N° d'ordre : 827

THÈSE

présentée à

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE

pour obtenir le titre de

DOCTEUR DE 3ème CYCLE

(Traitement de l'Information)

par

Jeannine LEGUY-CORDELLIER

TRANSDUCTIONS RATIONNELLES DECROISSANTES ET SUBSTITUTION

Soutenue le Mardi 24 juin 1980 devant la Commission d'Examen

Membres du Jury	MM. G. JACOB	Président
	M. LATTEUX	Rapporteur
	J. BEAUQUIER	Examinateur
	J. BERSTEL	Examinateur
	L. BOASSON	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

Il m'est agréable d'exprimer ma reconnaissance envers Gérard JACOB pour l'honneur et le plaisir qu'il me fait en présidant le Jury.

Je remercie Joffroy BEAUQUIER de s'être intéressé à mon travail et d'avoir accepté de faire partie du Jury.

Je remercie Jean BERSTEL de l'intérêt qu'il a montré pour mes travaux. Ses encouragements et ses nombreux conseils m'ont été très précieux.

Je remercie Luc BOASSON dont les remarques et critiques ont souvent été à l'origine de nouveaux développements.

Je remercie particulièrement Michel LATTEUX qui a eu le mérite de me faire apprécier le difficile domaine qu'est la théorie des langages. Le temps qu'il m'a consacré, les conseils qu'il m'a donnés, la confiance qu'il m'a témoignée ont été des atouts précieux sans lesquels ce travail n'aurait pu aboutir.

Je suis reconnaissante envers les membres de l'Equipe d'Informatique Théorique de Lille d'avoir su créer une ambiance à la fois gaie et stimulante, et plus particulièrement envers Bernard LEGUY qui, à divers titres, m'a aidée dans la réalisation de ce travail.

Je remercie le sécrétariat du Laboratoire d'Informatique de Lille et plus précisément Madame Patricia CARON, Madame et Monsieur DEBOCK qui ont su, fidèles à leur réputation, réaliser un travail de qualité.

à Emmanuel

à Eve

à mes parents

TABLE DES MATIÈRES

* * *

* * *

INTRODUCTION

La précision indispensable dans la communication avec une machine a conduit l'informaticien à introduire la notion de langage formel, c'est à dire de langage pour lequel il correspond une définition précise et rigoureuse. Il est donc naturel de constater une collaboration souvent étroite entre les concepteurs de langages de programmation et les théoriciens des langages formels. On trouve dans certains "vieux" langages des restrictions (limitation à trois niveaux d'imbrication de parenthèses en Fortran II par exemple) dues à des problèmes théoriques non résolus à l'époque de leur définition. De façon symétrique bien des problèmes de transformations de grammaires, par exemple, n'auraient pas été étudiés s'il n'y avait eu des problèmes pratiques sousjacents. Cet échange est très fructueux dans bien des domaines tels que l'analyse syntaxique, la compilation, la programmation ...

Pour définir un langage infini, comme pour tout autre ensemble infini d'ailleurs, il y a plusieurs moyens de procéder. On peut, par exemple utiliser une caractérisation des mots du langage (reconnaissance), il est possible également de mettre en relief une propriété du processus de génération (grammaire) ou encore de définir ce langage à partir de langages déjà connus à l'aide de certaines opérations.

Reconnaître de façon automatique les mots d'un langage a très vite été une préoccupation des chercheurs d'où est née la notion d'automate. Parmi ces machines, celles qui possèdent un nombre fini d'états et qui peuvent avoir la possibilité de mémoriser un nombre non borné d'informations, ont particulièrement été étudiées ce qui a permis de dégager différents types d'automates dont les automates à pile de mémoire et les automates

sans mémoire annexe . Les machines d'états fini
(sans mémoire annexe) ont des applications pratiques (en programmation par exemple) évidentes, et Reg, la famille des langages réguliers reconnus par ce type de machine, est une famille fondamentale de langages.

Très vite également les grammaires formelles ont fait l'objet d'études intensives. Chomsky a dégagé une célèbre hiérarchie en distinguant quatre types de grammaire dont, en particulier, les grammaires algébriques et les grammaires linéaires à droite correspondant aux langages rationnels. C'est à dire aux langages obtenus à partir des langages finis, par application des opérations classiques d'union, de produit et d'étoile.

Kleene fut le premier à rapprocher ces points de vue en démontrant l'identité entre les langages rationnels et les langages réguliers.

S'il n'est pas impossible, compte tenu des développements technologiques et, en particulier du parallèlisme, de voir se développer des langages de programmation possédant une certaine "commutativité" qui refléterait une indépendance quant à l'ordre d'écriture, à l'heure actuelle les langages algébriques correspondent à un bon modèle des langages de programmation (Algol, Pascal). L'étude de la famille Alg, des langages algébriques puise d'autres motivations dans des domaines différents tels que la théorie des séries formelles ou encore la théorie des schémas de programmes (arbres ...).

La transduction rationnelle, outil fondamental en théorie des langages fut définie par Elgot et Mezei [11]. L'équivalence entre transduction rationnelle d'une part et homomorphisme, homomorphisme inverse, intersection avec un langage rationnel d'autre part a été montrée par Nivat [28]. Cette caractérisation est à l'origine d'un grand nombre de travaux. Une synthèse des résultats concernant tant l'outil que ses applications a été réalisée par Berstel.[6].

Ceci a conduit la théorie des langages à s'enrichir du concept de famille agréable de langages [16]. Un cône rationnel est une famille de langages fermée par transduction rationnelle, une famille agréable de langage est un cône rationnel fermé par union produit et étoile. Ainsi ce nouveau concept permet, par une vue plus globale, de dégager des propriétés communes à certaines familles, comme par exemple des propriétés de clôture. Cette notion a servi à unifier des résultats et à mettre en évidence des propriétés structurelles de certains types de langages.

L'opérateur de substitution est certainement l'opérateur le plus naturel en Informatique. Quoi de plus naturel en effet que de remplacer, dans un programme écrit dans un langage de programmation un nom de variable par un autre , un nom par un groupe d'ordres (procédure) ... ! La richesse et la puissance de cet opérateur explique certainement pourquoi tant de chercheurs l'ont étudié. Après avoir substitué à une lettre une lettre, à une lettre un mot, on a substitué à une lettre un langage. L'opérateur a ensuite été étendu aux familles de langages et peut être considéré comme un constructeur de familles dont certaines propriétés sont héritées des familles initiales. L'autre outil de construction de familles de langages qu'est la transduction rationnelle a très vite été rapproché de l'opérateur de substitution et ce par l'intermédiaire de la substitution syntaxique.

A partir de la famille {L} constituée du seul langage L, on peut définir le cône rationnel principal, noté C(L), engendré par L. C(L) est la plus petite famille de langages contenant L et fermée par transduction rationnelle. Plus précisément, compte tenu de la composition des transductions rationnelles [11], C(L) est la famille des langages que l'on peut obtenir à partir de L par une transduction rationnelle.

En imposant certaines restrictions aux transductions rationnelles nous obtenons un outil plus manipulable mais moins puissant. Le but, poursuivi ici, est de tenter de déterminer des propriétés des langages L pour lesquels la famille de langages contenant L et fermée par transduction rationnelle "restreinte" reste identique à C(L). Les transductions rationnelles fidèles (tout mot est l'image d'un nombre fini de mots) ont permis de définir les langages effaçables, langages pour lesquels il y a identité entre C(L) et $C^{f}(L)$, la plus petite famille contenant L et fermée par transduction rationnelle fidèle. De nombreuses études [7], [8], [15], [33] ont montré que, par exemple tout générateur de Alg, la famille des langages algébriques, est effaçable, de même pour tout langage défini sur une lettre. De façon symétrique, on définit les transductions rationnelles d'image finie (l'ensemble des images d'un mot est fini). Ces transductions n'ont pas la puissance des transductions rationnelles quelconques, et d'un langage L on ne pourra généralement pas atteindre, par transduction d'image finie, tous les éléments de C(L). Ceci dépendra de la structure du langage L. On appelle langage d'image finie un langage vérifiant justement l'égalité entre C(L) et $C^{if}(L)$, la plus petite famille contenant L et fermée par transduction rationnelle d'image finie. Un langage d'image finie est,en quelques sortes, un langage suffisamment "fort" pour compenser la perte de puissance des transductions rationnelles d'image finie.

Un cas particulier des transductions d'image finie dont le rôle dans cette thèse est important est constitué par les transductions rationnelles décroissantes (la longueur d'un mot est supérieure ou égale aux longueurs des images de ce mot). De façon analogue à ce qui a été fait précédemment on définit $C^d(L)$, la plus petite famille contenant L et fermée par transduction rationnelle décroissante. Les transductions rationnelles bifidèles (à la fois fidèles et d'image finie), les seules à être utilisables sur les ensembles de mots infinis [29], vont, quant à elles, nous conduire à l'étude des langages bifidèles c'est à dire des langages vérifiant l'égalité de C(L) et de $C^{bf}(L)$. La plus petite famille contenant L et fermée par transduction rationnelle bifidèle. Nous savons, depuis [23] que tout générateur de Alg est bifidèle, de même pour tout générateur de Ocl (les langages algébriques à un compteur). Par contre cette propriété n'est pas vérifiée pour Lin [23].

Le premier chapitre est destiné à fixer la terminologie utilisée dans les chapitres suivants, y sont rassemblés tous les résultats connus que nous utiliserons tout au long de cette thèse.

Comment une propriété sur une transduction rationnelle est-elle répercutée dans sa caractérisation en terme de bimorphisme et réciproquement ? De telles recherche ont déjà été réalisées par Boasson et Nivat [8] pour les transductions rationnelles fidèles ou d'image finie, par Eilenberg [10] pour les transductions rationnelles conservant les longueurs. Le chapitre II s'inscrit dans cette problématique et permet d'apporter une caractérisation des transductions rationnelles décroissantes à savoir :

τ est une transduction rationnelle décroissante si et seulement si elle peut être mise sous la forme $\tau = g \circ (\cap R) \circ h^{-1}$ où R est un langage rationnel défini sur un alphabet Z, h un homomorphisme strictement alphabétique de Z^* dans X^* et g un homomorphisme alphabétique de Z^* dans Y^*.

Notons que cette caractérisation nous permet de retrouver immédiatement, comme cas particulier, le résultat d'Eilenberg cité plus haut et aussi par symétrie, de caractériser les transductions rationnelles croissantes. Nous terminons ce chapitre en donnant quelques lemmes de factorisation de transduction rationnelle.

Dans le chapitre III nous montrons quelques propriétés des langages d'image finie et bifidèles. En particulier si L est un langage d'image finie, nous montrons qu'il domine, de façon décroissante tout élément de C(L) (c'est à dire $C^{if}(L) = C(L) \Rightarrow C^{d}(L) = C^{if}(L) = C(L)$). Cette propriété, ajoutée au fait que les transductions rationnelles décroissantes sont d'utilisation agréable, permet d'envisager des simplifications de démonstration chaque fois qu'interviendront des langages d'image finie. En particulier pour Lin (la famille des langages linéaires), pour Ocl (la famille des langages à un compteur) ou pour Alg qui sont tous trois des cônes principaux dont tous les générateurs sont des langages d'image finie.
Il était démontré dans [33] que tout langage défini sur un alphabet d'une seule lettre était effaçable. Par symétrie, on peut se demander s'il est également d'image finie. Une réponse négative est apportée à cette question et nous démontrons, au contraire, que, parmi ces langages, seuls les langages rationnels sont d'image finie. Nous donnons dans cette partie un exemple de langage vérifiant $C^{d}(L) = C^{if}(L) \subsetneq C(L)$.

Le langage $L \uparrow c^*$ (où $\uparrow$ est la substitution syntaxique définie par S. Greibach) formé des mots de $L \subseteq X^*$ dans lesquels on a intercalé un nombre quelconque de c ($c \notin X$) intervient dans une caractérisation des langages d'image finie. Ceci nous amène dans le chapitre IV à étudier les substitutions dans des cônes rationnels décroissants. Nous donnons les relations qui lient le cône rationnel (fidèle) engendré par un langage L quelconque et le cône rationnel décroissant (fidèle) engendré par L ce qui nous permet de retrouver très facilement le résultat classique [17] :

$\forall L \subseteq X^*$, $\forall L' \subseteq X'^*$ avec $X \cap X' = \emptyset$ alors

$C(L) \square C(L') = C(L \uparrow (L' c)^*)$.

Un langage L est complet par substitution si et seulement si pour tout langage L' défini sur un alphabet disjoint de celui de L $C(L) \square C(L') = C(L \uparrow L')$. Un des résultats principaux de ce chapitre est de montrer qu'il y a identité entre les langages complets par substitution et les langages d'image finie, ce qui permet d'énoncer :

$C(L) \square C(L') = C(L \uparrow L') \iff$ L est d'image finie ou L' engendre une FAL.

Pour arriver à ce résultat nous utilisons une version améliorée du lemme de substitution de S. Greibach due à M. Latteux.

La notion de langage très fortement effaçable, que nous introduisons dans ce chapitre va nous permettre de définir de nouvelles classes de langages (algébriques) non effaçables.

Nous abordons, dans le chapitre V, une étude systématique des relations entre les différents "sous-cônes particuliers" d'un cône rationnel principal. Nous montrons que la notion de langage très fortement effaçable et la notion de langage d'image finie permettent d'affiner les différences entre les langages définis sur un alphabet d'une seule lettre ce que ne permettait pas la notion de langage effaçable.

CHAPITRE I

PRELIMINAIRES

A : NOTATIONS

B : AUTOMATES, TRANSDUCTEURS ET APPLICATIONS (SOUS-) SÉQUENTIELLES

C : TRANSDUCTIONS RATIONNELLES

D : FAMILLES DE LANGAGES

E : SUBSTITUTIONS

A : NOTATIONS

Un **alphabet** X est un ensemble fini dont les éléments sont appelés **lettres**. Les éléments de X^*, le monoïde libre engendré par X, sont des **mots**. Toute partie (finie) de X^* est un **langage** (fini). Pour tout mot w de X^* on notera $|w|$ le nombre de lettres du mot w (encore appelé longueur de w) et $|w|_x$ le nombre d'occurrences de la lettre x dans le mot w :

$$|w| = \sum_{x \in X} |w|_x$$

naturellement le **mot vide**, noté ε, est le seul mot de longueur nulle.

Si E_1 et E_2 sont deux sous-ensembles de E, nous noterons $E_2 \backslash E_1$ l'ensemble des éléments de E_2 qui ne sont pas dans E_1 :

$$E_2 \backslash E_1 = \{e \in E \;/\; e \in E_2 \text{ et } e \notin E_1\}$$

Soient L_1 et L_2 deux langages inclus dans X^* :

l'union $L_1 \cup L_2$ est l'ensemble des mots de X^* qui appartiennent à L_1 ou à L_2

$$L_1 \cup L_2 = \{w \in X^* \;/\; w \in L_1 \text{ ou } w \in L_2\}$$

le produit $L_1 \, L_2$ est le langage de X^* constitué des mots de L_1 concatenés (à droite) à un mot de L_2

$$L_1 \, L_2 = \{w_1 \, w_2 \in X^* \;/\; w_1 \in L_1 \text{ et } w_2 \in L_2\}$$

l'étoile L^* du langage L est :

$$L^* = \bigcup_{n \geq 0} L^n \text{ avec } L^0 = \{\varepsilon\} \text{ et } L^{n+1} = L^n \, L = L \, L^n$$

A partir de ces trois opérations sur les langages on définit une famille de langages contenus dans X^* dont le rôle est fondamental en théorie des langages :

<u>*Définition*</u> *: La famille des langages rationnels de* X^**, notée* <u>$Rat(X^*)$</u> *est la plus petite famille de langages de* X^* *qui contienne les langages finis et qui soit close par union produit et étoile.*

Soit L un langage défini sur un alphabet X, F(L), FG(L), FD(L) sont respectivement l'ensemble des facteurs, des facteurs gauches, des facteurs droits de mots de L :

$F(L) = \{w \in X^* \ / \ \exists\ x, y \in X^* \text{ et } x\,w\,y \in L\}$

$FG(L) = \{w \in X^* \ / \ \exists\ y \in X^* \text{ et } w\,y \in L\}$

$FD(L) = \{w \in X^* \ / \ \exists\ y \in X^* \text{ et } y\,w \in L\}$

<u>*Définition*</u> *: Soient* L *un langage contenu dans* X^**,* w *un mot de* L*. On dira que* u *est facteur itérant de* w *dans* L *si et seulement si :*

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

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

Cette définition nous permet d'énoncer un résultat de base :

<u>*Lemme de l'étoile*</u> [OGDEN] *: Si* R *est un langage rationnel de* Z^**, il existe un entier* n *tel que tout mot* w*, de longueur supérieur à* n*, admette un facteur itérant dans* R.

Le **quotient à droite** (resp. **à gauche**) du langages L_1 par le langage L_2, noté $L_1 L_2^{-1}$ (resp. $L_2^{-1} L_1$) est l'ensemble :

$$L_1 L_2^{-1} = \{w \in X \ / \ \exists\ w_2 \in L_2 \text{ tel que } w\, w_2 \in L_1\}.$$

(resp. $L_2^{-1} L_1 = \{w \in X \ / \ \exists\ w_2 \in L_2 \text{ tel que } w_2\, w \in L_1\}$).

La notation L^+ sera quelque fois utilisée dans ce travail, elle désigne $\bigcup_{n\geq 1} L^n = LL^* = L^* L$.

Soient X^* et Y^* deux monoïdes libres, un **homomorphisme** h de X^* dans Y^* est une application vérifiant :

$$h(w_1\, w_2) = h(w_1)\, h(w_2) \text{ pour tous } w_1, w_2 \in X^*$$

et donc $h(\varepsilon) = \varepsilon$.

Un homomorphisme h de X^* dans Y^* sera :

.**stricte** si et seulement si $h(X) \subset Y^+$

.**alphabétique** si et seulement si $h(X) \subset Y \cup \{\varepsilon\}$

.**strictement alphabétique** si et seulement si $h(X) \subset Y$

.**k-limité** sur L avec $L \subset Y^*$ si et seulement si $L \cap X^* X_o^k X^* = \emptyset$ où $X_o = \{x \in X \ / \ h(x) = \varepsilon\}$

.**limité** sur L si et seulement s'il existe k tel que h soit k-limité sur L.

Soient X et Y deux alphabets disjoints la **projection** Π_X (resp. Π_Y) de $(X \cup Y)^*$ dans X^* (resp. Y) est l'homomorphisme alphabétique défini par $\Pi_X(z) = z$ si $z \in X$ et $\Pi_X(z) = \varepsilon$ si $z \in Y$ (resp. $\Pi_y(z) = \varepsilon$ si $z \in X$ et $\Pi_Y(z) = z$ si $z \in Y$).

Lemme : Si h *est un homomorphisme alphabétique alors :*

$$h^{-1}(w_1\ w_2) = h^{-1}(w_1)\ h^{-1}(w_2)$$

$$h^{-1}(w_1^+) = (h(w_1))^+$$

$\mathbb{N}_+$ dénote l'ensemble des entiers positifs, $\mathbb{N} = \mathbb{N}_+ \cup \{0\}$ et $\mathbb{N}^k = \{(x_1, \ldots, x_k) \ / \ \forall\ i\ [1, k],\ x_i \in \mathbb{N}\}$. Un ensemble $B \subseteq \mathbb{N}^k$ est un **ensemble linéaire** s'il existe $x_o \in \mathbb{N}^k$, un ensemble fini $P = \{x_1, \ldots, y_n\} \subseteq \mathbb{N}^k$ appelé ensemble de périodes tels que

$$B = \{x_o + \sum_{i=1}^{n} \lambda_i\ x_i \ / \ \lambda_i \in \mathbb{N}\ \forall\ i \in [1, n]\}$$

un ensemble **semi-linéaire** inclus dans $\mathbb{N}^k$ est une union finie d'ensembles linéaires inclus dans $\mathbb{N}^k$.

Soit $X = \{x_1, \ldots, x_n\}$ un ensemble ordonné, la **fonction de Parikh**, notée Ψ, est la fonction de X^* à valeur dans $\mathbb{N}^k$ définie par :

$$\forall\ w \in X^*;\ \Psi(w) = (|w|_{x_1}, \ldots, |w|_{x_k})$$

Rappelons enfin que l'image par la fonciton de Parikh d'un langage algébrique est un ensemble semi-linéaire.

B : AUTOMATES, TRANSDUCTEURS ET APPLICATIONS (SOUS-)SÉQUENTIELLES

Définition : Un **automate d'état fini déterministe** A *est un 5-uple* (X, Q, q_o, δ, F) *ou* X *est un alphabet d'entrée,* Q *un ensemble fini d'états,* q_o *un élément particulier de* Q *appelé état initial,* δ *une fonction partielle de* $Q \times X \to Q$ *que l'on étend à* $Q \times X^*$ *en posant, pour tout mot* f *de* X^* *et pour toute lettre* x *de* X :

$$\delta(q, \varepsilon) = q$$

$$\delta(q, fx) = \delta(\delta(q, f), x)$$

δ *est appelée fonction de transition*

F *est un sous-ensemble de* Q*, tout élément de* F *est un état terminal de l'automate.*

Remarque : Il nous arrivera d'utiliser le symbole $*$ pour dénoter la fonction de transition de l'automate et alors $q' = \delta(q, w)$ sera noté $q' = q * w$.

Définition : Un mot w *de* X^* *est* **reconnu** *par l'automate* $A = (X, Q, q_o, \delta, F)$ *si et seulement si* $\delta(q_o, w) \in F$.

Définition : Un langage L *de* X^* **est régulier** *si et seulement si il existe un automate* A *tel que l'ensemble des mots reconnus par* A *soit identique à* L*. On dira que* A *reconnaît* L.

Définition : $\mathrm{Reg}(X^*)$ *dénote la famille des langages réguliers définis sur* X^*.

Le théorème de Kleene, outil de base en théorie des langages, lie la famille $Reg(X^*)$ des langages réguliers de X^* et la famille $Rat(X^*)$ des langages rationnels de X^*.

Théorème de Kleene : Pour tout alphabet X,

$$Rat\ (X^*) = Reg\ (X^*)$$

Définition : L'automate minimal qui reconnait R *sera, parmi les automates qui reconnaissent* L *celui qui possède le moins d'états (il est unique à une numérotation des états près).*

La notion de transducteur séquentiel va généraliser la notion d'automate.

Définitions : Un transducteur séquentiel T *est un 6-uple* $(\Sigma, \Delta, Q, q_o, \delta, \lambda)$ *où* Σ *est un alphabet d'entrée,* Δ *un alphabet de sortie,* Q, q_o *et* δ *ont la même définition que précédemment,* λ *est une fonction partielle de* $Q \times \Sigma \rightarrow \Delta^*$ *que l'on étend à* $Q \times \Sigma^*$ *en posant, pour tout mot de* Σ^*, *pour toute lettre* x *de* Σ :

$$\lambda(q, \varepsilon) = \varepsilon\ ;\ \lambda(q, fx) = \lambda(q, f)\ \lambda(\delta(q, f), x)$$

λ *est appelée fonction de sortie (cf. Eilenberg [10], Berstel [6]).*

Une application γ *de* Σ^* *dans* Δ^* *est* séquentielle *s'il existe un transducteur qui la réalise c'est à dire si pour tout mot* w *de* Σ^* $\gamma(w)=\lambda(q_o,w)$.

Introduite en 1977 par M.P. Schützenberger [33], la notion de transducteur sous-séquentiel est largement utilisée dans ce travail.

<u>*Définitions*</u>*: Un transducteur* **sous-séquentiel** *est un couple* (T, s) *formé d'un transducteur séquentiel* T *défini comme ci-dessus et d'une fonction partielle* s *de* Q *dans* Δ^*.

Une application γ *est* **sous-séquentielle** *s'il existe un transducteur sous-séquentiel qui la réalise c'est à dire tel que pour tout mot de* Σ^*, $\gamma(w) = \lambda(q_0, w)\, s(\delta(q_0, w))$.

C : TRANSDUCTIONS RATIONNELLES

Soient X^* et Y^* deux monoïdes libres, $X^* \times Y^*$ est un monoïde dont le produit est défini par ; $(x, y)\,(x', y') = (xx', yy')$ pour tous couples (x, y), (x', y') de $X^* \times Y^*$ l'élément neutre $(\varepsilon, \varepsilon)$ sera noté ε Rat $(X^* \times Y^*)$ est la famille des langages rationnels du monoïde $X^* \times Y^*$.

<u>*Définition*</u> *: Une* transduction τ *de* X^* *dans* Y^* *est une application de* X^* *dans l'ensemble* $P(Y^*)$ *des parties de* Y^*.

Cette définition peut être étendue à $P(X^*)$ par

$$\forall\, L \subseteq X^* \quad \tau(L) = \bigcup_{w \in L} \tau(w)$$

le graphe de τ est l'ensemble ;

$$\hat{\tau} = \{(w, w') \in X^* \times Y^* \;/\; w' \in \tau(w)\}$$

nous utiliserons également les notations

$$\tau(w) = \{w' \in Y^* \;/\; (w, w') \in \hat{\tau}\}$$

$$\tau^{-1}(w) = \{w' \in X^* \;/\; (w', w) \in \hat{\tau}\}$$

Le concept de tranductions rationnelles, due à Elgot et Mezei [11], est apparu en 1965.

<u>*Définition*</u> *: Une transduction* τ *de* X^* *dans* Y^* *est* rationnelle *si et seulement si* $\hat{\tau}$ *est une partie rationnelle de* Rat$(X^* \times Y^*)$

M. Nivat [28], en caractérisant les transductions rationnelles en terme de bimorphisme, a fait de cette notion un outil de base :

<u>*Théorème de Nivat*</u> [28] *: Une transduction* τ *de* X^* *dans* Y^* *est rationnelle si et seulement s'il existe un alphabet* Z, *un langage rationnel* R *défini sur* Z, *deux homomorphismes (alphabétiques)* h *de* Z^* *dans* X^*, g de Z^* *dans* Y^* *tels que*

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

Si $L \subseteq X^*$, $\tau(L) = \bigcup_{w \in L} \tau(w) = \bigcup_{w \in L} \{w' \in Y^* \; / \; (w, w') \in \hat{\tau}\}$

Or le couple (w, w') appartient au graphe $\hat{\tau}$ de τ si et seulement s'il existe un mot r de R vérifiant à la fois $w = h(r)$ et $g(r) = w'$ donc $r \in h^{-1}(w) \cap R$ et $w' \in g(h^{-1}(w) \cap R)$

$$\tau(L) = \bigcup_{w \in L} g(h^{-1}(w) \cap R) = g(h^{-1}(L) \cap R)$$

Ce qui permet d'écrire, en utilisant les notations d'Eilenberg [10], $\tau = g \circ (\cap R) \circ h^{-1}$

$$X^* \xrightarrow{h^{-1}} Z^* \xrightarrow{\cap R} Z^* \xrightarrow{g} Y^*$$

A partir de maintenant, nous supposerons, ce qui n'est pas une restriction, que toute lettre de Z ne peut pas avoir des images par h et g vides simultanément.

Une variante de ce théorème, due à Eilenberg [10], sera utilisée dans cette thèse :

<u>*Théorème*</u> [10] : *Soient* X *et* Y *deux alphabets disjoints. Une transduction* τ *de* Y^* *dans* Y^* *est rationnelle si et seulement s'il existe un langage rationnel* R *contenu dans* $(X \cup Y)^*$ *vérifiant :*

$$\hat{\tau} = \{(\Pi_X(w), \Pi_Y(w)) \ / \ w \in R\}$$

ou encore $\tau = \pi_Y \circ (\cap R) \circ \Pi_X^{-1}$

On appelera cette forme, la **forme normale** de τ.

Une tranduction rationnelle τ de X^* dans Y^* sera dite :

. **fonctionnelle** si $\forall w \in X^*$ $\tau(w)$ a au plus un élément

. **d'image finie** si $\forall w \in X^*$ $\tau(w)$ est un ensemble fini

. **fidèle** si $\forall w \in Y$ $\tau^{-1}(w)$ est un ensemble fini

. **bifidèle** si τ est à la fois fidèle et d'image finie

. **décroissante** si $\forall w \in X^*$, $\forall w' \in \tau(w)$ $|w'| \leq |w|$

. **croissante** si $\forall w \in X^*$, $\forall w' \in \tau(w)$ $|w'| \geq |w|$

. **préservant les longueurs** si $\forall w \in X^*$, $\forall w' \in \tau(w)$ $|w'| = |w|$

. **décroissante fidèle** si elle est à la fois décroissante et fidèle.

<u>*Théorème d'Elgot et Mezei*</u> [11] *: La composée de deux transductions rationnelles (resp. fidèle, d'image finie, décroissante, bifidèle, décroissante fidèle, préservant les longueurs, croissante) est une transduction rationnelle (resp. fidèle, d'image finie, décroissante bifidèle, préservant les longueurs, croissante).*

Comme pour les transductions rationnelles quelconques, on peut caractériser les transductions rationnelles fidèles, d'image finie et bifidèles :

<u>*Théorème de Boasson-Nivat*</u> [*8*] *: Une transduction rationnelle* τ *de* X^* *dans* Y^* *est rationnelle fidèle (resp. d'image finie, bifidèle) si et seulement s'il existe un alphabet* Z*, deux homomorphismes alphabétiques* h *et* g *de* Z^* *dans* X^* *et* Y^* *respectivement et un langage rationnel* $R \subseteq Z^*$ *tels que*

$$\forall\, w \in X^*,\ \tau(w) = g(h^{-1}(w) \cap R)$$

et g *est limité sur* R *(resp.* h *est limité sur* R*,* h *et* g *sont tous les deux limités sur* R*).*

D : FAMILLES DE LANGAGES

Une **famille de langages** $(\Sigma, \mathcal{L})$ est un couple constitué d'un ensemble infini dénombrable Σ et d'une famille non vide de parties $\mathcal{L}$ de Σ^* vérifiant :

$\forall$ L $\in \mathcal{L}$, il existe un ensemble fini X_L de Σ tel que $L \subseteq X_L^*$.

Généralement l'ensemble Σ est sous entendu et le couple $(\Sigma, \mathcal{L})$ est noté $\mathcal{L}$.

Parmi les familles de langages, deux types de familles vont particulièrement nous intéresser. Il s'agit des cônes rationnels d'une part et des familles agréables de langages d'autre part.

Définition : *$\mathcal{L}$, une famille de langages, est un* **cône rationnel** *(resp. cône rationnel* **fidèle**, *cône rationnel* **d'image finie**, *cône rationnel* **bifidèle**, *cône rationnel* **décroissant**, *cône rationnel* **décroissant fidèle**) *si et seulement si elle est fermée par transduction rationnelle (resp. transduction rationnelle fidèle, transduction rationnelle d'image finie, transduction rationnelle bifidèle, transduction rationnelle décroissante, transduction décroissante fidèle).*

Remarque : Un cône rationnel fidèle (resp. bifidèle) n'est généralement pas un cône rationnel (resp. rationnel d'image finie). Par contre tout cône rationnel resp. rationnel d'image finie) est un cône rationnel fidèle (resp. bifidèle).

On notera $C(L)$ (resp. $C^{f}(L)$, $C^{if}(L)$, $C^{bf}(L)$, $C^{d}(L)$, $C^{df}(L)$) le plus petit cône rationnel (resp. rationnel fidèle, rationnel d'image finie, rationnel bifidèle, rationnel décroissant, rationnel décroissant fidèle) qui contient L. En fait, compte tenu du théorème de composition des transductions rationnelles, $C(L) = \{L' \ / \ \exists\ L \in L$, il existe une transduction rationnelle telle que $L' = \tau(L)\}$.

Un cône rationnel C sera dit **principal** s'il existe un langage L appelé générateur, tel que

$$C(\{L\}) = C$$

Par abus de notation nous écrirons $C(L)$ à la place de $C(\{L\})$.

De façon naturelle, la notion de principalité est étendue aux cônes rationnels particuliers.

En conséquence ;

$C(L) = \{L' \ / \ \exists\ \tau$ transduction rationnelle telle que $L' = \tau(L)\}$

$C^{f}(L) = \{L' \ / \ \exists\ \tau$ transduction rationnelle fidèle telle que $L' = \tau(L)\}$

$C^{if}(L) = \{L' \ / \ \exists\ \tau$ transduction rationnelle d'image finie telle que $L' = \tau(L)\}$

$C^{d}(L) = \{L' \ / \ \exists\ \tau$ transduction rationnelle décroissante telle que $L' = \tau(L)\}$

$C^{bf}(L) = \{L' \ / \ \exists\ \tau$ transduction rationnelle bifidèle telle que $L' = \tau(L)\}$

$C^{df}(L) = \{L' \ / \ \exists\ \tau$ transduction rationnelle décroissante fidèle telle que $L' = \tau(L)\}$

Définition : L *est effaçable si et seulement si* $C(L) = C^{f}(L)$.

De façon symétrique nous pouvons maintenant définir les langages d'image finie et les langages bifidèles, langages qui sont au centre de l'étude ici entreprise :

<u>Définition</u> : *Soit* L *un langage :*

L *est d'image finie si et seulement si* $C(L) = C^{if}(L)$

L *est bifidèle si et seulement si* $C(L) = C^{bf}(L)$

Un cône rationnel n'est en général pas fermé pour les opérations union, produit et étoile c'est la raison pour laquelle on a été amené à définir les familles agréables de langages ou FAL.

<u>Définition</u> : *Une famille de langages* $\mathcal{L}$ *est une* FAL *(fidèle) si et seulement si* $\mathcal{L}$ *est un cône rationnel (fidèle) fermé par union, produit et étoile.*

<u>Lemme</u> : *Tout cône principal est fermé par union.*

Parmi les familles agréables de langages, objets de nombreuses études, nous citerons la famille Rat des langages rationnels déjà mentionnée dans ce travail, la famille Alg des langages algébriques, la famille des langages bornés ou encore la famille des langages commutatifs. Nous allons rappeler des résultats concernant les familles Rat et Alg.

<u>Proposition</u> : *Pour tout langage* L *non rationnel* Rat $\subsetneq C^{f}(L) \subseteq C(L)$

<u>Proposition</u> : *Tout langage rationnel est générateur (fidèle) de* Rat.

Les moyens de définir la famille des langages Algébriques Alg encore appelée famille de langages à contexte libre ne manquent pas tant les études entreprises sur cette famille sont nombreuses. Nous choisissons le moyen cadrant le mieux avec ce travail :

Les langages de Dyck restreints D'^{*}_{n},"langages de parenthèses" peuvent être definis de la façon suivante
soit $X_n = \{x_1, x_2, \dots, x_n\}$, $\bar{X}_n = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$.

Soit δ_n la congruence de $(X \cup \bar{X})^*$ engendrée par

$\forall\ i \subset [1, n]\quad x_i\ \bar{x}_i \equiv \varepsilon$

<u>*Définition*</u> *: Le langage de Dyck* D'^{*}_{n} *restreint est la classe de* ε *dans la congruence* δ_n.

Nous sommes alors en mesure de définir Alg ;

<u>*Définition*</u> *:* Alg *est le cône rationnel engendré par* D'^{*}_{2}.

<u>Remarque</u> = D'^{*}_{2} n'est pas le seul générateur de Alg.

En fait ce résultat a été affiné par L. Boasson - M. Nivat et par S. Ginsburg - J. Goldstine - S. Greibach.

<u>*Théorème [8] et [14]*</u> *; Tout générateur de* Alg *est un générateur fidèle.*

Résultat encore amélioré par M. Latteux.

<u>Théorème</u> [23] : *Tout générateur de* Alg *est un générateur bifidèle.*

D'où la conclusion, pour tout langage L tel que C(L) = Alg on a les relations

$$\text{Alg} = C(L) = C^{f}(L) = C^{if}(L) = C^{bf}(L)$$

Deux sous familles de Alg ont une place tout à fait particulières du fait des propriétés intéressantes qu'elles possèdent. Elles ont fait l'objet de nombreuses études. Il s'agit de la famille des langages linéaires Lin d'une part et de la famille des langages à un compteur Ocl d'autre part.

Une façon agréable de définir Lin est d'utiliser le langage Sym = $\{w \bar{w}^R \,/\, w \in \{a, b\}^*\}$ où $\bar{w}^R$ est l'image miroir de w dans laquelle on a "barré" toutes les lettres de w alors Lin = C(Sym).

<u>Proposition</u> [23] Lin = C(Sym) = C^{if}(Sym) = C^{f}(Sym)

Mais, contrairement à Alg nous savons que tout générateur de Lin n'est pas bifidèle. En effet :

<u>Proposition</u> [23] C^{bf}(Sym) $\subsetneq$ Lin.

La famille des langages à un compteur notée Ocl, est quant à elle le cône rationnel engendré par le langage de Dyck restreint sur une lettre D'^{*}_{1}. Il a été montré que :

<u>Proposition</u> [24] : *Tout langage générateur de* Ocl *est générateur bifidèle de* Ocl.

E : SUBSTITUTION

La transduction rationnelle est un moyen de construire des familles de langages, la substitution en est un autre, ce qui en fait un outil privilégié.

Définition : Soit s, *un homomorphisme de* X^* *dans* $P(Y^*)$, *soit* $\mathcal{L}$ *une famille de langages. Nous dirons que* s *est une* $\mathcal{L}$*-substitution si et seulement si* s(x) $\in \mathcal{L}$ *pour tout* x *de* X, s *est une* $\mathcal{L}$*-substitution propre si c'est une* $\mathcal{L}$*-substitution et si* s(x) *ne contient pas le mot vide*, pour tout x *de* X.

s étant un homomorphisme, on est en droit de parler de substitution de mots et plus généralement de parler de substitution d'un langage $L \subseteq X^*$ en posant :

$$s(L) = \bigcup_{w \in L} s(w).$$

Il semble naturel d'étendre les substitutions aux familles de langages ;

Définition : $\mathcal{L} \,\square\, \mathcal{L}' = \{s(L) \,/\, L \in \mathcal{L}$ *et* s *est une* $\mathcal{L}'$*-substitution propre*$\}$

Quoique légèrement différente de la définition classique cette définition lui devient équivalente dès que $\mathcal{L}$ est fermée par transductions décroissantes (voir pages suivantes). Par contre si $\mathcal{L}$ n'est pas clos par homomorphisme effaçant cette définition est plus adaptée et permet d'affiner les résultats.

Possédant deux opérateurs sur les familles de langages l'un unaire (la transduction rationnelle), l'autre binaire (la substitution), il est intéressant de faire le "lien" entre ces deux outils. L'opérateur syntaxique défini par S. Greibach, que nous noterons $\uparrow$, va faire le "pont" entre les deux notions puisqu'on va pouvoir factoriser, suivant certaines hypothèses, toute substitution en la composition d'une transduction rationnelle et de l'opérateur syntaxique.

<u>*Définition de l'opérateur syntaxique* $\uparrow$</u> *: Si* L *et* L' *sont deux langages,* L $\uparrow$ L' *sera le langage obtenu en insérant entre chaque lettre d'un mot de* L *un mot quelconque de* L' *;*

$L \uparrow L' = \{x_1\, w_1\, x_2\, w_2 \ldots x_n\, w_n \,/\, x_1\, x_2 \ldots x_n \in L \,;\, w_i \in L'$
et x_i *est une lettre,* $\forall\, i \in [1, n]\}$

Le résultat suivant est fondamental, Il dit que la substitution d'un cône rationnel principal dans un autre cône rationnel principal est encore un cône rationnel principal ;

<u>*Théorème*</u> [17] *: Pour tout couple de langages* (L_1, L_2)*, définis sur des alphabets disjoints* $C(L_1) \square C(L_2) = C(L_1 \uparrow (L_2\, c)^*$

Les lemmes suivants sont connus sous le nom de lemme de substitution. La première version [17], due à S. Greibach est très utilisée et constitue un résultat de base :

<u>*Lemme de substitution (Greibach)*</u> *: Soient* L_1 *et* L_2 *des langages définis sur des alphabets disjoints, soient* C_1 *et* C_2 *deux cônes rationnels fermés par union si* $L_1 \uparrow L_2 \in C_1 \square C_2$ *alors* $L_1 \in C_1$ *ou* $L_2 \in C_2$.

J. Beauquier a montré [2] que l'hypothèse de fermeture par union des cônes C_1 et C_2 n'était pas nécessaire. Par contre il ne peut être étendu aux cônes fidèles [4]. C'est essentiellement la version la plus puissante due à M. Latteux [25] que nous utiliserons dans cette thèse.

<u>Lemme de substitution (Latteux)</u> : *Soient* $\mathcal{L}_1$ *un cône rationnel bifidèle,* $\mathcal{L}_2$ *un cône rationnel et les langages* $L_1 \subseteq X_1^*$, $L_2 \subseteq X_2^*$ *avec* $X_1 \cap X_2 = \emptyset$.
Alors : $L_1 \uparrow L_2 \in \mathcal{L}_1 \square \mathcal{L}_2$ *implique* $L_1 \in \mathcal{L}_1$ *ou* $(L_2\, C)^* \in \mathcal{L}_2$.

<u>Définition</u> : *Une famille* $\mathcal{L}$ *est fermée par* $\mathcal{L}'$*-substitution si et seulement si* $\mathcal{L} \square \mathcal{L}' = \mathcal{L}$.

<u>Définition</u> : *Une famille* $\mathcal{L}$ *est fermée par substitution si et seulement si* $\mathcal{L} \square \mathcal{L} = \mathcal{L}$.

Parmi les familles déjà définies Rat et Alg sont fermées par substitution alors que Lin et Ocl ne le sont pas. La famille des langages quasi-rationnels notée QR, clôture par substitution de la famille des langages linéaires $\mathcal{L}in$ est contenue strictement dans Alg.

On notera $\mathcal{F}(L)$ la famille agréable de langages engendrée par L c'est à dire la fermeture pour les opérations union produit et étoile de C(L) ce qui s'écrit, en terme de substitution :

$\mathcal{F}(L) = \text{Rat} \square C(L)$

Nous nous servirons, dans le chapitre IV du résultat :

<u>Lemme</u> : *Si* $\mathcal{L}_1$ *est une famille de langages close par transduction rationnelle décroissante alors* $\mathcal{L}_1 \square \mathcal{L}_2$ *est clos par homomorphisme effaçant et est donc*

égal à $\{s(L_1) \ / \ L_1 \in \mathcal{L}_1$, s *est une* $\mathcal{L}_2$*-substitution*$\}$.

<u>Démonstration</u> : Tout élément L de $\mathcal{L}_1 \square \mathcal{L}_2$ est de la forme $L = s(L_1)$ où $L_1 \in \mathcal{L}_1$ et s est une $\mathcal{L}_2$-substitution propre. Appelons X et X_1 les alphabets respectifs sur lesquels sont définis L et L_1. Soit h un homomorphisme défini sur X^*. Si H_o désigne l'ensemble des lettres de X effacées par h nous pouvons construire :

$$s_- : X_1^* \to X_1^*$$

$$\text{où } \forall y \in X_1 \text{ si } s(y) \cap H_o^* \neq \emptyset \ s_-(y) = \{\varepsilon, y\}$$

$$\text{si } s(y) \cap H_o^* = \emptyset \ s_-(y) = y$$

$$s_+ : X_1^* \to X^* \text{ où}$$

$$s_+(y) = h \circ s(y) \setminus \{\varepsilon\}$$

s_- est une substitution rationnelle effaçante c'est donc une transduction rationnelle décroissante.

s_+ est une substitution propre.

Montrons que $h \circ s(L_1) = s_+ \circ s_- (L_1)$;

Pour toute lettre y de X_1 deux cas peuvent se présenter :

<u>1er cas</u> : $s(y) \cap H_o^* \neq \emptyset$

$h \circ s(y) = h((s(y) \setminus H_o^*) \cup s(y) \cap H_o^*) = (h \circ s(y) \setminus \{\varepsilon\}) \cup \{\varepsilon\} = s_+(\{y, \varepsilon\})$

$= s_+ \circ s_-(y)$

<u>2ème cas</u> : $s(y) \cap H_o^* = \emptyset$

$h \circ s(y) = s_+(y) = s_+ \circ s_-(y)$

$L'_1 = s_-(L_1) \in \mathcal{L}_1$ puisque $\mathcal{L}_1$ est fermée par transduction rationnelle décroissante et $L = s_+ \circ s_- (L_1) \in \mathcal{L}_1 \square \mathcal{L}_2$

$\square$

Rappelons, pour terminer, deux théorèmes que nous utiliserons dans le chapitre III :

<u>*Théorème*</u> [*13*] : *Soit* L *un langage contenu dans* a^* *et soit* t *un entier strictement positif. Alors* $L((a^t)^*)^{-1} = \{w \in a^* \; / \; \exists \; i \in N$ *tel que* $q \; a^{it} \in L\}$ *est un langage rationnel.*

<u>*Théorème*</u> [*21*] : *Soit* L *un langage contenu dans* $a^* \, b^*$. *Pour tout couple d'entiers* i *et* j *strictement positifs le langage* $L' = ((a^i)^*)^{-1} \, (L((b^j)^*)^{-1}) = \{w \in a^* \, b^* \; / \; \exists \; s, t \in \mathbb{N}$ *tels que* $a^{is} \, w \, b^{jt} \in L\}$ *est un langage rationnel.*

CHAPITRE II

TRANSDUCTIONS RATIONNELLES DECROISSANTES

A : DÉCIDABILITÉ

B : CARACTÉRISATIONS

C : DÉCOMPOSITIONS

Comme nous l'avons déjà vu, une transduction rationnelle est décroissante si et seulement si pour tout mot, la longueur d'un élément quelconque de son image par cette transduction est inférieure ou égale à la longueur de ce dernier.

Le but de ce chapitre est multiple. Après avoir montré qu'il est décidable de savoir si une transduction rationnelle est décroissante, nous abordons le résultat essentiel de ce chapitre, à savoir une caractérisation des transductions rationnelles décroissantes. Des lemmes de factorisation de transductions rationnelles quelconques en transductions rationnelles particulières termineront cette partie, factorisation qui se révèlent être des outils pratiques dans la suite de ce travail.

A : DÉCIDABILITÉ

Proposition II - 1 : *Il est décidable de savoir si une transduction τ de X^* dans Y^* est décroissante.*

Démonstration : Nous supposerons, ce qui ne nuit en rien à la généralité de la démonstration, que les alphabets X et Y sont disjoints. Soit $\pi_Y \circ (\cap R) \circ \pi_X^{-1}$ la forme normale de τ. On définit alors l'homomorphisme t de $(X \cup Y)^*$ dans $\{a, b\}^*$ par $t(x) = a$ pour tout x de X et $t(y) = b$ pour tout y de Y. Naturellement τ est décroissante si et seulement si pour tout mot w de R, t(w) ne contient pas plus de b que de a.
Soit D le langage $D_1^* \sqcup b^+$, c'est à dire le langage défini sur $\{a, b\}$ tel que tout mot contient plus de b que de a. La transduction τ est décroissante si et seulement si $t(R) \cap D$ est vide. D est algébrique, t(R) est rationnel, $t(R) \cap D$ est donc algébrique. Comme il est décidable de savoir si un langage algébrique est vide ou pas, nous obtenons bien le résultat.

□

Il est possible d'avoir un résultat plus précis :

Proposition II - 2 : *Il est décidable de savoir si une transduction rationnelle τ de X^* dans Y^* est décroissante sur un langage algébrique quelconque L de X^*.*

Démonstration : Nous supposons que $X \cap Y = \emptyset$ et nous appelons $\pi_Y \circ (\cap R) \circ \pi_X^{-1}$ la forme normale de τ. Soit t l'homomorphisme de $(X \cup Y)^*$ dans $\{a, b\}^*$ tel que $t(x) = a$ pour tout x de X et $t(y) = b$ pour tout y de Y.

τ est décroissante sur L si pour tout mot w de $L' = \mathcal{R} \cap \pi_X^{-1}(L)$ $|t(w)|_a \geq |t(w)|_b$, ce qui se traduit par $t(w) \in D_1^* \sqcup a^*$, ce qui signifie que $t(L') \subseteq D_1^* \sqcup a^*$. τ est décroissante sur L si et seulement si $L'' = t(L') \cap D_1^* \sqcup b^+$ est vide ou, ce qui est équivalent, l'image de L'' par la fonction de Parikh est vide.

$\Psi(L'') = \Psi(t(L') \cap D_1^* \sqcup b^+)$ mais, puisque $D_1^* \sqcup b^+$ est un langage commutatif, $\Psi(L'') = \Psi(t(L') \cap \Psi^{-1} \circ \Psi\,(D_1^* \sqcup b^+)) = \Psi(t(L')) \cap \Psi(D_1^* \sqcup b^+)$. Or, L est algébrique, $t(L')$ est algébrique et $\Psi(D_1^* \sqcup b^+)$ est un ensemble semi-linéaire et comme on peut effectivement construire l'intersection de deux ensembles linéaires, on peut décider si $\Psi(L'')$ est vide.

$\square$

B : CARACTÉRISATIONS

Le résultat principal de ce chapitre est le suivant :

Théorème II - 3 : Une transduction rationnelle τ *de* X^* *dans* Y^* *est décroissante si et seulement s'il existe un langage rationnel* $R \subseteq Z^*$*, deux homomorphismes alphabétiques* h *et* g*,* h *de* Z^* *dans* X^**,* g *de* Z^* *dans* Y^**,* h *strictement alphabétique tels que* $\tau = g \circ (\cap R) \circ h^{-1}$.

Notons qu'une autre formulation en terme de monoïde est possible. Si M représente le monoïde $\{(u, v) \in X^* \times Y^* / |u| \geq |v|\}$ alors :

Théorème II - 4 : $\mathrm{Rat}(X^* \times Y^*) \cap M = \mathrm{Rat}(M)$

Du thèorème II-3, on déduit facilement le résultat suivant, dû à Eilenberg :

Théorème II - 5 [*10*] *: Une transduction rationnelle* τ *de* X^* *dans* Y^* *conserve les longueurs si et seulement s'il existe un alphabet* Z*, un langage rationnel* $R \subseteq Z^*$*, deux homomorphismes* h *et* g *strictement alphabétiques respectivement de* Z *dans* X^* *et* Y^* *tels que* $\tau = g \circ (\cap R) \circ h^{-1}$.

En effet, si on suppose que τ conserve les longueurs, τ est décroissante, son graphe peut donc se mettre sous la forme $\hat{\tau} = \{(h(w), g(w)) / w \in R\}$ où h est strictement alphabétique et g alphabétique. Mais alors, pour tout mot w de R, $|w| = |h(w)| = |g(w)|$ encore, puisque g est alphabétique, pour toute lettre z de Z, $|g(z)| = |h(z)| = 1$.

□

<u>Preuve du théorème II - 3</u>

Nous supposerons, sans nuire à la généralité de la démonstration, que X et Y sont des alphabets disjoints. Afin de démontrer le théorème II-3, nous considérons la propriété (P) suivante, qui s'avère être essentielle :

<u>Définition II - 6</u> : Une partie R *de* $(X \cup Y)^*$ *vérifie (P) si et seulement s'il existe un entier* $k > 0$ *tel que, pour tout* w *de* $F(R)$ *on ait*
$|w|_Y \leq |w|_X + k$.

Notons que si $\tau = \Pi_Y \circ (\cap R) \circ \Pi_X^{-1}$ est une transduction rationnelle en forme normale alors τ est décroissante si et seulement si, pour tout mot w de R, $|w|_X \geq |w|_Y$. Nous pouvons alors énoncer le lemme de base de la démonstration :

<u>Lemme II- 7</u> : Si $R \subset (X \cup Y)^*$ *est un langage rationnel et* $|w|_X \geq |w|_Y$ *pour tout mot* w *de* R*, alors* R *vérifie la propriété (P).*

<u>Démonstration</u> : Soit k le nombre d'états de l'automate minimal qui reconnaît R et supposons, en raisonnant par l'absurde, que l'on puisse trouver un élément w de F(R) vérifiant l'inégalité $|w|_Y > |w|_X + k$. Il existe x et y, éléments de $(X \cup Y)^*$ tels que xwy appartienne à R. Le mot w peut se mettre sous la forme $w = w_1 \ldots w_k$ où, pour tout i de $[1, k]$, $|w_i|_Y > |w_i|_X + 1$ et comme $|w| > k$, il existe i et $j \in [1, k]$ tels que $x\, w_1 \ldots w_i\, (w_{i+1} \ldots w_j)^*$ $w_{j+1} \ldots w_k\, y \subseteq R$. Posons $l = |xwy|_X - |xwy|_Y$, par hypothèse $l \geq 1$. Mais le mot $w' = x\, w_1 \ldots w_i\, (w_{i+1} \ldots w_j)^{l+1}\, w_{j+1} \ldots w_k\, y \in R$ et vérifie $|w'|_X - |w'|_Y < 0$ ce qui est contraire à l'hypothèse.

$\square$

Nous déduisons immédiatement :

Corollaire II - 8 : Si $\tau = \Pi_Y \circ (\cap R) \circ \Pi_X^{-1}$ *est une transduction rationnelle décroissante en forme normale alors* R *vérifie (P).*

Soient $Z = X \times (Y \cup \{a\})$ où $a \notin X \cup Y$, α et β les deux morphismes alphabétiques de Z^* dans X^* et Y^* respectivement, définis pour tout couple (x, v) de Z par :

$$\alpha(x, v) = x \text{ et } \beta(x, v) = \begin{cases} y \text{ si } v = y \in Y \\ \varepsilon \text{ si } v = a \end{cases}$$

Notons que α est strictement alphabétique.

Lemme II - 9 : Soit $R \subset (X \cup Y)^*$ *un langage rationnel tel que pour tout mot* w *de* R, $|w|_X \geq |w|_Y$. *Il existe une application sous séquentielle* γ *de* $(X \cup Y)^*$ *dans* Z *telle que :*

$$\text{pour tout mot } w \text{ de } R, \quad \Pi_X(w) = \alpha \circ \gamma(w)$$
$$\Pi_Y(w) = \beta \circ \gamma(w).$$

Démonstration : Avant de donner la preuve formelle, donnons l'idée qui nous a conduit à cette preuve et dont l'essentiel est la construction d'un transducteur sous-séquentiel T. Un mot de R ayant plus de lettres dans X que de lettres dans Y, à toute lettre de X on associe une lettre de Y. Le transducteur aura pour mission de construire des paires constituées d'une lettre dans X et d'une lettre dans Y, plus, éventuellement, un excédent en lettres de X. Le transducteur séquentiel à construire lit le mot w de la gauche vers la droite. Afin de rendre son comportement plus clair, nous considérons, dans une première approche, qu'il travaille sur deux files, une X-file (ne contenant que des lettres de X) et une Y-file (ne contenant que des lettres de Y). Chaque lettre lue va être rangée dans la file correspondante et si les deux files sont non vides, les premiers éléments de chacune des deux files sont groupés par paires émises en sortie.

Pour qu'un transducteur rationnel puisse réaliser ce travail, il convient que la taille des files soit bornée par un nombre indépendant du mot en traitement. On pourra alors simuler les files à l'aide des états du transducteur.

Le langage rationnel R, d'après le lemme II-7, vérifie la propriété (P) ce qui nous assure que la taille maximale de la Y-file est k où k est l'entier du lemme II - 7. Par contre, nous n'avons aucune garantie quant à la taille de la X-file. Aussi décidons-nous que, dès que la taille de la X-file atteint k, le premier élément de la X-file soit sorti pour former une paire avec le symbole spéciale "a". La taille maximale de la X-file sera donc de k.

A la fin de la lecture de w, seule la X-file peut ne pas être vide. Elle sera alors vidée en regroupant ses éléments avec le symbole spécial "a".

Le transducteur T sera défini par :

$$T = (\Sigma, \Delta, Q, (\varepsilon, \varepsilon), \delta, \lambda)$$

où

- $\Sigma = X \cup Y$ est l'alphabet d'entrée,
- $\Delta = X \times (Y \cup \{a\})$ est l'alphabet de sortie,
- $Q = ((\bigcup_{i=o}^{k} X^i) \times \{\varepsilon\}) \cup (\{\varepsilon\} \times (\bigcup_{i=o}^{k} Y^i))$ est l'ensemble des états.

Les états de Q sont donc de la forme (u, ε) avec $u \in X^*$ et $|u| \leq k$ où (ε, v) avec $v \in Y^*$ et $|v| \leq k$.

- $(\varepsilon, \varepsilon)$ est l'état initial,
- δ et λ, applications partielles de $Q \times \Sigma^*$ dans Q et Δ, sont respectivement les fonctions de transition et de sortie définies par :

état	lettre lue	λ	δ	
(u, ε) où $\|u\| < k$	$x \in X$	ε	(ux, ε)	rangement dans les "files"
(ε, v) où $\|v\| < k$	$y \in Y$	ε	(ε, vy)	
(u, ε) $\|u\| = k$ $u = x'u'$ et $x' \in X$	$x \in X$	(x', a)	$(u'x, \varepsilon)$	sortie du premier de la file avec a, entrée du dernier lu
(u, ε) $u = x'u'$ et $x' \in X$	$y \in Y$	(x', y)	(u', ε)	ou "vide"
(ε, v) $v = y'v'$ et $y' \in Y$	$x \in X$	(x, y')	(ε, v')	

La fonction partielle s de Q dans Z est définie par :

$\forall\, u = u_1 \ldots u_n \in X^*$ avec pour tout $i \in [1, n]$, $u_i \in X$

$s(u, \varepsilon) = (u_1, a) \ldots (u_n, a)$,

l'application sous-séquentielle γ recherchée est réalisée par le transducteur sous-séquentiel (T, s).

Pour montrer les relations entre Π_X, α, γ et Π_Y, β, γ nous avons besoin du lemme suivant :

Lemme II - 10 : Pour tout mot w de $FG(R)$, l'état $(u, v) = \delta((\varepsilon, \varepsilon), w)$ vérifie :

i) $\Pi_X(w) = \alpha \circ \lambda((\varepsilon, \varepsilon), w)\, u,$

ii) $\Pi_Y(w) = \beta \circ \lambda((\varepsilon, \varepsilon), w)\, v,$

iii) $|v| \leq \sup \{|w'|_Y - |w'|_X \,/\, w' \in FD(w)\},$

iv) pour tout w'' *tel que* $ww'' \in R$, $|v| \leq |w''|$

Démonstration : Elle se fait par récurrence sur la longueur des facteurs gauches de R. En effet, les relations sont trivialement vérifiées si $|w| = 0$, car alors $w = u = v = \varepsilon$.

Supposons les relations vérifiées pour les facteurs gauches de longueur inférieure ou égale à n. Posons $w = ft$ avec $|f| = n$ et t est une lettre alors $(u, v) = \delta(\delta((\varepsilon, \varepsilon), f), t)$, $\lambda((\varepsilon, \varepsilon), w) = \lambda((\varepsilon, \varepsilon), f)\, \lambda(\delta((\varepsilon, \varepsilon), f), t)$.

Par hypothèse de récurrence, nous avons les relations :

$(u', v') = \delta((\varepsilon, \varepsilon), f) \in Q$ en raison de iii) et du lemme II-7 qui assure que $l(v') \leq k$,

i) $\Pi_X(f) = \alpha \circ \lambda((\varepsilon, \varepsilon), f)\, u',$

ii) $\Pi_Y(f) = \beta \circ \lambda((\varepsilon, \varepsilon), f)\, v',$

iii) $|v'| \leq \sup \{|f'|_Y - |f'|_X \,/\, f' \in FD(f)\},$

iv) $\forall f''$ tel que $ff'' \in R$ $|f''| \geq |v'|$.

Nous allons examiner un par un les 5 cas possibles :

1) $t \in X$

a) (u', v') vérifie $|u'| < k$ et $v' = \varepsilon$,

$(u, v) = \delta((u', \varepsilon), t) = (u' t, \varepsilon) \in Q.$

De plus $\lambda((u', v'), t) = \varepsilon$

$$\begin{aligned} \text{i) } \Pi_X(w) &= \Pi_X(f)\, \Pi_X(t) = \alpha \circ \lambda((\varepsilon, \varepsilon), f)\, u'\, t \\ &= \alpha \circ \lambda((\varepsilon, \varepsilon), f)\, \lambda((u', v'), t)\, u'\, t' \\ &= \alpha \circ \lambda((\varepsilon, \varepsilon), w)\, u, \end{aligned}$$

ii) $\Pi_Y(w) = \Pi_Y(f)\ \Pi_Y(t) = \beta \circ \lambda((\varepsilon, \varepsilon), f)\ \varepsilon$

$= \beta \circ \lambda((\varepsilon, \varepsilon), f)\ \lambda((u', v'), t)\ \varepsilon$

$= \beta \circ \lambda((\varepsilon, \varepsilon), w)\ v.$

iii) est bien vérifiée car $|v| = |v'| = 0.$

iv) comme $|v| = 0$, pour tout mot w'' vérifiant $ww'' \in R$

$|w''| \geq 0 = |v|$

b) (u', v') est de la forme (u', ε) avec $|u'| = k$ alors $u'=x'u''$ et $x' \in X$

$(u, v) = \delta((u', \varepsilon), t) = (u''\ t, \varepsilon) \in Q$

de plus $\lambda((u', \varepsilon), t) = (x', a)$ ou encore

$\alpha \circ \lambda((u', \varepsilon), t) = x'$

$\beta \circ \lambda((u', \varepsilon), t) = \varepsilon$

i) $\Pi_X(w) = \Pi_X(f)\ \Pi_X(t) = \alpha \circ \lambda((\varepsilon, \varepsilon), f)\ u'\ t$

$= \alpha \circ \lambda((\varepsilon, \varepsilon), f)\ x'\ u''\ t = \alpha \circ \lambda((\varepsilon,\varepsilon),f)\ \alpha \circ \lambda((u',\varepsilon),t) u''t$

$= \alpha \circ \lambda((\varepsilon, \varepsilon), w)\ u$

ii) $\Pi_Y(w) = \Pi_Y(f)\ \Pi_Y(t) = \beta \circ \lambda((\varepsilon, \varepsilon), f)$

$= \beta \circ \lambda((\varepsilon, \varepsilon), f)\ \beta \circ \lambda((u', \varepsilon), t)$

$= \beta \circ \lambda((\varepsilon, \varepsilon), w)\ v$

$\left.\begin{matrix}\text{iii)}\\ \text{iv)}\end{matrix}\right\}$ $|v| = |v'| = 0$. Alors iii) et iv) se démontrent comme dans le cas 1-a.

c) (u', v') est de la forme (ε, v') avec $v' = y'\ v''$ et $y' \in Y.$

$(u, v) = \delta((\varepsilon, v'), t) = (\varepsilon, v'') \in Q.$

de plus $\lambda((\varepsilon, v'), t) = (t, y')$ ou encore

$\alpha \circ \lambda((\varepsilon, v'), t) = t,$

$\beta \circ \lambda((\varepsilon, v'), t) = y'.$

i) $\Pi_X(w) = \Pi_X(f)\ \Pi_X(t) = \alpha \circ \lambda((\varepsilon,\ \varepsilon),\ f)t$

$= \alpha \circ \lambda((\varepsilon,\ \varepsilon),\ f)\ \alpha \circ \lambda((\varepsilon,\ v'),\ t)$

$= \alpha \circ \lambda((\varepsilon,\ \varepsilon),\ w)\ u.$

ii) $\Pi_Y(w) = \Pi_Y(f)\ \Pi_Y(t) = \beta \circ \lambda((\varepsilon,\ \varepsilon),\ f)\ v'$

$= \beta \circ \lambda((\varepsilon,\varepsilon),f)\ y'v'' = \beta \circ \lambda((\varepsilon,\varepsilon),f)\ \beta \circ \lambda((\varepsilon,v'),t)v''$

$= \beta \circ \lambda((\varepsilon,\ \varepsilon),\ w)\ v.$

iii) est évidemment vérifiée car $|v| = |v'| - 1.$

iv) soit w'' un mot quelconque vérifiant $w\ w'' \in R.$

$w = ft$, par hypothèse de récurrence $|tw| \geq |v'|$

$|w''| = |tw''| - 1 \geq |v'| - 1 = |v|.$

2) $t \in Y$

a) $(u',\ v')$ est de la forme $(\varepsilon,\ v')$ où $|y| < k$

$(u,\ v) = \delta((\varepsilon,\ v'),\ t) = (\varepsilon,\ v't).$

de plus $\lambda((\varepsilon,\ v'),\ t) = \varepsilon.$

i) $\Pi_X(w) = \Pi_X(f) = \alpha \circ \lambda((\varepsilon,\ \varepsilon),\ f)\ \alpha \circ \lambda((\varepsilon,\ v'),\ t)$

$= \alpha \circ \lambda((\varepsilon,\ \varepsilon),\ w)\ u.$

ii) $\Pi_Y(w) = \Pi_Y(f)\ \Pi_Y(t) = \beta \circ \lambda((\varepsilon,\ \varepsilon),\ f)\ v'\ t$

$= \beta \circ \lambda((\varepsilon,\ \varepsilon),\ f)\ \beta \circ \lambda((\varepsilon,\ v'),\ t)\ v't$

$= \beta \circ \lambda((\varepsilon,\ \varepsilon),\ w)\ v.$

iii) $|v| = |v'| + 1,$

$\sup\{|w'|_Y - |w'|_X / w' \in FD(w)\} = \sup\{|f'|_Y - |f'|_X / f' \in FD(f)\} + 1$

puisque $t \in Y$

d'où $|v| \leq \sup\{|w'|_Y - |w'|_X\ /\ w' \in FD(w)\}.$

iv) Soit w'' un mot vérifiant $ww'' \in R.$

$w = ft$, f peut se décomposer en $f = f_o\ f_1$ avec

$\delta((\varepsilon,\ \varepsilon),\ f_o) = (\varepsilon,\ \varepsilon)$ et $\delta((\varepsilon,\ \varepsilon),\ f_1) = (\varepsilon,\ v').$

On supposera, ce qui ne nuit pas à la généralité de la démonstration, qu'aucune lettre de f_1 ne nous conduit à l'état $(\varepsilon, \varepsilon)$ en d'autres termes

$|f_1|_Y - |f_1|_X = |v'|$. Mais $f_1 \in FG(R)$ on peut lui appliquer l'hypothèse de récurrence, ce qui nous assure que $|v'| \leq |tw''|$.

$|f_1 t|_Y - |f_1 t|_X = |v|$ mais $f_1\ tw'' \in R$

donc $|f_1\ tw''|_X - |f_1\ tw''|_Y \geq 0$

$|f_1 t|_X - |f_1 t|_Y + |w''|_X - |w''|_Y = - |v| + |w''|_X - |w''|_Y \geq 0$

ce qui se traduit par $|w''|_X - |w''|_Y \geq |v|$ et à fortiori $|w''| > |v|$.

b) (u', v') est de la forme (u', ε) avec $u' = x'\ u''$ et $x' \in X$

$(u, v) = \delta((u', \varepsilon), t) = (u'', \varepsilon) \in Q$,

de plus $\lambda((u', \varepsilon), t) = (x', t)$ ou encore

$\alpha \circ \lambda((u', \varepsilon), t) = x'$,

$\beta \circ \lambda((u', \varepsilon), t) = t$,

i) $\Pi_X(w) = \Pi_X(f)\ \Pi_X(t)$

$= \alpha \circ \lambda((\varepsilon, \varepsilon), f)\ u' = \alpha \circ \lambda((\varepsilon, \varepsilon), f)\ x'\ u''$

$= \alpha \circ \lambda((\varepsilon, \varepsilon), f)\ \alpha \circ \lambda((u', \varepsilon), t)\ u''$

$= \alpha \circ \lambda((\varepsilon, \varepsilon), w)\ u$.

ii) $\Pi_Y(w) = \Pi_Y(f)\ \Pi_Y(t)$

$= \beta \circ \lambda((\varepsilon, \varepsilon), f)\ t$

$= \beta \circ \lambda((\varepsilon, \varepsilon), f)\ \beta \circ \lambda((u', \varepsilon), t)$

$= \beta \circ \lambda((\varepsilon, \varepsilon), w)\ v$.

$\left.\begin{matrix}\text{iii)}\\ \text{iv)}\end{matrix}\right\}$ $|v| = |v'| = 0$. La démonstration est la même que dans le cas 1-a.

Reprenons la démonstration du lemme II-9.

Pour tout mot w de R, les relations du lemme II-10 sont vérifiées. On peut donc écrire que

$\Pi_Y(w) = \beta \circ \lambda((\varepsilon, \varepsilon), w)\ v$.

Mais d'après iv) $|v| = |\varepsilon| = 0$, ce qui implique que $v = \varepsilon$ et $\delta((\varepsilon, \varepsilon), w) = (u, \varepsilon)$.

Par définition $\gamma(w) = \lambda((\varepsilon, \varepsilon), w)\ s(w, \varepsilon)$. En posant $u = u_1 \ldots u_n$ avec $u_i \in X$, $\forall i \in [1, n]$ $s(u, \varepsilon) = (u_1, a) \ldots (u_n, a)$ donc $\alpha \circ s(u, \varepsilon) = u_1 \ldots u_n = u$ et $\beta \circ s(u, \varepsilon) = \varepsilon$.

$\alpha \circ \gamma(w) = \alpha \circ \lambda((\varepsilon, \varepsilon), w)\ u = \Pi_X(w)$ d'après le lemme II-10 et

$\beta \circ \gamma(w) = \alpha \circ \lambda((\varepsilon, \varepsilon), w)\ \varepsilon = \alpha \circ \lambda((\varepsilon, \varepsilon), w)\ v = \Pi_Y(w)$ d'après le lemme II-10.

□

Preuve du théorème II - 3

La condition est suffisante. Réciproquement, soit τ une transduction rationnelle décroissante de X^* dans Y^*. On peut, sans nuire à la généralité de la démonstration, supposer que X et Y sont des alphabets disjoints. τ s'écrit, en forme normale, $\tau = \Pi_Y \circ (\cap R) \circ \Pi_X^{-1}$ mais alors R vérifie (P) (corollaire II-8), on peut appliquer le lemme II - 9 ce qui permet d'écrire τ sous la forme $\tau = \beta \circ (\cap\ \gamma(R)) \circ \alpha^{-1}$ où α et β sont des homomorphismes alphabétiques et α est strictement alphabétique.

□

La construction utilisée dans le lemme II - 9 peut être faite sur la première factorisation de $\tau = g \circ (\cap R) \circ h^{-1}$, Ce qui nous conduit au résultat :

Proposition II - 11 : *Si $\tau = g \circ (\ R) \circ h^{-1}$ (où g et h sont des homomorphismes alphabétiques) est une transduction rationnelle décroissante de X^* dans Y^*, il existe une application sous-séquentielle γ, un homomorphisme strictement alphabétique h', un homomorphisme alphabétique g' tels que :*

$$\tau = g' \circ (\cap\ \gamma(R)) \circ h'^{-1}$$

<u>Démonstration</u> : Supposons, dans un premier temps que les alphabets X et Y soient disjoints

Définissons l'homomorphisme de $Z^* \to (X \cup Y)^*$ (où Z est l'alphabet sur lequel est défini le langage rationnel R) par

$$\forall z \in Z \quad \theta(z) = h(z)g(z)$$

alors pour tout mot $w \in R$ on a les relations

$$h(w) = \Pi_X(\theta(w))$$

$$g(w) = \Pi_Y(\theta(w))$$

dont on déduit $\hat{\tau} = \{(h(w), g(w)) \ / \ w \in R\} = \{(\Pi_X(w), \Pi_Y(w)) \ / w \in \theta\ (R)\}$.

d'après le lemme précédent il existe g', h', γ' où γ' est une application sous-séquentielle, h' un homomorphisme strictement alphabétique, g' un homomorphisme alphabétique tels que

$$\{(\Pi_Y(w), \Pi_Y(w)) \ / \ w \in \theta(R)\} = \{(h'(w), g'(w)) \ / \ w \in \gamma(\theta(R))\}$$

il suffit de choisir $\gamma = \gamma' \circ \theta$.

Supposons, maintenant, que $X \cap Y \neq \emptyset$; il existe une transduction rationnelle τ' de X^* dans $\bar{Y}^*$ avec $X \cap \bar{Y} = \emptyset$, il existe un homomorphisme strictement alphabétique t tels que

$$\tau = t \circ \tau'$$

la construction précédente peut être faite sur τ', comme t est alphabétique, τ possède bien la décomposition voulue.

Puisque τ est croissante si et seulement si τ^{-1} est décroissante, le théorème II-3 nous permet d'avoir une caractérisation des transductions rationnelles croissantes :

Théorème II - 12 : Une transduction rationnelle τ *de* X^* *dans* Y^* *est croissante si et seulement si il existe un langage rationnel* $R \subseteq Z^*$, *deux homomorphismes alphabétiques* h *et* g, h *de* Z^* *dans* X^*, g *de* Z^* *dans* Y^*, g *strictement alphabétique tels que* $\tau = g \circ (\cap R) \circ h^{-1}$.

Remarque : Toute transduction rationnelle décroissante (resp. croissante) est d'image finie (resp. fidèle).

C : DÉCOMPOSITIONS

Lemme II - 13 : Pour toute transduction rationnelle τ (resp. rationnelle fidèle τ_f, rationnelle d'image finie τ_{if}, rationnelle bifidèle τ_{bf}) de X^* dans Y^*, il existe une transduction rationnelle décroissante τ_d (resp. décroissante fidèle τ_{df}, décroissante τ_d, décroissante fidèle τ_{df}), une transduction rationnelle croissante τ_c (resp. croissante τ_c, croissante d'image finie τ_{cif}, croissante d'image finie τ_{cif}) telles que :

$$\tau = \tau_d \circ \tau_c$$

$$(\text{resp. } \tau_f = \tau_{df} \cdot \tau_c,\ \tau_{if} = \tau_d \cdot \tau_{cif},\ \tau_{bf} = \tau_{df} \cdot \tau_{cif}).$$

Démonstration : D'après le théorème de Nivat, τ peut se mettre sous la forme $\tau = g \circ (\cap R) \circ h^{-1}$ où R est un langage rationnel de Z^*, g et h deux homomorphismes alphabétiques. Il suffit alors de choisir :

$$\tau_d = g \quad \text{et} \quad \tau_c = id \circ (\cap R) \circ h^{-1}.$$

(resp. τ peut se décomposer également en $\tau_d = g \circ (\cap R) \circ id$ et $\tau_c = id \circ (\cap R) \circ h^{-1}$. τ_d est décroissante car g est alphabétique et id est strictement alphabétique si en plus g est limite sur R, τ_d est décroissante fidèle). τ_c est croissante car h est alphabétique et id strictement alphabétique. Si, en plus, h est limite sur R, τ_c est croissante d'image finie.

$\square$

Afin de montrer la factorisation symétrique c'est à dire : Toute transduction rationnelle est la composition d'une transduction rationnelle décroissante suivie d'une transduction rationnelle croissante, nous utiliserons le résultat suivant, obtenu en utilisant la représentation matricielle des transductions rationnelles due à M. Nivat et dont la démonstration se trouve dans le livre de J. Berstel [6] :

<u>*Lemme II - 14*</u> [6] : *Une transduction* τ *de* X^* *dans* Y^* *est rationnelle si et seulement s'il existe un alphabet* Z, *un homomorphisme strictement alphabétique* h *de* Z^* *dans* X^*, *une substitution rationnelle* s *de* Z^* *dans* Y^* *et un langage rationnel* $K \subseteq Z^*$ *tels que*

$$\tau(w) = s(h^{-1}(w) \cap R).$$

Avant d'aborder la décomposition au niveau des transductions remarquons que :

<u>*Lemme II - 15*</u> : *Toute substitution* s *de* X^* *dans* Y^* *peut être décomposée en*

$$s = s_+ \circ s_-$$

où s_+ *est une substitution propre,*

s_- *est une substitution décroissante.*

<u>Démonstration</u> : Pour toute lettre x de X, on définit la substitution décroissante :

$$s_- : x \to \{\varepsilon, \tilde{x}\} \quad \text{si } \varepsilon \in s(x)$$

$$x \to \tilde{x} \quad \text{si } \varepsilon \notin s(x)$$

la substitution ci-dessous est croissante

$$s_+ : \tilde{x} \to s(x) \setminus \{\varepsilon\} \quad \forall \tilde{x}$$

montrons que $s = s_+ \circ s_-$

1) x vérifie $\varepsilon \notin s(x)$

$$s(x) = s_+(\tilde{x}) = s_+(s_-(x)) = s_+ \circ s_- (x)$$

2) x vérifie $\varepsilon \in s(x)$

$$s(x) = (s(x) \setminus \{\varepsilon\}) \cup \{\varepsilon\} = s_+(\tilde{x}) \cup s_+ (\varepsilon) = s_+ (\{\tilde{x},\varepsilon\})$$

$$= s_+ (s_- (x)) = s_+ \circ s_- (x)$$

□

<u>Proposition II - 16</u> : Pour toute transduction rationnelle τ, il existe une transduction rationnelle décroissante τ_d, une substitution rationnelle propre s_+ telle que

$$\tau = s_+ \circ \tau_d .$$

<u>Démonstration</u> : D'après le lemme II - 14, τ peut se mettre sous la forme $\tau = s \circ (\cap R) \circ h^{-1}$ où s est une substitution rationnelle, R est un langage rationnel défini sur l'alphabet Z, h est un homomorphisme strictement alphabétique. Mais le lemme II - 15 permet de décomposer s en $s = s_+ \circ s_-$ où s_+ est une substitution rationnelle propre et s une substitution rationnelle décroissante ce qui s'écrit :

$$\tau = s_+ \circ s_- \circ (\cap R) \circ h^{-1}$$

Comme h est strictement alphabétique, il suffit de choisir $\tau_d = s_- \circ (\cap R) \circ h^{-1}$ pour obtenir le résultat.

□

On peut utiliser ce résultat pour montrer que :

<u>Proposition II - 17</u> : Une transduction rationnelle τ_f (resp. τ_{if}, τ_{bf}) de X^* dans Y^* est fidèle (resp. d'image finie, bifidèle) si et seulement si il existe une transduction rationnelle décroissante fidèle τ_{df} (resp. décroissante τ_d, décroissante fidèle τ_{df}), une substitution rationnelle propre s_+ (resp. une substitution finie propre s_{f+}) telles que

$$\tau_f = s_+ \circ \tau_{df}$$

$$(\text{resp. } \tau_{if} = s_{f+} \circ \tau_d,\ \tau_{bf} = s_{f+} \circ \tau_{df})$$

<u>Démonstration</u> : Comme toute substitution rationnelle est une transduction rationnelle, que toute substitution rationnelle propre est une transduction rationnelle croissante donc fidèle, que toute substitution propre finie est une transduction rationnelle bifidèle, que la composition de deux transductions fidèles (resp. d'image finie, bifidèles) est une transduction fidèle (resp. d'image finie, bifidèle), on en déduit la condition suffisante du corollaire.

* Montrons que τ_f fidèle implique que τ_f peut se décomposer en $s_+ \circ \tau_{df}$. Remarquons d'abord que l'on peut supposer $s_+(w) \neq \emptyset$ quel que soit w. La transduction τ_f est rationnelle, elle peut se mettre sous la forme $\tau_f = s_+ \circ \tau_d$. Supposons τ_d non fidèle, il existe un mot w tel que $\tau_d^{-1}(w)$ ne soit pas finie, mais $s_+(w)$ existe et est non vide, soit $w_1 \in s_+(w)$, $w \in s_+^{-1}(w_1)$ et $\tau_d^{-1}(s_+^{-1}(w_1)) = \tau_f^{-1}(w_1)$ n'est pas finie, ce qui est contraire à l'hypothèse.

* Montrons que τ_{if} d'image finie implique que τ_{if} peut se décomposer en $s_{f_+} \circ \tau_d$ La transduction τ_{if} peut s'écrire $\tau_{if}=s_+ \circ \tau_d$. Supposons que s_+ ne soit pas finie c'est à dire, il existe une lettre z telle que $s_+(z)$ soit un ensemble infini.

1) z n'occure dans aucun mot image par τ_d d'un mot quelconque de X^* alors z n'intervient pas dans la transduction τ on peut l'abandonner.
2) Il existe $w_1 \in X^*$ tel que z occurre dans un mot de son image par τ_d alors $\tau_{if}(w_1) = s_+ \circ \tau_d(w_1)$ est un ensemble infini ce qui est impossible.

* De ces deux démonstrations on déduit que τ_{bf} bifidèle implique que τ_{bf} peut se décomposer en $s_{f+} \circ \tau_{df}$.

□

Puisque toute substitution rationnelle propre (resp. toute substitution finie propre) est une transduction rationnelle croissante (resp. croissante d'image finie), on en déduit le théorème de factorisation suivant :

<u>*Théorème II - 18*</u> *: Une transduction τ (resp. τ_f, τ_{if}, τ_{bf}) est rationnelle (resp. rationnelle fidèle, rationnelle d'image finie, rationnelle bifidèle) si et seulement il existe une transduction rationnelle décroissante τ_d (resp. décroissante fidèle τ_{df}, décroissante τ_d, décroissante fidèle τ_{df}), une tranduction rationnelle croissante τ_c (resp. croissante τ_c, croissante d'image finie τ_{cil}, croissante d'image finie τ_{cil}) telles que*

$$\tau = \tau_c \circ \tau_d$$

(*resp.* $\tau_f = \tau_c \circ \tau_{df}$, $\tau_{if} = \tau_{cif} \circ \tau_d$, $\tau_{bf} = \tau_{cif} \circ \tau_{df}$)

CHAPITRE III

LANGAGES D'IMAGE FINIE ET LANGAGES BIFIDELES

A : PROPRIÉTÉS ET CARACTÉRISATIONS DES LANGAGES D'IMAGE FINIE

B : EXEMPLES, CONTRE EXEMPLES DE LANGAGES D'IMAGE FINIE ET BIFIDÈLES

Définition III-1 : Un langage L *est d'image finie (resp. bifidèle) si et seulement si* C(L), *la plus petite famille contenant* L *et close par transduction rationnelle, est identique à* $C^{if}(L)$ *(resp.* $C^{bf}(L)$*), la plus petite famille contenant* L *et close par transduction rationnelle d'image finie (resp. transduction rationnelle bifidèle).*

Ce chapitre est destiné à isoler des propriétés des langages d'image finie et des langages bifidèles. Après avoir rassemblé un certain nombre de propriétés de Rat vis à vis de $C^{df}(L)$, $C^{bf}(L)$, $C^{d}(L)$, $C^{if}(L)$ suivant les propriétés de L et réciproquement, nous donnons une caractérisation des langages d'image finie et des langages bifidèles, caractérisation faisant intervenir l'opérateur syntaxique. Une caractérisation plus fine sera donnée pour les langages d'image finie et bornés sur une et deux lettres. Nous montrons ensuite que si L est un langage d'image finie alors $C^{d}(L) = C(L)$ c'est à dire que tout langage de C(L) peut être atteint à partir de L par transductions rationnelle décroissante, cette propriété et les résultats du chapitre II permettent d'affirmer que si L est un langage d'image finie, (resp. bifidèle) pour tout langage L' de C(L), il existe un langage rationnel R, un homomorphisme h, strictement alphabétique , un homomorphisme (resp. k-limite) g tels que $L' = g(h^{-1}(L) \cap R)$. En particulier comme tout générateur de Alg, la famille des langages algébriques, est bifidèle ce résultat peut être utilisé pour tout langage algébrique.

Alors que l'égalité $C^{if}(L) = C(L)$ implique $C^{d}(L) = C^{if}(L)$ nous montrons, en étudiant le langage $C_1 = \{a^n b^n \ / \ n \geq 0\}$ que $C^{d}(L) = C^{if}(L)$ n'implique pas $C^{if}(L) = C(L)$, mais, par contre, C_1 possède une propriété intéressante que ne possède pas forcément les langages d'image finie à savoir $C^{d}(L) = C^{df}(L)$, propriété que nous utiliserons dans le chapitre IV de cette thèse pour déterminer des classes de langages effaçables.

Nous terminons cette partie en montrant qu'il existe des langages L pour lesquels le cône décroissant engendré par L est strictement inclus dans le cône d'image finie engendré par L.

A : PROPRIÉTÉS ET CARACTÉRISATIONS DES LANGAGES D'IMAGE FINIE

Lemme III-2 : L est un langage fini si et seulement si $C^{bf}(L) = \mathrm{Fin}$.

Démonstration : La condition est suffisante puisque L est contenu dans $C^{bf}(L) = \mathrm{Fin}$.

Réciproquement, si L est un langage fini, il peut se mettre sous la forme $L = \{w_1, w_2, \ldots, w_p\}$. Si τ est une transduction rationnelle d'image finie alors $\tau(L) = \bigcup_{i=1}^{p} \tau(w_i)$ est une union finie d'ensembles finis on en déduit que $C^{bf}(L) \subseteq C^{if}(L) \subseteq \mathrm{Fin}$.

Il nous reste à montrer que la famille Fin est contenue dans $C^{df}(L)$. Pour ce faire, choisissons L' un langage fini quelconque. L' peut s'écrire $L' = \{w'_1, w'_2, \ldots, w'_n\}$. Soit $A = \{a_1, \ldots, a_n\}$ un alphabet, les homomorphismes h et g définis par $h(a_i) = w_i$ et $g(a_i) = w'_i$, quel que soit $i \in [1, n]$ alors

$$g(h^{-1}(L) \cap A) = g(A) = L'.$$

La transduction $\tau = g(\cap A) \circ h^{-1}$ est bifidèle donc L' appartient à $C^{bf}(L)$ d'où on conclue

$$\mathrm{Fin} \subseteq C^{bf}(L) \subseteq C^{if}(L) \subseteq \mathrm{Fin}. \ \square$$

Nous en déduisons immédiatement que seuls les langages infinis pourront être d'image finie en particulier :

Lemme III-3 : L est un langage rationnel infini si et seulement si $C^{df}(L) = \mathrm{Rat}$.

Démonstration : Supposons d'abord que $C^{df}(L)$ = Rat. Alors L est contenu dans $C^{df}(L)$ est rationnel. En raison du lemme précédent L ne peut pas être fini.

Réciproquement, supposons que L soit un langage rationnel infini contenu dans X^*. Comme $C^{df}(L) \subseteq C(L)$ = Rat, il suffit de montrer que Rat $\subseteq C^{df}(L)$. Choisissons donc R, un langage rationnel quelconque contenu dans Z^* et définissons h_a l'homomorphisme de X^* dans $\{a\}^*$ par :

pour tout x de X, $h_a(x) = a$

$L' = h_a(L)$ est un langage rationnel infini contenu dans a^*. Il existe donc deux entiers s et t, t strictement positif, tels que $L' \cap a^s(a^t)^*$ soit égal à $a^s(a^t)^*$. On en déduit que $a^s(a^t)^* \in C^{df}(L)$. Mais, l'entier t est strictement positif ce qui implique que $a^* \in C^{df}(a^s(a^t)^*)$.

soit $h' : Z^* \to \{a^*\}$ tel que $\forall z \in Z$, $h'(z) = a$, alors $R = R \cap h'^{-1}(a^*)$ appartient à $C^{df}(a^*)$ donc

$\forall R \in$ Rat : $R \in C^{df}(a^*) \subseteq C^{df}(a^s(a^t)^*) \subseteq C^{df}(L) \subseteq$ Rat. □

Comme Rat $= C^{df}(L) \subseteq C^{bf}(L) \subseteq C(L)$ = Rat on déduit le résultat :

<u>*Corollaire III-4*</u> *: Tout langage rationnel infini est bifidèle.*

<u>*Lemme III-5*</u> *: Si* L *est un langage infini,* Rat $\subseteq C^d(L)$.

<u>Démonstration</u> : Pour tout langage infini $L' \subseteq a^*$, $FG(L') = a^*$. Pour tout langage infini L, $h_a(L)$ (où h_a est l'homomorphisme qui, à toute lettre, associe a) est un langage infini contenu dans a^*, comme Rat $= C^d(a^*)$ alors :

$$\text{Rat} = C^d(a^*) = C^d(FG(h(L)) \subseteq C^d(L). \ \square$$

Remarque : Si L est un langage algébrique infini alors Rat $\subseteq C^{df}(L)$ puisqu'alors $h_a(L)$ est un langage algébrique contenu dans a^*, il est donc rationnel et d'après le lemme III-3 :

$$\text{Rat} = C^{df}(h_a(L)) \subseteq C^{df}(L).$$

Si L n'est pas algébrique, cette inclusion n'est pas vérifiée par tout langage infini, il faut une condition supplémentaire sur les longueurs des mots de L :

Lemme III-6 : Pour tout langage L de X^*, les trois propriétés suivantes sont équivalentes :

i) Rat $\subseteq C^{bf}(L)$

ii) Rat $\subseteq C^{df}(L)$

iii) Il existe $k \in \mathbb{N}$ *vérifiant* $\forall w \in L, \exists w' \in L$ *tels que*

$|w| < |w'| \leq k(|w| + 1)$

Démonstration

** i) => ii) c'est à dire Rat $\subseteq C^{bf}(L)$ implique Rat $\subseteq C^{df}(L)$.

Soit R un langage rationnel contenu dans Y^* et soit c une nouvelle lettre. Définissons h_c l'homomorphisme $(Y \cup \{c\})^*$ dans Y^* par, pour toute lettre y de Y, $h_c(y) = y$ et $h_c(c) = \varepsilon$, bien entendu $h_c^{-1}(R)$ appartient à Rat donc à $C^{bf}(L)$. Les lemmes III-9 et III-13 permettent de conclure que $h_c^{-1}(R)$ est dans $C^{df}(L)$ et de l'égalité $R = h_c^{-1}(R) \cap Y^*$ on déduit que R est dans $C^{df}(L)$.

** ii) => iii) c'est à dire Rat $\subseteq C^{df}(L)$ implique qu'il existe un entier k vérifiant : $\forall w \in L, \exists w' \in L$ tel que $|w| < |w'| \leq k(|w| + 1)$.

D'après le lemme III-3, $C^{df}(c^*) = \text{Rat}$, il existe une transduction rationnelle décroissante fidèle $\tau = g \circ (\cap R) \circ h^{-1}$ vérifiant $c^* = \tau(L)$ où h est un homomorphisme strictement alphabétique, g est un homomorphisme alphabétique k'-limité sur R (cf. chapitre II). Soit $k = k' + 1$. Pour tout mot w de L, $c^{|w|+1}$ appartient à c^*, il existe donc u dans R tel que $c^{|w|+1} = g(u)$ et $h(u) \in L$. Or

$$|w| + 1 = |g(u)| \leq |u| = |h(u)|$$

et puisque g est k'-limite.

$$|h(v)| = |u| \leq (k'+1)\ |g(u)| \leq k(|w| + 1)$$

il suffit alors de choisir $w' = h(u)$ pour obtenir

$$|w| < |w'| \leq k(|w| + 1)$$

** iii) => i) c'est à dire : il existe un entier k vérifiant $\forall\ w \in L$, $\exists\ w' \in L$ tels que $|w| < |w'| \leq k(|w| + 1)$ implique $Rat \subseteq C^{bf}(L)$.

Appelons ℓ la longueur du mot de L le plus court. Soit θ l'homomorphisme de X^* dans $\{c\}^*$ qui remplace toute lettre de X par c.

Soit I tel que $L'=\theta(L) = \{c^i / i \in I\}$ et pour tout élément i de I, $I_i =]i, k(i + 1)]$. De l'hypothèse on déduit facilement que pour tout intervalle I_i, il exite un entier $j > i$ tel que $I_i \cap I_j \neq \emptyset$, autrement dit :

$$\bigcup_{i \in I}]i, k(i+1)] = \{n \in \mathbb{N}\ /\ n \geq \ell\}$$

Choisissons un entier n supérieur à ℓ et appelons i le plus grand élément de I tel que $n \in I_i$. Il existe $j \in I$ tel que $j \in I_i$, i étant le plus grand élément de I tel que $n \in I_i$, de l'inégalité $j > i$, on déduit que $j \geq n$ ou encore $n \leq j \leq k(i+1) \leq kn$ car $i + 1 \leq n$, en conclusion, pour tout $n \geq \ell$, il existe $j \in I$ tel que $n \leq j \leq kn$. On est alors en mesure de construire la transduction décroissante fidèle $\tau = g \circ (\cap R) \circ h^{-1}$ de la façon suivante : h et g sont des homomorphismes de $\{a, b\}^*$ dans c^* définis par

$$h : h(a) = c,\ h(b) = c$$

$$g : g(a) = \varepsilon,\ g(b) = c$$

Si $R_o = \bigcup_{i=o,\dots,k-1} a^i b$ alors R est égal à $R_o^* \cap g^{-1}(c^\ell c^*)$

Montrons, maintenant que $c^{\ell} c^{*} = g(h^{-1}(L') \cap R)$

* De façon évidente $g(h^{-1}(L') \cap R)$ est contenu dans $c^{\ell} c^{*}$.

* Soit $w \in c^{\ell} c^{*}$ c'est à dire $w = c^{p}$ avec $p \geq \ell$

$g^{-1}(w) \cap R = \{a^{i_1} b\, a^{i_2} b \ldots a^{i_p} b \;/\; i_j \in [0, k-1]$ et $j = 1, \ldots, p\}$

$h(g^{-1}(w) \cap R) = \{c^{i} \;/\; i \in [p, kp]\}$

mais p étant plus grand que ℓ, on a vu qu'il existait $j \in I$ tel que $j \in [p, kp]$ donc $c^{j} \in h(g^{-1}(w) \cap R)$. On peut donc trouver dans R un mot w' vérifiant $c^{j} = h(w')$ et $g(w') = w \cdot c^{j}$ est élément de L', $w \in g(h^{-1}(L') \cap R)$

$$\mathrm{Rat} = C^{df}(c^{\ell} c^{*}) \subseteq C^{df}(L') \subseteq C^{bf}(L). \;\square$$

Tout cône rationnel principal est fermé par union, on montre, de façon analogue que :

<u>*Lemme III-7*</u> *: Le cône rationnel d'image finie* $C^{if}(L)$ *(resp. décroissant* $C^{d}(L)$*, bifidèle* $C^{bf}(L)$*, décroissant fidèle* $C^{df}(L)$*) est fermé par union.*

<u>*Lemme III-8*</u> : $\forall\, L, L'\;\; C^{df}(L') = C^{df}(L) \Rightarrow C^{d}(L') = C^{d}(L)$

<u>Démonstration</u> : Une transduction décroissante fidèle est décroissante, la composée de deux transductions décroissantes est décroissante, on en déduit que $C^{d}(C^{df}(L)) = C^{d}(L)$. $C^{d}(C^{df}(L)) = C^{d}(C^{df}(L'))$ implique le résultat.$\square$

Nous allons, à partir de maintenant, très souvent utiliser le langage $L \uparrow c^{*}$, construit à partir de L en insérant entre deux lettres de mots de L un nombre quelconque de c. Plus précisément

$L \subseteq X^{*}$, $c \notin X$

$L \uparrow c^{*} = \{x_1 c^{n_1} x_2 c^{n_2} \ldots x_p c^{n_p} \;/\; x_1 x_2 \ldots x_p \in L$ et $\forall\, i \in [1, p]\; n_i \geq 0,\; x_i \in X\}$

Avant d'aborder les propositions III-10 et 11, constamment utilisées dans la suite de ce travail et qui constituent des résultats de base, commençons par une remarque :

Lemme III-9 : Pour tout langage $L \neq \{\varepsilon\}$,

$$C^{df}(L \uparrow c^*) = C^{df}(c^* L \uparrow c^*)$$

Démonstration :

* $L \uparrow c^*$ est égal à $(X c^*)^* \cap (c^* L \uparrow c^*)$ ce qui implique, naturellement, que $C^{df}(L \uparrow c^*)$ est contenu dans $C^{df}(c^* L \uparrow c^*)$.

* Réciproquement, pour montrer que $c^* L \uparrow c^*$ appartient à $C^d(L \uparrow c^*)$, il y a lieu de distinguer deux cas :

1er cas : Le langage L ne contient pas le mot vide.

Prenons $L' = L \uparrow c^* \cap X c^+ (X c^*)^*$, bien entendu, L' appartient à $C^{df}(L \uparrow c^*)$. Nous allons construire une transduction rationnelle décroissante fidèle τ telle que $c^+ L \uparrow c^* = \tau(L')$.

Soit $Z = X \cup \bar{X} \cup \tilde{X} \cup \{c_x \ / \ x \in X\} \cup \{c\}$ avec $\bar{X} = \{\bar{x} \ / \ x \in X\}$ et $\tilde{X} = \{\tilde{x} \ / \ x \in X\}$. Le langage R défini sur Z par $R = \bigcup_{x \in X} \bar{x}\, \tilde{x}^*$ est rationnel.

On définit alors les homomorphismes h et g de Z^* dans $(X \cup \{c\})^*$ par :

$$\begin{array}{lll} \forall x \in X & h(\bar{x}) = x & g(\bar{x}) = c \\ & h(\tilde{x}) = c & g(\tilde{x}) = c \\ & h(c_x) = c & g(c_x) = x \\ & h(x) = x & g(x) = x \\ & h(c) = c & g(c) = c \end{array}$$

La transduction $\tau = g \circ (\cap R) \circ h^{-1}$ conserve les longueurs car h et g sont strictement alphabétiques [10], elle est, à fortiori, décroissante fidèle.

Montrons que $c^+ L \uparrow c^* = \tau(L') = g(h^{-1}(L') \cap R)$.

- Soit w, un mot de $g(h^{-1}(L') \cap R)$.

Il existe $w \in R$ vérifiant $h(w') \in L'$ et $g(w') = w$. Le mot w' est dans R, ceci implique qu'il existe x dans X, un entier $k \geq 0$, un mot w" de $\{c, X\}^*$ vérifiant $w' = \bar{x}\, \tilde{x}^k\, c_x\, w''$. Mais h(w') appartient à L', ce qui signifie que $h(w') = h(\bar{x})\, (h(\tilde{x}))^k\, h(c_x)\, h(w'') = x\, c^{k+1}\, w''$ appartient à L', ou encore, $x\, w'' \in L \uparrow c^*$ mais $w = g(w') = g(\bar{x})\, (g(\tilde{x}))^k\, g(c_x)\, g(w'') = c^{k+1}\, x\, w'' \in c^+ L \uparrow c^*$.

- Soit w, un mot de $c^+ L \uparrow c^*$.

$w = c^{k_o} x_1 c^{k_1} x_2 \ldots x_n c^{k_n}$ où $x_1 \ldots x_n$ est un mot de L et pour tout $i \geq 1$, $k_i \geq 0$, $x_i \in X$ et $k_o \geq 1$.

Le mot $w' = \bar{x}_1\, \tilde{x}_1^{k_o - 1}\, c_{x_1}\, c^{k_1}\, x_2 \ldots x_n\, c^{k_n}$ est dans R et vérifie $h(w') = x_1 c^{k_o} c^{k_1} x_2 \ldots x_n c^{k_n} \in L''$ et $g(w') = w$. En d'autres termes $w \in g(h^{-1}(L') \cap R)$.

De façon évidente $c^* L \uparrow c^* \in C^{df}(c^+ L \uparrow c^*)$

<u>2nd cas</u> : Le langage L contient le mot vide.

Posons $L' = L \setminus \varepsilon$. Alors $c^* L \uparrow c^* = c^* \cup c^* L' \uparrow c^*$ or, nous venons de le voir, $c^* L' \uparrow c^* \in C^{df}(L' \uparrow c^*)$ et $c^* \in C^{df}(c^*)$ d'où :

$$c^* L \uparrow c^* \in C^{df}(L' \uparrow c^*) \cup C^{df}(c^*) = C^{df}(L \uparrow c^*)$$

<u>Lemme III-10</u> : Pour tout langage $L \subseteq X^*$, ***on a les relations***

$$C(L) = C^{d}(L \uparrow c^*) = C^{if}(L \uparrow c^*)$$

$$C^{f}(L) = C^{df}(L \uparrow c^*) = C^{bf}(L \uparrow c^*).$$

<u>Démonstration</u> : Soit $L' \subseteq X'^*$ un élément de C(L). D'après le théorème de Nivat, il existe un alphabet Z, un langage rationnel $R \subseteq Z^*$, deux homomorphismes

alphabétiques h de Z^* dans X^* et g de Z^* dans X'^* vérifiant $L' = g(h^{-1}(L) \cap R)$. Si H_1 désigne l'ensemble $\{z \in Z \ / \ h(z) \neq \varepsilon\}$ alors on peut définir, à partir de h, h' de Z^* dans $(X \cup \{c\})^*$ par : $h'(x) = h(x)$ si $x \in H_1$ et $h'(x) = c$ sinon. Alors L' est égal à $g(h'^{-1}(c^* \ L \uparrow c^*) \cap R)$, ou encore puisque h' est strictement alphabétique, d'après le théorème II-3 :

$$L' \in C^d(c^* \ L \uparrow c^*)$$

mais, en utilisant le lemme précédent,

$$L' \in C^d(c^* \ L \uparrow c^*) = C^d(L \uparrow c^*)$$

Choisissons, maintenant, L' dans $C^f(L)$ le même raisonnement peut être refait avec, comme hypothèse supplémentaire, g k-limité sur R, ce qui nous conduit à

$$L' \in C^{df}(c^* \ L \uparrow c^*) = C^{df}(L \uparrow c^*) \quad \square$$

Ce lemme permet d'avoir une caractérisation très simple et très utile des langages d'image finie.

<u>Proposition III-12</u> : L est un langage d'image finie (resp. $C^f(L) = C^{bf}(L)$*) si et seulement si* $L \uparrow c^* \in C^{if}(L)$ *(resp.* $L \uparrow c^* \in C^{bf}(L)$*).*

<u>Démonstration</u>

$C^{if}(L) = C(L) \Rightarrow L \uparrow c^* \in C(L) = C^{if}(L)$

Réciproquement, $L \uparrow c^* \in C^{if}(L)$ implique $C(L) = C^{if}(L \uparrow c^*) \subseteq C^{if}(L)$.

La démonstration est la même pour $C^{bf}(L)$. $\square$

Une transduction rationnelle décroissante est, bien entendu, une tranduction rationnelle d'image finie. Pour un langage quelconque L nous avons donc les inclusions :

$$C^d(L) \subseteq C^{if}(L) \subseteq C(L)$$

Nous allons montrer que l'égalité entre $C^{if}(L)$ et $C(L)$ implique obligatoirement l'égalité entre $C^{d}(L)$ et $C(L)$.

Lemme III-13 : Soient L *et* L' *deux langages. Si* $L \uparrow c^*$ *appartient à* $C^{if}(L')$ *(resp.* $C^{bf}(L')$*) alors* $L \uparrow c^*$ *appartient à* $C^{d}(L')$ *(resp.* $C^{df}(L')$*)*

Démonstration : Soient X et X' les alphabets respectifs de L et L'. L'idée de la démonstration consiste à partir de τ, transduction rationnelle d'image finie telle que $L \uparrow c^* = \tau(L')$, à construire une transduction τ' qui par rapport à τ va supprimer un certain nombre de c afin de rendre τ' décroissante tout en gardant la possiblilité d'avoir entre deux lettres de X un nombre quelconque de c.

Si $L \uparrow c^*$ appartient à $C^{if}(L')$. D'après le théorème de Boassont-Nivat,il existe un langage rationnel $R \subseteq Z^*$, deux homomorphismes alphabétiques h et g, h, k-limité sur R, de Z^* dans X'^* et g de Z^* dans $(X \cup \{\varepsilon\})^*$ vérifiant $L \uparrow c^* = g(h^{-1}(L') \cap R)$.

A partir de $\tau = g \circ (\cap R) \circ h^{-1}$ construisons $\tau' = g' \circ (\cap R') \circ h'^{-1}$ vérifiant d'une part $|h'(w)| \geq |g'(w)|$ et ce quel que soit $w \in Z^*$ et d'autre part $L \uparrow c^* = g'(h'^{-1}(L') \cap R')$.

Construction de R'

Sélectionnons dans R les mots dont les images par g appartiennent à $(X c^* c^k)^*$ (c'est à dire entre 2 lettres de X il y a au moins k occurrences de c).

$$R_1 = R \cap g^{-1}(X c^* c^k)^*.$$

On va maintenant découper tout mot de R_1 en "tranches" de k lettres avec bien entendu un "résidu" de longueur $\leq k$. Pour réaliser ce découpage nous allons définir deux nouveaux alphabets

$$Z_1 = \underbrace{Z \times Z \times \ldots \times Z}_{k\text{-fois}} \qquad Z_2 = \bigcup_{i=1}^{k-1} \underbrace{Z \times Z \times \ldots \times Z}_{i\text{-fois}}$$

Soit t l'homomorphisme de $Z' = Z_1 \cup Z_2$ dans Z^* défini par :

$$t(x_1, \ldots, x_n) = x_1 \ldots x_n \quad \forall (x_1, \ldots, x_n) \in Z'$$

Le langage rationnel annoncé est

$$R' = t^{-1}(R_1) \cap Z_1^*(Z_2 \cup \{\varepsilon\})$$

Avant de donner les définitions de g' et h' nous pouvons faire quelques constatations :

Remarque 1

$$t(R') = R_1 \cap t(Z_1^*(Z_2 \cup \{\varepsilon\})) = R_1$$

Remarque 2

Toute lettre z de Z' a pour image par t un mot de longueur supérieure ou égale à k donc $|h(t(z))| \geq 1$.

Remarque 3

Toute lettre z de Z' apparaissant dans un mot R' possède une image par g qui peut être

- soit vide
- soit de la forme $c^i x c^j$ où $x \in X$
- soit contenue entièrement dans c^+

Construction de h' et g'

Compte tenu de ces remarques on définit les homomorphismes h' de Z'^* dans X'^* et g'^* de Z'^* dans $(X \cup \{\varepsilon\})$:

$\forall\ z = (z_1, \ldots, z_n) \in Z'$

$h'(z) = h(z_1 \ldots z_n)$ ou encore $h' = h \circ t$

$g'(z) = x$ si $g(z_1 \ldots z_n)$ est de la forme $c^i\, x\, c^j$

$\quad = c$ si $g(z_1 \ldots z_n)$ appartient à c^+

$\quad = \varepsilon$ si $g(z_1 \ldots z_n) = \varepsilon$

$g'(z)$ et $g(z_1 \ldots z_n)$ ne diffèrent que par le nombre d'occurrences de la lettre c.

Montrons que $L \uparrow c^* = g'(h'^{-1}(L') \cap R')$

* $g'(h'^{-1}(L') \cap R') = g \circ t(t^{-1} \circ h^{-1}(L') \cap R')$

$$= g(h^{-1}(L') \cap t(R'))$$

$$= g(h^{-1}(L') \cap R_1)$$

$$\subseteq g(h^{-1}(L') \cap R)$$

$$= L \uparrow c^*$$

* Afin de montrer l'inclusion dans l'autre sens, considérons $L_k = L \uparrow c^* \cap (X\, c^k\, c^*)^*$. Appelons h_c l'homomorphisme qui efface la lettre c et qui est l'identité sur X. Comme $c \notin X$, $h_c(L_k) = L$ et par construction $L_k = g(h^{-1}(L) \cap R_1)$. Soit $w \in L_k$, il existe $u \in R_1$ vérifiant $h(u) \in L'$ et $w = g(u)$. On peut découper u en $u_1 \ldots u_n\, u_{n+1}$ où $\forall\ i \in [1, n]$ $u_i = u_{i,1} \ldots u_{i,k}$ avec $u_{ij} \in Z\ \forall\ j \in [1, k]$ et $u_{n+1} = u_{n+1,1} \ldots u_{n+1,\ell}$ avec $u_{n+1,j} \in Z\ \forall\ j \in [1, k]$ et $\ell < k$. Mais alors le mot $u' = (u_{1,1}, \ldots, u_{1,k}) \ldots (u_{n,1}, \ldots, u_{n,k})(u_{n+1,1}, \ldots, u_{n+1,\ell})$ est élément de R' et vérifie :

$$h'(u') = h(t(u')) = h(u) \in L'$$

et, puisque $g'(u')$ et $g(u)$ ne diffèrent que par le nombre d'occurrences de la lettre c, on peut écrire que

$$h_c(g'(u')) = h_c(g(u)) = h_c(w)$$

ce qui signifie que $h_c(w) \in h_c \circ g'(h'^{-1}(L') \cap R')$ et ce pour tout mot w de L_k. C'est à dire

$$L = h_c(L_k) \subseteq h_c \circ g'(h'^{-1}(L') \cap R')$$

$$L \uparrow c^* = h_c^{-1}(L) \cap (X\, c^*)^* \subset g'(h'^{-1}(L') \cap R')$$

Pour toute lettre z de Z', $|h'(z)| \geq 1$ et $|g'(z)| = 1$. La transduction $\tau' = g' \circ (\cap R') \circ h'^{-1}$ est donc décroissante, et de conclure que $L \uparrow c^*$ appartient bien à $C^d(L')$.

(Resp. si au départ g est limité sur R, alors g' garde cette propriété sur R'). □

En remplaçant L' par L dans l'énoncé du lemme précédent on obtient une propriété très intéressante des langages d'image finie :

<u>*Proposition III-14*</u> *:* L *est un langage d'image finie (resp. bifidèle) si et seulement si* $C(L) = C^d(L)$ *(resp.* $C^{f}(L) = C^{df}(L)$ *et* L ***est effaçable).***

<u>Démonstration</u> : De la chaîne d'inclusion $C^d(L) \subseteq C^{if}(L) \subseteq C(L)$ (resp. $C^{df}(L) \subseteq C^{bf}(L) \subseteq C(L)$) on déduit la condition suffisante.

Réciproquement, si L est d'image finie, $L \uparrow c^*$ qui appartient à $C(L)$ appartient également à $C^{if}(L)$. Le lemme précédent implique que $L \uparrow c^*$ appartient à $C^d(L)$ c'est à dire $C^d(L \uparrow c^*) \subseteq C^d(L)$, ce qui permet d'écrire que $C(L) = C^d(L \uparrow c\ ) \subseteq C^d(L)$. (Resp. se démontre de façon analogue). □

Notons qu'un langage bifidèle est fidèle et d'image finie, la réciproque n'est pas vraie puisque le langage symétrique sur deux lettres Sym et à la fois fidèle et d'image finie mais pas bifidèle

$$C(L) = C^{if}(L) = C^f(L) \not\Rightarrow C^{bf}(L) = C(L)$$

Ces résultats peuvent être regroupés en

<u>*Théorème III-15*</u> *: Soit* L *un langage défini sur un alphabet* X, *les trois propriétés sont équivalentes.*

i) L *est d'image finie*

ii) $L \uparrow c^* \in C^{if}(L)$

iii) $C(L) = C^{d}(L)$

de même, si L *est un langage effaçable* $\subseteq X^*$, *les trois propriétés sont équivalentes :*

i) L *est un langage bifidèle*

ii) $L \uparrow c^* \in C^{bf}(L)$

iii) $C(L) = C^{df}(L)$

B : EXEMPLES, CONTRE-EXEMPLES DE LANGAGES D'IMAGE FINIE ET BIFIDÈLES

Les résultats de L. Boasson - M. Nivat [8], S. Ginsburg - J. Goldstine - S. Greibach [14] sur les générateurs fidèles, ceux de M. Latteux [23] sur les générateurs bifidèles permettent d'énoncer ;

Tout générateur de Alg est un langage bifidèle [23]

Tout générateur de Ocl (la famille des langages à un compteur) est un langage bifidèle [24]

Tout générateur de Lin (la famille des langages linéaires) est à la fois d'image finie et effaçable mais non bifidèle [23].

C'est à dire :

$$C^{bf}(\mathrm{Sym}) \subsetneq C^{d}(\mathrm{Sym}) = C^{if}(\mathrm{Sym}) = C^{f}(\mathrm{Sym}) = C(\mathrm{Sym}) \text{ [23]}$$

On sait, d'autre part, que tout langage défini sur une lettre est effaçable [34].

Par contre nous allons démontrer que pour qu'un langage défini sur une lettre soit d'image finie il faut (et il suffit) qu'il soit rationnel infini. Cette propriété est également vérifiée par les langages bornés sur deux lettres, ce qui permet d'affirmer que $C_1 = \{a^n\, b^n \ / \ n \geq 0\}$ n'est pas un langage d'image finie.

Cependant, l'étude du langage C_1 permettra de montrer que

$$\mathrm{Rat} \subsetneq C^{df}(C_1) = C^{bf}(C_1) = C^{d}(C_1) = C^{if}(C_1) \subsetneq C^{f}(C_1) = C(C_1)$$

or, on a vu que l'égalité $C(L) = C^{if}(L)$ implique $C^{if}(L) = C^{d}(L)$, l'étude de C_1 permet de montrer que la réciproque est fausse

$$C^{if}(L) = C^{d}(L) \not\Rightarrow C(L) = C^{if}(L)$$

A la vue de ces résultats on peut se demander si les cônes décroissants et d'image finie engendrés par un langage L quelconque ne sont pas toujours identiques. L'étude du langage $\{a^{n!} / n \geq 0\}$ permet de répondre négativement à cette question.

<u>*Proposition III-16*</u> *: Soit* L *un langage contenu dans* $a^* b^*$. L *est d'image finie si et seulement si* L *est un langage rationnel infini.*

<u>Démonstration</u> : Supposons L d'image finie alors $L' = c^* L c^* \in C(L) = C^d(L)$. En vertu de la caractérisation des transductions rationnelles décroissantes du chapitre II, L' peut se mettre sous la forme $L' = g(h^{-1}(L) \cap R)$; où R est un langage rationnel contenu dans Z^*, g un homomorphisme alphabétique de Z^* dans $\{a, b, c\}^*$ et h un homomorphisme strictement alphabétique de Z^* dans $\{a, b\}^*$.

Soient $A = \{z \in Z / g(z) = a\}$

$B = \{z \in Z / g(z) = b\}$

$C = \{z \in Z / g(z) = c\}$

L' est encore égal à $g(h^{-1}(L) \cap R')$ où $R' = R \cap C^* A^* B^* C^* \cap h^{-1}(a^* b^*)$ puisque h est strictement alphabétique. R' est donc une union finie de langages de la forme $C_i A_i B_i C'_i$ où A_i, B_i, C_i, C'_i sont des langages rationnels et $A_i \subseteq A^*$ $B_i \subseteq B^*$, $C_i \in C^*$ et $C'_i \subseteq C^*$. Puisque tout langage rationnel est une union finie de langages rationnels dont l'image, par la fonction de Parikh, est un ensemble linéaire on peut choisir directement les C_i et C'_i tels que $\Psi(C_i)$ et $\Psi(C'_i)$ soient des ensembles linéaires.

$$R' = \bigcup_{i=1}^{n} C_i A_i B_i C'_i$$

Numérotons les C_i et C'_i de telle manière que les C_i infinis aient un numéro de 1 à m, les C'_i infinis aient un numéro de 1 à m'.

Soit $m_1 = \inf(m, m')$

$$L' = g(h^{-1}(L) \cap (\bigcup_{i=1}^{n} C_i\, A_i\, B_i\, C'_i))$$
$$= \bigcup_{i=1}^{n} g(h^{-1}(L) \cap C_i\, A_i\, B_i\, C'_i)$$

Si h_c est l'homomorphisme qui efface les c, bien évidemment,

$h_c(\bigcup_{i=1}^{m_1} (g(h(L) \cap C_i\, A_i\, B_i\, C'_i))$ est contenu dans $h_c(L')$ donc dans L.

Montrons que tout mot de L appartient à $h_c(\bigcup_{i=1}^{m_1} (g(h^{-1}(L) \cap C_i\, A_i\, B_i\, C'_i))$.

Soit b_1 la longueur du mot le plus long de $\bigcup_{i=m+1}^{n} C_i$ et soit b_2 la longueur du mot le plus long de $\bigcup_{i=m'+1}^{n} C'_i$ le mot $w' = c^{b_1+1}\, w\, c^{b_2+1} \in$ L' $\forall\, w \in L$. De la définition de b_1, b_2 et m_1, w' ne peut pas être l'image par g d'un mot de $C_i\, A_i\, B_i\, C'_i$ pour $i > m_1$. Il existe donc $i \leq m_1$ tel que $w' \in g(h^{-1}(L) \cap C_i\, A_i\, B_i\, C'_i)$ et $w = h_c(w')$ appartient à $h_c(\bigcup_{i=1}^{m_1} g(h^{-1}(L) \cap C_i\, A_i\, B_i\, C'_i)$ d'où on déduit que

$$L = \bigcup_{i=1}^{m_1} (h_c \circ g(h^{-1}(L) \cap C_i\, A_i\, B_i\, C'_i)$$

où C_i, C'_i sont des ensembles infinis et $\Psi(C_i)$, $\Psi(C'_i)$ des ensembles linéaires et ce pour tout $i \in [1, m_1]$

Mais $h_c \circ g(h^{-1}(L) \cap C_i\, A_i\, B_i\, C'_i$

$$= \underbrace{g(h^{-1}((h(C_i))^{-1}\ \underbrace{(L(h(C'_i))^{-1}}_{\text{quotient à droite par } h(C'_i)}}_{\text{quotient à gauche par } h(C_i)})) \cap A_i\, B_i).$$

Pour conclure que L est rationnel, il suffit de montrer que $(h(C_i))^{-1}$ $(L(h(C'_i))^{-1})$ est rationnel quel que soit i.

1) <u>Si $h(C_i) \subseteq a^*$ et $h(C'_i) \subseteq b^*$</u>

Il existe un entier t_i strictement positif vérifiant $h(C_i) = a^{r_i}(a^{t_i})^*$ car C_i est infini, que l'image de C_i par la fonction de Parikh $\Psi(C_i)$ est un ensemble linéaire, et enfin h est un homomorphisme strictement alphabétique.

De la même façon, on montre qu'il existe un entier t'_i strictement positif vérifiant $h(C'_i) = h(C'_i)\,(b^{t'_i})^*$

2) Si $h(C_i) \subseteq a^*$ et $h(C'_i) \subseteq a^* b^*$

(Le cas $h(C_i) \subseteq a^* b^*$ et $h(C'_i) \subseteq b^*$ se démontre de façon analogue).

Alors, comme précédemment $h(C_i) = h(C_i)\ (a^{t_i})^*$ avec $t_i > 0$. $h(C'_i)$ est de la forme $h(C'_i) = a^{r_i}(a^{\theta_i})^*\ b^{r'_i}(b^{t'_i})^*$ car $\Psi(C'_i)$ ensemble linéaire implique, puisque h est strictement alphabétique, que $\Psi(h(C'_i))$ est linéaire.

Mais $h(C'_i)$ est infini donc soit θ_i, soit t'_i est strictement positif .

- Supposons que $t'_i = 0$. Alors

$$L(h(C'_i))^{-1} \subseteq a^* \text{ et } ((a^{t_i})^*)^{-1}\ (L(h(C'_i))^{-1}) \text{ est un langage rationnel}$$

[13] et $(h(C_i))^{-1}(L(h(C'_i))^{-1})$ est rationnel

- Supposons $t'_i > 0$, et là encore $h(C'_i)$ peut se mettre sous la forme $h(C'_i) = h(C'_i)\ (b^{t'_i})^*$. Dans ce cas $h(C_i) = h(C_i)\ (a^{t_i})^*$ avec $t_i > 0$ et $h(C'_i) = h(C'_i)\ (b^{t'_i})^*$ et on s'est ramené au cas 1).

Et alors dans ces deux cas :

$$(h(C_i))^{-1}(L(h(C'_i))^{-1}) = (h(C_i))^{-1}(((a^{t_i})^*)^{-1}\ ((L((b^{t'_i})^*)^{-1})(h(C'_i))^{-1}))$$

Or, d'après [21], puisque $L \subseteq a^* b^*$, le langage $((a^{t_i})^*)^{-1}(L((b^{t'_i})^*)^{-1})$ est un langage rationnel d'où on déduit naturellement que le langage $(h(C_i))^{-1}$ $(L(h(C'_i))^{-1})$ l'est aussi.

Réciproquement, si L est un langage rationnel infini, il est bifidèle (corollaire III-4) donc d'image finie. □

Corollaire III-7 : Un langage $L \subseteq a^*$ *est d'image finie si et seulement si* L *est un langage rationnel infini.*

Conjecture : Tout langage borné L est d'image finie si et seulement si L est un langage rationnel infini.

Corollaire III-18 : Le langage $C_1 = \{a^n b^n \ / \ n \geq 0\}$ *n'est pas d'image finie.*

Par contre le langage C_1 vérifie un certain nombre de propriétés intéressantes que nous utiliserons dans le chapitre V.

Posons $C'_1 = C_1 \setminus \varepsilon$ alors :

Lemme III-19 : Le langage C'_1 *vérifie :*

i) $C^d(C'_1) = C^{if}(C'_1)$

ii) $C^{df}(C'_1) = C^{bf}(C'_1)$

Démonstration : Nous verrons, dans le chapitre V , que, pour un langage quelconque L, l'égalité $C^{df}(L) = C^{bf}(L)$ implique $C^d(L) = C^{if}(L)$, il suffit donc de montrer ii).

Soit $L \subseteq X^*$ un langage de $C^{bf}(C'_1)$. D'après le théorème de Boasson-Nivat, il existe une transduction rationnelle d'image finie $\tau = g \circ (\cap R) \circ h^{-1}$ où R est un langage rationnel de Z^*, h un homomorphisme alphabétique k-limité sur R et g un homomorphisme alphabétique k'-limité sur R telle que $L = \tau(C'_1)$.

Nous allons construire à partir de τ, une transduction rationnelle décroissante τ' vérifiant, naturellement $L = \tau'(C'_1)$

1) Construction de $\tau' = g' \circ (\cap R') \circ h'^{-1}$

Soit $Z_o = \{z \in Z \ / \ h(z) = \varepsilon\}$, $Z_1 = Z \setminus Z_o$. Posons $\Sigma = Z_o \cup \{d\}$ où d est une nouvelle lettre, $\bar{\Sigma} = \{\bar{x} \ / \ x \in \Sigma\}$ et $\tilde{\Sigma} = \{\tilde{x} \ / \ x \in \Sigma\}$.

Soit θ_e l'homomorphisme de $(Z_1 \cup \Sigma \cup \bar{\Sigma} \cup \tilde{\Sigma})^*$ dans $(\Sigma \cup \Sigma_1)^*$ qui efface les "marques" c'est à dire ;

$$\theta(x) = \begin{cases} x \text{ si } x \in \Sigma \cup Z_1 \\ a \text{ si } x = \tilde{a} \\ a \text{ si } x = \bar{a} \end{cases}$$

soit h_d l'homomorphisme de $(\Sigma \cup Z_1)^*$ dans $(Z_o \cup Z_1)^*$ qui efface d c'est à dire :

$h_d(x) = x \quad \forall x \neq d$

$h_d(d) = \varepsilon$

Alors $R' = \theta^{-1}(h_d^{-1}(R)) \cap \Sigma^k(Z_1\ \Sigma^k)^*\ \bar{\Sigma}^k(Z_1\ \tilde{\Sigma}^k)^*$

h' : $\forall x \in \Sigma$, $h(x) = a$

$\forall x \in \bar{\Sigma} \cup \tilde{\Sigma}$, $h(x) = b$

g' ; $g'(d) = \varepsilon$

$g'(x) = g(x) \quad \forall x \neq d$

La transduction τ' est décroissante fidèle puisque h' est strictement alphabétique et que g' est k''-limité sur R' où $k'' = \sup(k, k')$.

2) <u>Validité de la construction</u>

Il faut montrer que $L = \tau'(C'_1)$, sachant que $L = g(h^{-1}(C'_1) \cap R)$

* Soit w un mot de $g'(h'^{-1}(C'_1) \cap R')$. Il existe dans R' un mot w' pouvant être mis sous la forme $w' = w_o\ z_1\ w_1 \ldots z_n\ w_n\ \bar{u}_o\ z'_1\ \tilde{u}_1\ z'_2 \ldots z'_m\ \tilde{u}_m$ avec $|w_i| = |\bar{u}_o| = |\tilde{u}_i|$ $\forall i \in [0, n]$, $\forall j \in [1, m]$ et z_i, $z'_j \in Z_1$, $\forall i \in [1, n]$ et $\forall j \in [1, m]$. Mais $h_d(\theta(w'))$ appartient à R, $h'(w')$ appartient à C'_1 et $w = g'(w')$.

$h'(w') = a^k\ h(z_1)\ a^k\ h(z_2) \ldots h(z_n)\ a^k\ b^k\ h(z'_1)\ b^k\ h(z'_2) \ldots h(z'_m)\ b^k$ est élément de C'_1 ce qui implique que

$\forall i \in [1, n]$ $h(z_i) = a$, $\forall j \in [1, m]$ $h(z'_j) = b$ donc que $n = m$, Mais $h(h_d(\theta(w_i))) = h(h_d(\theta(\bar{u}_o))) = h(h_d(\theta(\tilde{u}_j)) = \varepsilon\ \forall\ i \in [0, n],\ \forall\ j \in [1, n]$ et $h(h_d(\theta(z))) = h(z)\ \forall\ z \in \{z_1,\dots,z_n, z'_1, \dots, z'_n\}$ ce qui permet d'écrire que $h(h_d(\theta(w')) = h(z_1 \dots z_n z'_1 \dots z'_n) \in C'_1$ et $g(h_d(\theta(w'))) = g'(w') = w$.

* Réciproquement, soit w un mot de $L = g(h^{-1}(C'_1) \cap R)$. On peut trouver dans R un mot w' vérifiant $g(w') = w$ et $h(w') \in C'_1$. w' appartient à R, il peut se mettre sous la forme $w' = w_o z_1 w_1 z_2 \dots z_n w_n$ avec $\forall\ i \in [1, n]\ z_i \in Z_1$ et $\forall\ i \in [0, n]\ w_i \in Z_o^*$. Mais, puisque h est k-limité, la longueur $k_i = |w_i|$ du mot w_i est inférieure à k. De l'appartenance de $h(w')$ à C_1, du fait que h est strictement alphabétique et que $h(w') = h(z_1 \dots z_n)$, on déduit que n est un nombre pair. Alors, en posant $p = n / 2$,

$w'' = w_o\, d^{k-k_o} z_1\, w_1\, d^{k-k_1} \dots z_p\, w_p\, d^{k-k_p}\, \tilde{d}^k\, z_{p+1}\, \tilde{w}_{p+1}\, \tilde{d}^{k-k_{p+1}} \dots z_n\, \tilde{w}_n\, d^{k-k_n}$

est élément de R'

$h'(w'') = a^k(a\, a^k)^p\, b^k(b\, b^k)^p$ appartient à C'_1.

Les mots w' et w'' ne diffèrent que par le nombre d'occurrences de la lettre d et par les marques '-' et '∿', donc $g'(w'') = g(w') = w$. En d'autres termes $w \in g'(h'^{-1}(C'_1) \cap R')$ □

<u>Remarque</u>

Nous avons vu, en début de chapitre, que pour tout langage algébrique L, on a l'inclusion $Rat \subseteq C^{df}(L)$. En particulier la famille Rat est contenue dans $C^{df}(C_1)$.

Pour montrer l'égalité entre $C^{bf}(C'_1)$ et $C^{if}(C'_1)$ nous avons besoin du résultat :

<u>Lemme III-20</u> *: Pour tout entier j, $C'_1((b^j)^*)^{-1}$, le quotient à droite de C'_1 par $(b^j)^*$ et $((a^j)^*)^{-1} C'_1$, le quotient à gauche de C'_1 par $(a^j)^*$ appartient à $C^{bf}(C'_1)$*

<u>Démonstration</u> : Les démonstrations concernant $C_1((b^j)^*)^{-1}$ d'une part et $((a^j)^*)^{-1}\ C'_1$ d'autre part sont analogues. Nous ne démontrerons que l'appartenance de $C'_1((b^j)^*)^{-1}$ à $C^{bf}(C'_1)$.

Si $j = 0$ alors $C'_1\ ((b^j)^*)^{-1} = C'_1 \in C^{bf}(C'_1)$

Si $j > 0$ alors $C'_1((b^j)^*)^{-1} = (C'_1(b^*)^{-1}) \cap (\bigcup_{i=1}^{j-1} a^i(a^j)^*\ (b^j)^*\ b^i)$

On en déduit que $C'_1((b^j)^*)^{-1} \in C^{bf}(C'_1(b^*)^{-1})$ appartient donc à $C^{bf}(C'_1)$.

Il suffit donc de montrer que $C'_1(b^*)^{-1}$ appartient à $C^{bf}(C'_1)$.

Soit $\tau = g \circ (\cap R) \circ h^{-1}$ la transduction bifidèle déterminée par ;

$R = a'\ a^*\ c^*\ b^* \cup a^*\ c^*\ b^*$

h ; $h(a') = h(a) = a$, $h(c) = h(b) = b$

g:$g(a) = g(c) = a$, $g(b) = b$, $g(a') = \varepsilon$

Montrons que $C'_1(b^*)^{-1} = \tau(C'_1)$:

- Soit w un mot de $C'_1(b^*)^{-1}$, il existe deux entiers n et k où $k \leq n$ tels que $w = a^n\ b^k$.

* si $n+k$ est pair, posons $n + k = 2m$

$w' = a^m\ c^{n-m}\ b^k$ est un mot de R, $g(w') = a^m\ a^{n-m}\ b^k = w$ et $h(w') = a^m\ b^{n-m+k} = a^m\ b^m$ est élément de C'_1.

* Si $n+k$ est impair, posons $n+k = 2m+1$

$w' = a'a^m\ c^{n-m}\ b^k$ est un mot de R, $g(w') = a^m\ a^{n-m}\ b^k = w$ et

$h(w') = a^{m+1}\ c^{n+k-m} = a^{m+1}\ b^{m+1}$ est élément de C'_1 donc $w \in g(h^{-1}(C'_1) \cap R)$.

- Soit w un mot de $g(h^{-1}(C'_1) \cap R)$, il existe w' un mot de R vérifiant $w = g(w')$ et $h(w') \in C'_1$.

Deux possibilités peuvent se présentées.

* $w' = a'\ a^n\ c^p\ b^k$; $h(w') \in C'_1$ implique $p+k = n+1$ alors

$g(w') = a^{n+p}\ b^k \in C'_1(b^*)^{-1}$ puisque $n+q \geq k$

* $w' = a^n\ c^p\ b^k$, $h(w') \in C'_1$ implique que $n = p+k$ alors

$g(w') = a^{n+p}\ b^k \in C'_1(b^*)^{-1}$ puisque $n+p \geq k$. $\square$

Ce lemme va nous permettre de montrer que

Lemme III-21 : $C^{bf}(C'_1) = C^{if}(C'_1)$.

Démonstration : Il suffit de montrer que $C^{if}(C'_1)$ est contenu dans $C^{bf}(C'_1)$. Soit $L \subseteq X^*$ un langage de $C^{if}(C'_1)$. Du lemme III-19 on déduit que L appartient à $C^d(C'_1)$, il existe donc une transduction décroissante $\tau = g \circ (\cap R) \circ h^{-1}$ vérifiant $\tau(C'_1) = L$ où, en utilisant le théorème II-3, R est un langage rationnel contenu dans Z^*, g est un homomorphisme alphabétique et h un homomorphisme strictement alphabétique.

Soient $A = \{z \in Z \ / \ h(z) = a\}$

$B = \{z \in Z \ / \ h(z) = b\}$

Alors, car h est strictement alphabétique, L est une union finie de langages de la forme $g(h^{-1}(C_1) \cap A_1 B_1)$ où A_1 est un langage rationnel inclus dans A et B_1 un langage rationnel inclus dans B. En fait, on peut, sans nuire à la généralité de la démonstration, supposer que $L = g(h^{-1}(C_1) \cap A_1 B_1)$.

Si on appelle M_1 l'automate d'états fini qui reconnaît A_1 et dont l'ensemble des états est Q_1, si on appelle M_2 l'automate d'états fini qui reconnaît B_1 et dont l'ensemble des états est Q_2, alors la construction faite par M. Latteux dans [21] peut ici être réutilisée ce qui nous conduit à dire que L appartient à $C^{bf}(C''_1)$ avec :

$$C''_1 = \bigcup_{\substack{X_1 \subset Q_1 \\ X_2 \subset Q_2}} (h(Z_{X_1}))^{-1} \, (C'_1((h(Z_{X_2}))^{-1})$$

où Z_{X_i} représente l'ensemble des mots de Z^*, d'image vide par g et qui conduisent d'un état quelconque de X_i vers l'état d'entrée de M_i.

Or $h(Z_{X_1})$ est contenu dans a^*, $h(Z_{X_2})$ dans b^*. Ce sont tous deux des langages rationnels, $h(Z_{X_1})$ est donc une union finie de langages de la forme $a^s(a^t)^*$ et $h(Z_{X_2})$ une union finie de langages de la forme $b^{s'}(a^{t'})^*$. Comme h est strictement alphabétique, au moins un t et un t' sont strictement positifs.

Pour simplifier la démonstration, nous supposerons que $h(Z_{X_1}) = a^s(a^t)^*$ et $h(Z_{X_2}) = b^{s'}(b^{t'})^*$.

Il reste à montrer que l'on peut passer de C'_1 à C''_1 par une transduction rationnelle bifidèle.

Eliminons le cas trivial où $t = t' = 0$, deux cas peuvent alors se présenter :

1) t et t' sont tous deux non nuls. Alors C''_1 est un langage rationnel infini [21] sur $\{a, b\}$ il appartient donc à $\mathcal{C}^{bf}(C'_1)$.

2) $t = 0$ (ou ce qui est équivalent $t' = 0$)
$C''_1 = (a^s)^{-1}\ (C'_1(b^s(b^t)^*)^{-1})$ et dans ce cas, c'est le lemme III-20 qui permet de conclure que $C''_1 \in \mathcal{C}^{bf}(C'_1)$.□

De ces lemmes nous pouvons déduire, puisque $\mathcal{C}^{df}(C'_1) = \mathcal{C}^{df}(C_1)$

<u>*Proposition III-22*</u> *:* $C_1 = \{a^n\ b^n\ /\ n \geq 0\}$ *vérifie :*
$\text{Rat} \subsetneq \mathcal{C}^{df}(C_1) = \mathcal{C}^{bf}(C_1) = \mathcal{C}^{d}(C_1) = \mathcal{C}^{if}(C_1) \subsetneq \mathcal{C}^{f}(C_1) = \mathcal{C}(C_1)$

Pour terminer ce chapitre nous allons montrer qu'il existe des langages dont le cône d'image finie contient strictement le cône décroissant.

<u>*Proposition III-23*</u> *:* $L = \{a^{n!} \ / \ n \geq 1\}$ ***vérifie***

$$C^d(L) \subsetneq C^{if}(L) \subsetneq C(L)$$

et Rat *n'est pas contenue dans* $C^{df}(L)$.

<u>Démonstration</u> : D'après le corollaire III-17, L n'est pas un langage d'image finie. Il est clair que $L' = \{(a\ b)^{n!} \ / \ n \geq 1\}$ appartient à $C^{if}(L)$ Supposons qu'il appartienne à $C^d(L)$. D'après la caractérisation des transductions décroissantes, il existe un langage rationnel $R \subseteq Z^*$ un homomorphisme h de Z^* dans $\{a\}^*$ strictement alphabétique, un homomorphisme g de Z^* dans $\{a, b\}^*$ alphabétique tels que :

$$L' = g(h^{-1}(L) \cap R)$$

Soit k le nombre d'états de l'automate qui reconnait R, et soit $*$ sa fonction de transition. Le mot $(a\ b)^{k!}$ appartient à L', il existe donc un mot w de R vérifiant $(a\ b)^{k!} = g(w)$ et $h(w) = L$. L'homomorphisme g étant strictement alphabétique $|w| \geq |g(w)| > k$. Il existe dans w un facteur itérant d'image non vide par g car, si ce n'était pas le cas, alors :

Soit $W = \{w \in R \ / \ w = w_1 \ldots w_x, \ q_i = q_{i-1} * w_i$ pour $i = 1, \ldots, n$
et $\forall\, i \neq j \quad q_i \neq q_j\}$

$\exists\, w' \in W$ tel que $|g(w')| = |g(w)| \leq |w'| < k$ ce qui est impossible.

w peut donc se mettre sous la forme

$w = u_1\ v\ u_2$ avec $1 \leq |v| \leq k$, $u_1\ v^*\ u_2 \in R$ et $|g(v)| \neq 0$

h étant strictement alphabétique $|w| = |h(w)|$

et il existe $p > k$ tel que $h(w) = a^{p!}$.

Pour tout $n > p$, la quantité

$q = 1 + \frac{p!}{|v|}$ $(n(n-1) \ldots (p+1) - 1)$ est entière puisque $|v| \leq k < p$,

le mot $w_q = u_1 x^q u_2$ est dans R et $h(w_q) = a^{n!}$ est dans L. $g(w)$ est, par conséquent, un élément de L', il existe donc $m \in \mathbb{N}$, $m > k$ tel que $g(w_q) = (ab)^{m!}$ et, puisque l'on a choisi une transduction décroissante $2m! \leq n!$ ou encore $m < n$.

$$\begin{aligned} |g(w_q)| &= |g(w)| + |v| \, (q-1) \\ &= 2k! + n! \frac{|g(v)|}{|v|} \\ &= \underbrace{2k!(1 + n(n-1) \ldots (k+1) \frac{|g(v)|}{|v|})}_{\substack{\text{quantité non divisible par un} \\ \text{nombre de l'intervalle } [k,\, n]}} \end{aligned}$$

Il n'existe donc pas d'entier m tel que $g(w_q) = 2m!$. $g(w')$ ne peut donc pas appartenir à L'.

D'autre part, $\forall$ k, il existe $w \in L$ tel que pour tout $w' \in L$, $|w'| > |w|$ implique $|w'| \geq k(|w| + 1)$ il suffit de choisir $w = a^{(k+1)!}$.

Donc, le lemme III-6 permet de conclure que la famille Rat n'est pas contenue dans $C^{df}(L)$. □

CHAPITRE IV

CONES RATIONNELS ET SUBSTITUTION

A : LEMMES PRÉLIMINAIRES

B : CÔNES DÉCROISSANTS ET CÔNES DÉCROISSANTS FIDÈLES

C : RELATIONS ENTRE $C^{df}(L)$, $C^{d}(L)$, $C^{f}(L)$ ET $C(L)$

D : UNE CARACTÉRISATION DES LANGAGES COMPLETS PAR SUBSTITUTION

E : LANGAGES TRÈS FORTEMENT EFFAÇABLES

F : UNE PROPRIÉTÉ DE LA FAMILLE DES LANGAGE QUASI-RATIONNELS

L'opérateur syntaxique, noté $\uparrow$, introduit par S. Greibach, a permis d'étudier plus commodément l'opérateur de substitution et ses conséquences quant aux familles de langages construites à partir de cet opérateur d'une part et des familles particulières mais fondamentales que sont les cônes rationnels d'autre part. On sait, en effet, que pour tout couple L et L' de langages, définis sur des alphabets disjoints, pour toute nouvelle lettre c, on a la relation :

(*) $C(L) \square C(L') = C(L \uparrow (L'c)^*)$ [17]

Il est facile de voir que l'inclusion $C(L \uparrow L') \subseteq C(L) \square C(L')$ est toujours vérifiée. Se pose alors la question de savoir quelles propriétés doivent posséder L (et-ou) L' pour que l'égalité soit vraie. Il est possible de répondre partiellement à cette question puisqu'il est connu que, si L' est un langage générateur d'une FAL, on a la relation :

$\forall L : C(L) \square C(L') = C(L \uparrow L')$.

Symétriquement, quelles conditions doit remplir le langage L pour avoir :

$\forall L', C(L) \square C(L') = C(L \uparrow L')$.

Cette relation, non vérifiée dans le cas général, a été montrée pour certains langages tels que D'^*_1 ou Sym [6] et [3]. Les démonstrations mises en oeuvre sont relativement difficiles et assez différents les unes des autres. Trouver une caractérisation de tels langages, que nous appelerons langages complets par substitution, paraît utile et permettrait de simplifier certaines démonstrations.

Nous allons, dans ce chapitre, montrer que la classe des langages d'image finie est identique à la classe des langages complets par substitution. Pour arriver à ce résultat, nous allons essayer de "découper", en terme de substitution, $C(L)$, le cône rationnel engendré par L et ce, indépendemment des propriétés du langage L (que nous supposerons non vide). Ce qui nous conduit à montrer que :

$$(**) \quad \forall L : C(L) = C^d(L) \,\square\, \mathrm{Rat}.$$

Relation qui est à rapprocher des résultats du chapitre II puisque nous avions alors montré que toute transduction rationnelle pouvait être décomposée en une transduction rationnelle décroissante suivie d'une substitution rationnelle. La caractérisation des transductions rationnelles décroissantes de ce même chapitre va faciliter grandement l'étude de $C^d(L)$ et, plus précisément, l'étude des substitutions d'un cône rationnel principal décroissant (fidèle) dans un cône rationnel principal décroissant (fidèle). Le résultat de base de cette section est :

$$C^d(L \uparrow L') = C^d(L) \,\square\, C^d(\# L')$$

où # est une nouvelle lettre.

Notons, enfin, que cette relation et la relation (**) permettent de retrouver très facilement la relation (*).

Nous terminerons ce chapitre en introduisant la notion de langage très fortement effaçable :

L est très fortement effaçable si et seulement si $C^d(L) = C^{df}(L)$

Nous possédons, à l'heure actuelle, peu d'exemples de langages algébriques non effaçables. S. Greibach et L. Boasson ont montré l'existence de tels langages dans [18] et [7]. La notion de langage très fortement effaçable permet de déterminer toute une classe de langages algébriques non effaçables.

Rappelons les résultats (les démonstrations de ces résultats sont dans le livre de J. Berstel [6]) :

Soit $\mathcal{L}$ une famille de langages,on appelle X_L l'alphabet sur lequel est défini L. On note $\# \mathcal{L} = \{c\, L \ / \ L \in \mathcal{L},\ c \notin X_L\}$.

Pour tout couple $(\mathcal{L}, \mathcal{L}')$ de familles de langages on a les relations :

1) $\mathcal{C}(\mathcal{L}') \square \mathcal{C}(\mathcal{L}) = \mathcal{C}(\mathcal{L}') \square \mathcal{F}(\mathcal{L})$

2) $\mathcal{L}' \square \mathcal{C}(\mathcal{L}) \subseteq \mathcal{C}(\mathcal{L}' \square \# \mathcal{L})$

3) $\mathcal{C}(\mathcal{L}' \square \mathcal{L}) \subseteq \mathcal{C}(\mathcal{L}') \square \mathcal{C}(\mathcal{L})$

4) $\mathcal{C}(\mathcal{L}') \square \mathcal{C}(\mathcal{L}) = \mathcal{C}(\# \mathcal{L}' \square \mathcal{F}(\mathcal{L}))$

5) $\mathcal{F}(\mathcal{L}) = \mathrm{Rat} \square \mathcal{C}(\mathcal{L})$

6) $\forall L \quad \mathcal{F}(L) = \mathcal{C}((Lc)^*)$

Nous avons alors un lemme fort utile faisant intervenir l'opérateur syntaxique dont nous reproduisons la démonstration.

Lemme IV-1 : Pour tout couple de langages définis sur des alphabets disjoints $\mathcal{C}(L \uparrow L') = \mathcal{C}(\{L\} \square (\# \{L'\}))$.

Démonstration : Supposons que $L \subseteq X^*$ et $L' \subseteq X'^*$, alors $L \uparrow L' = s(L)$ où $\forall x \quad s(x) = x\, L'$, s est donc une $\#\{L'\}$-substitution c'est à dire $s(L) \in \{L\} \square (\# \{L'\})$ d'où l'inclusion $\mathcal{C}(L \uparrow L') \subseteq \mathcal{C}(\{L\} \square (\# \{L'\}))$. Choisissons, maintenant, L" dans $\{L\} \square (\# \{L'\})$ et défini sur l'alphabet X". L" peut se mettre sous la forme L" = s(L) où s est une $\# \{L'\}$-substitution.

Pour toute lettre x de X, $s(x) = c_x L'$ avec $c_x \in X'' \setminus X'$. X" est donc égal à $\{c_x \mid x \in X\} \cup X'$. Soit h, l'homomorphisme de X''^* dans $(X' \cup X)^*$ défini par :

$$\forall x \in X' : h(x) = x$$
$$h(c_x) = x$$

Montrons que $h(L'') = L \uparrow L'$:

Soit w, un mot de h(L"). Il existe w' appartenant à L" vérifiant $w = h(w')$. Etant donnée la forme de L", on peut trouver un mot w" dans L tel que $w' \in s(w'')$. Posons $w'' = x_1 \ldots x_n$, alors $s(w'') = c_{x_1} L' c_{x_2} L' \ldots c_{x_n} L'$. $w = h(w') \in h(s(w''))$ appartient à $x_1 L' \ldots x_n L'$ qui est contenu dans $L \uparrow L'$ puisque $x_1 \ldots x_n \in L$.

Réciproquement, soit $w \in L \uparrow L'$ avec $w = x_1 w_1 \ldots x_n w_n$ où $x_1 \ldots x_n \in L$ et $\forall i\ w_i \in L'$. Le mot $w' = c_{x_1} w_1 \ldots c_{x_n} w_n$ appartient à $s(x_1 \ldots x_n) \subset s(L)$ et w est égal à h(w') d'où on déduit naturellement que $w \in h(s\ \ (L)) \subset h(L'')$.

De l'égalité $h(L'') = L \uparrow L'$ on déduit que $L'' \subset \mathcal{C}(L \uparrow L')$.□

Nous pouvons alors rappeler la démonstration de la relation (*) du début de ce chapitre.

__Théorème IV-2__ : $\mathcal{C}(L) \,\square\, \mathcal{C}(L') = \mathcal{C}(L \uparrow (L'c)^*)$.

<u>Démonstration</u> :

$$\begin{aligned}
\mathcal{C}(L \uparrow (L'c)^*) &= \mathcal{C}(\ \{L\} \,\square\, (\# \{(Lc)^*\}) && \text{(lemme IV-1)}\\
&\subseteq \mathcal{C}(\ \{L\}) \square\, \mathcal{C}(\# \{L'c\}) && \text{(relation 3))}\\
&= \mathcal{C}(L) \quad \square\, \mathcal{F}(L') && \text{(relation 6))}\\
&= \mathcal{C}(L) \,\square\, \mathcal{C}(L') && \text{(relation 1))}
\end{aligned}$$

Réciproquement :

$$C(L) \,\square\, C(L') = C((\#\{L\}) \,\square\, F(L')) \qquad \text{(relation 4))}$$

$$= C((\#\{L\}) \,\square\, C((L'c)^*)) \qquad \text{(relation 6))}$$

$$\subseteq C(C((\#\{L\}) \,\square\, (\#\{(L'c)^*\})) \qquad \text{(relation 2))}$$

$$= C((\#\{L\}) \,\square\, (\#\{(L'c)^*\})$$

$$= C(\{L\} \,\square\, (\#\{(L'c)^*\}))$$

$$= C(L \uparrow (L'c)^*) \qquad \text{(lemme IV-1)}$$

Afin de déterminer les langages vérifiant (**) nous allons, à l'aide de la caractérisation des transductions rationnelles décroissantes du chapitre II, montrer que $C^d(L) \,\square\, C^d(\# \; L') = C^d(L \uparrow L')$.

Notons que dans la démonstration du lemme IV-1 la transduction τ construite est bifidèle. La relation reste donc vraie en remplaçant C par C^d ou C^{df}. Malheureusement les autres relations (1) → 5)) ne restent pas toutes vérifiées, c'est la raison pour laquelle un certain nombre de propriétés vont devoir être démontrées.

A : LEMMES PRÉLIMINAIRES

Lemme IV-4 : Soient A *et* A' *deux langages définis sur l'alphabet* X *et* B *et* B' *deux langages définis sur l'alphabet* Y*. Si* $X \cap Y = \emptyset$ *et* $A \cap A' \neq \emptyset$ *on a la relation :*

$$(A \cap A') \uparrow (B \cap B') = (A \uparrow B) \cap (A' \uparrow B')$$

Démonstration : Soit w un mot de $(A \cap A') \uparrow (B \cap B')$, w peut s'écrire :
$w = x_1 w_1 x_2 w_2 \ldots x_n w_n$ avec $x_1 \ldots x_n \in A \cap A'$ et $\forall$ i $x_i \in X$, $w_i \in B \cap B'$.
On déduit naturellement que $w \in A \uparrow B$ et à $A' \uparrow B'$.

Réciproquement, soit w un mot de $(A \uparrow B) \cap (A' \uparrow B')$.
$w \in A \uparrow B \Rightarrow w = x_1 w_1 \ldots x_n w_n$ et $x_1 \ldots x_n \in A$
$\forall$ i $x_i \in X$ et $w_i \in B \subset Y^*$
$w \in A' \uparrow B' \Rightarrow w = y_1 u_1 \ldots y_p u_p$ et $y_1 \ldots y_p \in A'$
et $\forall$ j $y_j \in X$ et $u_j \in B' \subset Y^*$
Comme les alphabets sont disjoints, on montre par récurrence sur k que
$\forall$ k $y_k = x_k$ et $w_k = u_k$ d'où on déduit que n = p et surtout que $x_1 \ldots x_n = y_1 \ldots y_p \in A \cap A'$ et $\forall$ k, $w_k = u_k \in B \cap B'$.□

Si h est un homomorphisme strictement alphabétique, il est facile de voir que pour tous langages L et L' définis sur des alphabets dijoints on a $h(L \uparrow L') = h(L) \uparrow h(L')$ nous allons montrer que :

Lemme IV-5 : Soit h *un homomorphisme strictement alphabétique de* X^* *dans* Y^*.
Pour tout couple de langages (L_1, L_2) *définis sur* Y^* *on a la relation :*

$$h^{-1}(L_1) \uparrow h^{-1}(L_2) = h^{-1}(L_1 \uparrow L_2)$$

Démonstration : Soit w un mot de $h^{-1}(L_1 \uparrow L_2)$, $w' = h(w)$ est donc élément de $L_1 \uparrow L_2$. w' peut s'écrire $w' = u_1 u'_1 \dots u_n u'_n$ avec $u_1 \dots u_n \in L_1$ et pour tout $i \in [1, n]$, $u_i \in X$ et $u'_i \in L_2$. Le mot w appartient à $h^{-1}(w')$ donc, puisque h est alphabétique, à $h^{-1}(u_1)\, h^{-1}(u'_1) \dots h^{-1}(u_n)\, h^{-1}(u'_n)$, de plus chaque $h^{-1}(u_i)$ est une lettre puisque h est strict, w est donc élément de $h^{-1}(L_1) \uparrow h^{-1}(L_2)$. Réciproquement, choisissons w appartenant à $h^{-1}(L_1) \uparrow h^{-1}(L_2)$. Le mot w peut se mettre sous la forme $w = u_1 u'_1 \dots u_n u'_n$ avec $u_1 \dots u_n \in h^{-1}(L_1)$ et pour tout i $u_i \in X$ et $u'_i \in h^{-1}(L_2)$. Mais $h(w)$ est égal à $h(u_1 u'_1 \dots u_n u'_n)$ et $h(u_1 \dots u_n) \in L_1$; or, h étant strictement alphabétique, quel que soit i $h(u_i) \in Y$ et $h(u'_i) \in L_2$ d'où on déduit que $h(w) = h(u_1)\, h(u'_1) \dots h(u_n)\, h(u'_n)$ est élément de $L_1 \uparrow L_2$ et w appartient à $h^{-1}(L_1 \uparrow L_2)$.$\square$

Lemme IV-6 : Soient $L_1 \subseteq X_1^*$ *et* $L_2 \subseteq X_2^*$ *deux langages avec* $X_1 \cap X_2 = \emptyset$ *et* # *une nouvelle lettre. Pour tout langage rationnel* R *on a*

$$(L_1 \uparrow L_2) \cap R \subseteq C^{df}(L_1) \,\square\, C^{df}(\# L_2).$$

Démonstration : Il s'agit d'une démonstration constructive. Remarquons que nous pouvons toujours nous ramener au cas où Z, l'alphabet sur lequel est défini R, est de la forme $Z = X_1 \cup X_2$.

1) Construction :

Soit $A = (Z, Q, q_o, *, F)$ l'automate d'états fini déterministe minimal qui reconnait R, où

- Q est l'ensemble des états
- q_o l'état initial
- $*$ la fonction de transition
- F l'ensemble des états terminaux.

Soit R' le langage rationnel défini sur l'alphabet $\Delta = Q \times X_1 \times Q$ par

$$R' = \{(q_o, a_1, q_1)(q_1, a_2, q_2) \ldots (q_{n-1}, a_n, q_n) \,/\, q_n \in F\}.$$

On définit l'homomorphisme ϕ de Δ^* dans Z^* par $\phi(q, a, q') = a$.

Alors, $L'_1 = \phi^{-1}(L_1) \cap R' \in C^{df}(L_1)$ puisque ϕ est strictement alphabétique.

Soit $R_{qq'} = \{w \in Z^* \,/\, q * w = q'\}$, nous pouvons alors définir la substitution s de Δ^* dans Z^* par $s((q, a, q')) = a\, L_2 \cap R_{qq'}$.

La substitution s est, bien évidemment, une $C^{df}(\# L_2)$-substitution.

Alors $(L_1 \uparrow L_2) \cap R = s(L'_1)$

2) Validité de la construction

* Soit w choisi dans $(L_1 \uparrow L_2) \cap R$. Le mot w appartient à R d'une part et peut se mettre sous la forme $w = x_1 w_1 \ldots x_n w_n$ d'autre part, avec $x_1 \ldots x_n$ dans L_1 et pour tout i, $w_i \in L_2$ et $x_i \in X_1$. Mais l'appartenance de w à R implique que $q_o * w \in F$. Posons $q_i = q_o * x_1 w_1 \ldots x_i w_i$, pour tout i de $[1, n]$. Ce qui peut encore s'écrire $q_i = q_{i-1} * x_i w_i$ et q_n est un état de sortie. Autrement dit $x_i w_i \in R_{q_{i-1} q_i}$ et, puisque $x_i \in X_1$ et $w_i \in L_2$, $x_i w_i$ appartient à $x_i L_2 \cap R_{q_{i-1}q_i} = s(q_{i-1}, x_i, q_i)$. Posons $w' = (q_o, x_1, q_1)(q_1, x_2, q_2) \ldots (q_{n-1}, x_n, q_n)$. Bien entendu $w \in s(w')$, mais $\Phi(w') = x_1 \ldots x_n \in L_1$ donc $w' \in \phi^{-1}(L_1) \cap R' = L'_1$ ou encore $w \in s(L'_1)$.

* Réciproquement, soit $w \in s(L'_1)$.

Il existe $w_1 \in L'_1$ tel que $w \in s(w_1)$

$w_1 \in \phi^{-1}(L_1) \cap R'$ si et seulement si $\phi(w_1) \in L_1$ et $w_1 \in R'$

$w_1 \in R'$, il est donc de la forme $w_1 = (q_o, x_1, q_1)(q_1, x_2, q_2) \ldots (q_{n-1}, x_n, q_n)$ avec $q_n \in F$ et $\phi(w_1) = x_1 \ldots x_n \in L_1$.

Donc $w \in s(w_1) = s(q_o, x_1, q_1)\, s(q_1, x_2, q_2) \ldots s(q_{n-1}, x_n, q_n)$

$$= (x_1 L_2 \cap R_{q_oq_1})(x_2 L_2 \cap R_{q_1q_2}) \ldots (x_n L_2 \cap R_{q_{n-1}q_n}) \text{ et}$$

$$w = x_1 w_1 x_2 w_2 \ldots x_n w_n$$

avec $\forall$ i, $w_i \in L_2$, ce qui peut s'écrire

$$q_o \xrightarrow{x_1\ w_1} q_1 \xrightarrow{x_2\ w_2} q_2 \cdots q_{n-1} \xrightarrow{x_n\ w_n} q_n \in F$$

c'est à dire $w \in R$ et $w \in L_1 \uparrow L_2$

$\Rightarrow w \in (L_1 \uparrow L_2) \cap R.\square$

De ce résultat, on déduit facilement que :

<u>Proposition IV-7</u> *: Si* L'_1 *appartient à* $C^d(L_1)$ *(resp.* $C^{df}(L_1)$*), si* L'_2 *appartient à* $C^d(L_2)$ *(resp.* $C^{df}(L_2)$*) alors, pour tout langage rationnel* R*, on a*

$$(L'_1 \uparrow L'_2) \cap R \in C^d(L_1) \square C^d(\# L_2)$$

(resp. $(L'_1 \uparrow L'_2) \cap R \in C^{df}(L_1) \square C^{df}(\# L_2)$*)*

où L'_1 *et* L'_2 *sont définis sur des alphabets disjoints.*

<u>Démonstration</u> : Si $L'_1 \in C^d(L_1)$ et $L'_2 \in C^d(L_2)$ d'après le lemme précédent $(L'_1 \uparrow L'_2) \cap R \in C^{df}(L'_1) \square C^{df}(\# L'_2) \subseteq C^d(L_1) \square C^d(\# L_2)$ (même démonstration pour C^{df}).$\square$

B : CÔNES DÉCROISSANTS ET CÔNES DÉCROISSANTS FIDÈLES

Le passage par les cônes rationnels décroissants pour étudier les substitutions de cônes rationnels dans cônes rationnels est naturel dans la mesure où les cônes décroissants possèdent des propriétés que n'ont pas, en général, les cônes rationnels. Prenons, en effet, deux langages L_1 et L_2 définis sur des alphabets disjoints. Soit L'_1 un langage appartenant à $C(L_1)$, on peut alors vérifier que $L'_1 \uparrow L_2$ n'est pas obligatoirement un élément de $C(L_1 \uparrow L_2)$. Par contre si on choisit L' dans $C^d(L_1)$ (resp. $C^{df}(L_1)$) alors, en utilisant la caractérisation des transductions rationnelles décroissantes et le lemme IV-5 on peut énoncer :

<u>*Lemme IV-8*</u> *: Si $L' \subseteq X'^*$ est un élément de $C^d(L_1)$ (resp. $C^{df}(L_1)$), alors quel que soit $L_2 \subseteq X_2^*$, avec $X_2 \cap X' = \emptyset$, $L' \uparrow L_2 \in C^d(L_1 \uparrow L_2)$ (resp. $C^{df}(L_1 \uparrow L_2)$).*

<u>Démonstration</u> : D'après le théorème II-3, L' appartient à $C^d(L_1)$ (resp. $C^{df}(L_1)$) si et seulement si $L' = g(h^{-1}(L_1) \cap R)$ où R est un langage rationnel de Z^*, h un homomorphisme strictement alphabétique, g est un homomorphisme alphabétique (resp. limité sur R).
Pour montrer que $L' \uparrow L_2$ appartient à $C^d(L_1 \uparrow L_2)$ on construit $\tau' = g' \circ (\cap R') \circ h'^{-1}$ avec

$$R' = R \uparrow X_2^*$$

(on supposera, ce qui est toujours possible, que Z et X_2 sont des alphabets disjoints).

h' et g' sont des homomorphismes définis sur $Z \cup X_2^*$ où h' est l'identité sur X_2^* et est égal à h sur Z^* g' est l'identité sur X_2^* et $g'(z) = g(z)$ si $g(z) \neq \varepsilon$ et $g'(z) = \#$ sinon, où # est une nouvelle lettre.

Les homomoprhismes h' et g' sont strictement alphabétiques, la transduction τ' conserve les longueurs. Elle est donc décroissante fidèle.

Appelons $L'' = g'(h^{-1}(L_1) \cap R)$. Alors, $L'' \uparrow L_2 = g'(h'^{-1}(L_1 \uparrow L_2) \cap R')$.

En effet,

$$
\begin{aligned}
L'' \uparrow L_2 &= g'(h^{-1}(L_1) \cap R) \uparrow L_2 \\
&= g'(h^{-1}(L_1) \cap R) \uparrow g'(L_2) \text{ par définition de } g' \\
&= g'((h^{-1}(L_1) \cap R) \uparrow L_2) \text{ car } g' \text{ est strictement alphabétique} \\
&= g'(h^{-1}(L_1) \cap R) \uparrow (L_2 \cap X_2^*)) \text{ car } L_2 \subseteq X_2^* \\
&= g'((h^{-1}(L_1) \uparrow L_2) \cap (R \uparrow X_2^*)) \text{ car } Z \cap X_2 = \emptyset \text{ et lemme IV-4} \\
&= g'((h^{-1}(L_1) \uparrow L_2) \cap R') \text{ par définition de } R \\
&= g'((h'^{-1}(L_1) \uparrow h'^{-1}(L_2)) \cap R') \text{ par définition de } h' \\
&= g'(h'^{-1}(L_1 \uparrow L_2) \cap R') \text{ lemme IV-6 et } L \text{ strictement alphabétique.}
\end{aligned}
$$

Le langage $L'' \uparrow L_2$ appartient donc à $C^{df}(L_1 \uparrow L_2)$.

Les langages L" et L' ne diffèrent que par les lettres #, c'est à dire :

soit $\theta : (X' \cup X_2 \cup \{\#\})^*$ dans $(X' \cup X_2)^*$ tel que

$$\theta(x) = x \text{ si } x \neq \# \text{ et } \theta(\#) = \varepsilon$$

$g = \theta \circ g''$ (resp. g limité sur R implique θ limite sur $g''(R)$ donc sur $L'' \uparrow L_2$) et $L = g(h^{-1}(L_1) \cap R) = \theta(L'')$.

Soit k la longueur du plus petit mot de L_2. Gardons dans $L'' \uparrow L_2$ les mots tels que la lettre # soit suivie de k lettres de X_2, c'est à dire

$$L'' \uparrow L_2 \cap ((X' X_2^*)^* (\# X_2^k)^*)^*$$

Il suffit, maintenant, d'effacer les lettres de X_2 comprises entre un # et un # ou une lettre de X' pour obtenir $L' \uparrow L_2$ (aux # près).

Ceci peut se faire facilement à l'aide d'une application séquentielle décroissante fidèle γ. D'où on déduit

$$L' \uparrow L_2 = \theta \circ \gamma(L'' \uparrow L_2 \cap ((X' \, X_2^*)^* \, (\# \, X_2^k)^*)^*)$$

donc $L'_1 \uparrow L_2 \in C^d(L'' \uparrow L_2) = C^d(L_1 \uparrow L_2)$

(resp. si θ est limité $L'_1 \uparrow L_2 \in C^{df}(L'' \uparrow L_2) = C^{df}(L_1 \uparrow L_2)$).□

<u>*Lemme IV-9*</u> *: Soit* $L \subseteq X^*$, *si s est une* $C^d(\# \, L)$*-substitution (resp.* $C^{df}(\# \, L)$*-substitution) définie sur* X' *avec* $X \cap X' = \emptyset$ *alors pour tout langage* $L' \subseteq X'^*$ $s(L') \in C^d(L' \uparrow L)$ *(resp.* $C^{df}(L' \uparrow L)$*).*

<u>Démonstration</u> : Posons $X' = \{x'_2, \ldots, x'_n\}$. Par définition, pour tout i, $s(x'_i) \in C^d(\# \, L) = C^d(x_i \, L)$. Il existe donc une transduction rationnelle décroissante $\tau_i = g_i \circ (\cap R_i) \circ h_i^{-1}$ telle que $s(x'_i) = \tau_i(x_i \, L_2)$. D'après le thèorème II-3, h_i peut être choisi strictement alphabétique et g_i alphabétique. On suppose que les langages rationnels R_i sont définis sur sur des alphabets Z_i, disjoints deux à deux. (Si ce n'est pas le cas on effectue des copies, ce qui est toujours possible).

h_i étant strictement alphabétique $h^{-1}(x_i \, L) \cap R_i = h^{-1}(x_i \, L) \cap R_i \cap h^{-1}(X') \, (h^{-1}(X))^*$ choisissons $R'_i = R_i \cap h^{-1}(X') \, (h^{-1}(X))^*$

Soient $R = (\bigcup_{i=1}^{n} R'_i)^* \subseteq (\bigcup_{i=1}^{n} Z_i)^*$

Les homomorphismes h et g définis sur Z^* par :

$h = h_i$ sur Z_i $\forall$ i

$g = g_i$ sur Z_i $\forall$ i

(l'image par h d'un mot de R'_i commence par une lettre de X' , cette lettre est x_i si l'image est dans $x_i \, L$).

Montrons que $s(L') = g(h^{-1}(L' \uparrow L) \cap R)$.

Soit $w \in s(L')$, $\exists$ $w' \in L'$ tel que $w \in s(w')$.

Posons $w' = x_{j_1} \ldots x_{j_{p-1}}$, w peut se décomposer en $w = w_1 \ldots w_p$ où
$\forall\, i\ w_i \in s(x_{j_i}) = g_{j_i}(h^{-1}_{j_i}(x_{j_i} L) \cap R'_{j_i})$.
w appartient donc à

$$g_{j_1}(h^{-1}_{j_1}(x_{j_1} L) \cap R'_{j_1}) \ldots g_{j_p}(h^{-1}_{j_p}(x_{j_p} L) \cap R'_{j_p})$$
$$= g((h^{-1}_{j_1}(x_{j_1} L) \cap R'_{j_1}) \ldots (h^{-1}_{j_p}(x_{j_p} L) \cap R'_{j_p})) \text{ par définition de } g$$
$$= g(h^{-1}_{j_1}(x_{j_1} L) \ldots h^{-1}_{j_p}(x_{j_p} L) \cap R'_{j_1} \ldots R'_{j_p})$$
$$= g(h^{-1}(x_{j_1} L \ldots x_{j_p} L) \cap R'_{j_1} \ldots R'_{j_p}) \text{ car } h \text{ est alphabétique}$$
$$\subseteq g(h^{-1}(L' \uparrow L) \cap R)$$

ou encore $w \in \tau(L' \uparrow L)$.

Réciproquement, soit $w \in \tau(L' \uparrow L) = g(h^{-1}(L' \uparrow L) \cap R)$. Il existe $w' \in R \cap h^{-1}(L' \uparrow L)$ tel que $w = g(w')$. $w' \in R$, il est donc de la forme $w' = z_{j_1} u_{j_1} \ldots z_{j_p} u_{j_p}$ avec $\forall\, i\ z_{j_i} u_{j_i} \in R'_{j_i}$, $z_{j_i} \in h^{-1}(X')$ et $u_{j_i} \in (h^{-1}(X))^*$ mais $h(w') = h(z_{j_1})\, h(u_{j_1}) \ldots h(z_{j_p})\, h(w_{j_p})$ appartient à $L \uparrow L'$ on déduit que $h(z_{j_1}) \ldots h(z_{j_p}) \in L$ et $\forall\, i, h(u_{j_i}) \in L'$

$$\left.\begin{array}{l} h(z_{j_i} u_{j_i}) \in h(z_{j_i})\, L \\ z_{j_i} u_{j_i} \in R'_{j_i} \end{array}\right\} \Rightarrow h(z_{j_i}) = x_{j_i}$$

$z_{j_i} u_{j_i}$ appartient donc à $h^{-1}(x_{j_i} L)$

$$\begin{aligned} w = g(w) &= g_{j_1}(z_{j_1} u_{j_1}) \ldots g_{j_p}(z_{j_p} u_{j_p}) \\ &\subset g_{j_1}(h^{-1}(x_{j_1} L) \cap R'_{j_1}) \ldots g_{j_p}(h^{-1}(x_{j_p} L) \cap h'_{j_p}) \\ &= s(x_{j_1}) \ldots s(x_{j_p}) \\ &\in s(L'). \ \square \end{aligned}$$

Ces lemmes vont permettre de démontrer la première proposition de base de ce chapitre :

Proposition IV-10 : Pour tous langages $L_1 \subseteq X_1^*$, $L_2 \subseteq X_2^*$ *avec* $X_1 \cap X_2 = \emptyset$, *on a les relations :*

$$C^d(L_1) \,\square\, C^d(\# L_2) = C^d(L_1 \uparrow L_2)$$
$$C^{df}(L_1) \,\square\, C^{df}(\# L_2) = C^{df}(L_1 \uparrow L_2)$$

Démonstration : Commençons par montrer l'inclusion $C^d(L_1) \,\square\, C^d(\# L_2) \subseteq C^d(L_1 \uparrow L_2)$ (resp. $C^{df}(L_1) \,\square\, C^{df}(\# L_2) \subseteq C^{df}(L_1 \uparrow L_2)$).
Tout langage L appartenant à la famille $C^d(L_1) \,\square\, C^d(\# L_2)$ peut se mettre sous la forme $L = s(L')$ où L' appartient à $C^d(L_2)$ et s est une $C^d(\# L_2)$-substitution, alors $L = s(L') \in C^d(L' \uparrow L_2)$ d'après le lemme IV-9 famille contenue dans $C^d(L_1 \uparrow L_2)$ d'après le lemme IV-8. La démonstration est la même pour le cône décroissant fidèle.

Réciproquement, choisissons L appartenant à $C^d(L_1 \uparrow L_2)$, (resp. $C^{df}(L_1 \uparrow L_2)$), L se met sous la forme $L = g(h^{-1}(L_1 \uparrow L_2) \cap R)$ avec (cf théorème II-3) R langage rationnel, g homomorphisme alphabétique, h homomorphisme strictement alphabétique. Mais puisque h est strictement alphabétique on peut utiliser le lemme IV-5 qui nous permet d'écrire $L = g((h^{-1}(L_1) \uparrow h^{-1}(L_2)) \cap R)$ mais, le lemme IV-6 permet d'affirmer que : $(h^{-1}(L_1) \uparrow h^{-1}(L_2)) \cap R \in C^{df}(h^{-1}(L_1)) \,\square\, C^{df}(\# h^{-1}(L_2))$ or, h strictement alphabétique implique que $h^{-1}(L_1) \in C^{df}(L_1) \subset C^d(L_1)$ et $\# h^{-1}(L_2) \in C^{df}(L_2) \subset C^d(\# L_2)$ et par composition des transductions rationnelles décroissantes et des transductions rationnelles décroissantes fidèles on obtient

$$(h^{-1}(L_1) \uparrow h^{-1}(L_2)) \cap R \in C^d(L_1) \,\square\, C^d(\# L_2)$$

(respectivement $(h^{-1}(L_1) \uparrow h^{-1}(L_2)) \cap R \in C^{df}(L_1) \,\square\, C^{df}(\# L_2)$)
pour terminer la démonstration il suffit de se souvenir (cf chapitre I) que $C^d(L_1) \,\square\, C^d(\# L_2)$ est clos par homomorphisme d'une part et que $C^{df}(L_1) \,\square\, C^{df}(\# L_2$ est clos par homomorphisme limité sur L.$\square$

C : RELATIONS ENTRE $C^{df}(L)$, $C^{d}(L)$, $C^{f}(L)$ ET $C(L)$

Dans les chapitres précédents nous avons démontré les inclusions :

$\forall$ L : $C^{d}(L) \subseteq C^{if}(L) \subseteq C(L)$

De plus, l'égalité des familles $C^{df}(L)$ et $C(L)$ implique l'égalité des familles $C^{d}(L)$ et $C^{if}(L)$ mais, la réciproque n'est pas vraie. Nous allons voir ce qui "manque" à $C^{d}(L)$ pour obtenir $C^{if}(L)$ et ce qui "manque" à $C^{d}(L)$ pour obtenir $C(L)$.

<u>*Lemme IV-11*</u> *: Pour tout langage* L*, on a les relations :*

$C^{if}(L) = C^{d}(L) \square \text{ Fin}$

$C^{bf}(L) = C^{df}(L) \square \text{ Fin}$

<u>Démonstration</u> : Commençons par montrer la première relation.
Puisque toute substitution finie propre est une transduction rationnelle d'image finie, que toute transduction décroissante est d'image finie, on en déduit l'inclusion $C^{d}(L) \square \text{ Fin} \subseteq C^{if}(L)$. Pour montrer l'inclusion dans l'autre sens il suffit d'utiliser la proposition II-17 qui affirme que toute transduction rationnelle d'image finie est la composée d'une substitution finie propre suivie d'une transduction rationnelle décroissante.
La deuxième relation se démontre de la même façon.□

<u>*Lemme IV-12*</u> *: Pour tout langage* L*, on a les relations*

$C(L) = C^{d}(L) \square \text{ Rat}$

$C^{f}(L) = C^{df}(L) \square \text{ Rat}$

<u>Démonstration</u> : D'après le lemme III-8, pour tout langage L

$C(L) = C^d(L \uparrow c^*)$ et $C^f(L) = C^{df}(L \uparrow c^*)$ alors

$$
\begin{aligned}
C(L) &= C^d(L \uparrow c^*) \\
&= C^d(L) \square C^d(\# c^*) \quad \text{proposition IV-10} \\
&= C^d(L) \square \mathrm{Rat} \quad \text{lemme III-3}
\end{aligned}
$$

et

$$
\begin{aligned}
C(L) &= C^{df}(L \uparrow c^*) \\
&= C^{df}(L) \square C^{df}(\# c^*) \quad \text{proposition IV-10} \\
&= C^{df}(L) \square \mathrm{Rat} \quad \text{Lemme III-3}
\end{aligned}
$$

<u>Remarque</u> : Ce résultat peut également être obtenu à partir de la proposition II-16 qui permet de factoriser toute transduction rationnelle en une substitution rationnelle propre suivie d'une transduction rationnelle décroissante.

<u>Proposition IV-13</u> : Pour tous langages L_1 *et* L_2 *définis sur des alphabets disjoints :*

$$
C(L_1 \uparrow L_2) = C^d(L_1) \square C(L_2)
$$
$$
C^f(L_1 \uparrow L_2) = C^{df}(L_1) \square C^f(L_2)
$$

<u>Démonstration</u> :

$$
\begin{aligned}
C(L_1 \uparrow L_2) &= C^d(L_1 \uparrow L_2) \square \mathrm{Rat} \quad \text{lemme IV-12} \\
&= C^d(L_1) \square C^d(\# L_2) \square \mathrm{Rat} \quad \text{Proposition IV-10} \\
&= C^d(L_1) \square C(\# L_2) \quad \text{lemme IV-12} \\
&= C^d(L_1) \square C(L_2)
\end{aligned}
$$

de même

$$
\begin{aligned}
C^f(L_1 \uparrow L_2) &= C^{df}(L_1 \uparrow L_2) \quad \mathrm{Rat} \quad \text{lemme IV-12} \\
&= C^{df}(L_1) \square C^{df}(\# L_2) \square \mathrm{Rat} \quad \text{proposition IV-10} \\
&= C^{df}(L_1) \square C^f(\# L_2) \quad \text{lemme IV-12} \\
&= C^{df}(L_1) \square C^f(L_2). \square
\end{aligned}
$$

Remarquons, enfin, que de la relation $C(L) = C^{d}(L) \square \mathrm{Rat}$ on déduit très facilement la relation $(*)$ à savoir $\forall$ L, L' définis sur des alphabets disjoints, c, une nouvelle lettre, $C(L) \square C(L') = C(L \uparrow (L'\,c)^{*})$, il suffit de remarquer que $L \uparrow (L'\,c)^{*}$ est rationnellement équivalent à $(L \uparrow c^{*}) \uparrow L'$.

Alors $C(L \uparrow (L'\,c)^{*}) = C((L \uparrow c^{*}) \uparrow L')$

$= C^{d}(L \uparrow c^{*}) \square C(L')$ proposition IV-13

$= C(L) \square C(L')$ lemme III-8

Il est à noter que le passage par les cônes rationnels décroissants simplifie notablement la démonstration de cette relation.

D : Une caractérisation des langages complets par substitution

Définition IV-14 : Un langage $L \subseteq X^*$ *est complet par substitution si et seulement si, pour tout langage* L'

$$A \in C(L) \Rightarrow A \uparrow L' \in C(L \uparrow L')$$

Remarquons que l'idée de rapprocher les langages d'image finie et les langages complets par substitution est assez naturelle dans la mesure où une des caractérisations des langages d'image finie fait intervenir l'opérateur syntaxique. Le résultat essentiel de cette section est de montrer l'identité des familles de langages d'image finie et de langages complets par substitution.

Commençons par montrer que la définition annoncée dans l'introduction de ce chapitre, quoique légèrement différente de la définition IV-14, lui est équivalente ;

Lemme IV-15 : L *est un langage complet par substitution si et seulement si pour tout langage* L' *défini sur un alphabet disjoint de celui de* L $C(L) \square C(L') = C(L \uparrow L')$.

Démonstration : De façon évidente nous avons l'inclusion $C(L \uparrow L') \subset C(L \uparrow (L'\, c)^*) = C(L) \square C(L')$. Réciproquement, on suppose L complet par substitution. Puisque $L \uparrow c^*$ appartient à $C(L)$, par définition, $\forall L'$ $(L \uparrow c^*) \uparrow L' \in C(L \uparrow L')$ mais $(L \uparrow c^*) \uparrow L'$ est rationnellement équivalent à $L \uparrow (L'\, c)^*$, on déduit l'inclusion $C(L \uparrow (L'\, c)^*) \subseteq C(L \uparrow L')$ ou encore $C(L) \square C(L') \subseteq C(L \uparrow L')$. $\square$

Rappelons la version de M. Latteux du lemme syntaxique de substitution.

__Proposition IV-16__ [25] : Soient $\mathcal{L}_1$ *un cône rationnel bifidèle,* $\mathcal{L}_2$ *un cône rationnel et les langages* $L_1 \subseteq X_1^*$, $L_2 \subseteq X_2^*$ *avec* $X_1 \cap X_2 = \emptyset$ *alors* $L_1 \uparrow L_2 \in \mathcal{L}_1 \square \mathcal{L}_2$ *implique* $L_1 \in \mathcal{L}_1$ *ou* $(L_2\ c)^* \in \mathcal{L}_2$.

Dans les mêmes conditions d'application ($X_1 \cap X_2 = \emptyset$), de cette proposition et des résultats de la section précédente on déduit :

__Proposition IV-17__ : Soient $L'_1 \subseteq X'_1$ *et* $L'_2 \subseteq X'_2$ *deux langages et* $X'_1 \cap X'_2 = \emptyset$ *alors :*
$L'_1 \uparrow L'_2 \in C(L_1 \uparrow L_2)$ *implique* $L'_1 \in C^{if}(L_1)$ *ou* $(L'_2\ c) \in C(L_2)$

__Démonstration__ : $L'_1 \uparrow L'_2 \in C(L_1 \uparrow L_2)$. D'après la proposition IV-13 $C(L_1 \uparrow L_2) = C^d(L_1) \square C(L_2) \subseteq C^{if}(L_1) \square C(L_2)$ or $C^{if}(L_1)$ est un cône rationnel bifidèle, ce qui implique bien, en vertu de la proposition précédente, que $L'_1 \in C^{if}(L_1)$ ou $(L'_2\ c) \in C(L_2)$. □

En particulier, si nous prenons $L'_1 = L_1$ et $L'_2 = (L_2\ c)^*$ nous obtenons :

__Corollaire IV-18__ : $L_1 \uparrow (L_2\ c)^* \in C(L_1 \uparrow L_2)$ *implique* L_1 *d'image finie ou* $(L_2\ c)^* \in C(L_2)$.

<u>Démonstration</u> : Le langage $L_1 \uparrow (L_2\ c)^*$ est rationnellement équivalent à $(L_1 \uparrow c^*) \uparrow L_2$ et l'appartenance de $L_1 \uparrow (L_2\ c)^*$ à $C(L_1 \uparrow L_2\ )$ implique celle de $(L_1 \uparrow c^*) \uparrow L_2$ à $C(L_1 \uparrow L_2)$. Nous pouvons alors utiliser la proposition précédente qui garantit soit l'appartenance de $L_1 \uparrow c^*$ à $C^{if}(L_1)$, soit l'appartenance de $(L_2\ c)^*$ à $C(L_2)$. Mais $L_1 \uparrow c^* \in C^{if}(L_1)$ est une caractérisation des langages d'image finie. $\square$

Nous sommes alors en mesure d'énoncer ;

<u>*Thèorème IV-19*</u> *: L est complet par substitution si et seulement si L est d'image finie.*

<u>Démonstration</u> ; Supposons, tout d'abord, L complet par substitution, $C(L \uparrow (L'\ c)^*) = C(L) \square C(L') = C(L \uparrow L')$, de ces égalités on déduit que le langage $L \uparrow (L'\ c)^*$ appartient à $C(L \uparrow L')$ et ce pour tout langage L'. Choisissons L' = Sym, le langage symétrique sur deux lettres. Nous avons donc $L \uparrow (\mathrm{Sym}\ c)^* \in C(L \uparrow \mathrm{Sym})$. Le corollaire précédent permet de conclure que L est d'image finie puisque Sym n'engendre pas une FAL.

Réciproquement, supposons L d'image finie, alors $C(L) = C^d(L)$. D'après la proposition IV-13 $C(L \uparrow L') = C^d(L) \square C(L') = C(L) \square C(L')$ ou encore, selon le lemme IV-15, L est complet par substitution. $\square$

On peut également montrer que :

<u>*Proposition IV-20*</u> *: Si C(L) est le cône rationnel engendré par le langage non rationnel L, C(L) clos par substitution implique que L est un langage d'image finie.*

Démonstration : Appelons X l'alphabet sur lequel est défini L. Si θ est l'homomorphisme défini sur X qui barre toutes les lettres de X. On appelle $\bar{L} = \theta(L)$. Soit $c \notin X$.

$L \uparrow c^* \in C(L)$ donc

$$\begin{aligned}(L \uparrow c^*) \uparrow \bar{L} &\in C(L) \,\square\, C(L) \\ &= C(L) \quad \text{(clôture par substitution)} \\ &= C^d(L) \,\square\, \text{Rat} \quad \text{(lemme IV-12)} \\ &\subseteq C^{if}(L) \,\square\, \text{Rat}\end{aligned}$$

Nous pouvons utiliser la proposition IV-6, alors

- soit $(L \uparrow c^*) \in C^{if}(L)$
- soit $(L\ a)^* \in \text{Rat}$ ce qui est impossible car L n'est pas rationnel

L est donc bien d'image finie. $\square$

La réciproque, L langage d'image finie implique C(L) clos par substitution, est fausse. Il suffit de choisir pour langage L le langage Sym.

E : LANGAGES TRÈS FORTEMENT EFFAÇABLES

Définition IV-21 : L ***est très fortement effaçable si et seulement si*** $C^d(L) = C^{df}(L)$.

Si L est très fortement effaçable, le lemme IV-12 et cette définition permettent d'écrire que $C^f(L) = C^{df}(L) \square \mathrm{Rat} = C^d(L) \square \mathrm{Rat} = C(L)$, ce qui permet d'énoncer que :

Proposition IV-22 : ***Si*** L ***est très fortement effaçable alors il est effaçable.***

La réciproque de cette proposition est fausse. Il suffit, pour s'en convaincre, de choisir Sym, langage pour lequel nous avons :

$$C^{df}(\mathrm{Sym}) \subsetneq C^f(\mathrm{Sym}) = C^{if}(\mathrm{Sym}) = C^d(\mathrm{Sym}) = C(\mathrm{Sym}) \quad [23]$$

Proposition IV-23 : ***Un langage*** L ***est bifidèle si et seulement s'il est d'image finie et fortement effaçable.***

Démonstration : Supposons que L soit d'image finie et fortement effaçable alors $C(L) = C^{if}(L) = C^d(L)$ et $C^d(L) = C^{df}(L)$ implique $C^{df}(L) = C^d(L) = C(L)$ c'est à dire $C(L) \subseteq C^{bf}(L) \subseteq C(L)$ d'où l'égalité $C(L) = C^{bf}(L)$.

Réciproquement, de $C^{bf}(L) = C(L)$ on déduit que

$C(L) = C^{bf}(L) \subseteq C^{if}(L) \subseteq C(L) \Rightarrow L$ d'image finie

et $C(L) = C^{bf}(L) \subseteq C^f(L) \subseteq C(L) \Rightarrow C^{bf}(L) = C^f(L) = C^d(L)$

Le langage $L \uparrow c^*$ appartient à $C(L)$ donc à $C^{bf}(L)$. D'après les résultats du chapitre III ceci entraine que $C^{df}(L) = C^{f}(L)$ d'où la conclusion $C^{df}(L) = C^{d}(L)$. L est donc très fortement effaçable. $\square$

Dans l'énoncé de la proposition IV-23 le "et" est nécessaire. En effet, on connait des langages d'image finie non bifidèle (Sym par exemple) et nous avons vu, dans le chapitre III de cette thèse, que le langage $C_1=\{a^n b^n \ / \ n \geq 0\}$ est un langage très fortement effaçable mais non bifidèle.

<u>Proposition IV-24</u> : *Si* L_1 *est un langage très fortement effaçable et si* L_2 *est effaçable alors* $L_1 \uparrow L_2$ *est effaçable.*

<u>Démonstration</u> :

$$
\begin{aligned}
C^{f}(L_1 \uparrow L_2) &= C^{df}(L_1) \ \square \ C^{f}(L_2) \text{(proposition IV-13)}\\
&= C^{d}(L_1) \ \square \ C(L_2) \text{ par hypothèse}\\
&= C(L_1 \uparrow L_2) \text{ (proposition IV-13).} \quad \square
\end{aligned}
$$

<u>Proposition IV-25</u> : *Soient* L_1 *et* L_2 *deux langages effaçables. Si* L_2 *engendre une* FAL, *alors* $L_1 \uparrow L_2$ *est effaçable.*

<u>Démonstration</u> : Par hypothèse $C^{f}(L_1) = C^{d}(L_1) = C(L_1)$ et

$C^{f}(L_2) = C(L_2) = \mathrm{Rat} \ \square \ C(L_2)$

$$
\begin{aligned}
C^{f}(L_1 \uparrow L_2) &= C^{df}(L_1) \ \square \ C^{f}(L_2) \quad \text{(proposition IV-13)}\\
&= C^{df}(L_1) \ \square \ \mathrm{Rat} \ \square \ C(L_2) \text{ par hypothèse}\\
&= C^{f}(L_1) \ \square \ C(L_2) \text{ (lemme IV-12)}\\
&= C(L_1) \ \square \ C(L_2) \quad \text{par hypothèse}\\
&\supseteq C(L_1 \uparrow L_2)
\end{aligned}
$$

d'où l'égalité $C^{f}(L_1 \uparrow L_2) = C(L_1 \uparrow L_2)$

De la proposition IV-16 et des relations $C^f(L_1 \uparrow L_2) = C^{df}(L_1) \square C^f(L_2) \subseteq C^{bf}(L_1) \square C(L_2)$ on déduit immédiatement.

<u>Proposition IV-26</u> : *Soient* $L_1 \subseteq X_1^*$, $L_2 \subseteq X_2^*$, $L'_1 \subseteq X'^*_1$, $L'_2 \subseteq X'^*_2$ *avec* $X_1 \cap X_2 = \emptyset$ *et* $X'_1 \cap X'_2 = \emptyset$, *c une nouvelle lettre. Alors* $L'_1 \uparrow L'_2 \in C^f(L_1 \uparrow L_2)$ *implique que soit* L'_1 *appartient à* $C^{bf}(L_1)$ *soit* $(L'_2\ c)^* \in C(L_2)$.

En particulier, nous obtenons en choisissant $L'_1 = L_1 \uparrow c^*$ et $L'_2 = L_2$:

<u>Proposition IV-27</u> : *Soient* $L_1 \subseteq X_1^*$ *et* $L_2 \subseteq X_2^*$ *avec* $X_1 \cap X_2 = \emptyset$. *Si* L_1 *est un langage effaçable et si* $L_1 \uparrow (L_2\ c)^* \in C^f(L_1 \uparrow L_2)$ *alors* L_1 *est un langage bifidèle ou* $(L_2\ c)^* \in C(L_2)$.

<u>Démonstration</u> : $(L_1 \uparrow c^*) \uparrow L_2$ est rationnellement équivalent à $L_1 \uparrow (L_2\ c)^*$. De l'hypothèse on déduit que $(L_1 \uparrow c^*) \uparrow L_2 \in C^f(L_1 \uparrow L_2)$ et la proposition précédente permet d'écrire que soit $(L_1 \uparrow c^*) \in C^{bf}(L_1)$ soit $(L_2\ c)^* \in C(L)$ mais $(L_1 \uparrow c^*) \in C^{bf}(L_1)$ implique (théorème III-11) $C^{df}(L_1) = C^f(L_1)$ si on suppose L_1 effaçable on obtient bien $C^{df}(L_1) = C(L_1)$. □

<u>Proposition IV-28</u> : *Soient* $L_1 \subseteq X_1^*$ *un langage d'image finie non bifidèle,* $L_2 \subseteq X_2^*$ *un langage n'engendrant par une FAL et* $X_1 \cap X_2 = \emptyset$ *alors* $L_1 \uparrow L_2$ *n'est pas effaçable.*

<u>Démonstration</u> : Choisissons L'_1 dans $C^d(L_1) \setminus C^{bf}(L_1)$ et L'_2 dans $\text{Rat} \square C(L_2) \setminus C(L_2)$. Le langage $L'_1 \uparrow L'_2$ appartient à $C^d(L_1) \square \text{Rat} \square C(L_2) = C(L_1) \square C(L_2)$ d'après le lemme IV-13 famille encore égale à $C(L_1 \uparrow L_2)$ car L_1

est un langage d'image finie. Le langage $L'_1 + L'_2$ est donc un élément de $C(L_1 + L_2)$. Supposons que $L'_1 + L'_2$ appartienne également à $C^f(L_1 + L_2)$ de la proposition IV-26 on déduit que :

- soit L'_1 appartient à $C^{bf}(L_1)$, ce qui est impossible
- soit $(L'_2\, c)^* \in C(L_2)$ mais $L'_2 \in C((L_2\, c)^*)$

ce qui implique $L'_2 \in C(L_2)$ ce qui est impossible également.
$L'_1 + L'_2$ ne peut pas appartenir à $C^f(L_1 + L_2)$ et

$$C^f(L_1 + L_2) \subsetneq C(L_1 + L_2)\ . \square$$

La proposition IV-25 (resp. IV-28) va permettre de construire des langages (algébriques) effaçables (resp. non effaçables).
Nous allons utiliser ces derniers résultats pour étudier la famille des langages quasi-rationnels.

F : UNE PROPRIÉTÉ DE LA FAMILLE DES LANGAGES QUASI-RATIONNELS :

La famille QR des langages quasi-rationnels peut être définie de plusieurs façons, la définition que nous allons utiliser est celle qui fait intervenir l'opérateur syntaxique.

soit θ un homomorphisme qui marque les lettres a, b, $\bar{a}$, $\bar{b}$

$$\theta(z) = \dot{z} \quad \forall\, z \in \{a, b, \bar{a}, \bar{b}\}$$

soit Ψ un homomorphisme qui démarque les lettres $\dot{a}$, $\dot{b}$, $\dot{\bar{a}}$, $\dot{\bar{b}}$

et laisse inchangées les autres lettres c'est à dire

$$\Psi(z) = \begin{cases} z \text{ si } z \in \{a, b, \bar{a}, \bar{b}\} \\ \\ y \text{ si } z = \dot{y} \end{cases}$$

Soit $\mathrm{Sym} = \{w\, \bar{w}^R \;/\; w \in \{a, b\}\}$

On définit la famille de langage $(L_i)_{i \in N}$ par

$L'_1 = \mathrm{Sym}$

$\forall\, i \geq 0 \quad L_{i-1} = \theta \circ \Psi(L'_{i-1})$ et $L_i = \mathrm{Sym} \uparrow L_{i-1}$

Alors, $QR = \bigcup_{i \geq 1} C(L_i)$ est une FAL non principale close par substitution.

Par contre :

<u>*Lemme IV-29*</u> [9] : *Pour tout entier* i, $C(L_i)$ *n'est pas une* FAL.

<u>*Proposition IV-30*</u> *: Pour tout entier* $i \geq 2$ L_i *n'est pas un langage effaçable, c'est à dire* $C^f(L_i)$ *est strictement contenu dans* $C(L_i)$.

<u>Démonstration</u> : Sym est un langage d'image finie non bifidèle et, pour tout i, $C(L_{i-1}) \subsetneq F(L_{i-1})$. Nous sommes donc dans les conditions d'application de la proposition IV-28 qui affirme que $L_i = \mathrm{Sym} \uparrow L_{i-1}$ n'est pas effaçable.□

Cependant, pour toute famille L telle que QR = $C(L)$ nous avons QR = $C^{bf}(L)$. C'est ce que nous allons montrer maintenant :

Définition : Une famille L de langage est uniformément fidèle (resp. d'image finie, resp. bifidèle) si et seulement si pour toute famille L' telle que $L = C(L')$ on a également $L = C^f(L')$ (resp. $L = C^{if}(L')$, resp. $L = C^{bf}(L')$).

Avant de montrer que QR est uniformément bifidèle nous avons besoin de quelques lemmes.

Lemme IV-31 : Pour tout $i \geq 1$ $C(L_i) \subseteq C^f(L_{2i-1})$

Démonstration : Elle se fait par récurrence sur i.

Si $i = 1$ alors, puisque L_1 est effaçable

$$C(L_1) \subseteq C^f(L_1)$$

Hypothèse de récurrence : $\forall\, p < i,\ C(L_p) \subseteq C^f(L_{2p-1})$

$C(L_i) = C(\mathrm{Sym} \uparrow L_{i-1})$

$= C(\mathrm{Sym}) \square C(L_{i-1})$ (Sym est un langage complet par substitution)

$= C^f(\mathrm{Sym}) \square C(L_{i-1})$ (Sym est effaçable)

$= C^{df}(\mathrm{Sym}) \square \mathrm{Rat} \square C(L_{i-1})$ (lemme IV-12)

$\subseteq C^{df}(\mathrm{Sym}) \square C^{df}(\mathrm{Sym}) \square C(L_{i-1})$ (Monotonie de $\square$ et lemme III- 6)

$\subseteq C^{df}(\mathrm{Sym}) \square C^{df}(\mathrm{Sym}) \square C^f(L_{2(i-1)-1})$ (hypothèse de récurrence)

$= C^{df}(\mathrm{Sym}) \square C^f(L_{2(i-1)})$ (lemme IV-13)

$= C^f(L_{2i-1})$ (lemme IV-13). $\square$

Les lemmes que nous allons utiliser et dont nous rappelons les énoncés sont dus à L. Boasson - M. Nivat et à M. Latteux.

Lemme IV-32 [8] : *Soit* L *un langage algébrique vérifiant* $L' \in C(L)$ *si* L' *est un* IRS-*langage alors* $L' \in C^f(L)$.

Lemme IV-33 [25] : *Soit* $\mathcal{L}$ *un cône rationnel fermé par* $\mathcal{L}'$*-substitution avec* Rat $\subsetneq \mathcal{L}'$. *Alors,* $\mathcal{L} = \mathcal{L}'' \,\square\,$ Rat *avec* $\mathcal{L}''$ *cône rationnel bifidèle implique* $\mathcal{L}'' = \mathcal{L}$.

Proposition IV-34 : QR *est un cône rationnel uniformément bifidèle.*

Démonstration : Soit $\mathcal{L}$ une famille génératrice de QR, c'est à dire telle que $C(\mathcal{L})$ = QR. Pour tout i, $L_i \subseteq C(\mathcal{L})$. De plus L_i est un IRS-langage on peut utiliser le lemme IV-32 qui implique que $L_i \in C^f(\mathcal{L})$ ou encore $C^f(L_i) \subseteq C^f(\mathcal{L})$. QR = $\bigcup_{i\geq 1} C(L_i)$, , soit L un langage quelconque de QR, il existe i tel que $L \in C(L_i)$, cône contenu, d'après le lemme IV-31, dans $C^f(L_{2i-1}) \subseteq C^f(\mathcal{L})$.

$$C(\mathcal{L}) = QR \subseteq C^f(\mathcal{L}) \Rightarrow C(\mathcal{L}) = C^f(\mathcal{L}).$$

Donc, QR = $C^f(\mathcal{L}) = C^{bf}(\mathcal{L}) \,\square\,$ Rat. Comme QR est clos par substitution, le lemme IV-33 entraine $C^{bf}(\mathcal{L})$ = QR et QR et bien uniformément bifidèle. $\square$

CHAPITRE V

SOUS-CONES PARTICULIERS DE $C(L)$

Nous appelons **sous-cônes particuliers** de $C(L)$ les familles $C^{f}(L)$, $C^{if}(L)$, $C^{d}(L)$, $C^{bf}(L)$, $C^{df}(L)$.

$\mathcal{L}$, $\mathcal{L}'$, $\mathcal{M}$ désignent des familles de langages. Si $\mathcal{H}(\mathcal{L})$ est $\{h(L) \;/\; L \in \mathcal{L}\}$, la notation $\mathcal{L} \xrightarrow{\mathcal{H}} \mathcal{L}'$ signifie que $\mathcal{L}' = \mathcal{H}(\mathcal{L})$. De même, si on adopte l'écriture $\mathcal{L} \,\square\, \mathcal{M} = \mathcal{L}\,\tilde{\mathcal{M}}$ (où l'opération $\square$ a le sens de la définition du chapitre I), la notation $\mathcal{L} \xrightarrow{\tilde{\mathcal{M}}} \mathcal{L}'$ signifie que $\mathcal{L}' = \mathcal{L}\,\tilde{\mathcal{M}}$. Nous pouvons schématiser les résultats des chapitres précédents par

$$
\begin{tikzcd}
 & & C(L) \arrow[loop right, "\mathcal{H}\ \overline{\mathrm{Rat}}"] & & \\
 & C^{f}(L) \arrow[ur, "\mathcal{H}"] \arrow[loop left, "\overline{\mathrm{Rat}}"] & & C^{if}(L) \arrow[ul, "\overline{\mathrm{Rat}}"'] \arrow[loop right, "\mathcal{H}\ \overline{\mathrm{Fin}}"] & \\
 & & C^{bf}(L) \arrow[ul, "\overline{\mathrm{Rat}}"'] \arrow[ur, "\mathcal{H}"] \arrow[loop below, "\overline{\mathrm{Fin}}"] & & C^{d}(L) \arrow[ul, "\overline{\mathrm{Fin}}"'] \arrow[loop right, "\mathcal{H}"] \\
 & & & C^{df}(L) \arrow[ul, "\overline{\mathrm{Fin}}"'] \arrow[ur, "\mathcal{H}"] &
\end{tikzcd}
$$

De ce schéma on déduit facilement :

Proposition V-1 : *Pour tout langage* L *nous avons*

$C^{bf}(L) = C^{if}(L) \Rightarrow C^{f}(L) = C(L)$

$C^{bf}(L) = C^{f}(L) \Rightarrow C^{if}(L) = C(L)$

$C^{f}(L) \subseteq C^{if}(L) \Rightarrow C^{if}(L) = C(L)$

$C^{if}(L) \subseteq C^{f}(L) \Rightarrow C^{f}(L) = C(L)$

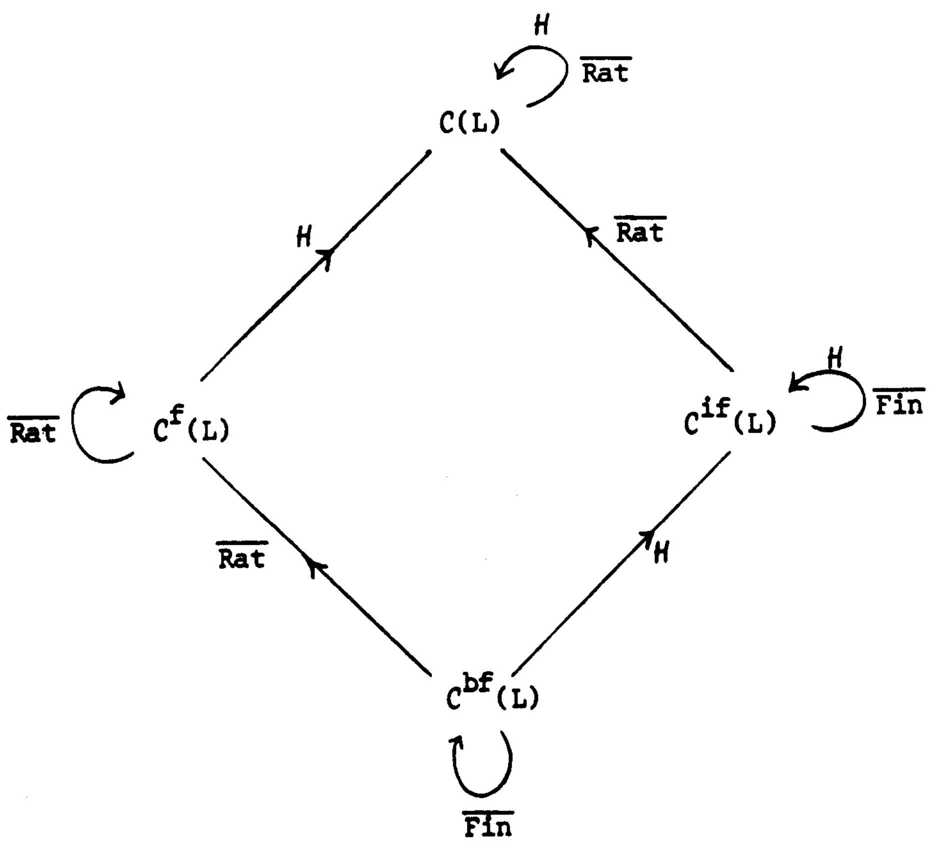
$C^{f}(L) = C^{if}(L) \Rightarrow C(L) = C^{f}(L) = C^{if}(L)$

A ces résultats on peut associer les implications du chapitre III à savoir :

$C^{if}(L) = C(L) \Rightarrow C^{d}(L) = C(L)$

$C^{bf}(L) = C^{f}(L) \Rightarrow C^{df}(L) = C^{f}(L)$

Nous allons maintenant nous attacher au schéma.



Nous savons, par définition même, que

$$\begin{array}{ccccc} & & C^{f}(L) & & \\ & \subseteq & & \subseteq & \\ C^{bf}(L) & & & & C(L) \\ & \subseteq & & \subseteq & \\ & & C^{if}(L & & \end{array}$$

Par exploration systématique, montrons qu'il existe des langages pour lesquels certaines inclusions sont des inclusions strictes et d'autres des égalités. Etant donnée la proposition V.1 nous avons à étudier sept cas différents.

1er cas	$\begin{array}{ccccc} & & C^{if}(L) & & \\ & \subsetneq & & \subsetneq & \\ C^{bf}(L) & & & & C(L) \\ & \subsetneq & & \subsetneq & \\ & & C^{if}(L) & & \end{array}$
2^{nd} cas	$C^{bf}(L) \subsetneq C^{if}(L) \subsetneq C^{f}(L) = C(L)$
3ème cas	$C^{bf}(L) \subsetneq C^{f}(L) \subsetneq C^{if}(L) = C(L)$
4ème cas	$C^{bf}(L) \subsetneq C^{f}(L) = C^{if}(L) = C(L)$
5ème cas	$C^{bf}(L) = C^{if}(L) \subsetneq C^{f}(L) = C(L)$
6ème cas	$C^{bf}(L) = C^{f}(L) \subsetneq C^{if}(L) = C(L)$
7ème cas	$C^{bf}(L) = C^{if}(L) = C^{f}(L) = C(L)$..

<u>Premier cas</u>

Le langage $B = \{a^{n_k} b^{n_k} \ldots a^{n_1} b^{n_1} x^p y^q z y^q x^p b^{m_1} a^{m_1} \ldots b^{m_k} a^{m_k} /$ $k \geq 1, n_i, m_i, p, q \geq 0\}$, dont L. Boasson a montré qu'il n'est pas effaçable [7] , n'est pas d'image finie. Pour obtenir ce résultat nous avons besoin du lemme V-2.

Auparavant notons que B est encore égal à $\bigcup_{k=1}^{\infty} B_k$ où $B_k = \{a^{n_k} b^{n_k} \ldots a^{n_1} b^{n_1} x^p y^q z y^q x^p b^{m_1} a^{m_1} \ldots b^{m_k} a^{m_k} / n_i, m_i,$ $p, q \geq 1\}$ ce qui permet d'écrire que $C^{if}(B) = \bigcup_{k=1}^{\infty} C^{if}(B_k)$

De la définition même des B_k, $\forall\, k' \geq k$, $B_{k'} \mathrel{\underline{\ncong}} B_k$

<u>*Lemme V-2*</u> *: Quels que soient les entiers* k *et* k' *tels que* $k' \geq k, B_k \uparrow c^*$ $\notin C^d(B_{k'})$.

<u>Démonstration</u> : Elle se fait par l'absurde en supposant qu'il existe $k' \geq k$ tel que $C^d(B_{k'})$ contienne $B_k \uparrow c^*$.

$B_{k'} = \{a^{n_{k'}} b^{n_{k'}} \ldots a^{n_{k+1}} b^{n_{k+1}} B_k b^{n_{k+1}} a^{n_{k+1}} \ldots b^{m_{k'}} a^{m_{k'}} / n_i, m_i \geq 1\}$

il peut donc se mettre sous la forme $B_k = A_{n_{k'}} \ldots A_{n_1} A A_{m_1} \ldots A_{m_{k'}}$

où les A_{n_i}, A_{m_i}, A sont des langages algébriques. $B_k \uparrow c^* \in C^d(B_{k'})$ si et seulement si $L_k \uparrow c^* = g(h^{-1}(B_{k'}) \cap R)$ où R est un langage rationnel de Σ^* h un homomorphisme strictement alphabétique, g un homomorphisme alphabétique.

Posons $A = \{\sigma \in \Sigma \;/\; h(\sigma) = a\}$

$B = \{\sigma \in \Sigma \;/\; h(\sigma) = b\}$

$X = \{\sigma \in \Sigma \;/\; h(\sigma) = x\}$

$Y = \{\sigma \in \Sigma \;/\; h(\sigma) = y\}$

$Z = \{\sigma \in \Sigma \;/\; h(\sigma) = z\}$

Etant donné que $B_{k'}$ est un langage borné, que h est strictement alphabétique on peut supposer que R est contenu dans $(A^+ B^+)^{k'} X^+ Y^+ Z Y^+ X^+(B^+ R^+)^{k'}$.
De plus, puisque tout rationnel est une union finie d'ensembles dont l'image, par la fonction de Parikh Ψ est un ensemble linéaire on peut supposer que

$$R = \bigcup_{i=1}^{n} R_i \text{ où}$$

$$R_i = R_{i,-2(k'+1)} R_{i,-2k'-1} \cdots R_{i,-4} R_{i,-3} R_{i,-2} R_{i,-1} R_{i,o} R_{i,1} R_{i,2} R_{i,3}$$

$$R_{i,4} \cdots R_{i,2k'+1}, R_{i,2(k'+1)}$$

où les $R_{i,j}$ sont des ensembles linéaires et où

$R_{i,o} \subseteq Z$

$R_{i,-1}, R_{i,1} \subseteq Y^+$

$R_{i,-2}, R_{i,2} \subseteq X^+$

$R_{i,-2j-1}, R_{i,2j+1} \subseteq B^+ \quad \forall j = 1, \ldots, k'$

$R_{i,-2(j+1)}, R_{i,2(j+1)} \subseteq A^+ \quad \forall j = 1, \ldots, k'$

En fait nous pouvons supposer que les $R_{i,j}$ $j \neq 0$ sont infinis, en effet pour $j \neq 0$, numérotons les $R_{i,j}$ de façon à ce que les $R_{i,j}$ infinis aient un numéro de 1 à n_j et posons $m = \inf\{n_j \mid j \in [-2(k'+1), 2(k'+1)], j \neq 0\}$.

Nous allons montrer que

$B_k = h_c(g(h^{-1}(B_{k'}) \cap (\bigcup_{i=1}^{m} R_i))$ où h_c est l'homomorphisme qui efface les lettres c et laisse inchangées les autres.

Bien évidemment $h_c(g(h^{-1}(B_{k'}) \cap (\bigcup_{i=1}^{m} R_i)) \subseteq h_c(B_k + c^*) = B_k$.

Réciproquement, montrons que tout mot w de B_k est dans $h_c(g(h^{-1}(B_{k'}) \cap (\bigcup_{i=1}^{m} R_i))$.

Soit ℓ_j la longueur du mot le plus long de $\bigcup_{i=n_j+1}^{n} R_{i,j}$, tout mot de B_k

est de la forme $a^{n_k} b^{n_k} \dots a^{n_1} b^{n_1} x^p y^q z y^q x^p b^{m_1} a^{m_1} \dots b^{m_k} a^{m_k}$ mais $w' = a^{n_k} c^{\ell_{-2(k+1)}} b^{n_k} c^{\ell_{-2k-1}} \dots a^{n_1} c^{\ell_{-4}} b^{n_1} c^{\ell_{-3}} x^p c^{\ell_{-2}} y^q c^{\ell_{-1}} z y^q c^{\ell_1} x^p c^{\ell_2} \dots b^{m_k} c^{\ell_{2k+1}} a^{m_k} c^{\ell_{2(k+1)}}$ est dans $B_k + c^*$ et il ne peut être l'image par g d'un mot de $\bigcup_{i=m+1}^{n} R_i$, il existe donc $i \leq m$ tel que $w' \in g(h^{-1}(B_{k'}) \cap R_i)$ et $w = h_c(w')$ appartient à $h_c(\bigcup_{i=1}^{m} g(h^{-1}(B_{k'}) \cap R_i)$.

Or pour tout j, $h(R_{ij})$ est un langage commutatif (L est commutatif si et seulement si $\Psi^{-1} \circ \Psi(L) = L'$, où Ψ est la fonction de Parikh).
Posons $R^o_{i,j} = \Psi^{-1} \circ \Psi(R_{i,j}) \cap Z_{\bar{c}}\, Z_c$ où $Z_c = \{\sigma \in \Sigma \ / \ g(\sigma) = \varepsilon \text{ ou } g(\sigma) = c\}$
et $Z_{\bar{c}} = \Sigma \setminus Z_c$
et $R^o = \bigcup_{i=1}^{m} R^o_i$ avec $R^o_i = R^o_{i,-2(k+1)} \dots R^o_{i,o} \dots R^o_{i,2(k+1)}$. Puisque les $\Psi(R_{i,j})$ sont linéaires et que les $R_{i,j}$ sont des langages rationnels infinis alors
$B_k = \bigcup_{i=1}^{m} h(g(h^{-1}(B_{k'})) \cap R_i) = \bigcup_{i=1}^{m} g(h^{-1}(B'_{k'}) \cap R^o_i)$
où $B'_{k'} = \{a^{n_k}((a^{t_{-(2k'+1)}})^*)^{-1} b^{n_k}((b^{t_{-2k'-1}})^*)^{-1} \dots x^p((x^{t_{-2}})^*)^{-1} y^q((y^{t_{-1}})^*)^{-1} z \ y^q((y^{t_1})^*)^{-1} x^p((x^{t_2})^*)^{-1} \dots b^{m_t}((b^{t_{2k'+1}})^*)^{-1} a^{m_t}((a^{t_{2(k'+1)}})^*)^{-1}$
$m_i, n_i, p, q \geq 0 \ i \in [1, k']\}$ où $t_j > 0 \ \forall j \in [-2(k'+1), 2(k'+1)], j \neq 0$.

Nous allons, maintenant, montrer que B'_k est un langage rationnel ce qui aménera la contradiction.
On peut se limiter au cas où tous les t_j sont égaux à 1.
$B'_{k'} = A'_{n_1} \dots A'_{n_{k'}} A' A'_{m_1} \dots A'_{m_k}$.
où $A'_{n_i} = \{a^n(a^*)^{-1} b^n(a^*)^{-1} / n \geq 1\}$, $A'_{m_i} = \{b^n(b^*)^{-1} a^n(a^*)^{-1} / n \geq 1\}$, sont des langages rationnels [13] et $A' = \{x^p(x^*)^{-1} y^q(y^*)^{-1} z \ y^q(y^*)^{-1} x^p(x^*)^{-1} / p, q \geq 1\}$ est un langage rationnel d'où on déduit naturellement que B'_1 est un langage rationnel et par conséquent B_1 est un langage rationnel ce qui est faux.

$\square$

De ce lemme on déduit immédiatement que, puisque $C^d(B) = \bigcup_{k\geq 1} C^d(B_k)$, quel que soit $k, B_k \rightarrow c^* \in C(B) \setminus C^d(B)$ ou encore $C^d(B) \subsetneq C(B)$.

D'où :

<u>Proposition V-3</u> : *Le langage algébrique* $L = \{a^{n_k} b^{n_k} \ldots a^{n_1} b^{n_1} x^p y^q z\ y^q x^p b^{m_1} a^{m_1} \ldots b^{m_k} a^{m_k} \,/\, k \geq 1\ n_i, m_i, q, q \geq 1\}$ *n'est ni effaçable ni d'image finie.*

D'autre part, en vertu de la proposition V-1 les inclusions de $C^{bf}(L)$ dans $C^{if}(L)$ et de $C^{bf}(L)$ dans $C^{if}(L)$ sont strictes.

Second cas

$$C^{bf}(L) \neq C^{if}(L) \neq C^f(L) = C(L)$$

Pour étudier ce cas nous allons travailler sur les langages définis sur un alphabet d'une seule lettre. M. Latteux a montré un résultat concernant les langages commutatifs dans [21], il est possible, sans modifier la démonstration qu'il propose, d'affiner l'énoncé :

<u>Proposition V-4</u> : *Un langage commutatif* L *vérifie* $C^{df}(L) = C^d(L)$ *si et seulement si pour tout langage rationnel* R, $L(R)^{-1}$, *le quotient à droite de* L *par* R, *appartient à* $C^{df}(L)$.

Cet énoncé va nous permettre de montrer que :

<u>Proposition V-5</u> : *Si* L *est un langage non algébrique défini sur un alphabet d'une lettre alors les trois propriétés sont équivalentes :*

i) L *est très fortement effaçable*

ii) Rat $\subseteq C^{df}(L)$

iii) $\exists\, k \in \mathbb{N}$ *vérifiant* $\forall\, w \in L$, $\exists\, w' \in L$ *tel que* $|w| < |w'| \leq k(|w| + 1)$

Démonstration : Puisque l'équivalence entre ii) et iii) a été montrée dans le chapitre II non ne démontrerons que l'équivalence entre i) et ii).

i) => ii) L est fortement effaçable si et seulement si $C^d(L) = C^{df}(L)$. L est commutatif, la proposition précédente affirme que pour tout langage rationnel R, $L(R)^{-1}$ appartient à $C^{df}(L)$. Le langage a^* est égal à $FG(L) = L(a^*)^{-1} \in C^{df}(L)$.

ii) => i) Pour tout i strictement positif $L(a^i)^{-1}$ est rationnel [13]. Comme tout langage rationnel R de a^* peut se mettre sous la forme d'une union finie de langages de la forme $a^r(a^t)^*$, on en déduit que $L(R)^{-1} \subset C^{df}(L)$ et la proposition V-4 permet de conclure que $C^d(L) = C^{df}(L)$.

Pour tout langage non algébrique contenu dans a^* tel que Rat $\subseteq C^{df}(L)$ on a la chaine :

$$\text{Rat} \subsetneq C^d(L) = C^{df}(L) \subseteq C^{if}(L) = C^{bf}(L) \subseteq C(L) = C^f(L).$$

En particulier le langage $L = \{a^{2^n} \ / \ n \geq 0\}$ est très fortement effaçable puisque la condition iii) est vérifiée en choisissant k = 2.

Par contre le langage $L = \{a^{n!} \ / \ n \geq 0\}$, déjà étudié au chapitre III n'est pas fortement effaçable et vérifie

$\text{Rat} \not\subset C^{df}(L) \subsetneq C^d(L) \subsetneq C^{if}(L) \subsetneq C^f(L) = C(L)$

or $C^{bf}(L) = C^{df}(L) \,\square\, \text{Fin} \subsetneq C^d(L) \,\square\, \text{Fin} = C^{if}(L) \subsetneq C^f(L) = C(L)$

Le langage L vérifie bien le second cas.

Troisième cas

$$C^{bf}(L) \subsetneq C^{f}(L) \subsetneq C^{if}(L) = C(L)$$

soit B le langage défini dans le premier cas dont on sait qu'il n'est ni effaçable ni d'image finie. $B \uparrow c^*$ n'est pas effaçable mais d'image finie. Choisissons $L = \text{Sym} \cup (B \uparrow c^*)$.

a) L est d'image finie car

$$L \uparrow c^* = \text{Sym} \uparrow d^* \cup (B \uparrow c^*) \uparrow d^* \in C(\text{Sym}) \cup C(B \uparrow c^*)$$
$$= C^{if}(\text{Sym}) \cup C^{if}(B \uparrow c^*) = C^{if}(L)$$

b) L vérifie $C^f(L) \subsetneq C^{if}(L)$

Pour montrer cette inclusion stricte nous allons prouver que le langage $C_1^* L_1 \notin C^f(L)$ où $L_1 = \{x^p y^q z y^q x^p \ / \ p \ ; \ q \geq 1\}$

supposons que $C_1^* L_1 \in C^f(L)$.

Remarque :

Notons que si un langage C est élément de $C^f(C_1 \cup C_2)$ alors il se décompose en $C' \cup C''$ où C' appartient à $C^f(C_1)$ et C" appartient à $C^f(C_2)$.

De cette remarque on détuit naturellement que $C_1 L_1^* = A_1 \cup A_2$ avec A_1 élément de $C^f(\text{Sym}) = \mathcal{L}in$ et A_2 élément de $C^f(B \uparrow c^*) = C^f(B)$. Mais il existe deux entiers n et p vérifiant $A_1 \cap a^n b^n a^p b^p \{a, b, x, y, z\}^* = \emptyset$.

Si ce n'était pas le cas le langage $\gamma(A_1)$ (où γ est l'application séquentielle qui efface toutes les lettres situées derrière la deuxième série de b) serait en fait égal à $C_1^2 = \{a^n b^n a^p b^p / n,p \geq 0\}$ ce qui signifirait que C_1^2 est élément de $\mathcal{C}(A_1) \subset \mathcal{L}in$, ce qui est impossible. Puisque $C_1^* L_1 = A_1 \cup A_2$, $A_2 \cap a^n b^n a^p b^p \{a, b, x, y, z\}^* = C_1^* L_1 \cap a^n b^n a^p b^p \{a, b, x, y, z\}^* = a^n b^n a^p b^p C_1^* L_1$ ce qui implique l'appartenance de $C_1^* L_1$ à $C(A_2)$ inclus lui-même dans $C^f(B)$

or L. Boasson a montré dans [7] que $C_1^* L_1$ n'appartient pas à $C^f(B)$ ce qui nous amène la contradiction.

c) <u>L vérifie $C^{bf}(L) \subsetneq C^f(L)$</u>

Soit Sym' le langage symétrique défini sur trois lettres.

$$\mathrm{Sym}' = \{w\, \bar{w}^R \;/\; w \in \{a, b, c\}^*\}$$

Nous allons prouver que le langage Sym' d^* n'appartient pas à $C^{bf}(L)$.
Comme Sym' d^* est un langage linéaire, il est dans $C^f(\mathrm{Sym}) \subset C^f(L)$.
Supposons qu'il soit également dans $C^{bf}(L)$. La remarque du b) est encore vraie si on remplace C^f par C^{bf} ce qui nous permet d'affirmer que
Sym' $d^* = L_1 \cup L_2$ avec $L_1 \in C^{bf}(\mathrm{Sym})$ et $L_2 \in C^{bf}(B + d^*) \subseteq C(B)$.
Etablissons, tout d'abord qu'il existe un mot w appartenant à $\{a, b\}^*$ tel que $w\, c\, \{a, b, c\}^* \{\bar{a}, \bar{b}, \bar{c}\}^* \bar{c}\, \bar{w}^R d^* \cap L_2 = \emptyset$
En effet, si ce n'est pas le cas alors :

Comme on peut effacer, par transduction rationnelle, dans tout mot de $\{a, b\}^* c \{a, b, c\}^* \{\bar{a}, \bar{b}, \bar{c}\}^* \bar{c} \{a, b\}^*$ le facteur qui commence par la première occurrence de la lettre c et se termine par la dernière occurrence de la lettre $\bar{c}$, ceci nous permet de conclure que Sym $\in C(B)$ famille incluse dans $F_\sigma(\mathrm{Bor})$, la plus petite FAL fermée par substitution et contenant la famille des langages bornés or il est connu [21] que Sym n'appartient pas à $F_\sigma(\mathrm{Bor})$ d'où la contradiction.
Nous pouvons donc écrire que $(w\, c\, \mathrm{Sym}'\, \bar{c}\, \bar{w}^R)\, d^* = L_1 \cap w\, c\, \{a, b, c\}^* \{\bar{a}, \bar{b}, \bar{c}\}^* \bar{c}\, \bar{w}^R d^*$ c'est à dire $(w\, c\, \mathrm{Sym}'\, \bar{c}\, \bar{w}^R)\; d^* \in C^{bf}(L_1) \subset C^{bf}(\mathrm{Sym})$.
Il est montré dans [23] que si L_1 et L_2 sont deux langages infinis définis sur des alphabets disjoints vérifiant $L_1 L_2 \in C^{bf}(\mathrm{Sym})$ alors L_1 et L_2 sont des langages rationnels. Ce résultat permet de conclure que $w\, c\, \mathrm{Sym}'\, \bar{c}\, \bar{w}^R$ est un langage rationnel ce qui est manifestement faux. Le langage L vérifie bien le troisième cas.

Quatrième Cas

$$C^{bf}(L) \subsetneq C^{f}(L) = C^{if}(L) = C(L)$$

il suffit de choisir L = Sym [23].

Cinquième Cas

$$C^{bf}(L) = C^{if}(L) \subsetneq C^{f}(L) = C(L)$$

il suffit de choisir $L = C_1$ (cf. chapitre III).

Sixième Cas

$$C^{bf}(L) = C^{f}(L) \subsetneq C^{if}(L) = C(L)$$

il suffit de choisir $L = L' \uparrow c^*$ où c est une nouvelle lettre et L' est un langage non effaçable.

Septième Cas

$$C^{bf}(L) = C^{if}(L) = C^{f}(L) = C(L)$$

il suffit de choisir pour L un générateur quelconque de ALg ou de Ocl.

Si maintenant on examine la place des sous-cônes $C^{df}(L)$ et $C^{d}(L)$ par rapport aux autres sous-cônes particuliers, on peut se demander quelles conditions doit remplir L pour avoir

$$C^{df}(L) = C^{bf}(L)$$

$$\text{et } C^{d}(L) = C^{if}(L)$$

Il est à remarquer que nous ne connaissons pas, pour l'instant, de langage algébrique pour lequel ces relations ne soient pas vérifiées.

Entre la notion de langage effaçable ($C(L) = C^f(L)$) et la notion de langage très fortement effaçable ($C^d(L) = C^{df}(L)$) il y a place pour la notion de langage fortement effaçable ($C^{if}(L) = C^{bf}(L)$) nous ne savons pas, pour l'instant, s'il existe des langages fortement effaçables non très fortement effaçables.

INDEX

(NOTATION)

I N D E X

(VOCABULAIRE)

BIBLIOGRAPHIE

[1] J.M. AUTEBERT, J. BEAUQUIER, L. BOASSON, M. NIVAT, *Quelques problèmes ouverts en théorie des langages algébriques* , R.A.I.R.O. Informatique Théorique 13 (1979) 363-379.

[2] J. BEAUQUIER, *A remark about a substitution property* (1978) à paraître dans Math. Syst. Theory.

[3] J. BEAUQUIER, *Substitutions de langages linéaires et à compteur* , n° 78-15 du LITP (1978).

[4] J. BEAUQUIER, *Substitutions of semi-AFL's* , 1979, soumis à Theorical Computer Science.

[5] J. BEAUQUIER, *Independance of Linear and One-Counter Generators* , in FCT'79, Budach, ed 1979 45-61.

[6] J. BERSTEL, *"Transductions and Contexte-free Languages"*, Teubner Verlag, 1978.

[7] L. BOASSON, *Langages algébriques, paires itérantes et transductions rationnelles* , Theorical Computer Science, 2 (1976) 209-223.

[8] L. BOASSON et M. NIVAT, *Sur diverses familles de langages fermées par transduction rationnelle* , Acta Informatica, 2 (1973) 180-188.

[9] J.P. CRESTIN, *Langages Quasirationnels* , dans Langages Algébriques, J.P. CRESTIN et M. NIVAT, eds, 1973, 123-166.

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

[11] C.C. ELGOT et J.F. MEZEI, *On Relations Defined by generalized finite automata* , I.B.M. J. Res. Dev. vol 9, (1965) 47-68.

[12] S. GINSBURG, *"The Mathematical Theory of Contex-free Languages"*, Mc Graw Hill, New-York, 1966.

[13] S. GINSBURG, *"Algebraic and Automata-Theoretic Properties of Formal Languages"*, North Holland, 1975.

[14] S. GINSBURG, J. GOLDSTINE et S.A. GREIBACH, *Uniformly Erasable AFL* , J. Comput. System Sc., vol 10, (1975) 165-182.

[15] S. GINSBURG, J. GOLDSTINE et S.A. GREIBACH, *Some Uniformly Erasable Families of Languages* , Theoretical Computer Science, Vol 2 (1976) 29-44.

[16] S. GINSBURG et S.A. GREIBACH, *Abstract Families of Languages* , In Memoirs of the Amer. Math. Soc., vol 87, 1969, 1-32.

[17] S.A. GREIBACH, *Syntaxtic Operators on Full semi-AFLs* , J. Comp. Syst. Sc. 6 (1972), 30-76.

[18] S.A. GREIBACH, *Erasing in Context-free AFLs* Information and Control, 21 (1972), 436-465.

[19] S.A. GREIBACH, *Erasable Contexte-free Languages* , Information and Control, 29, (1975), 301-326.

[20] G. JACOB, *Représentation et Substitutions Matricielles dans la Théorie Algébrique des Transductions* , Thèse Sc. Math., Université Paris VII, Paris 1975.

[21] M. LATTEUX, *Langages Commutatifs* , Thèse Sc. Math, Université de Lille I, Lille 1978.

[22] M. LATTEUX, *Une Note sur le Lemme Syntaxique de Greibach* , Publication n° 106 du Laboratoire de Calcul de l'Université de Lille I (1978).

[23] M. LATTEUX, *Sur les Générateurs Algébriques et Linéaires* (1979) A paraître dans Acta-Informatica.

[24] M. LATTEUX, *Quelques Propriétés des Langages à un Compteur* , Publication de l'Equipe Lilloise d'Informatique Théorique, IT 17.79 (1979).

[25] M. LATTEUX, *A propos du Lemme de Substitution* , Publication de l'Equipe Lilloise d'Informatique Théorique, IT 21.80 (1980).

[26] J. LEGUY, *Transductions Rationnelles Décroissantes* (1979) à paraître dans R.A.I.R.O. Informatique Théorique.

[27] J. LEGUY, *Languages d'Image Finie et Langages Bifidèles* , Publication de l'Equipe Lilloise d'Informatique Théorique, IT 16.79 (1979).

[28] M. NIVAT, *Transductions des Langages de Chomsky* , Ann. Inst. Fourier (Grenoble), 18 (1967) 339-455.

[29] M. NIVAT, à paraître

[30] R.J. PARIKH, *Languages Generating Devices* , M.I.T. Res. Lab. Electron. Quart. Prog. Rept. 60 (1961)

[31] R.J. PARIKH, *On Contexte-free Languages* , JACM 13 (1966), 570-581.

[32] M.P. SCHÜTZENBERGER *Sur une variante des fonctions séquentielles* TCS, 4, (1977), 47-57

[33] J.S. ULLIAN, *Three Theorems concerning Principal AFL* , J. Comp. Syst. Sc. 5, (1971), 304-314.