminus BC_n$. Nous pouvons, alors, énoncer :

<u>Proposition III.34</u> : *Pour tout n $\in \mathbb{N}_+$, la famille BC_n est incluse strictement dans la famille BC_{n+1}.*

D - Un langage commutatif minimal

Quand on étudie les transductions rationnelles, un problème intéressant et toujours non résolu se pose : existe-t-il un langage minimal, c'est-à-dire un langage L tel que tout langage non rationnel de C(L) soit un générateur de ce cône ? Nous allons montrer, dans cette section, l'existence d'un tel langage pour la famille des langages commutatifs. Plus précisément, nous établissons qu'il n'existe pas de langage commutatif L vérifiant $\mathrm{Rat} \subsetneq C(L) \subsetneq C(\bar{D}_1^*)$ où $\bar{D}_1^* = \{a,b\}^* \setminus D_1^*$. Pour obtenir ce résultat, nous donnerons, d'abord, une caractérisation compléte des langages commutatifs de $C(\bar{D}_1^*)$ (lemme III.35).

Comme tout langage non rationnel inclus dans a^* est rationnellement incomparable avec $\bar{D}_1^*$, il est clair que ce langage n'est pas un langage commutatif minimum pour la transduction rationnelle. Par contre, nous conjecturons que tout PSL-langage commutatif non rationnel domine rationnellement le langage $\bar{D}_1^*$. Ce résultat serait d'une très grande utilité puisqu'on aurait, alors, pour tout PSL-cône rationnel $\mathcal{L}$: $\mathcal{L}$ contient un langage commutatif non rationnel si et seulement si $\bar{D}_1^*$ appartient à $\mathcal{L}$. En particulier, il suffirait de montrer que $\bar{D}_1^*$ n'appartient pas à $F_\sigma(\mathcal{L}in)$ pour montrer que tout langage commutatif quasi-rationnel est rationnel.

<u>Lemme III.35</u> : *Soit L un langage commutatif inclus dans* $\{a_1,\ldots,a_k\}^*$. *Alors* $L \in C(\bar{D}_1^*)$ *ssi* $L = \bigcup_{i=1}^{n} (L_i \cap X_i)$ *où* $\forall\, i \in \{1,\ldots,n\}$, X_i *est un langage rationnel commutatif et* $L_i = \Psi^{-1}(\{x \in \mathbb{N}^k \,/\, x(u_i)^T \neq \beta_i\})$ *avec* $\beta_i \in \mathbb{Z}$ *et* $(u_i)^T$ *désignant le vecteur transposé de* $u_i \in \mathbb{Z}^k$.

<u>Démonstration</u> : Si $L \in C(\bar{D}_1^*)$, $L' = L \cap a_1^* \ldots a_k^* \in C(\bar{D}_1^*)$ et vérifie $\Psi(L) = \Psi(L')$. Il existe, alors, un alphabet Y, un langage rationnel $R \subseteq Y^*$ et deux homomorphismes alphabétiques h et g tels que $L' = g(h^{-1}(\bar{D}_1^*) \cap R)$ avec $g(R) \subseteq a_1^* \ldots a_k^*$. Pour tout $i \in \{1,\ldots,k\}$, notons $Y_i = \{y \in Y \,/\, g(y) \in \{a_i, \varepsilon\}\}$. Alors, R est une union finie de produits $R_1 \ldots R_k$ où, $\forall\, i \in \{1,\ldots,k\}$, R_i est inclus dans Y_i^* et $\Psi(R_i)$ est un ensemble linéaire.

Considérons un tel produit et le langage $L'_1 = g(h^{-1}(\bar{D}_1^*) \cap R_1 \ldots R_k)$. Comme $\Psi(R_1)$ est un ensemble linéaire, il existe $u_0, u_1,\ldots,u_p \in Y_1^*$ tels que $R'_1 = u_0 u_1^* \ldots u_p^*$ vérifie $c(R_1) = c(R'_1)$ et comme $h^{-1}(\bar{D}_1^*)$ et $g(R_1) \subseteq a_1^*$ sont des langages commutatifs, on obtient $L'_1 = g(h^{-1}(\bar{D}_1^*) \cap R'_1 R_2 \ldots R_k)$. Pour tout mot $w \in \{a,b\}^*$, notons $d(w) = \ell_a(w) - \ell_b(w)$ et montrons que le langage $R''_1 = \{x \in R'_1 \,/\, \exists\, x' \in R'_1$ tel que $g(x') = g(x)$ et $d(h(x)) \neq d(h(x'))\}$ est un langage rationnel. En effet, posons $V = \{(i_1,\ldots,i_p) \in \mathbb{N}^p \,/\, u_0 u_1^{i_1} \ldots u_p^{i_p} \in R''_1\}$ et V_m l'ensemble des éléments minimaux de V ($v \in V$, $v' \in V_m$ et $v \leq v' \Rightarrow v = v'$). Comme $V_m \subseteq \mathbb{N}^p$ ne contient que des éléments incomparables, V_m est fini ([24], corollaire 5.4.1). D'autre part, pour tout $v \in V_m$, $L_v = \{u_0 u_1^{i_1} \ldots u_p^{i_p} \,/\, (i_1,\ldots,i_p) \geq v\}$ est un langage rationnel et il est facile de vérifier que $(j_1,\ldots,j_p) \geq v$ entraine $(j_1,\ldots,j_p) \in V$. Donc $R''_1 = \bigcup_{v \in V_m} L_v$ est un langage rationnel.

Montrons, maintenant, que $L''_1 = g(h^{-1}(\bar{D}_1^*) \cap R''_1 R_2 \dots R_k)$ est égal à $g(R''_1 R_2 \dots R_k)$. Il est clair que L''_1 est inclus dans $g(R''_1 \dots R_k)$. Réciproquement, prenons $w = g(x z)$ avec $x \in R''_1$ et $z \in R_2 \dots R_k$. Par construction de R''_1, il existe $x' \in R''_1$ tel que $g(x') = g(x)$ et $d(h(x')) \neq d(h(x))$ ce qui entraine que l'un des deux mots $h(x z)$, $h(x' z)$ appartient à $\bar{D}_1^*$ et $w = g(x z) = g(x' z) \in L''_1$. Donc $L''_1 = g(R''_1 \dots R_k)$ est un langage rationnel. Alors, $L'_1 = L''_1 \cup L'''_1$ avec $L'''_1 = g(h^{-1}(\bar{D}_1^*) \cap R'''_1 R_2 \dots R_k)$ où $R'''_1 = R'_1 \setminus R''_1$ est un langage rationnel qui vérifie :

$$(*)\ \forall\, x, x' \in R'''_1,\ g(x) = g(x') \Rightarrow d(h(x)) = d(h(x')).$$

Le langage R'''_1 est une union finie de langages rationnels dont les images par la fonction de Parikh sont des ensembles linéaires. Considérons l'un de ces langages. En raisonnant de la même façon que ci-dessus, on peut supposer qu'il est égal à $v_o\, v_1^* \dots v_q^*$ où $\forall\, i \in \{0,\dots,q\}$, $v_i \in Y_1^*$. Pour tout $i \in \{0,\dots,q\}$, posons $t_i = \ell(g(v_i))$ et $s_i = d(h(v_i))$. Pour tout $i, j \in \{1,\dots,q\}$, $s_i\, t_j = s_j\, t_i$. En effet, dans le cas contraire, considérons les mots $x = v_o\, v_i^{t_j}$, $x' = v_o\, v_j^{t_i}$ avec $s_i\, t_j \neq s_j\, t_i$. Alors $g(x) = g(x') = a_1^{t'}$ avec $t' = t_o + t_i\, t_j$, mais par contre $d(h(x)) = s_o + s_i\, t_j \neq s_o + s_j\, t_i = d(h(x'))$, d'où la contradiction avec la propriété $(*)$, puisque x et x' appartiennent à R'''_1. De même si $t_i = 0$, v_o et $v_o\, v_i$ ont la même image par g et la propriété $(*)$ entraine $s_i = 0$. Posons, alors, $Q = \{i \in \{1,\dots,q\} \ /\ t_i \neq 0\}$. Si Q est vide, prenons $U_1 = \{v_o\}$. Dans le cas contraire, soit t le plus grand commun diviseur de l'ensemble $\{t_i \ /\ i \in Q\}$. Il existe $\lambda_1,\dots,\lambda_q \in \mathbb{Z}$ tels que $t = \sum_{i=1}^{q} \lambda_i\, t_i$. Alors, $s = \sum_{i=1}^{q} \lambda_i\, s_i$ vérifie : $\forall\, i \in \{1,\dots,q\}$, $s\, t_i = s_i\, t$. De plus, il existe un ensemble fini $I \subseteq \mathbb{N}$ et $i_o \in \mathbb{N}$ tels que $\{\sum_{i=1}^{q} \alpha_i\, t_i \ /\ \alpha_i \in \mathbb{N}\} = \{\alpha t \mid \alpha \in I\} \cup \{(i_o + \alpha)t \ /\ \alpha \in \mathbb{N}\}$. Donc, pour tout $x \in v_1^* \dots v_q^*$, il existe un entier α tel que $\ell(g(x)) = \alpha t$ et $d(h(x)) = \alpha s$ avec $\alpha \in I \cup \{i \ /\ i \geq i_o\}$. Prenons alors, $v \in Y_1^*$ tel que $g(v) = a_1^t$ et $d(h(v)) = s$ et posons $U_1 = \{v_o\, v^i \ /\ i \in I\} \cup v_o\, v^{i_o}\, v^*$. Il est clair que U_1 est un langage rationnel qui vérifie : $g(h^{-1}(\bar{D}_1^*) \cap U_1 R_2 \dots R_k) = g(h^{-1}(\bar{D}_1^*) \cap v_o\, v_1^* \dots v_q^* R_2 \dots R_k)$.

En recommençant les mêmes constructions pour $R_2,\dots,R_k$, on peut, ainsi, montrer que L' est égal à l'union finie de langages rationnels et de langages qui peuvent se mettre sous la forme $g(h^{-1}(\bar{D}_1^*) \cap x_1\, y_1^* \dots x_k\, y_k^*)$ avec $\forall\, i \in \{1,\dots,k\}$, $g(x_i\, y_i^*) \subseteq a_i^*$ et $g(y_i) = \varepsilon \Rightarrow y_i = \varepsilon$. Soit S un tel langage. Pour tout $i \in \{1,\dots,k\}$, posons $m'_i = d(h(x_i))$, $m_i = d(h(y_i))$, $n'_i = \ell(g(x_i))$ et $n_i = \ell(g(y_i))$. Il est, alors, facile de vérifier que $\Psi(S)$ est égal à $\{(n'_1 + \lambda_1\, n_1,\dots,n'_k + \lambda_k\, n_k) \ /\ \lambda_i \in \mathbb{N},\ \sum_{j=1}^{k} (m'_j + \lambda_j\, m_j) \neq 0\}$. Posons $Q' = \{i \in \{1,\dots,k\} \ /\ n_i > 0\}$. Si Q' est vide, $y_1 = \dots = y_k = \varepsilon$ et $g(h^{-1}(\bar{D}_1^*) \cap x_1 \dots x_k)$ est un langage fini, donc rationnel. Si Q' est non vide, soit r le plus petit commun multiple des éléments de Q' et considérons le vecteur $z = (z_1,\dots,z_k) \in \mathbb{Z}^k$ avec $\forall\, i \in \{1,\dots,k\}$, $z_i = r\, m_i / n_i$ si $n_i > 0$, r sinon. Comme le produit scalaire $(n'_1 + \lambda_1\, n_1,\dots,n'_k + \lambda_k\, n_k)\, z^T$ est égal à $\sum_{j=1}^{k} n'_j\, z_j + r \sum_{j=1}^{k} \lambda_j\, m_j$, on en déduit que $S = a_1^{n'_1}\, (a_1^{n_1})^* \dots a_k^{n'_k}\, (a_k^{n_k})^* \cap \Psi^{-1}(\{x \in \mathbb{N}^k \ /\ x\, z^T \neq z_o\})$ avec $z_o = \sum_{j=1}^{k} (n'_j\, z_j - r\, m'_j)$. Enfin, si on prend $e = (1,\dots,1) \in \mathbb{N}^k$, le langage $\Psi^{-1}(\{x \in \mathbb{N}^k \ /\ x\, e^T \neq -1\}$ est égal à $\{a_1,\dots,a_k\}^*$ et pour tout langage rationnel X, $X = X \cap \Psi^{-1}(\{x \in \mathbb{N}^k \ /\ x\, e^T \neq -1\})$. Donc

$L' = \bigcup_{i=1}^{n} (X'_i \cap L_i)$ où $\forall\ i \in \{1,\ldots,n\}$, X'_i est un langage rationnel inclus dans $a_1^* \ldots a_k^*$ et $L_i = \Psi^{-1}(\{x \in \mathbb{N}^k \ /\ x(u_i)^T \neq \beta_i\})$ et $L = c(L') = \bigcup_{i=1}^{n} (X_i \cap L_i)$ où $X_i = c(X'_i)$ est un langage rationnel commutatif.

Pour montrer la réciproque, il suffit d'établir que pour tout $z = (z_1,\ldots,z_k) \in \mathbb{Z}^k$ et $z_o \in \mathbb{Z}$, le langage $L = \Psi^{-1}(\{x \in \mathbb{N}^k \ /\ x\, z^T \neq z_o\})$ appartient à $C(\bar{D}_1^*)$. Il est clair que $L' = \{w \in \{a,b\}^* \ /\ \ell_a(w) - \ell_b(w) \neq z_o\}$ est rationnellement équivalent à $\bar{D}_1^*$. Donc $C(\bar{D}_1^*)$ contient $L = h^{-1}(L')$ où l'homomorphisme h est défini sur $\{a_1,\ldots,a_k\}$ par $h(a_i) = a^{z_i}$ si $z_i \geq 0$, b^{-z_i} sinon.□

Nous allons utiliser cette caractérisation des langages commutatifs appartenant à $C(\bar{D}_1^*)$ pour montrer le résultat principal de cette section :

<u>Proposition III.36</u> : *Tout langage L commutatif non rationnel, appartenant à $C(\bar{D}_1^*)$ est rationnellement équivalent à $\bar{D}_1^*$.*

<u>Démonstration</u> : Soit $L \subseteq \{a_1,\ldots,a_k\}^*$ un langage commutatif non rationnel appartenant à $C(\bar{D}_1^*)$. Nous allons raisonner par recurrence sur k. Si k = 2, le résultat se déduit immédiatement de la proposition III.25. Supposons donc k supérieur à 2.

D'après le lemme précédent, $L' = L \cap a_1^* \ldots a_k^*$ peut s'écrire $\bigcup_{i=1}^{n} (L_i \cap X'_i)$ où X'_i est un langage rationnel inclus dans $a_1^* \ldots a_k^*$ et $L_i = \Psi^{-1}(\{x \in \mathbb{N}^k \ /\ x(u_i)^T \neq \beta_i\})$ avec $u_i \in \mathbb{Z}^k$ et $\beta_i \in \mathbb{Z}$. On peut supposer, sans nuire à la généralité de la démonstration que pour i, $j \in \{1,\ldots,n\}$, soit $X'_i \cap X'_j = \emptyset$, soit $X'_i = X'_j$. Tout rationnel $R \subseteq a_1^* \ldots a_k^*$ peut s'écrire $\bigcup_{i=1}^{t} R_{i,1} \ldots R_{i,k}$ où $\forall\ i \in \{1,\ldots,t\}$, $\forall\ j \in \{1,\ldots,k\}$, $R_{i,j}$ est un langage rationnel inclus dans a_j^*. On peut supposer que, soit $R_{i,j} = R_{i',j}$, soit $R_{i,j} \cap R_{i',j} = \emptyset$. Enfin, comme tout langage rationnel $A \subseteq a^*$ s'écrit $\{a^i \ /\ i \in F\} \cap \{a^{t_o+j+\lambda d} \ /\ j \in J, \lambda \in \mathbb{N}\}$ avec $F \subseteq \{0,\ldots,t_o - 1\}$ $J \subseteq \{0,\ldots,d - 1\}$, t_o, $d \in \mathbb{N}$, on en déduit que tout langage rationnel $R \subseteq a_1^* \ldots a_n^*$ est une union finie disjointe de langages rationnels dont l'image par la fonction de Parikh est un ensemble linéaire propre, c'est-à-dire de la forme $x_1\, y_1^* \ldots x_k\, y_k^*$ avec $\forall\ i \in \{1,\ldots,k\}$, x_i, $y_i \in a_i^*$.

Il existe, donc, des langages rationnels $Y_1,\ldots,Y_q$, disjoints deux à deux, dont les images par Ψ sont des ensembles linéaires et tels que tout X'_i soit une union de tels langages. Comme L' est non rationnel et que $L' = L' \cap (Y_1 \cup \ldots \cup Y_q) = \bigcup_{i=1}^{q} (L' \cap Y_i)$, il existe i tel que $L' \cap Y_i$ soit non rationnel. Pour tout $j \in \{1,\ldots,n\}$, soit $Y_i \cap X_j = \emptyset$, soit $Y_i \subseteq X_j$, donc il existe $I = \{j \in \{1,\ldots,n\} \ /\ Y_i \subseteq X_j\}$ tel que $L' \cap Y_i = L'' \cap Y_i$ avec $L'' = \bigcup_{i \in I} L_i$. Il existe x_1, $y_1 \in a_1^*,\ldots,x_k$, $y_k \in a_k^*$ tels que $Y_i = x_1\, y_1^* \ldots x_k\, y_k^*$. Considérons les homomorphismes h et g définis sur $\{a_1, a'_1,\ldots,a_k, a'_k\}$ par : $\forall\ i \in \{1,\ldots,k\}$, $h(a_i) = y_i$, $h(a'_i) = x_i$, $g(a_i) = a_i$ et $g(a'_i) = \varepsilon$, ainsi que le langage

$K = g(h^{-1}(L'' \cap Y_i) \cap a'_1\, a_1^* \ldots a'_k\, a_k^*) = \{a_1^{i_1} \ldots a_k^{i_k} \,/\, x_1\, y_1^{i_1} \ldots x_k\, y_k^{i_k} \in L''\}$. Comme $L'' \cap Y_i$ est égal à $h(g^{-1}(K) \cap a'_1\, a_1^* \ldots a'_k\, a_k^*)$, K est un langage non rationnel. Posons, pour tout $i \in \{1,\ldots,k\}$, $s_i = \ell(x_i)$ et $t_i = \ell(y_i)$. Alors, $K = \bigcup_{i \in I} K_i$ avec $\Psi(K_i) = \{(i_1,\ldots,i_k) \,/\, (s_1 + i_1\, t_1,\ldots, s_k + i_k\, t_k)\,(u_i)^T \neq \beta_i\} = \{x \in \mathbb{N}^k \,/\, x(u'_i)^T \neq \beta'_i\}$ où $u_i = (u_{i,1},\ldots,u_{i,k})$, $u'_i = (t_1\, u_{i,1},\ldots,t_k\, u_{i,k})$ et $\beta'_i = \beta_i - \sum_{j=1}^{k} s_j\, u_{i,j}$. Supposons $I = \{1,\ldots,p\}$ et considérons la matrice A à k lignes et p colonnes où la $i^{\text{ème}}$ colonne, notée A^i, est égale à $(u'_i)^T$ ainsi que le vecteur $\beta' = (\beta'_1,\ldots,\beta'_p)$. Alors le langage $\bar{K} = a_1^* \ldots a_k^* \setminus K$ vérifie : $\Psi(\bar{K}) = \{x \in \mathbb{N}^k \,/\, x\, A = \beta'\}$. Distinguons plusieurs cas :

A) Le langage $\bar{K}$ posséde un facteur itérant. Il existe $j \in \{1,\ldots,k\}$, $t \in \mathbb{N}_+$, w_1, $w_2 \in \{a_1,\ldots,a_k\}^*$ tels que $w_1 (a_j^t)^*\, w_2 \subseteq \bar{K}$ ce qui entraine que la $j^{\text{ème}}$ ligne de la matrice A, notée A_j ne contient que des zéros. Donc $\bar{K} = \text{shuf}\,(h_j(\bar{K}), a_j^*) \cap a_1^* \ldots a_k^*$ où h_j est l'homomorphisme défini sur $\{a_1,\ldots,a_k\}$ par $h_j(a_i) = \varepsilon$ si $i = j$, a_i sinon. De même $K = \text{shuf}(h_j(K), a_j^*) \cap a_1^* \ldots a_k^*$ ce qui entraine que $h_j(K)$ est un langage non rationnel inclus dans $a_1^* \ldots a_{j-1}^*\, a_{j+1}^* \ldots a_k^*$ qui appartient à $C(L)$. La proposition II.14 implique, alors, $c(h_j(K)) \in C(L)$ et comme $c(h_j(K))$ est défini sur un alphabet de k-1 lettres, l'hypothèse de recurrence s'applique et $\bar{D}_1^* \in C(c(h_j(K))) \subseteq C(L)$.

B) Plaçons nous, maintenant, dans le cas où le langage $\bar{K}$ ne posséde pas de facteurs itérants et considérons le rang r de la matrice A. Comme $\bar{K}$ est infini, r est inférieur à k. Distinguons, alors, deux cas :

i) Si r est inférieur à k-1, montrons que $S = \{x \in \mathbb{N}^k \,/\, x\, A = 0\}$ posséde un élément $z = (z_1,\ldots,z_k)$ tel que $\exists\, i, j \in \{1,\ldots,k\}$, $z_i = 0$, $z_j > 0$. En effet, comme $\bar{K}$ est infini, S posséde un élément $y = (y_1,\ldots,y_k)$ non nul. Supposons que, pour tout $j \in \{1,\ldots,k\}$, $y_j \neq 0$. Alors, comme $S' = \{x \in \mathbb{Z}^k \,/\, x\, A = 0\}$ est un espace vectoriel de dimension supérieure à 1, il existe dans S' un élément $y' = (y'_1,\ldots,y'_k)$ linéairement indépendant de y. Soit $i_o \in \{1,\ldots,k\}$ tel que $\forall\, j \in \{1,\ldots,k\}$, $y'_{i_o}\, y_j - y_{i_o}\, y'_j \geq 0$. Alors, $z = y'_{i_o}\, y - y_{i_o}\, y'$ appartient à S, est non nul, puisque $y_{i_o} \neq 0$ et y, y' sont linéairement indépendants, et vérifie $z_{i_o} = 0$. Prenons $x = (x_1,\ldots,x_k) \in \Psi(\bar{K})$ et posons $\bar{K}' = \bar{K} \cap a_1^* \ldots a_{i_o}^{x_{i_o}} \ldots a_k^*$. Comme $\Psi(\bar{K}')$ contient l'ensemble $\{x + \lambda z \,/\, \lambda \in \mathbb{N}\}$, $\bar{K}'$ est infini et donc non rationnel puisque $\bar{K}$ ne contient aucun langage rationnel infini. Donc $K'' = K \cap a_1^* \ldots a_{i_o}^{x_{i_o}} \ldots a_k^* = a_1^* \ldots a_{i_o}^{x_{i_o}} \ldots a_k^* \setminus \bar{K}'$ est un langage non rationnel appartenant à $C(L)$ et comme $K'' = \text{shuf}(h_{i_o}(K''), a_{i_o}^{x_{i_o}}) \cap a_1^* \ldots a_k^*$ où $h_{i_o}(a_j) = \varepsilon$ si $j = i_o$, a_j sinon, l'hypothèse de récurrence permet de conclure en utilisant le même raisonnement que dans le cas A.

ii) Si r est égal à k-1, en triangularisant la matrice A, on peut trouver α_1, α_2, $\alpha_3 \in \mathbb{Z}$ tels que $\alpha_1 \neq 0$ ou $\alpha_2 \neq 0$ et $\forall\, x = (x_1,\ldots,x_k) \in \Psi(\bar{K})$, $\alpha_1\, x_1 + \alpha_2\, x_2 = \alpha_3$. Supposons $\alpha_1 \neq 0$. Soit R un langage rationnel inclus dans $\{a_1, a_2\}^*$ tel que $\Psi(R) = \{(x_1, x_2) \in \mathbb{N}^2 \,/\, \alpha_1\, x_1 + \alpha_2\, x_2 = \alpha_3\}$. Le langage $H = h_1(c(K) \cap R\, a_3^* \ldots a_k^*)$ avec $h_1(a_i) = \varepsilon$ si $i = 1$, a_i sinon, est inclus $a_2^* \ldots a_k^*$ et appartient à $C(L)$.

Si H est un langage non rationnel, l'hypothèse de récurrence permet de conclure. Il ne nous reste donc plus qu'à examiner le cas où H est un langage rationnel. Etablissons, d'abord, l'égalité entre $\bar{H} = a_2^* \dots a_k^* \setminus H$ et $h_1(\bar{K}) \cup R' a_3^* \dots a_k^*$ où R' est égal à $\{a_2^{i_2} \ / \ \nexists\, j$ tel que $(j, i_2) \in \Psi(R)\}$.

Prenons $y = a_2^{i_2} \dots a_k^{i_k} \in \bar{H}$. Alors $\forall j \in \mathbb{N}$, soit $a_1^j y \in \bar{K}$, soit $(j, i_2) \notin \Psi(R)$. S'il n'existe pas de $j \in \mathbb{N}$ tel que $(j, i_2) \in \Psi(R)$, alors $y \in R' a_3^* \dots a_k^*$. Dans le cas contraire, il existe $i \in \mathbb{N}$ tel que $(i, i_2) \in \Psi(R)$ et donc $a_1^i y \in \bar{K}$ et $y \in h_1(\bar{K})$.

Réciproquement, prenons $y = a_2^{i_2} \dots a_k^{i_k} \in R' a_3^* \dots a_k^*$ et supposons que y appartienne à H. Alors, $y = h_1(y')$ avec $y' \in R\, a_3^* \dots a_k^*$ donc il existe $j \in \mathbb{N}$ tel que $(j, i_2) \in \Psi(R)$, d'où la contradiction. Si $y \in h_1(\bar{K})$, il existe $j_o \in \mathbb{N}$ tel que $a_1^{j_o} y \in \bar{K}$ ce qui implique, en particulier, $\alpha_1 j_o + \alpha_2 i_2 = \alpha_3$. Comme α_1 est non nul, $j \neq j_o$ implique $\alpha_1 j + \alpha_2 i_2 \neq \alpha_1 j_o + \alpha_2 i_1 = \alpha_3$. Donc, $\forall j \in \mathbb{N}$, soit $a_1^j y \in \bar{K}$, soit $(j, i_2) \notin \Psi(R)$ ce qui entraine $y \in \bar{H}$.

Il est facile de vérifier que $\Psi(R') = \{x_2 \in \mathbb{N} \ / \ \alpha_3 - \alpha_2 x_2 \in \{\lambda_1 \alpha_1 \ / \ \lambda_1 \in \mathbb{N}\}\}$ est un ensemble semi-linéaire ce qui implique que $R' \subseteq a_1^*$ est rationnel ainsi que $R' a_3^* \dots a_k^*$. D'autre part, il est clair que $h_1(\bar{K})$ et $R' a_3^* \dots a_k^*$ sont disjoints, donc si H est rationnel, $\bar{H}$ et $h_1(\bar{K})$ sont aussi rationnels. Si $h_1(\bar{K})$ est fini, $\bar{K}$ est égal à $\bigcup_{w \in h_1(\bar{K})} (((\bar{K} \ / \ w) \cap a_1^*) w)$ est rationnel puisque pour tout $w \in h_1(\bar{K})$, $(\bar{K} \ / \ w) \cap a_1^*$ est un PSL-langage inclus dans a_1^*. Enfin, si $h_1(\bar{K})$ est un rationnel infini, il existe $i \in \{2,\dots,k\}$, $t \in \mathbb{N}_+$, u, v tels que $u(a_i^t)^* v \subseteq h_1(\bar{K})$, donc $\bar{K} \cap a_1^* u(a_i^t)^* v$ est infini ce qui implique qu'il existe un élément $z = (z_1,\dots,z_k) \in \mathbb{N}^k$ tel que $z A = 0$, $z_i > 0$ et $\forall j \in \{2,\dots,k\} \setminus \{i\}$, $z_j = 0$. Alors, le raisonnement utilisé au i) permet de conclure.□

Ce résultat est à opposer à celui de Berstel et Boasson [7] qui ont montré l'existence d'une suite infinie décroissante de cônes rationnels engendrés par des langages bornés appartenant à $C(\bar{D}_1^* \cap a^* b^*)$.

Supposons, maintenant, que la famille des langages commutatifs non rationnels posséde un plus petit élément L, vis-à-vis de la transduction rationnelle. Alors, L est un langage commutatif non rationnel qui vérifie $\forall L' \in \mathrm{Com} \setminus \mathrm{Rat}$, $L \in C(L')$. En particulier, $L \in C(\bar{D}_1^*)$ ce qui entraine, d'après la proposition précédente, $C(L) = C(\bar{D}_1^*)$. Prenons un langage $L' \subseteq a^*$, non rationnel. Durieux a montré dans [18] que $\bar{D}_1^* \notin C(L')$ donc $L \notin C(L')$, d'où la contradiction et nous pouvons énoncer :

<u>Corollaire III.37</u> : *La famille des langages commutatifs non rationnels ne possède pas de plus petit élément pour la transduction rationnelle.*

En fait, les résultats de la section E du chapitre II, permettent de préciser ce corollaire en montrant que pour tout langage commutatif non rationnel L, il existe un langage commutatif L' rationnellement incomparable avec L. En effet, si L appartient à $c(\mathrm{Rat}_k)$, il suffit de prendre $L' = \{a_1,\dots,a_{k+1}\}^* \setminus c(\ a_1^n a_2^{n^2} \dots a_{k+1}^{n^{k+1}} \ / \ n \geq 0\})$. Comme L' n'est pas un PSL-langage, $L' \notin C(L) \subseteq \mathcal{L}_\Psi$ et d'après la proposition II.70

$L \notin C(L')$ qui ne contient pas de langages non rationnels de $c(Rat_k)$. Dans le cas où L appartient à $Com_k \setminus c(Rat_k)$, il suffit de prendre $L' = E'_{k+1} = c((a_1 \ldots a_{k+1})^*)$. La proposition II.66 implique que $L' \notin C(L) \subseteq C(Com_k)$ et comme L n'est pas un PSL-langage $L \notin C(L')$.

Si nous considérons, maintenant, la famille des PSL-langages commutatifs non rationnels, nous ne pouvons plus montrer cette propriété. Au contraire, la technique utilisée pour la démonstration de la proposition III.36 pourrait, semble-t-il, servir pour démontrer :

<u>Conjecture</u> : *Tout PSL-langage commutatif non rationnel est rationnellement supérieur au langage* $\bar{D}_1^*$.

Comme il est assez facile de montrer que $\bar{D}_1^* \notin \mathcal{L}in$, on pourrait, alors, en déduire que tout langage linéaire commutatif est rationnel. De même, il suffirait de montrer que $\bar{D}_1^* \notin F_\sigma(\mathcal{L}in)$ pour démontrer :

<u>Conjecture</u> : *Tout langage quasi-rationnel commutatif est rationnel.*

E - LANGAGES SANS PRODUIT

Nous dirons qu'un cône rationnel $\mathcal{L}$ est **sans produit** s'il ne contient pas le produit de deux langages non rationnels définis sur des alphabets disjoints, c'est-à-dire si $L_1 L_2 \in \mathcal{L}$ avec $L_1 \subseteq T_1^*$, $L_2 \subseteq T_2^*$ et $T_1 \cap T_2 = \emptyset$ implique soit $L_1 \in$ Rat, soit $L_2 \in$ Rat. Un exemple de cône rationnel sans produit est fourni par $\mathcal{L}in$ (cf. [38]).

Un langage L est sans produit si $C(L)$, le cône rationnel engendré par L est sans produit.

Nous allons, d'abord, définir les langages indécomposables et montrer que tout langage indécomposable est un langage sans produit (la réciproque est fausse). Il est très facile de vérifier que les langages 1-bornés sont indécomposables. Nous retrouvons, ainsi, un résultat démontré par Ginsburg dans [25]. De même, on montre aisément que Init (D'^*_1), l'ensemble des facteurs gauches des mots de D'^*_1, le langage de semi-Dyck sur une lettre, est un langage indécomposable. Nous obtenons le même résultat, à l'aide de quelques lemmes intermédiaires, pour D_1^*. Comme D'^*_1 n'est pas un langage sans produit (proposition III.49), nous apportons, ainsi, une nouvelle distinction entre les langages D_1^* et D'^*_1. Enfin, nous considérons les langages commutatifs inclus dans D_1^*. Ces langages ne sont pas tous indécomposables, mais en généralisant les lemmes qui permettent de montrer que D_1^* est indécomposable , nous obtenons le résultat principal de cette section : Tout langage commutatif inclus dans D_1^* est sans produit.

Un langage rationnel L sera dit **simple** si $\Psi(L)$ est un ensemble linéaire. Du théorème de Kleene, nous pouvons déduire immédiatement :

Lemme III.38 : *Tout langage rationnel est une union finie de langages rationnels simples.*

Définition : *Un langage* L *est* **indécomposable** *si pour tout langage rationnel simple* R, R', *l'un des deux langages* $R \cap (L / R')$, $R' \cap (R \setminus L)$ *est un langage rationnel.*

Proposition III.39 : *Tout langage indécomposable est sans produit.*

Démonstration : Soient $L \subseteq T^*$ un langage indécomposable, $L_1 \subseteq T_1^*$, $L_2 \subseteq T_2^*$ avec $T_1 \cap T_2 = \emptyset$ et $L_1 L_2 \in C(L)$. Il existe un alphabet Y, un langage rationnel $R \subseteq Y^*$ et deux homomorphismes alphabétiques h et g tels que $L_1 L_2 = g(h^{-1}(L) \cap R)$ [64]. Pour $i \in \{1, 2\}$, posons $Y_i = \{a \in Y \mid g(a) \in T_i^*\}$. Comme $L_1 L_2$ est inclus dans $T_1^* T_2^*$, $L_1 L_2 = g(h^{-1}(L) \cap R')$ où $R' = R \cap Y_1^* Y_2^*$. Il est facile de vérifier que R' peut s'écrire $\bigcup_{j=1}^{k} R'_j R''_j$ où $\forall j \in \{1,\dots,k\}$, R'_j, R''_j sont des langages rationnels vérifiant $R'_j \subseteq Y_1^*$ et $R''_j \subseteq Y_2^*$. On pourra supposer d'après le lemme précédent que $\Psi(R'_j)$ et $\Psi(R''_j)$ sont des ensembles linéaires. Pour $i \in \{1, 2\}$, l'homomorphisme h_i est défini sur $T_1 \cup T_2$ par : $h_i(a) = a$ si $a \in T_i$, ε sinon. Pour tout $j \in \{1,\dots,k\}$, posons $V_j = g(h^{-1}(L) \cap R'_j R''_j)$ et montrons que l'un des deux langages $h_1(V_j)$, $h_2(V_j)$ est rationnel.

Comme $g(R'_j) \subseteq T_1^*$ et $g(R''_j) \subseteq T_2^*$, $h_1 \circ g(R'_j) = g(R'_j)$, $h_1 \circ g(R''_j) = \{\varepsilon\}$ et $h_1(V_j) = g(R'_j \cap (h^{-1}(L) / R''_j)) = g(R'_j \cap h^{-1}(L / h(R''_j))) = g(R'_j \cap V'_j)$ avec $V'_j = h^{-1}(h(R'_j) \cap (L / h(R''_j)))$. De même, $h_2(V_j) = g(R''_j \cap V''_j)$ avec $V''_j = h^{-1}(h(R''_j) \cap (h(R'_j) \setminus L))$. Comme $h(R'_j)$ et $h(R''_j)$ sont des langages rationnels simples, nous avons par hypothèse V'_j où V''_j rationnel et donc l'un des deux langages $h_1(V_j)$, $h_2(V_j)$ est rationnel.

Pour $i = 1, 2$, posons $K_i = \{j \in \{1,\dots,k\} \ / \ h_i(V_j)$ est rationnel$\}$. D'après la propriété ci-dessus, $K_1 \cup K_2 = \{1,\dots,k\}$ et comme, pour $i = 1, 2$, $h_i(L_1 L_2) = L_i$, $h_i(V_j) \subseteq L_i$, $\forall j \in \{1,\dots,k\}$. Nous obtenons alors $L_1 L_2 = R_1 L'_2 \cup L'_1 R_2$ où pour $i = 1, 2$, $R_i = \bigcup_{j \in K_i} h_i(V_j)$ est un langage rationnel et L'_i est égal à $\bigcup_{j \in K_{i'}} h_i(V_j)$ avec $i' = 3-i$. Il est, alors, clair que L_1 ou L_2 doit être un langage rationnel.□

Cette proposition va nous permettre de retrouver immédiatement une propriété des langages définis sur un alphabet d'une seule lettre :

Corollaire III.40 [25] *Tout langage $L \subseteq a^*$ est sans produit.*

Démonstration : Considérons deux langages rationnels simples R et R' inclus dans a^*. Si R est fini, $R \cap (L / R')$ est fini donc rationnel. Si R est infini, il existe $j \in \mathbb{N}_+$ tel que $R = R(a^j)^*$. Alors $R \setminus L$ est égal à $R \setminus ((a^j)^* \setminus L)$ et d'après le lemme II.38, $L' = (a^j)^* \setminus L$ est un langage rationnel ainsi que $R' \cap (R \setminus L) = R' \cap (R \setminus L')$. Le langage L est donc indécomposable et d'après la proposition précédente, il est sans produit.□

Considérons, maintenant, l'homomorphisme d, de $\{a, b\}^*$ dans $\mathbb{Z}$ muni de l'addition, défini par $d(a) = 1$ et $d(b) = -1$. Alors, pour tout $w \in \{a, b\}^*$, $d(w) = \ell_a(w) - \ell_b(w)$.

Le langage de semi-Dyck sur une lettre, D'^*_1 est égal à la classe du mot vide ε dans la congruence engendrée par $ab = \varepsilon$. La proposition III.39 va nous permettre de montrer que Init $(D'^*_1) = \{w \in \{a, b\}^* \ / \ w = w_1 w_2 \Rightarrow d(w_1) \geq 0\}$ est un langage sans produit.

Corollaire III.41 : *Les langages Init (D'^*_1) et Term $(D'^*_1) = \{w \in \{a, b\}^* \ / \ w = w_1 w_2 \Rightarrow d(w_2) \leq 0\}$ sont des langages sans produit.*

Démonstration : Soient L = Init (D'^*_1), R, R', deux langages rationnels simples inclus dans $\{a, b\}^*$. S'il existe $x, u, y \in \{a, b\}^*$ tels que $x u \in L$, $x u^* y \subseteq R$ et $d(u) > 0$, alors $R \setminus L = \{a, b\}^*$ et $R' \cap (R \setminus L) = R'$ est un langage rationnel. Dans le cas contraire, $w \in$ Init $(R \cap L)$ implique $d(w) \leq q$ où q est le nombre d'états de l'automate d'état fini qui reconnait R. Donc, pour tout $i \in \mathbb{N}$, le langage $R_i = \{w \in R \cap L \ / \ d(w) = i\}$ est un langage rationnel. Posons, alors, i_o le plus petit i tel que $a^i R' \cap L$ soit non vide. Nous avons $L / R' = L / b^{i_o} \subseteq$ Init $(L) = L$ et nous en déduisons que

$R \cap (L / R') = R_{i_o}$ est un langage rationnel.

Un raisonnement symétrique permet de montrer que Term (D'^*_1) est indécomposable donc sans produit.□

La proposition III.39 va nous servir, aussi, pour montrer que le langage D^*_1 est sans produit, mais nous aurons besoin de quelques résultats intermédiaires :

Lemme III.42 : *Pour tout $x \in \mathbb{Z}$ et tout $y \in \mathbb{N}_+$, $d^{-1}(x + \mathbb{Z}y)$ est un langage rationnel.*

Démonstration : Pour tout $i \in \{0,\dots,y-1\}$, posons R_i = shuf $(a^t(a^y)^*, b^i(b^y)^*)$ où t est le reste dans la division entière de x + i par y. Il est facile de vérifier que $d^{-1}(x + \mathbb{Z}y)$ est égal à $\bigcup_{i=0}^{y-1} R_i$ et, comme la famille des langages rationnels est close pour l'opération "shuffle", $d^{-1}(x + \mathbb{Z}y)$ est bien un langage rationnel.□

Pour tout $A \subseteq \mathbb{Z}$, notons $-A$, l'ensemble $\{-x / x \in A\}$.

Lemme III.43 : *Pour tout langage $L \subseteq T^*$, D^*_1/L est égal à $d^{-1}(-d(L))$.*

Démonstration : Prenons $u \in D^*_1/L$. Il existe $v \in L$ tel que $uv \in D^*_1$. Donc $d(u) + d(v) = 0$, $d(u) \in -d(L)$ et $u \in d^{-1}(-d(L))$. Réciproquement si $u \in d^{-1}(-d(L))$, il existe $v \in L$ tel que $d(uv) = d(u) + d(v) = 0$. Donc $uv \in D^*_1 = d^{-1}(0)$ et $u \in D^*_1/L$.□

Lemme III.44 : *Soit R un langage rationnel simple inclus dans $\{a, b\}^*$. Si R possède deux facteurs itérants u et v tels que $d(u)\, d(v) < 0$, alors D^*_1/R est un langage rationnel.*

Démonstration : Comme $\Psi(R)$ est un ensemble linéaire, il existe $x_o,\dots,x_k \in \mathbb{Z}$ tels que $d(R) = \{x_o + \sum_{i=1}^{k} \lambda_i x_i / \lambda_i \in \mathbb{N}\}$. Comme R possède un facteur itérant appartenant à $d^{-1}(\mathbb{N}_+)$, il existe $i \in \{1,\dots,k\}$ tel que $x_i \in \mathbb{N}_+$. De même, il existe $j \in \{1,\dots,k\}$ tel que x_j soit négatif. Nous pouvons en déduire que $d(R)$ est égal à $x_o + \mathbb{Z}y$ où y est le plus grand commun diviseur de $|x_1|,\dots,|x_k|$. Et d'après les lemmes précédents $D^*_1/R = d^{-1}(-x_o + \mathbb{Z}y)$ est un langage rationnel.□

Lemme III.45 : *Soit R un langage rationnel inclus dans $\{a, b\}^*$. Si tout facteur itérant de R appartient à $d^{-1}(\mathbb{N})$, alors, pour tout $k \in \mathbb{Z}$, les langages $R_1 = \{x \in R / d(x) = k\}$, $R_2 = \{x \in R / d(x) > k\}$ et $R_3 = \{x \in R / d(x) < k\}$ sont rationnels.*

Démonstration : Soient $M = (Q, T, f, q_o, F)$ un automate d'état fini qui reconnait R et n le nombre d'états de M. Prenons $Q' = Q \times \{-n,\dots,k+n\}$, $q'_o = (q_o, 0)$ et définissons f' sur $Q' \times T$ par : $f'((q, i), x) = (f(q, x), i + d(x))$ si $-n < i < k + n$, $(f(q, x), i)$ sinon. Pour j = 1, 2, 3, posons

$M'_j = (Q', T, f', q'_o, F'_j)$ où $F'_1 = F \times \{k\}$, $F'_2 = F \times \{k+1,\dots,k+n\}$ et $F'_3 = F \times \{-n+1,\dots,k-1\}$. Comme tout facteur itérant de R appartient à $d^{-1}(\mathbb{N})$, pour toute factorisation $w_1\ w_2\ w_3$ d'un mot de R, $d(w_2)$ est supérieur à $-n$. Nous pouvons en déduire que pour $j = 1, 2, 3$, le langage reconnu par M'_j est égal à R_j.□

Lemme III.46 : *Le langage D_1^* est indécomposable.*

Démonstration : Soient R et R' deux langages rationnels simples. Comme D_1^* est un langage commutatif, nous pouvons remplacer $R \setminus D_1^*$ par D_1^* / R qui lui est égal. Distinguons plusieurs cas :

i) Si R (resp. R') vérifie l'hypothèse du lemme III.44, D_1^* / R (resp. D_1^* / R') est un langage rationnel ainsi que $R' \cap (D_1^* / R)$ (resp. $R \cap (D_1^* / R')$).

ii) Si tous les facteurs itérants de R et tous ceux de R' appartiennent à $d^{-1}(\mathbb{N})$, il existe, alors, $x_o \in \mathbb{Z}$, $x_1,\dots,x_k \in \mathbb{Z}$, $x_1,\dots,x_k \in \mathbb{N}$ tels que $d(R') = \{x_o + \sum_{i=1}^{k} \lambda_i\, x_i \ /\ \lambda_i \in \mathbb{N}\}$. On peut donc trouver un ensemble fini $F \subseteq \mathbb{Z}$, $z \in \mathbb{Z}$ tels que $d(R') = F \cup (z + \mathbb{N}y)$ où y est le plus grand commun diviseur de $x_1,\dots,x_k$. D'après le lemme III.43, D_1^* / R' est égal à $d^{-1}(-d(R')) = d^{-1}(F') \cup d^{-1}(z' + \mathbb{N}\, y')$ où $F' = -F$, $z' = -z$ et $y' = -y$. Le langage $R \cap d^{-1}(F') = \bigcup_{t \in F'} \{x \in R \ /\ d(x) = t\}$ est rationnel d'après le lemme III.45. Quant à $R \cap d^{-1}(z' + \mathbb{N}y')$, il est égal à $\{x \in R \ /\ d(x) < z' + 1\} \cap d^{-1}(z' + \mathbb{Z}y)$ qui est un langage rationnel d'après les lemmes III.42 et III.45. Donc $R \cap (D_1^* / R') = (R \cap d^{-1}(F')) \cup (R \cap d^{-1}(z' + \mathbb{N}y'))$ est un langage rationnel.

iii) Si tous les facteurs itérants de R appartiennent à $d^{-1}(\mathbb{N})$ et tous ceux de R' appartiennent à $d^{-1}(-\mathbb{N})$, on montre de la même façon que $d(R') = F \cup (z + \mathbb{N}y)$ où F est un ensemble fini de $\mathbb{Z}$, $z = -z' \in \mathbb{Z}$ et $y' = -y \in \mathbb{N}$. La seule différence avec le ii) est que y est ≤ 0. Donc $R \cap d^{-1}(z' + \mathbb{N}y')$ est égal à $\{x \in R \ /\ d(x) > z' - 1\} \cap d^{-1}(z' + \mathbb{Z}y')$. Mais, d'après les lemmes III.42 et III.45, c'est encore un langage rationnel.

La démonstration s'achéve en remarquant que tous les autres cas possibles sont symétriques de l'un des cas étudiés ci-dessus.□

Nous en déduisons immédiatement :

Proposition III.47 : *Le langage de Dyck sur une lettre, D_1^*, est un langage sans produit.*

L'opération "chevron" qui à tout langage L fait correspondre $\langle L\rangle = \{a^n\, x\, b^n \ /\ x \in L,\ n \geq 0\}$, a été définie et étudiée dans [13] et [14]. En utilisant la commutativité de D_1^*, nous pouvons déduire de la proposition précédente :

Corollaire III.48 : *Soient* L *un langage inclus dans* T_1^*, a *et* b *deux symboles n'appartenant pas à* T_1. *Alors* $\langle L\rangle = \{a^n\ x\ b^n \ / \ x \in L,\ n \geq 0\}$ *appartient à* $C(D_1^*)$ *si et seulement si* L *est un langage rationnel.*

Démonstration : Si $\langle L\rangle \in C(D_1^*)$, il existe un alphabet Y, un langage rationnel $R \subseteq Y^*$ et deux homomorphismes alphabétiques h et g tels que $\langle L\rangle = g(h^{-1}(D_1^*) \cap R)$. Posons $A = \{x \in Y \ / \ g(x) \in \{a, \varepsilon\}\}$, $B = \{x \in Y \ / \ g(x) \in \{b, \varepsilon\}\}$ et $Y_1 = \{x \in Y \ / \ g(x) \in T_1 \cup \{\varepsilon\}\}$. Comme $\langle L\rangle \subseteq a^*\ T_1^*\ b^*$, $\langle L\rangle = g(h^{-1}(D_1^*) \cap R')$ où $R' = R \cap A^*\ Y_1^*\ B^*$. Donc R' peut s'écrire $\bigcup_{i=1}^{k} R'_i\ R''_i\ R'''_i$ où R'_i, R''_i, R'''_i sont des langages rationnels inclus respectivement dans A^*, Y_1^* et B^*. Comme $h^{-1}(D_1^*)$ est un langage commutatif, il est facile de vérifier que $L\ \{a^n\ b^n \ / \ n \geq 0\}$ est égal à $g(h^{-1}(D_1^*) \cap R'')$ où $R'' = \bigcup_{i=1}^{k} R''_i\ R'_i\ R'''_i$ et d'après la proposition précédente, L est rationnel.

Réciproquement si L est un langage rationnel, il est clair que $\langle L\rangle \in C(D_1^*)$.□

Remarquons que la propriété d'indécomposabilité n'est pas conservée par transduction rationnelle. Considérons, en effet, le langage $C_1 = \{a^n\ b^n \ / \ n \geq 0\} = D_1^* \cap a^*\ b^*$ qui appartient à $C(D_1^*)$ et les langages rationnels simples $R = R' = a^*\ b^*$. Alors, les langages $R \cap (L \ / \ R') = \{a^n\ b^p \ / \ n \geq p\}$ et $R' \cap (R \setminus L) = \{a^n\ b^p \ / \ n \leq p\}$ sont non rationnels.

D'autre part, la proposition suivante, montre l'existence d'un langage $L \subseteq D_1^*$ qui n'est pas sans produit et apporte une nouvelle distinction entre les langages de Dyck et de semi-Dyck sur une lettre :

Proposition III.49 : *Le langage de semi-Dyck sur une lettre,* D'^*_1, *domine rationnellement le produit* $L_1\ L_2$ *avec* $L_1 = \{a^n\ b^p \ / \ n \geq p\}$ *et* $L_2 = \{c^s\ d^t \ / \ s \leq t\}$.

Démonstration : Considérons l'alphabet $Y = \{b_1, b_2, \ldots, b_6\}$ et définissons l'homomorphisme h sur Y par $h(b_1) = h(b_6) = \varepsilon$, $h(b_2) = h(b_4) = a_1$ et $h(b_3) = h(b_5) = b$. De même l'homomorphisme g de Y^* dans $\{a, b, c, d\}^*$ est défini sur Y par $g(b_1) = g(b_2) = a$, $g(b_5) = g(b_6) = d$, $g(b_3) = b$ et $g(b_4) = c$. Il est, alors, facile de véfifier l'égalité entre $L_1\ L_2$ et $g(h^{-1}(D'^*_1) \cap b_1^*\ b_2^*\ b_3^*\ b_4^*\ b_5^*\ b_6^*)$.□

Considérons, maintenant, un langage commutatif L inclus dans D_1^*. Il existe, alors, $A \subseteq \mathbb{N}$ tel que $L = \{w \in D_1^* \ / \ \ell_a(w) \in A\}$. Le langage L n'est indécomposable que si L est un PSL-langage. En effet, considérons les langages rationnels simples $R = a^*$ et $R' = b^*$. Si L n'est pas un PSL-langage, les langages $R \cap (L \ / \ R') = \{a^i \ / \ i \in A\}$ et $R' \cap (R \setminus L) = \{b^i \ / \ i \in A\}$ ne sont pas rationnels. Donc, la proposition III.39 ne peut pas être utilisée pour les langages commutatifs inclus dans D_1^*. Par contre, en employant certains des résultats intermédiaires de cette section, nous allons pouvoir étendre la proposition III.47 à tous les langages commutatifs inclus dans D_1^*. Montrons d'abord que le lemme III.44 reste vérifié pour tout langage commutatif inclus dans D_1^* :

Lemme III.50 : *Soient L un langage commutatif inclus dans D_1^* et R un langage rationnel simple inclus dans $\{a, b\}^*$. Si R possède deux facteurs itérants u et v tels que $d(u)\, d(v) < 0$, alors L / R est un langage rationnel.*

Démonstration : Comme R est un langage rationnel simple, il existe $x_o \in \mathbb{N}^2$ et une partie finie $P \subseteq \mathbb{N}^2$ tels que $\Psi(R) = L(x_o ; P)$. Si u est un facteur itérant de R, il existe x, y $\in \{a, b\}^*$ tels que $x\, u^*\, y \subseteq R$, donc $\Psi(R) \cap L(\Psi(x\, y) ; \{\Psi(u)\})$ est infini et il existe $\lambda \in \mathbb{N}_+$ tel que $\lambda\, \Psi(u) \in L(0 ; P)$([24] lemme 5.4.5.) ce qui implique $(u^\lambda)^*\, c(R) = c(R)$. De même, il existe $\mu \in \mathbb{N}_+$ tel que $(v^\mu)^* c(R) = c(R)$ ce qui implique $(u^\lambda)^*\, (v^\mu)^*\, c(R) = c(R)$. Posons $w = u^{\lambda i}\, v^{\mu j}$ avec $i = |d(v^\mu)|$ et $j = |d(u^\lambda)|$. Alors $c(w) = c(a^t\, b^t)$ avec $t = \ell_a(w) \in \mathbb{N}_+$. Comme L est commutatif et que $c(R) = (a^t b^t)^*\, c(R)$, nous avons $L / R = L / c(R) = L / ((a^t\, b^t)^*\, c(R)) = (L / (a^t\, b^t)^*) / R$. Pour tout $s \in \{0,\dots,t-1\}$, posons $A_s = \{w \in \{a, b\}^* / \ell_a(w) = s(\text{mod. } t)\}$, $L_s = L \cap A_s$, $R_s = R \cap A_s$ et $I = \{s \in \{0,\dots,t-1\} / L_s$ est infini$\}$. Il est clair que $L_s / (a\, b)^* = D_1^*$ pour tout $s \in I$, donc $L / (a^t\, b^t)^* = \bigcup_{s\in\{0,\dots,t-1\}} ((L_s / (ab)^* \cap A_s) = F \cup (\bigcup_{s\in I} (D_1^* \cap A_s))$ où F est un ensemble fini. Il nous reste donc, à montrer que, pour $s \in I$, $(D_1^* \cap A_s) / R$ est un langage rationnel. Il est facile de vérifier que $(D_1^* \cap A_s) / R = \bigcup_{(p,q)\in B} ((D_1^* / R_p) \cap R_q)$ où $B = \{(p,q) / p, q \in \{0,\dots,t-1\}, p + q = s(\text{mod. } t)\}$ et comme R_p est une union finie de rationnels simples qui possédent tous u^t et v^t comme facteurs itérants, le résultat se déduit du lemme III.44.□

Nous utiliserons, aussi, le résultat suivant dont la démonstration est assez semblable à celles du lemme précédent et du lemme III.46 :

Lemme III.51 : *Soient L un langage commutatif inclus dans D_1^*, R, R' deux langages rationnels simples inclus dans $\{a, b\}^*$. Si R' (resp. R) possède un facteur itérant u tel que $d(u) = 0$, l'un des deux langages $R' \cap (L / R)$, $R \cap (L / R')$ est rationnel.*

Démonstration : Si R vérifie l'hypothèse du lemme précédent, L/R est un langage rationnel ainsi que le langage $R' \cap (L / R)$. Nous pouvons, donc, supposer, par exemple, que tout facteur itérant de R appartient à $d^{-1}(\mathbb{N})$. Comme R' posséde un facteur itérant appartenant à $d^{-1}(\{0\})$ et que $\Psi(R')$ est un ensemble linéaire, il existe $t \in \mathbb{N}_+$ tel que $c(R') = (a^t\, b^t)^*\, c(R')$, donc il existe $I \subseteq \{0,\dots,t-1\}$ et un langage fini F' tels que $L/R' = F' \cup (\bigcup_{s\in I} (D_1^* \cap A_s) / R')$ avec $A_s = \{w \in \{a, b\}^* / \ell_a(w) = s(\text{mod. } t)\}$ (cf. démonstration précédente). Posons $B = \{(p, q) / p, q \in \{0,\dots,t-1\}, p+q = s(\text{mod. } t)\}$ et $\forall\, p \in \{0,\dots,t-1\}$, $R'_p = R' \cap A_p$. Comme $(D_1^* \cap A_s) / R' = \bigcup_{(p,q)\in B} ((D_1^* / R'_p) \cap R'_q)$, il nous suffit de montrer que $R \cap (D_1^* / R'_p)$ est un langage rationnel, $\forall\, p \in \{0,\dots,t-1\}$. Le langage R'_p étant une union finie de langages rationnels simples, nous obtenons ce résultat en raisonnant de la même façon que pour la démonstration du lemme III.46.□

Nous sommes, maintenant, en mesure de démontrer le résultat principal de cette section :

Proposition III.52 : *Tout langage commutatif L, inclus dans D_1^* est sans produit.*

Démonstration : Supposons qu'il existe $L_1 \subseteq T_1^*$, $L_2 \subseteq T_2^*$ tels que $L_1 L_2 \in C(L)$ avec $T_1 \cap T_2 = \emptyset$. Alors, il existe un alphabet Δ, un langage rationnel $R \subseteq \Delta^*$ et deux homomorphismes alphabétiques g et h tels que $L_1 L_2 = g(h^{-1}(L) \cap R)$, $g(x)\, h(x) = \varepsilon$ implique $x = \varepsilon$ et où R est une union finie de produit $R_1 R_2$ avec $\forall i \in \{1, 2\}$, R_i est un langage rationnel simple qui vérifie $g(R_i) \subseteq T_i^*$. Considérons le langage $L' = g(h^{-1}(L) \cap R_1 . R_2)$ et distinguons plusieurs cas :

i) Si $h(R_1)$ (resp. $h(R_2)$) posséde deux facteurs itérants u et v tels que $d(u)\, d(v) < 0$, comme $h(R_1)$ et $h(R_2)$ sont des langages rationnels simples, le lemme III.50 permet de montrer, par un raisonnement analogue à celui de la proposition III.39, l'existence de deux langages rationnels $A_1 \subseteq L_1$, $A_2 \subseteq L_2$ tels que L' soit inclus dans $A_1 L_2 \cup L_1 A_2$.

ii) Si $h(R_1)$ (resp. $h(R_2)$) posséde un facteur itérant appartenant à $d^{-1}(\{0\})$, le lemme III.51 permet d'obtenir le même résultat.

iii) Si tout facteur itérant u de $h(R_1 R_2)$ appartient à $d^{-1}(\mathbb{N}_+)$ (resp. $d^{-1}(-\mathbb{N}_+)$), il est facile de vérifier que $L \cap h(R_1 R_2)$ est fini. Donc, $h^{-1}(L) \cap R_1 R_2 = h^{-1}(L \cap h(R_1 R_2)) \cap R_1 R_2$ est un langage rationnel ainsi que L'.

iv) Si pour tout facteur itérant u de $h(R_1)$ et v de $h(R_2)$, on a $d(u) > 0$ et $d(v) < 0$ (resp. $d(u) < 0$ et $d(v) > 0$), posons $c(R_1) = c(x_o\, x_1^* \dots x_p^*)$ et $c(R_2) = c(y_o\, y_1^* \dots y_q^*)$. Supposons qu'il existe $j \in \{1,\dots,q\}$ tel que $g(y_j) = \varepsilon$. Alors, comme $h^{-1}(L)$ est un langage commutatif, $L' = g(h^{-1}(L) \cap R_1 R_2) = g(h^{-1}(L) \cap R'_1 R_2)$ avec $R'_1 = R_1 y_j^*$ et nous nous retrouvons le cas i). Examinons, maintenant, le cas où $g(y_j) \neq \varepsilon$, $\forall j \in \{1,\dots,q\}$, c'est-à-dire le cas où g est limité sur R_2. Alors, pour tout $w_2 \in L_2$, le langage $R_2 \cap g^{-1}(w_2)$ est fini. Pour tout $y \in R_2 \cap g^{-1}(w_2)$, $L \cap h(R_1\, y)$ est fini, donc $L_y = h^{-1}(L) \cap R_1\, y = h^{-1}(L \cap h(R_1\, y)) \cap R_1\, y$ est un langage rationnel ainsi que le langage $L' \cap T_1^*\, w_2$ qui est égal à $\{w \in g(L_y) \ /\ y \in R_2 \cap g^{-1}(w_2)\}$. Ceci entraine que pour tout $w_2 \in L_2$, $T_1^* \cap (L' / w_2)$ est un langage rationnel.

Comme R est une union finie de produit $R_1 R_2$, nous pouvons déduire des considérations précédentes que $L_1 L_2$ est égal à $A\, L_2 \cup L_1\, B \cup L''$ où A et B sont des langages rationnels inclus respectivement dans L_1 et L_2 et où L" vérifie : $\forall w_2 \in L_2$, $T_1^* \cap (L'' / w_2)$ est un langage rationnel. Supposons que L_1 et L_2 soient des langages non rationnels. Alors $A \subsetneq L_1$, $B \subsetneq L_2$ et il existe $w_2 \in L_2 \setminus B$, donc $L_1\, w_2 \subseteq A\, L_2 \cup L''$ ce qui entraine $L_1 = A \cup (T_1^* \cap (L'' / w_2))$ est un langage rationnel, d'où la contradiction.□

F - Sur les langages algébriques commutatifs.

Les résultats concernant les langages algébriques commutatifs sont assez rares. Par contre de nombreux problèmes restent ouverts. En particulier, on ne sait pas caractériser les langages rationnels R tels que c(R) est un langage algébrique . De même, on a toujours pas démontré la conjecture de Fliess [23] qui affirme qu'un langage commutatif $L \subseteq \{a_1,\dots,a_n\}^*$ est algébrique si et seulement si, pour toute permutation π de $\{1,\dots,n\}$, le langage $L \cap a^*_{\pi(1)} \dots a^*_{\pi(n)}$ est algébrique. La conjecture que nous avons énoncée à la section D de ce chapitre concerne les langages quasi-rationnels commutatifs et donc, aussi les langages algébriques commutatifs.

Il existe cependant des résultats concernant les langages algébriques déterministes commutatifs [74] . D'autre part, il est bien connu (cf. [24]) que $c(Rat_2)$ est inclus dans Alg, c'est-à-dire que tout PSL-langage commutatif défini sur un alphabet de deux lettres est algébrique. En fait, la proposition III.25, affirme, plus précisément, que $c(Rat_2)$ est inclus dans $C(D_1^*)$. Perrot [71] puis Maurer [63] et Oshiba [68] ont montré qu'un langage commutatif $L \subseteq \{a_1, a_2, a_3\}^*$ est algébrique si et seulement si $L \cap a_1^* a_2^* a_3^*$ est un langage algébrique. Nous allons, dans cette section, établir l'existence de tels langages n'appartenant pas à $C(D_1^*)$. Pour obtenir ce résultat nous montrerons d'abord une propriété des prépaires itérantes (cf. [8]) pour les langages appartenant à $C(D_1^*)$. L'autre résultat de base de cette section ne concerne les langages commutatifs que par l'un de ses corollaires. Nous montrons (proposition III.56) que si L_1 et L_2 sont des langages définis sur des alphabets disjoints, shuf (L_1, L_2) appartient à Alg implique L_1 ou L_2 rationnel. Nous en déduisons que Rat est le seul cône rationnel clos par intersection contenu dans Alg et, en utilisant la proposition II.15, que tout langage algébrique commutatif est sans produit.

Commençons, maintenant, par construire un langage algébrique appartenant à $c(Rat_3)$. Pour cela, examinons une famille particulière de grammaires algébriques.

Pour tout $n \in \mathbb{N}_+$, notons $T_n = \{a_1,\dots,a_n\}$ et pour tout $A \subseteq T_n^2$ tel que $\forall\, i \in \{1,\dots,n\}$, a_i occurre au moins dans un mot de A, G_A est la grammaire algébrique $(\{S\}, T_n, P_A, S)$ où l'ensemble des règles est égal à

$$P_A = \{S \to \varepsilon\} \cup \{S \to SS\} \cup \{S \to a_i\, S\, a_j \ /\ a_i\, a_j \in A\}.$$

Caractérisons, maintenant, les grammaires G_A qui engendrent un langage commutatif $L_A = L(G_A)$:

Lemme III.53 : *Les quatre propriétés suivantes sont équivalentes :*

i) L_A *est commutatif,*

ii) $L_A = c(A^*)$,

iii) $c(A^2) \subseteq L_A$,

iv) A *est commutatif et* $\forall\, i, j, k \in \{1,\dots,n\}$, $a_i\, a_j \in A$ *implique soit* $a_i\, a_k \in A$, *soit* $a_j\, a_k \in A$.

Démonstration : Il est évident que ii) implique i) et que i) implique iii). L'équivalence entre iii) et iv) se fait en examinant cas par cas des mots de longueur 4. Enfin, on montre facilement par récurrence que iii) entraine ii).□

En particulier, si nous prenons $A = \{a_i\ a_j \ /\ i, j \in \{1,\dots,n\},\ i \neq j\}$, la propriété iv) est vérifiée, le langage L_A est commutatif et il est alors facile de vérifier que $L_A = L_n$ où $\forall\, n \geq 2$, L_n désigne le langage $\{w \in T_n^* \ /\ \ell(w) = 2p$ et $\forall\, i \in \{1,\dots,n\},\ \ell_{a_i}(w) \leq p\}$. Montrons, que pour tout entier n supérieur à 2, L_n n'appartient pas à $C(D_1^*)$. Comme, pour tout n, $j \in \mathbb{N}$, L_n est égal à $L_{n+j} \cap T_n^*$, il suffit de montrer que $L_3 \notin C(D_1^*)$. Pour obtenir ce résultat, nous utiliserons la notion de paire itérante qui a été introduite et étudiée par Boasson [8] et qui s'est révélée très fructueuse dans l'étude des langages algébriques ([4], [9], [10]).

Définition [8] : *Soient* L *un langage défini sur un alphabet* T *et* u, x, w, y, v *des mots appartenant à* T^*. *Alors* $\pi = (u,x,w,y,v)$ *est une* paire itérante *dans* L *si* $x\,y \neq \varepsilon$ *et si* $\forall\, n \in \mathbb{N}_+$, $u\, x^n\, w\, y^n\, v \in L$.

Définition [8] : *La paire itérante* π *est dite* très stricte *s'il existe* $k \in \mathbb{N}$ *tel que* $\forall\, m \in \mathbb{N}$ *le nombre d'éléments des ensembles* $\{n \in \mathbb{N} \ /\ u\, x^n\, w\, y^m\, v \in L\}$ *et* $\{n \in \mathbb{N} \ /\ u\, x^m\, w\, y^n\, v \in L\}$ *soient bornés par* k.

Définition : *Nous dirons que* $\pi = (u, x, w, y, v)$ *est une* prépaire itérante très stricte *dans* L *si il existe* $s, t \in \mathbb{N}$ *tels que* (u, x^s, w, y^t, v) *soit une paire itérante très stricte dans* L.

Etablissons, maintenant, une propriété des prépaires itérantes très strictes dans un langage appartenant à $C(D_1^*)$:

Proposition III.54 : *Soient* $L \subseteq T^*$ *un langage appartenant à* $C(D_1^*)$. *Alors, pour tout* $x, y \in T^+$, *il existe* $k \in \mathbb{N}$ *vérifiant : Si* (u, x, w, y, v) *est une prépaire itérante très stricte dans* L, *il existe deux entiers* k_1, k_2 *inférieurs à* k *tels que* $u\, x^{k_1}\, w\, y^{k_2}\, v \in L$.

Démonstration : Prenons $L \in C(D_1^*)$. Il existe alors un alphabet Y, un langage rationnel $R \subseteq Y^*$ et deux homomorphismes alphabétiques h et g tels que $L = g(h^{-1}(D_1^*) \cap R)$. Comme D_1^* est effaçable (proposition III.15), on peut supposer que g est limité sur R. Soit $M = (Q, Y, f, q_o, F)$ un automate d'état fini qui reconnait R. Pour tout $q, q' \in Q$, notons $R_{q,q'} = \{z \in Y^* \ /\ f(q, z) = q'\}$, $R_{q,q'}(x) = R_{q,q'} \cap g^{-1}(x)$ et $R_{q,q'}(y) = R_{q,q'} \cap g^{-1}(y)$. Soient s et t les entiers tels que (u, x^s, w, y^t, v) soient une paire itérante très stricte dans L. Il est facile de vérifier qu'il existe $u', w', v' \in Y^*$, $q, q', q_1, q'_1 \in Q$ tels que $g(h^{-1}(D_1^*) \cap u'\, R_{q,q'}(x)\, w'\, R_{q_1,q'_1}(y)\, v') \cap \{u\, x^{sn}\, w\, y^{tn}\, v \ /\ n \in \mathbb{N}_+\}$ soit infini avec $g(u') = u$, $g(w') = w$, $g(v') = v$, $f(q_o, u') = q$, $f(q', w') = q_1$ et $f(q'_1, v') \in F$. Tout langage rationnel est union finie de langages rationnels simples, donc il exsite $R_1 \subseteq R_{q,q'}(x)$, $R_2 \subseteq R_{q_1,q'_1}(y)$, $R'_1 = x_o\, x_1^* \dots x_p^*$, $R'_2 = y_o\, y_1^* \dots y_r^*$, les x_i, y_j appartenant à Y^*, tels que $c(R_1) = c(R'_1)$, $c(R_2) = c(R'_2)$ et $g(h^{-1}(D_1^*) \cap u'\, R_1\, w'\, R_2\, v') \cap \{u\, x^{sn}\, w\, y^{tn}\, v \ /\ n \in \mathbb{N}_+\}$ soit infini. Comme $c(x_o) \cap R_1 \neq \emptyset$ et que $g(R_1) \subseteq x^*$, il existe $z \in c(x_o)$ tel que

$g(z) \in x^*$ et on peut supposer que $x_o = z$. Pour tout $i \in \{1,\dots,p\}$, $c(x_o\ x_i) \cap R_1 \neq \emptyset$, il existe $z_i \in c(x_i)$ tel que $g(z_i) \in x^*$ et on peut supposer que pour tout $j \in \{0,\dots,r\}$, $g(y_j) \in y^*$. Pour tout mot $f \in \{a, b\}^*$, notons $d(f) = \ell_a(w) - \ell_b(w)$. Alors, comme $c(R_1) = c(R'_1)$, pour tout $z_1 \in R_1$ il existe $z'_1 \in R'_1$ tel que $c(z_1) = c(z'_1)$, ce qui entraine $d(h(z_1)) = d(h(z'_1))$ et $g(z_1) = g(z'_1)$ puisque $g(z_1)$ et $g(z'_1)$ sont inclus dans x^*. On en déduit, donc, $B = g(h^{-1}(D_1^*) \cap u'\ R'_1\ w'\ R'_2\ v') = g(h^{-1}(D_1^*) \cap u'\ R_1\ w'\ R_2\ v')$. Pour tout $i \in \{0,\dots,p\}$, notons $s_i = d(h(x_i))$ et soit $t_i \in \mathbb{N}$ tel que $g(x_i) = x^{t_i}$. De même, pour tout $j \in \{0,\dots,r\}$, on a $s'_j = d(h(y_j))$ et $g(y_j) = y^{t'_j}$. Comme $B \cap \{u\ x^{sn}\ w\ y^{tn}\ v\ /\ n \in \mathbb{N}_+\}$ est infini, il existe $n_1, n_2 \in \mathbb{N}_+$, $n_1 < n_2$, $\alpha_1, \alpha_2 \in R'_1$, $\beta_1, \beta_2 \in R'_2$ tels que $\forall\ i \in \{1, 2\}$, $g(h^{-1}(D_1^*) \cap u'\ \alpha_i\ w'\ \beta_i\ v') = u\ x^{sn_i}\ w\ y^{tn_i}\ v$ avec $\alpha_1 = x_o\ x_1^{i_1} \dots x_p^{i_p}$, $\alpha_2 = x_o\ x_1^{i'_1} \dots x_p^{i'_p}$, $\beta_1 = y_o\ y_1^{j_1} \dots y_r^{j_r}$, $\beta_2 = y_o\ y_1^{j'_1} \dots y_r^{j'_r}$ et $i_m \leq i'_m$, $\forall\ m \in \{1,\dots,p\}$, $j_m \leq j'_m$, $\forall\ m \in \{1,\dots,r\}$. Il existe, donc $x' = x_1^{i'_1-i_1} \dots x_p^{i'_p-i_p}$ et $y' = y_1^{j'_1-j_1} \dots y_r^{j'_r-j_r}$ tels que $g(x') = x^{s(n_2-n_1)}$, $g(y') = y^{t(n_2-n_1)}$ et $d(h(x')) + d(h(y')) = 0$. Clairement, comme g est limité sur R, donc sur R_1 et R_2, $\forall\ i \in \{1,\dots,p\}$, $t_i \neq 0$. Nous en déduisons que pour tout $x'' \in x_1^* \dots x_p^* \setminus \{\varepsilon\}$, $d(h(x'')) \neq 0$; en effet, dans le cas contraire, la paire itérante (u, x^s, w, y^t, v) ne serait pas très stricte. Nous pouvons, alors supposer, par exemple, que $s_i \in \mathbb{N}_+$, $\forall\ i \in \{1,\dots,p\}$. Supposons, maintenant, qu'il existe $i, i' \in \{1,\dots,p\}$ tels que $s_i\ t_{i'} \neq t_i\ s_{i'}$. Montrons qu'alors pour tout $\lambda \in \mathbb{N}_+$, il existe plus de λ valeurs de j tel que $u\ x^{sj}\ w\ y^{t\gamma}\ v \in L$ avec $\gamma = n_1 + (n_2 - n_1)\ \lambda\ s\ s_i\ s'_i$. En effet, posons $\delta = d(h(x'))$ et pour tout $\mu \in \{0,\dots,\lambda\}$, prenons $x'(\mu) \in x_o\ x_1^* \dots x_p^*$ tel que $c(x'(\mu)) = c(\alpha_1\ x_i^{\mu\ s\ s'_i\ \delta}\ x'_i{}^{(\lambda-\mu)\ s\ s_i\ \delta})$. Alors, $d(h(x'(\mu))) = d(h(\alpha_1)) + \lambda\ s\ s_i\ s'_i\ \delta$ et $g(x'(\mu)) = (x^s)^{m(\mu)}$ avec $m(\mu) = n_1 + \mu\ \delta\ t_i\ s'_i + (\lambda - \mu)\ \delta\ t'_i\ s_i$. Donc, pour tout $\mu \in \{0,\dots,\lambda\}$, $u'\ x'(\mu)\ w'\ y''\ v' \in h^{-1}(D_1^*) \cap u'\ R'_1\ w'\ R'_2\ v'$ où $y'' \in y_o\ y_1^* \dots y_r^*$ et vérifie $c(y'') = c(\beta_1\ y'^{\lambda\ s\ s_i\ s'_i})$ ce qui implique $u\ x^{sm(\mu)}\ w\ y^{t\gamma}\ v \in L$. Comme $\mu \neq \mu'$ entraine $m(\mu) \neq m(\mu')$, nous avons le résultat et la paire itérante (u, x^s, w, y^t, v) n'est pas très stricte d'où la contradiction.

Il existe, donc, un nombre rationnel positif r_1 tel que $r_1 = \frac{s_1}{t_1} = \dots = \frac{s_p}{t_p}$. De la même façon, on peut montrer qu'il existe un nombre rationnel négatif $r_2 = \frac{s'_1}{t'_1} = \dots = \frac{s'_r}{t'_r}$. Alors, comme $g(x') = x^{s(n_2-n_1)}$ et $d(h(x')) = \delta = -\ d(h(y'))$, nous obtenons $\delta = s\ r_1(n_2 - n_1) = -\ t\ r_2(n_2 - n_1)$ et $s\ r_1 + t\ r_2 = 0$. Soient α'_1 et β'_1 tels que $\alpha_1 = x_o\ \alpha'_1$ et $\beta_1 = y_o\ \beta'_1$. Comme $g(\alpha_1) = x^{sn_1}$ et $g(\beta_1) = y^{tn_1}$, nous avons $g(\alpha'_1) = x^{sn_1-t_o}$, $g(\beta'_1) = y^{tn_1-t'_o}$, $d(h(\alpha_1)) = s_o + r_1(sn_1-t_o)$ et $d(h(\beta_1)) = s'_o + r_2(tn_1 - t'_o)$. Comme $r_1(sn_1 - t_o) = s_1\ i_1 + \dots + s_p\ i_p$ et $r_2(tn_1 - t'_o) = s'_1\ j_1 + \dots + s'_r\ j_r$, $r_1(sn_1 - t_o) + r_2(tn_1 - t'_o) = n_1(r_1\ s + r_2\ t) - r_1\ t_o - r_2\ t'_o = -\ r_1\ t_o - r_2\ t'_o$ peut s'écrire $s_1\ \gamma_1 + \dots + s_p\ \gamma_p + s'_1\ \gamma'_1 + \dots\ s'_r\ \gamma'_r$ où les γ_i et $\gamma'_j \in \mathbb{N}$ et peuvent être bornés en fonction de x et y. Posons, alors, $k_1 = t_o + t_1\ \gamma_1 + \dots + t_p\ \gamma_p$, $k_2 = t'_o + t'_1\ \gamma'_1 + \dots + t'_r\ \gamma'_r$ et montrons que $u\ x^{k_1}\ w\ y^{k_2}\ v \in L$. Considérons $z'_1 = x_o\ x_1^{\gamma_1} \dots x_p^{\gamma_p}$ et $z'_2 = y_o\ y_1^{\gamma'_1} \dots y_r^{\gamma'_r}$. Alors $d(h(z'_1\ z'_2)) = s_o + s'_o - r_1\ t_o - r_2\ t'_o = s_o + s'_o + r_1(sn_1 - t_o) + r_2(tn_1 - t'_o)$ et $d(h(u'\ z'_1\ w'\ z'_2\ v')) = d(h(u'\ \alpha_1\ w'\ \beta_1\ v')) = 0$. Donc $u'\ z'_1\ w'\ z'_2\ v' \in h^{-1}(D_1^*) \cap u'\ R'_1\ w'\ R'_2\ v'$ et $u\ x^{k_1}\ w\ y^{k_2}\ v = g(u'\ z'_1\ w'\ z'_2\ v') \in L$. De plus, la construction de k_1 et k_2 montre qu'on peut trouver pour k_1 et k_2 une borne qui ne dépend que de x et y. □

Nous pouvons, maintenant, démontrer facilement notre premier résultat sur les langages algébriques commutatifs : le langage de Dyck sur une lettre, D_1^*, ne domine pas rationnellement tous les langages algébriques commutatifs. Considérons, en effet, le langage L_3 défini au début de cette section. Supposons que $L_3 \in C(D_1^*)$, et soit k la borne de la proposition précédente pour $L = L_3$, $x = a_2$ et $y = a_3$. Pour tout $n \in \mathbb{N}_+$, $a_1^{2k} a_2^{nk} a_3^{nk} \in L_3$ et comme $a_1^{2k} a_2^{i} a_3^{j} \in L$ implique $-2k \leq i - j \leq 2k$, $(a_1^{2k}, a_2, \varepsilon, a_3, \varepsilon)$ est une prépaire itérante très stricte dans L_3 et d'après la proposition précédente, il existe des entiers positifs k_1 et k_2 inférieurs à k tels que $a_1^{2k} a_2^{k_1} a_3^{k_2} \in L_3$ d'où la contradiction, puisque $k_1 + k_2 < 2k$. Donc L_3 n'appartient pas à $C(D_1^*)$. Par contre D_1^* est rationnellement équivalent à $L_3 \cap T_2^*$, donc $D_1^* \in C(L_3)$ et :

Proposition III.55 : *Il existe un langage algébrique commutatif L tel que* $C(D_1^*) \subsetneq C(L)$

Montrons maintenant que Rat est le seul cône rationnel (fidèle) clos par intersection et contenu dans la famille des langages algébriques. En fait nous allons établir une propriété plus fine qui nous permettra d'étendre aux langages commutatifs algébriques un résultat déjà montré pour le langage D_1^*.

Proposition III.56 : *Soient* $L_1 \subseteq T_1^*$, $L_2 \subseteq T_2^*$ *des langages algébriques avec* $T_1 \cap T_2 = \emptyset$. *Alors* $L = shuf(L_1, L_2)$ *est algébrique ssi l'un des deux langages* L_1, L_2 *est rationnel.*

Démonstration : Si $L_2 \in$ Rat, $L = \text{shuf}(L_1, L_2) \in C(L_1)$ et L est algébrique. De même si L_1 est rationnel, L est algébrique.

Considérons les alphabets $T'_1 = \{a'_1 / a_1 \in T_1\}$, $T'_2 = \{a'_2 / a_2 \in T_2\}$ tels que T_1, T_2, T'_1, T'_2 soient disjoints deux à deux et les homomorphismes h_1, h_2 définis respectivement sur T'_1 et T'_2 par $h_1(a'_1) = a_1$, $\forall a'_1 \in T'_1$ et $h_2(a'_2) = a_2$, $\forall a'_2 \in T'_2$. Le langage $L' = \{x_1 x_2 x'_1 x'_2 / x_i \in T_i^+, x'_i \in T'^+_i, x_i h_i(x'_i) \in L_i, \forall i \in \{1,2\}\}$ est égal à $h^{-1}(L) \cap T_1^+ T_2^+ T'^+_1 T'^+_2$ où h est l'homomorphisme défini sur $T = T_1 \cup T_2 \cup T'_1 \cup T'_2$ par $h(a) = a$ si $a \in T_1 \cup T_2$, $h_1(a)$ si $a \in T'_2$. Donc $L' \in C(L)$ et si L est algébrique, L' est aussi algébrique et il existe une grammaire algébrique réduite $G = (N, T, P, S)$ qui engendre L'. Comme $L' \subseteq TT^+$, on peut supposer que G est sous forme de Greibach double, c'est-à-dire que $P \subseteq N \times T V^* T$ avec $V = N \cup T$ (cf. [50], [64]). Notons, alors, $P_1 = P \cap (N \times T_1 V^* T'_1)$ et $P_2 = P \cap (N \times T_2 V^* T'_2)$. Comme L' est inclus dans $T_1^* T_2^* T'^*_1 T'^*_2$, aucune dérivation $S \overset{*}{\Rightarrow} w \in L'$ ne peut utiliser des règles de P_1 et des règles de P_2, donc $L(G) = L(G_1) \cup L(G_2)$ où $\forall i \in \{1,2\}$, G_i est la grammaire $(N, T, P \setminus P_i, S)$. Pour $i = 1, 2$, soit g_i l'homomorphisme défini sur V par $g_i(x) = \varepsilon$ si $x \in T_i \cup T'_i$, x sinon et posons $Y_i = g_2(L(G_i))$, $Y'_i = g_1(L(G_i))$. Montrons que, soit $Y_1 = g_2(L')$, soit $Y'_2 = g_1(L')$. Supposons, en effet, que $Y'_2 \neq g_1(L')$. Comme $Y'_2 \subseteq g_1(L')$, c'est qu'il existe $z_2 z'_2 \in g_1(L') \setminus Y'_2$ avec $z_2 \in T_2^+$ et $z'_2 \in T'^+_2$. Alors pour tout $z_1 z'_1 \in g_2(L')$ avec $z_1 \in T_1^+$ et $z'_1 \in T'^+_1$, $z = z_1 z_2 z'_1 z'_2 \in L' = L(G_1) \cup L(G_2)$ et $z \notin L(G_2)$, donc $z \in L(G_1)$ et $z_1 z'_1 = g_2(z) \in Y_1$ ce qui entraine $Y_1 = g_2(L')$.

Supposons, alors, que Y_1 soit égal à $g_2(L')$ (le cas $Y'_2 = g_1(L')$ se traiterait de façon symétrique) et montrons, que L_1 est un langage rationnel. Pour obtenir ce résultat, nous allons contruire une grammaire algébrique qui engendre Y_1 et qui soit aussi "proche" que possible de la grammaire G. Posons $H = \{(B \to g_2(z)\} / \{B \to z\} \in P \setminus P_1\}$ et soit la grammaire algébrique $G'_1 = (N', T_1 \cup T_3, P', S)$ construite à partir de la grammaire $(N, T_1 \cup T_3, H, S)$ en appliquant l'algorithme d'élimination des règles donnant le mot vide. Comme $Y_1 = g_2(L(G_1))$ est inclus dans $T_1^+ T'^+_1$, il est clair que $L(G'_1) = Y_1$. Pout tout $z \in V' = N' \cup T_1 \cup T_3$, notons $L(G'_1, z) = \{w \in (T_1 \cup T_3)^* / z \overset{*}{\underset{G'_1}{\Longrightarrow}} w\}$. Alors N' se partitionne en $N'_1 \cup N'_2 \cup N'_3$ avec $N'_1 = \{B \in N' / L(G'_1, B) \subseteq T_1^+\}$, $N'_2 = \{B \in N' / L(G'_1, B) \subseteq T'^+_1\}$ et $N'_3 = \{B \in N' / L(G'_1, B) \cap T_1^+ T'^+_1 \neq \emptyset\}$. Notons $P'_o = \{\{B \to z\} \in P' / z \in (T_1 \cup N'_1)^* (T'_1 \cup N'_2)^*\}$ et pour tout règle $\pi = \{B \to z_1 z_2\} \in P'_o$ avec $z_1 \in (T_1 \cup N'_1)^*$ et $z_2 \in (T'_1 \cup N'_2)^*$, posons $L_\pi = \{w L(G'_1, z_1) / S \overset{*}{\underset{G'_1}{\Longrightarrow}} w B$ avec $w \in T_1^*\}$ et $L'_\pi = L(G'_1, z_2)$. Montrons que $L(G'_1) = \bigcup_{\pi \in P'_o} L_\pi L'_\pi$.

Prenons $\pi = \{B \to z_1 z_2\} \in P'_o$ avec $z_1 \in (T_1 \cup N'_1)^*$ et $z_2 \in (T'_1 \cup N'_2)^*$, $w_1 \in L_\pi$ et $w_2 \in L'_\pi$. Alors $w_1 = w'_1 w''_1$ avec $S \overset{*}{\underset{G'_1}{\Longrightarrow}} w'_1 B$, $z_1 \overset{*}{\underset{G'_1}{\Longrightarrow}} w''_1$ et $z_2 \overset{*}{\underset{G'_1}{\Longrightarrow}} w_2$, donc $S \overset{*}{\underset{G'_1}{\Longrightarrow}} w'_1 B \underset{\pi}{\Longrightarrow} w'_1 z_1 z_2 \overset{*}{\underset{G'_1}{\Longrightarrow}} w'_1 w''_1 w_2 = w_1 w_2 \in L(G'_1)$ et $L_\pi L'_\pi \subseteq L(G'_1)$.

Réciproquement, prenons $w \in L(G'_1)$. Comme $S \overset{*}{\underset{G'_1}{\Longrightarrow}} w \in T_1^+ T'^+_1$, $S \in N'_3$. Alors, on peut écrire la dérivation $S = y_o \Rightarrow y_1 \Rightarrow \dots \Rightarrow y_n \Rightarrow w$. Soit t le plus petit $i \in \{0,\dots,n\}$ tel que y_i ne contienne aucun élément de N'_3. Alors y_{t-1} et y_t se factorisent en $y_{t-1} = u B v$, $y_t = u z_1 z_2 v$ avec $B \in N'_3$, $u z_1 \in (T_1 \cup N'_1)^*$ et $z_2 v \in (T'_1 \cup N'_2)^*$. Par construction de la grammaire G'_1, pour toute règle $\{B \to z\} \in P$, il existe dans G_1, une dérivation $S \overset{*}{\Rightarrow} z'$ avec $g_2(z') = z$. Donc il existe $u', v' \in V^*$ tels que $S \overset{*}{\underset{G_1}{\Longrightarrow}} u' B v'$, $g_2(u') = u$ et $g_2(v') = v$. Alors, comme $B \in N'_3$ et que la grammaire G_1 ne contient aucune règle de P_1, il existe $z' \in T_1 V^* T'_2$ tel que $\{B \to z'\} \in P$ et comme G est une grammaire réduite $v' \overset{*}{\underset{G}{\Rightarrow}} w' \in T^*$ implique $w' \in T'^*_2$, donc $v \overset{*}{\underset{G'_1}{\Longrightarrow}} w' \in (T_1 \cup T'_1)^*$ entraine $w' = \varepsilon$ et comme G'_1 ne contient pas de règle donnant le mot vide, $v = \varepsilon$. On peut, alors, factoriser w en $w_1 w'_1 w_2$ tels que $u \overset{*}{\underset{G'_1}{\Longrightarrow}} w_1$, $z_1 \overset{*}{\underset{G'_1}{\Longrightarrow}} w'_1$, $z_2 \overset{*}{\underset{G'_1}{\Longrightarrow}} w_2$ et comme $S \overset{*}{\underset{G'_1}{\Longrightarrow}} u B \overset{*}{\underset{G'_1}{\Longrightarrow}} w_1 B$ et que $\pi = \{B \to z_1 z_2\} \in P'_o$, on obtient $w_1 w'_1 \in L_\pi$, $w_2 \in L'_\pi$, $w \in L_\pi L'_\pi$ et $L(G'_1) = \bigcup_{\pi \in P'_o} L_\pi L'_\pi$.

Nous avons, donc, $g_2(L') = \bigcup_{\pi \in P'_o} L_\pi L'_\pi$ où P'_o est un ensemble fini, $L_\pi \subseteq T_1^+$ et $L'_\pi \subseteq T'^+_1$. Alors, pour tout mot $w \in L_1$, de longueur supérieure à 1, et toute factorisation $w = w_1 w'_1$ avec $w_1, w'_1 \neq \varepsilon$, il existe $\pi \in P'_o$ tel que $w_1 \in L_\pi$ et $w'_1 \in h_1(L'_\pi)$. Pour tout $x \in T_1^*$, soit $D(x) = \{z / x z \in L_1\}$, l'ensemble des contextes à droite de x pour L_1. Alors, $D(\varepsilon) = L_1$ et pour $x \neq \varepsilon$, $D(x) = \{z \in h_1(L'_\pi) / x \in L_\pi\} \cup E$ où E est égal à $\{\varepsilon\}$ si $x \in L_1$, $\emptyset$ sinon. Donc $\{D(x) / x \in T_1^*\}$ est fini et L_1 est un langage rationnel.□

En fait, la démonstration précédente reste valable si nous remplaçons dans l'énoncé de la proposition, shuf (L_1, L_2) par $i_2(L_1, L_2)$ l'intercalation de rang 2 que S. Greibach a définie dans [42] par

$$i_2(L_1, L_2) = \{x_1 y_1 x_2 y_2 x_3 / x_1, x_2, x_3 \in T_1^*, y_1, y_2 \in T_2^*, x_1 x_2 x_3 \in L_1, y_1 y_2 \in L_2\}.$$

Alors, pour tout langage L non rationnel, si nous prenons une recopie $\bar{L}$ de L sur un alphabet disjoint, la proposition précédente implique que $i_2(L, \bar{L})$ et shuf $(L, \bar{L})$ ne sont pas des langages algébriques. Comme un cône rationnel est clos par intersection ssi il est clos pour l'opération "shuffle", nous obtenons :

Corollaire III.57 : *La famille des langages rationnels, Rat, est le seul cône rationnel (fidèle) clos par intersection contenu dans la famille des langages algébriques.*

Pour toute famille de langages $\mathcal{L}$, notons $\mathcal{C}^f_\cup(\mathcal{L})$, le plus petit cône rationnel fidèle contenant $\mathcal{L}$, clos par union, différence avec un langage rationnel et donc par intersection. La famille $\mathcal{C}^f_\cup(C_1)$ avec $C_1 = \{a^n b^n \ / \ n \geq 0\}$ est la famille des langages correspondant aux prédicats rudimentaires (cf. [51]). Comme $D_1^* = \text{shuf}(C_1, C_1^R)$ (lemme III.11), que $D'^*_1 = D_1^* \setminus D_1^* \, b \, \{a, b\}^*$ et que D'^*_2, le langage de semi-Dyck sur deux lettres s'exprime au moyen d'opérations booléennes à partir de langages appartenant à $\mathcal{C}(D'^*_1)$ [80], $\text{Alg} = \mathcal{C}(D'^*_2)$ est inclus dans $\mathcal{C}^f_\cup(C_1)$ ([51], [80]), ainsi que la famille $\mathcal{C}^f_\cap(\text{Alg})$ étudiée par Book et Greibach dans [16]. On ne sait pas si ce résultat reste vrai si on remplace C_1 par un langage quelconque non rationnel. Par contre, le corollaire précédent permet d'affirmer que $\mathcal{C}^f_\cup(L) \subseteq \text{Alg}$ implique $L \in \text{Rat}$.

Prenons, maintenant, un langage algébrique commutatif L et supposons que $\mathcal{C}(L)$ contienne le produit $L_1 L_2$ avec $L_1 \subseteq T_1^*$, $L_2 \subseteq T_2^*$, $T_1 \cap T_2 = \emptyset$. Alors, d'après la proposition II.15, $L' = \text{shuf}(L_1, L_2) \in \mathcal{C}(L)$ et L' est un langage algébrique ce qui entraine, d'après la proposition III.56 que l'un des deux langages L_1, L_2 est rationnel. Nous obtenons, alors, pour tous les langages algébriques commutatifs une propriété démontrée pour D_1^* dans la section précédente par une méthode tout à fait différente :

Corollaire III.58 : *Tout langage algébrique commutatif est sans produit.*

Cette propriété est aussi vérifiée pour les langages linéaires [38]. En utilisant l'opérateur crochet, défini par Boasson, Crestin et Nivat [13], nous allons pouvoir étendre ce corollaire à une famille plus large de langages algébriques, contenant en particulier les langages linéaires.

Pour tout langage L, posons $[L] = \{w \, x \, w^R \ / \ w \in \{a, b\}^*, \ x \in L\}$ où w^R est l'image miroir de w.

Une famille de langages $\mathcal{L}$ est close par crochet si $L \in \mathcal{L}$ implique $[L] \in \mathcal{L}$.

Pour toute famille de langages $\mathcal{L}$, $[\mathcal{L}]$ désignera le plus petit cône rationnel contenant $\mathcal{L}$, clos par union et crochet. Ainsi, [Rat] = Lin la famille des langages linéaires. Rappelons un résultat que nous avons montré dans [56].

Proposition III.59 : *Soient* L_1 *et* L_2 *deux langages non rationnels définis sur des alphabets disjoints et* $\mathcal{L}$ *un cône rationnel clos par union. Alors* $L_1 L_2 \in [\mathcal{L}]$ *implique* $L_1 L_2 \in \mathcal{L}$.

Cette proposition implique, en particulier, que si L est un langage sans produit, [L] est aussi un langage sans produit et donc :

<u>Corollaire III.60</u> : *Pour tout langage algébrique commutatif* L, $[L] = \{w\, x\, w^R \mid x \in L,\ w \in \{a, b\}^*\}$ *est un langage sans produit.*

B I B L I O G R A P H I E

[1] A. ARNOLD et M. LATTEUX, "*Récursivité et cônes rationnels fermés par intersection*", (1976), soumis à Calcolo.

[2] J.M. AUTEBERT, "*Non-principalité du cylindre des langages à compteur*", (1977) à paraître dans Math. Syst. Theory.

[3] B.S. BAKER et R.V. BOOK, "*Reversal-bounded Multipushdownmachines*", J. Comp. Syst. Sc. 8 (1974) 315-332.

[4] J. BEAUQUIER, "*Contribution à l'étude de la complexité structurelle des langages algébriques*", Thèse Sc. Math., Université de Paris VI, Paris 1977.

[5] J. BERSTEL, "*Transductions and Context-free languages*", Teubner Verlag, 1978.

[6] J. BERSTEL, "*Une hiérarchie des parties rationnelles de* $\mathbb{N}^2$", Math. Syst. Theory 7 (1973), 114-137.

[7] J. BERSTEL et L. BOASSON, "*Une suite décroissante de Cônes Rationnels*", in Loeckx (Ed.), Automata, Languages and Programming, 2nd Colloquium, Saarbrücken, Springer Verlag 1974, 383-397.

[8] L. BOASSON, "*Paires Itérantes et Langages Algébriques*", Thèse Sc. Math. Université de Paris VII, Paris 1974.

[9] L. BOASSON, "*Two iteration theorems for some families of languages*", J. Comp. Syst. Sc. 7 (1973), 583-596.

[10] L. BOASSON, "*Langages algébriques, paires itérantes et transductions rationnelles*", Theoretical Computer Sciences 2 (1976), 209-223.

[11] L. BOASSON, "*Classification of the context-free languages*", in : Gruska (Ed), Mathematical Foundations of Computer Science, 6ème Symposium, Tatranska Lomnica 1977, 34-43.

[12] L. BOASSON, B. COURCELLE et M. NIVAT, "*A new Complexity measure for languages*", in Proceedings of a Conference on theoretical Computer Science, Waterloo (1977), 130-138.

[13] L. BOASSON, J.P. CRESTIN et M. NIVAT, "*Familles de langages translatables et fermées par crochet*", Acta Informatica 2 (1973), 383-393.

[14] L. BOASSON et M. NIVAT, "*Sur diverses familles de langages fermées par transduction rationnelle*", Acta Informatica, 2 (1973) 180-188.

[15] R.V. BOOK, "*On Languages with a Certain Prefix Property*", Math. Syst. theory 10 (1977), 229-237.

[16] R.V. BOOK et S. GREIBACH, "*Quasi-Realtime Languages*", Math. Syst. Theory 4 (1970), 97-111.

[17] R.V. BOOK, M. NIVAT et M. PATERSON, "*Intersection of Linear Context-Free Languages and Reversal-Bounded Multipushdown Machines*", Procedings of the sixth annual ACM Symposium on Theory of Computing, 1974, 290-296.

[18] J.L. DURIEUX, "*Sur l'image, par une transduction rationnelle, des mots sur une lettre*", R.A.I.R.O., R-2 (1975), 25-37.

[19] S. EILENBERG, Communication au Congrès international des mathématiciens, Nice, 1970.

[20] S. EILENBERG et M.P. SCHÜTZENBERGER,"*Rational Sets in Commutative Monoïds*", J. Algebra 13 (1969), 173-191.

[21] C.C. ELGOT et J.F. MEZEI, "*On relations defined by generalized finite automata*", IBM J. of Res. and Dev. 9 (1962), 47-68.

[22] M.J. FISCHER et A.L. ROSENBERG, "*Real-time solutions of the origin-crossing problem*", Math. Syst. Theory 2 (1968), 257-263.

[23] M. FLIESS, Communication personnelle.

[24] S. GINSBURG, "*The Mathematical theory of Context-free languages*", Mc Graw-Hill, New-York, 1966.

[25] S. GINSBURG, "*Algebraic and automata-theoretic properties of formal languages*", North-Holland Publishing company 1975.

[26] S. GINSBURG, J. GOLDSTINE et S. GREIBACH, "*Uniformly erasable AFL*", J. Comp. Syst. Sc. 10 (1975), 165-182.

[27] S. GINSBURG, J. GOLDSTINE et S. GREIBACH, "*Some Uniformly Erasable families of Languages*", Theoretical Computer Science 2 (1976), 29-44.

[28] S. GINSBURG et S. GREIBACH, "*Abstract Families of Languages*", Memoirs of the American Mathematical Society 87 (1969), 1-32.

[29] S. GINSBURG et S. GREIBACH "*Principal AFL*", J. Comp. Syst. Sc. 4 (1970), 308-338.

[30] S. GINSBURG et S. GREIBACH, "*Multitape AFA*" JACM 19 (1972), 193-221.

[31] S. GINSBURG et E.H. SPANIER, "*Bounded Algol-Like Languages*", Transductions of the American Math. Soc. 113 (1964), 333-368.

[32] S. GINSBURG et E.H. SPANIER, "*Bounded Regular Sets*", Proceedings of the American Math. Soc. 17 (1966), 1043-1049.

[33] S. GINSBURG et E.H. SPANIER, "*AFL with the semilinear Property*", J. Comp. Syst. Sc. 5 (1971), 365-396.

[34] S. GINSBURG et E.H. SPANIER, "*On Incomparable Abstract Families of Languages (AFL)*", J. Comp. Syst. Sc. 9 (1974), 88-108.

[35] J. GOLDSTINE, "*Abstract Families of Languages generated by Bounded-Languages*", Ph. D. Thesis, University of California at Berkeley, 1970.

[36] J. GOLDSTINE, "*Substitution and Bounded Languages*", J. Comp. Syst. Sc. 6 (1972), 9-29.

[37] J. GOLDSTINE, "*Bounded AFL's*", J. Comp. Syst. Sc. 12 (1976), 399-419

[38] S. GREIBACH, "*The Unsolvability of the Recognition of Linear Context-Free Languages*", JACM 13 (1966) 582-587.

[39] S. GREIBACH, "*An Infinite Hierarchy of context-free Languages*", JACM 16 (1969) 91-106.

[40] S. GREIBACH, "*Chains of Full AFL's*", Math. Syst. theory 4 (1970), 231-242.

[41] S. GREIBACH, "*Erasing in Context-free AFLs*", Information and Control 21 (1972), 436-465.

[42] S. GREIBACH, "*Simple Syntactic Operators on Full Semi-AFL's*" J. Comp. Syst. Sc. 6 (1972),30-76.

[43] S. GREIBACH, "*One Counter Language and the IRS Condition*", J. Comp. Syst. Sc. 10 (1975), 237-247.

[44] S. GREIBACH, "*Erasable Context-free languages*", Information and Control 29 (1975), 301-326.

[45] S. GREIBACH, "*the strong independence of substitution and homomorphic repliation*", (1977) à paraître.

[46] J. HARTMANIS et J.E. HOPCROFT, "*What makes some Language theory Problems Undecidable*", J. Comp. Syst. Sc. 4 (1970), 368-376.

[47] G.T. HERMAN et G. ROZENGERG, "*Developmental Systems and Languages*", North Holland Publishing Company, Amsterdam (1975).

[48] O.H. IBARRA, "*A note on Semilinear Sets and Bounded Reversal Multihead Push-down Automata*", Information Processing Letters 3 (1974), 25-28.

[49] R. ITO,"*Every Semilinear Set is a Finite Union of Disjoint Linear Sets*", J. Comp. Syst. Sc.3(1969)221-231.

[50] G. JACOB, "*Représentation et Substitutions Matricielles dans la Théorie Algébrique des Transductions*", Thèse Sc. Math., Univ. Paris VII, Paris 1975.

[51] N. JONES, "*Context-free Languages and Rudimentary Attributes*", Math. Syst. Theory 3 (1969), 102-109.

[52] M. LATTEUX, "*Sur les Semi-linéaires-langages bornés*", Publication du Laboratoire de Calcul de L'université de Lille I, N° 60, 1975.

[53] M. LATTEUX, "*Intersections de langages algébriques bornés*", (1976) soumis à Acta Informatica.

[54] M. LATTEUX, "*Cônes Rationnels Commutativement Clos*", R.A.I.R.O. Informatique théorique 11(1977) 29-51.

[55] M. LATTEUX, "*Produit dans le cône rationnel engendré par* D_1^*", Theoretical Computer Science 5 (1977) 129-134.

[56] M. LATTEUX, "*Substitutions dans les EDTOL-Systèmes ultralinéaires*", (1976) soumis à Information and Control.

[57] M. LATTEUX, "*Cônes rationnels commutatifs*", (1977) soumis à J. Comp. Syst. Sc.

[58] M. LATTEUX, "*Une note sur la propriété de préfixe*", (1977) à paraitre dans Math. Syst. Theory.

[59] M. LATTEUX, "*Mots infinis et langages commutatifs*", (1977) à paraître dans R.A.I.R.O. Informatique théorique .

[60] M. LATTEUX, "*EDTOL-systèmes ultralinéaires et opérateurs associés*", Publication du laboratoire de calcul de l'Université de Lille I, n° 100, 1977.

[61] L.Y. LIU et P. WEINER, "*A Characterization of Semilinear Sets*", J. Comp. Syst. Sc. 4 (1970) 299-307.

[62] L.Y. LIU et P. WEINER, "*An Infinite Hierarchy of Intersections of Context-Free Languages*", Math. Syst. Theory 7 (1973), 185-192.

[63] H.A. MAURER, "*The solution of a Problem by Ginsburg*", Information Processing letters 1 (1971) 7-10.

[64] M. NIVAT, "*Transductions des langages de Chomsky*", Ann. Inst. Fourier (Grenoble) 18 (1968), 339-455.

[65] M. NIVAT, "*Opérateurs sur les familles de langages*", Rapport de recherche IRIA N° 106, 1975.

[66] M. NIVAT, "*Mots Infinis engendrés par une grammaire algébrique*", R.A.I.R.O. Informatique théorique 11 (1977), 311-327.

[67] W. OGDEN, "*A hepful result for proving inherent ambiguity*", Math. Syst. Theory 2 (1967), 191-194.

[68] I. OSHIBA, "*On permutting letters of words in context-free Languages*", Information and Control 20 (1972), 405-409.

[69] R.J. PARIKH, "*Languages Generating Devices*", M.I.T. Res. Lab. Electron. Quart. Prog. Rept. 60 (1961) 199-212.

[70] R.J. PARIKH, "*On context-free Languages*",JACM 13 (1966), 570-581.

[71] J.F. PERROT, "*Sur la fermeture commutative des C-langages*", C.R. Acad. Sci. Paris, Sér. A, 265 (1967) 597-600.

[72] F. RODRIGUEZ, "*Familles de langages fermées par crochet et crochet ouvert*", Lectures Notes in Computer Science, Springer Verlag, Heidelberg, 48 (1977), 154-168.

[73] B. ROVAN, "*Proving containment of bounded AFL*", J. Comp. Syst. Sc. 11 (1975), 1-55.

[74] J. SAKAROVITCH, "*On deterministic commutative context-free languages*", in Proceedings of a Conference on Theoretical Computer Science, Waterloo (1977), 88-93.

[75] A. SALOMAA, "*Theorems on the representation of events in Moore-Automata*", Ann. Univ. Turku, Ser. AI 69 (1964).

[76] A. SALOMAA, "*Formal languages*", Academic Press, New-York, 1973.

[77] J.M. STEYAERT, Communication personnelle.

[78] I.H. SUDBOROUGH, "*Bounded-Reversal multihead finite automata languages*", Information and control 25 (1974), 317-328.

[79] J.S. ULLIAN, "*Three Theorems Concerning Principal AFL*", J. Comp. Syst. Sc 5 (1971) 304-314.

[80] C. WRATHALL, "*Characterizations of the Dyck Sets*", R.A.I.R.O. Informatique théorique 11 (1977), 53-62.