

AVANT-PROPOS

Qui trop embrasse, mal étreint.

Adage populaire

*On ne peut faire l'amour avec toutes les femmes,
mais il faut essayer.*

Autre adage

Longtemps, je me suis imaginé le bonheur de rédiger cet avant-propos. J'en roulais par avance des phrases entières, pour me donner du cœur à l'ouvrage. Maintenant que j'y suis rendu, je comprends que la tâche n'est pas moins ardue que le cœur de l'ouvrage. Comment justifier la rédaction d'un livre sur la théorie des automates? encore un! et si épais! Justifier? on peut toujours rêver; présenter peut-être.

Étoile de la recherche en informatique dans les années soixante, chapitre obligé de l'enseignement de la discipline dans les années soixante-dix quatre-vingt, la théorie des automates semble avoir disparu des estrades et des amphithéâtres. Et pourtant, on la retrouve, explicitement ou implicitement, au cœur ou en prémisses de nombre des sujets de l'informatique qui font la mode ou l'actualité. Pour tenter une explication, je dirai que la théorie des automates est *l'algèbre linéaire* de l'informatique. Ceci est à double entente. Au sens propre, la théorie des automates *est* de l'algèbre linéaire ou peut être vue comme telle, théorie des matrices à coefficients dans des algèbres idoines. Mais c'est plutôt le sens figuré qui m'intéresse. La théorie des automates comme connaissance de base, fondamentale, connue de tous et utilisée par tous, qui fait partie du « paysage intellectuel » depuis si longtemps qu'on ne l'y remarquerait plus. Et pourtant, elle y est, elle le structure, elle l'organise; la connaître permet de s'y orienter.

Les automates finis sont le modèle de machine le plus simple, si simple qu'il prend des formes, qu'il apparaît dans des contextes, qu'il se cache dans des utilisations aussi nombreux que divers. Je ne décrirai pas les avatars du modèle *ni les applications* de la théorie des automates — tout au plus en évoquerai-je quelques-uns. Je voudrais présenter cette théorie *pour elle-même* et je vais chercher à rendre sa richesse. Mais si ce livre est épais, ce n'est pas seulement, ce n'est pas tellement parce que les automates finis posent tant de problèmes, donnent lieu à tant de résultats — dont je suis d'ailleurs loin de rendre compte en totalité — mais parce que j'ai voulu donner de chaque propriété l'explication la plus directe sans renoncer à la replacer

ensuite dans le cadre le plus général. Les propriétés simples seront démontrées de façon simple puis interprétées comme des cas particuliers de propositions globales exprimées dans des formulations plus abstraites, qu'elles aideront à comprendre. C'est pourquoi, dans la première partie de cet ouvrage, organisée autour des notions de *rationalité* et de *reconnaissabilité*, je déroule trois fois la même histoire, chaque fois avec un point de vue, avec un bagage théorique, différents. La matière prend du relief sous des éclairages croisés.

La seconde partie traite des relations entre mots réalisées par automate fini. Ce sujet est exemplaire de la théorie des automates, de la variété de ses méthodes comme de celle de ses champs d'application. Les automates « avec sortie » sont susceptibles d'une présentation très élémentaire et pourtant certaines de leurs propriétés relèvent des méthodes les plus algébriques. Leur étude illustre l'utilité de chacun des aspects de la théorie qui auront été développés dans la première partie.

Carte

Après le chapitre 0 qui regroupe les définitions des structures qui seront utilisées tout au long du livre, le chapitre I présente une « théorie naïve » des automates finis, celle qui est enseignée dans tous les ouvrages et qui commence — et souvent finit — avec la version élémentaire du théorème dit « de Kleene ». Je me suis interdit d'y utiliser tout autre structure que celle du monoïde libre : pas le moindre morphisme, pas le plus petit semigroupe fini. J'y présente la notion d'expression rationnelle et aborde, prolongement naturel du théorème de Kleene, le problème de la transformation d'une expression en un automate soit, pour prendre une formulation plus vendeuse, la recherche d'un motif dans un texte. Élémentaire ne signifie pas simpliste, et cette théorie est déjà foisonnante. J'ai inclus dans ce chapitre deux belles propriétés combinatoires : la version nécessaire et suffisante du « lemme de l'étoile » et la preuve que la « hauteur d'étoile » d'un langage rationnel peut être arbitrairement grande.

Le chapitre II reprend le sujet avec pour guide l'algèbre — c'est-à-dire d'abord la notion de *morphisme* (de monoïdes, d'automates). La première conséquence de ce point de vue est la distinction entre *action* et *automate*, entre partie *reconnaissable* et partie *rationnelle* d'un monoïde quelconque, ce qui éclaire d'un jour nouveau et la notion d'automate et le théorème de Kleene. L'horizon s'ouvre si largement qu'il faut faire des choix et, en particulier, je n'aborde pas la théorie dite « des variétés » déjà traitée dans plusieurs ouvrages. En revanche, j'ai développé deux points plus originaux : la notion de morphisme d'automates d'abord, qui me permet de définir ce que j'appelle « le revêtement de Schützenberger » d'un automate et qui me servira à plusieurs reprises dans la suite, et la définition de ce que j'appelle « l'automate universel » d'un langage ensuite, qui est la reprise d'une construction due à Conway et qui permet, entre autres, de donner à la fin du chapitre une nouvelle présentation du théorème de McNaughton sur la hauteur d'étoile des langages à groupe. Une section est consacrée à la structure de *bon ordre partiel*, structure fondamentale pour toute l'informatique théorique et qui sera utilisée ici en deux points cruciaux.

J'étudie ensuite la famille des parties rationnelles dans deux structures : *le groupe libre* et *le monoïde commutatif libre*. Dans les deux cas, cette famille est une algèbre de Boole effective et cela explique, au moins en partie, que c'est souvent dans un tel cadre que les automates finis permettent de décrire le comportement de processus dont l'ensemble des états est infini (comme les réseaux de Petri, les automates temporisés, *etc.*) Le groupe libre est aussi la structure algébrique sous-jacente au comportement des *automates à pile*, tous sujets qui ne font pas partie de cet ouvrage.

Le chapitre III reprend encore une fois le sujet à la base. Il s'agit moins de généraliser encore que d'ajouter une dimension avec la prise en compte de la *multiplicité* des calculs : les « langages (formels) » deviennent des *séries formelles*, les actions des représentations (matricielles). L'épaisseur ainsi donnée aux résultats en fait mieux comprendre la nature profonde, les rattache à des domaines mathématiques plus classiques, aux méthodes puissantes. Même si je cherche à donner les énoncés les plus généraux et traite des séries sur les *monoïdes gradués*, c'est le cas des séries sur le monoïde libre, à coefficients dans un corps, qui reste le plus riche et dans lequel il est possible de présenter la théorie de la réduction des représentations (due à Schützenberger).

Les deux derniers chapitres sont consacrés aux relations et fonctions réalisées par automate fini. Le chapitre IV porte sur les relations en général, étudiées d'abord de façon « naïve » — c'est-à-dire en utilisant *a minima* les résultats des chapitres II et III — puis de façon générale, ce qui pose le délicat problème de la définition et du traitement des relations avec multiplicité. À la place de la solution — due à Jacob et reprise par les autres auteurs qui traitent du sujet — de se restreindre aux seules relations dites « régulées », je propose de restreindre la multiplicité aux semi-anneaux que j'appelle « raisonnables », ce qui n'élimine aucun des semi-anneaux usuels et suffit à assurer les deux théorèmes « pivot » de la théorie. Le chapitre se poursuit avec le problème de l'équivalence, équivalence indécidable même dans le cas où l'alphabet de sortie est unaire (théorème dû à Ibarra et Lisovik, indépendamment) mais décidable si on prend en compte les multiplicités, résultat dû à Harju et Karhumäki. Ce dernier repose sur la construction du corps des séries dites « de Malcev-Neumann », résultat dont je m'attache à donner une preuve complète. Deux familles de relations sont ensuite décrites, pour la variété des situations où elles apparaissent : les relations *déterministes* et, surtout, les relations *synchronisées*.

Le chapitre V traite des relations fonctionnelles réalisées par automate fini. Les deux hypothèses ensemble, fonctionnalité et rationalité, donnent en se croisant des résultats de structure remarquables. En particulier, le théorème d'Elgot et Mezei, établi à partir de la construction du revêtement de Schützenberger. Le chapitre se termine sur l'étude des transducteurs séquentiels — remis d'actualité par les chercheurs dans le domaine du traitement computationnel des langues naturelles — et sur la caractérisation, due à Choffrut, des fonctions séquentielles.

Un livre se décrit aussi par ce qu'il ne contient pas. Je ne traite pas des *arbres*, ni donc des automates d'arbres, même si une bonne partie de la théorie s'y étend naturellement et si ces objets sont fort usuels en informatique. Je ne traite pas non plus des *mots infinis* auxquels mes collègues et amis Dominique Perrin et Jean-Eric Pin viennent de consacrer un livre entier. Ces choix ont été délibérés. Je regrette en revanche — mais il faut bien s'arrêter d'écrire à un moment — de n'avoir pas traité des liens entre *logique* et automates finis (et donc, plus accessoirement, des automates alternants), chapitre qui trouverait une place naturelle dans la première partie.

La théorie des variétés déjà mentionnée, l'étude des automates à pile, celle des automates à multiplicité dans des semi-anneaux de type $(\max, +)$, les relations entre systèmes de numération et automates finis sont autant de sujets que je n'ai pas traités non plus, parfois évoqués ou effleurés à l'occasion d'un exemple ou d'un exercice, et qui pourraient être développés sur les bases des notions mises en place dans cet ouvrage. S'il a une suite, ils en constitueront sans doute certains chapitres.

*

Je ne voudrais pas terminer cette présentation sans mentionner tout ce que ce livre doit à l'enseignement et aux écrits du professeur Marcel-Paul Schützenberger, maître subjugant, esprit universel et paradoxal, qui n'eut point de modèle, et reste inimitable. J'ai puisé à beaucoup d'autres sources, j'ai bénéficié des suggestions et avis de nombreux collègues, je reconnais aussi volontiers la filiation de ce livre avec le traité de Samuel Eilenberg : *Automata, Languages and Machines*¹, mais l'influence de Schützenberger est d'une autre nature. Il me semble que, au moins implicitement, Schützenberger a développé dans ses articles, de *A remark on finite transducers* (1961) à *Une propriété de Hankel des relations rationnelles* (1977), une manière de voir les automates — il aurait peut-être dit une « Automatenanschauung » en prenant l'air de se moquer de vous — qu'il illustrait dans ses conférences et surtout dans ses conversations : mettre les automates au centre du motif, y ramener les autres concepts, exprimer les divers énoncés dans ce cadre, traiter enfin de ces automates principalement par leur représentation matricielle. De plus, Schützenberger assignait explicitement² des objectifs à cette théorie algébrique des automates : classifier les problèmes, dégager les concepts, unifier les arguments, appuyer ces derniers sur les résultats fondamentaux des mathématiques. Je pense avoir essayé d'adopter ces objectifs ; le lecteur jugera dans quelle mesure j'ai pu m'en approcher.

Je dois reconnaître enfin que j'ai fait la part belle, essentiellement dans les sections de complément, aux références à mes propres travaux. Si je ne le fais pas pour moi, qui le fera ? et si je ne le fais pas maintenant, quand le ferai-je ? Le lecteur conviendra quand même que je ne me suis pas restreint, tant s'en faut, à mes seuls

¹Qui ramène d'ailleurs aussi à Schützenberger puisqu'Eilenberg termine sa préface en lui rendant grâce ainsi : « every phase of the development was endlessly discussed with him. »

²Dans sa conférence au congrès de l'IFIP, tenu à New York en 1965.

travaux. Et inversement, ces travaux correspondent aux questions qui m'ont attiré pendant la très longue préparation de cet ouvrage.

Légende

Il n'est pas de carte sans légende, protocole qu'il peut être utile d'assimiler avant d'entreprendre le voyage.

De la numérotation L'unité de base est la *section*, subdivision du chapitre, elle-même partagée en *sous-sections* qui sont parfois divisées en *paragraphes*.

L'étiquette de chaque item numéroté est formée du numéro de sa section et de son numéro d'ordre dans sa section. Les propositions, lemmes et corollaires forment une seule classe pour cet ordre, les autres items, théorèmes, définitions, et propriétés, remarques, exemples, exercices, figures et équations sont chacun une classe à eux seuls³. Selon un usage courant, les items sont référencés par leur étiquette à l'intérieur de leur chapitre, par leur étiquette préfixée par le numéro de leur chapitre, écrit en romain comme ci-dessus, quand ils sont appelés depuis un autre chapitre.

La division en sections de chaque chapitre est reproduite à l'intérieur de la section des solutions des exercices. L'étiquette des rares items qui y sont identifiés est préfixée des initiales SE.

Des limites Les énoncés des théorèmes, propositions, lemmes et corollaires, ceux des définitions également, sont en italique. La fin des remarques et des exemples est marquée par un \square en fin de ligne.

Les preuves des énoncés sont ouvertes par un « **Démonstration.** » et leur fin est marquée par un « \blacksquare » en fin de ligne. Quelques preuves, trop courtes, trop simples pour être introduites aussi pompeusement, sont simplement encadrées par deux \blacksquare .

Quand un énoncé n'est pas suivi d'une preuve, qu'il soit la conclusion d'un raisonnement déjà explicité, ou que cette preuve soit si simple, ou renvoyée en exercice, cet énoncé lui-même est ponctué d'un \blacksquare .

Des notes Les notes en marge explicitent les énoncés auxquels il est fait référence dans le texte. Les notes en bas de pages sont plutôt réservées aux commentaires sur la terminologie et les notations adoptées. À la fin de chaque chapitre sont rassemblées les notes « historiques » et les renvois aux références bibliographiques.

Des exercices J'ai mis des exercices à la fin de la plupart des sous-sections et j'ai donné la solution, ou la réponse, à une bonne partie d'entre eux, marqués dans ce

³Ce choix, comme tous les choix, se discute et sera critiqué au motif qu'une numérotation unique facilite la recherche; la critique tombe, puisque les notes en marge renvoient à la page de chaque item cité.

cas par un point en marge. Il y a en gros trois sortes d'exercices (qui ne sont pas repérés comme tels).

Les premiers sont des exercices *pour s'exercer*. Je suis persuadé qu'on n'a compris un énoncé que si on est capable de faire, dans un cas particulier, les calculs qui correspondent à sa preuve dans le cas général et qu'inversement une telle démarche aide puissamment à ladite compréhension. Par ailleurs, j'ai profité en maintes occasions de ces « gammes » proposées au lecteur pour anticiper sur des calculs ou des exemples à venir.

Une seconde catégorie d'exercices consiste en la preuve, laissée au lecteur, de certains énoncés. Je les ai indiqués systématiquement à la fois pour inciter le lecteur à les faire et aussi parce qu'une partie est donnée en solution et qu'il fallait bien les référencer.

Les exercices de la troisième famille permettent d'ouvrir sur des compléments que je n'ai pas voulu traiter, ou pas voulu traiter avec le même soin que le texte principal. Quand ils sont corrigés, ils font partie intégrante du chapitre — c'est la raison pour laquelle les solutions ne sont pas regroupées à la fin de l'ouvrage. Dans plus d'un cas, leur contenu est appelé dans un chapitre suivant.

Remerciements

Il en est des livres (scientifiques) comme des expéditions ; même si on est seul à ramer, on n'arriverait pas au port, et surtout on ne se lancerait pas dans l'aventure si on n'était pas assuré d'une aide extérieure.

Les vivres, ... J'ai eu la chance d'appartenir au Laboratoire Informatique Théorique et Programmation qui pendant 25 ans a été l'un des centres de recherche les plus actifs au monde dans le domaine de la théorie des automates, et tous les aspects de cette théorie que j'ai cherché à rendre y étaient représentés. Non seulement mes collègues du laboratoire ont été des interlocuteurs précieux mais le rayonnement du LITP a permis que j'y rencontre d'autres collègues venus du monde entier.

Depuis que j'ai entrepris la rédaction de cet ouvrage, je n'ai pas cessé de m'adresser à eux pour demander conseil, éclairer un point, expliquer une difficulté, obtenir une référence. Pour leur accueil, leur patience, leur savoir, leurs réponses, pour leurs interrogations et leurs encouragements aussi, je voudrais remercier chaleureusement : F. Bassino, M.-P. Béal, J. Berstel, L. Boasson, J. Brzozowski, Ch. Choffrut, M. Crochemore, Z. Esik, Ch. Frougny, P. Gastin, J. von zur Gathen, R. Gilman, S. Grigorieff, P.-C. Héam, J. van der Hoeven, J. Karhumäki, J. Mairesse, A. Mandel, S. Margolis, Y. Métivier, M. Nivat, K. Peeva, M. Pelletier, D. Perrin, J.-E. Pin, Ch. Prieur, A. Restivo, G. Sénizergues, I. Simon, J. Siquier, M. Tanaka, M. Waldschmidt et P. Weil.

Plusieurs collègues ont pris la peine de relire attentivement une première version sortie en mars 2001. Leurs remarques et commentaires m'ont été une aide précieuse et un puissant encouragement. Que A. Arnold, J. Berstel, J. Brzozowski, S. Crespi-

*Mes excuses à ceux
que j'aurais pu
oublier.*

Reghizzi, G. Duchamp, G. Jacob, J. Justin, J. Néraud, B. Petazzoni, Ch. Reutenauer, P. Simonet, W. Thomas, R. Tijdeman et F. Yvon trouvent ici l'expression de ma reconnaissance.

Je voudrais enfin mentionner tout spécialement O. Carton et V. Bruyère qui m'ont fait l'amitié de me confier des manuscrits alors non publiés, et Sylvain Lombardy, qui fut mon étudiant en thèse, et avec qui j'ai travaillé en constante interaction ces quatre dernières années.

Ce livre est aussi nourri de l'expérience acquise dans les cours que j'ai pu donner pendant toutes ces années, dans le cadre du DEA qui s'est appelé successivement : *Langages, Algorithmes et Programmation, Informatique Théorique, Calcul et Programmation* et enfin *Algorithmique*. J'ai eu également l'occasion d'enseigner cette matière à l'étranger, à l'invitation des professeurs G. Hotz à Sarrebruck, A. Restivo à Palerme, I. Simon à São Paulo, J. Karhumäki à Turku et N. Sabadini à Côme — que j'ai plaisir à remercier ici — et enfin à l'ENST. Mes élèves de DEA des deux dernières années m'ont grandement aidé à corriger et à améliorer la première version ; je suis heureux de remercier particulièrement : L. Danthony, B. Durak, L. Fousse, S. Oudot, et Pham Dong Hieu.

Selon la formule consacrée, les erreurs et coquilles qui restent — et je sais, statistiquement, qu'il y en a encore beaucoup — sont toutes de mon seul fait. Je suis par avance reconnaissant aux lecteurs qui voudront bien me signaler celles qu'ils auront pu remarquer⁴. J'essaierai de tenir à jour une liste d'errata, accessible à partir de ma page⁵ « oueb ».

l'équipement, ... Devant son clavier et son écran, le scientifique est devenu à la fois l'auteur, le dactylographe et le typographe de son texte. Je ne suis pas de ceux qui se réjouissent de cette évolution mais puisque cela est, sans discussion ou révolte possibles, autant remercier ceux qui ont permis que je dispose, gracieusement, d'outils de qualité.

Au premier rang, D. Knuth pour $\text{T}_{\text{E}}\text{X}$ et L. Lamport pour $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$. J'ai utilisé le remarquable éditeur de textes Alpha de P. Keleher, la chaîne de compilation CMacTeX de T. Kiffe et le diabolique paquetage PSTrick de T. van Zandt pour les figures.

Mais il n'est pas vrai que cela marche tout seul ! Je suis très redevable à tous ceux qui, tout au long de ces années, m'ont guidé dans les choix des logiciels, m'ont fourni les bonnes versions, m'ont sorti des impasses et des plantages. Je remercie M. Bovani, R. Card, A. Demaille, P. Gastin, J.-L. Mounier, S. Tardieu, et M. Zeitoun, pour leur aide et leur patience et adresse une pensée spéciale à mon vieil ami Éric Picheral pour sa totale disponibilité.

⁴En m'envoyant par exemple un courriel à sakarovitch@enst.fr.

⁵www.enst.fr/~jsaka.

Une version très préliminaire des chapitres 0 et I est parue dans un manuel⁶ édité par les éditions Ellipses que je remercie de m'avoir laissé libre de la reprendre. J'ai plaisir enfin à remercier Jean-Philippe Moreux des éditions Vuibert pour sa patience et sa compétence face à mes retards d'auteur torturé, à mes lubies d'apprenti typographe et aux démons de L^AT_EX que je n'ai cessé de réveiller.

... **et le couvert.** B. Robinet et J.-P. Tubach m'ont accueilli chaleureusement voici six ans à l'ENST, au sein du laboratoire Traitement et Communication de l'Information; les directeurs successifs du département Informatique et Réseaux, U. Finger et M. Riguidel, ont fait en sorte que je n'y manque de rien; je les en remercie tous très sincèrement.

Je voudrais enfin remercier le Centre national de la recherche scientifique, institution remarquable, souvent décriée en France et que le monde entier nous envie. Sans la liberté d'esprit que mon statut de chercheur m'a donnée, je ne pense pas que j'aurais pu travailler comme j'estimais nécessaire de le faire la matière de ce texte. En manière de retour, j'espère que j'aurai pu, si cet ouvrage emporte la faveur de mes pairs et pour finir enfin sur un ton moins austère, montrer que la recherche n'est pas du temps perdu.

J. S.

Paris, septembre 2003

⁶P. BELLOT ET J. SAKAROVITCH. *Logique et automates*. Ellipses, 1998.

SOMMAIRE

AVANT-PROPOS

PROLOGUE LA MACHINE À DIVISER DE M. PASCAL

CHAPITRE 0 STRUCTURES FONDAMENTALES

Première partie : Les trois stades de la rationalité

CHAPITRE I LA MACHINE LA PLUS SIMPLE ...

Automates sur un monoïde libre

CHAPITRE II PUISSANCE DE L'ALGÈBRE

Automates sur un monoïde quelconque

CHAPITRE III PERTINENCE DE L'ÉNUMÉRATION

Automates avec multiplicité

Deuxième partie : La rationalité dans les relations

CHAPITRE IV RICHESSE DES TRANSDUCTEURS

Automates réalisant une relation entre mots

CHAPITRE V SIMPLICITÉ DES TRANSDUCTEURS FONCTIONNELS

Automates réalisant une fonction de mots

BIBLIOGRAPHIE, INDEX & TABLE

