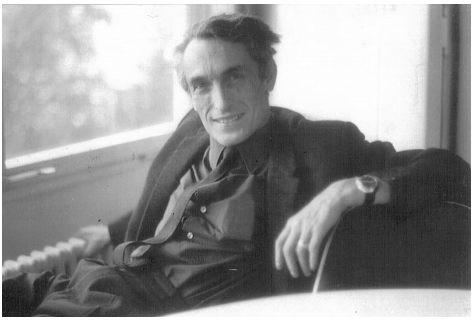
Une introduction à la contribution scientifique de Marcel-Paul Schützenberger

Robert Cori [1] et Dominique Perrin [2]

Il y a 20 ans, au cours de l'été 1996, la maladie finissait par prendre le dessus sur la rayonnante intelligence de Marcel-Paul Schützenberger. Ainsi disparaissait un homme d'une culture exceptionnelle.



Marcel-Paul Schützenberger (1920 –1996)

Sa mort mettait fin à une production scientifique hors du commun qui avait débuté dès 1943 par des travaux d'algèbre, suivis ensuite par de nombreuses publications en médecine et biologie basées sur les règles et méthodes de la statistique mathématique. Par la suite, ses contributions majeures sont en Informatique et en Mathématique Combinatoire.

Un colloque, organisé par ses élèves pour honorer sa mémoire, s'est tenu à Bordeaux au cours du mois de mars dernier. Les contributions de plus de 30 chercheurs junior ou senior ont montré combien ses travaux restent d'actualité. Elles ont montré comment ses idées se sont répandues dans le monde entier et ont aussi mis en avant les très nombreux résultats obtenus par les jeunes qui suivent les voies qu'il a tracées.

1. LaBRI, université de Bordeaux.
2. IGM, université Paris Est.

Nous nous proposons de présenter ici, en tentant d'être accessibles au plus grand nombre, ses principales contributions contenues dans près de 300 publications parues entre 1943 et 2000. Celles-ci ont été réunies, analysées ; elles sont accessibles sur le site : `http://www-igm.univ-mlv.fr/~berstel/Mps/index.html`.

Elles forment un ensemble de 12 volumes dans lesquels elles sont classées par ordre de parution. Une analyse succincte de ces publications apparaît dans l'introduction de chacun des volumes. Cet ensemble de 12 introductions nous a été particulièrement utile pour rédiger ce texte. Nous avons choisi de présenter ici l'ensemble des contributions par thèmes plutôt que par date de parution.

Une des grandes convictions de Schützenberger (souvent abrégé en MPS dans la suite) dès le début des années 60, c'était qu'un développement considérable des outils informatiques ne faisait que débuter. Il nécessitait des concepts mathématiques nouveaux qu'il fallait construire. Les mathématiques de l'époque se focalisaient sur les applications à la mécanique et à la physique. Il fallait donc construire de nouveaux objets mathématiques délaissés par les écoles mathématiques de l'époque. On peut aujourd'hui constater que le développement de l'informatique lui a donné profondément raison, comme l'a dit André Lichnerovicz :

> « *Schützenberger avait le génie de la "préscience" – du génie tout court : il a su dégager très tôt les fondements algébriques et combinatoires qui doivent sous-tendre tout le champ de l'informatique.* »

Les mots

Qui aurait pu penser que la structure mathématique dont l'objet de base est une suite de lettres, et dont l'opération principale est la concaténation de telles suites, pouvait donner lieu à des questions difficiles et permettre d'obtenir des théorèmes profonds qui seraient utilisés par la suite dans de nombreux domaines ?

En effet la structure algébrique sur les mots est celle de monoïde : un ensemble doté d'une opération dont la seule propriété est d'être associative. Pas grand chose à tirer de cette structure, dirait tout mathématicien formé par Bourbaki. Pourtant les travaux de Schützenberger et de nombreux algébristes lui ont donné un lustre certain ces dernières années. La définition d'une structure algébrique par des générateurs et des relations a ainsi rendu centrale la notion de mot, montrant la nature visionnaire de l'approche de Schützenberger. Sa collaboration avec R. C. Lyndon pour la résolution d'équations dans le groupe libre fait partie de ces élégants résultats qui mêlent combinatoire et algèbre (voir [21] par exemple).

Les questions algorithmiques sur la recherche de facteur dans un mot ont constitué un des problèmes récurrents en bioinformatique. Les nombreux algorithmes largement utilisés actuellement font appel à des constructions que lui et ses élèves ont introduites.

Un exemple frappant de l'influence que les travaux de MPS en combinatoire des mots ont eu par la suite est la *conjecture de Dejean*. Dans sa thèse, préparée sous la direction de MPS, Françoise Dejean a formulé une fascinante conjecture, qui a été résolue plus de 40 ans plus tard (à la suite d'une série de travaux dont nous donnons ci-dessous en détail la liste pour témoigner de l'effort collectif qui a été réalisé à cette occasion).

En 1906 et 1912, le mathématicien norvégien Axel Thue a donné un exemple d'un mot infini sur deux lettres ne contenant pas de cube (un cube étant la répétition de trois facteurs identiques, comme 'abaabaaba' par exemple) et d'un mot infini sur trois lettres ne contenant pas de carré (voir [19]). Ce type de question, envisagé par Thue comme un pur défi pour l'esprit, se trouve aujourd'hui fréquemment étudié, en particulier en bioinformatique où les carrés dans les mots, connus sous le nom de « *tandem repeats* », jouent un rôle important pour l'analyse des chaînes moléculaires.

Plus généralement, on peut essayer d'éviter des *puissances rationnelles*, et pas seulement des puissances entières. On dit pour cela qu'un mot w a exposant p/q si $w = x^k y$ où y est un préfixe de x et que la longueur de w est p/q fois la longueur de x. Par exemple, le mot 'entente' a exposant $7/3$ (avec x='ent', $k = 2$ et y='e'). La conjecture de Dejean concerne le plus grand exposant possible e tel qu'il n'existe aucun mot infini sur un alphabet à k lettres qui évite les puissances d'exposant e ou plus. Cet exposant est noté RT(k). Le sigle RT signifie « seuil de répétition » et provient de l'anglais « *repetition threshold* ». Le résultat de Thue peut se reformuler en $\mathrm{RT}(2) = 2$, car tout mot assez long sur un alphabet à deux lettres contient un carré, et il existe des mots infinis (le mot de Thue-Morse par exemple) dont aucun facteur n'est une puissance d'exposant strictement supérieur à 2. En 1972, Françoise Dejean [12] démontre que $\mathrm{RT}(3) = 7/4$. En d'autres termes, sur un alphabet à trois lettres, tout mot assez long (et Françoise Dejean a vérifié que tout mot de longueur 39 est de cette nature) contient une répétition d'exposant 7/4 ou plus, et il existe un mot infini sur trois lettres dont aucun facteur n'a un exposant strictement supérieur à $7/4$. Dans le même travail, elle conjecture que $RT(4) = 7/5$ (résultat prouvé par Pansiot en 1984) et que $RT(k) = k/(k-1)$ pour $k \geq 5$. C'est cette dernière affirmation qui est la conjecture de Dejean, démontrée en 2009, qui devient ainsi le théorème de Dejean. En résumé, on a :

Le seuil de répétition $RT(k)$ prend les valeurs suivantes :

$$
\begin{aligned}
&RT(2) = 2,\\
&RT(3) = 7/4,\\
&RT(4) = 7/5,\\
&RT(k) = k/(k-1), \text{ pour } k \geq 5.
\end{aligned}
$$

Un mot infini qui atteint ce seuil de répétition est appelé mot de Dejean. Le premier résultat, après celui de Françoise Dejean, est dû à Jean-Jacques Pansiot qui montre que $RT(4) = 7/5$. Moulin Ollagnier a prouvé la conjecture pour $5 \leq k \leq 11$;

Mohammad-Noori et Currie l'ont prouvé pour $12 \leq k \leq 14$. Une véritable percée est intervenue quand Carpi a prouvé la conjecture pour tout $k \geq 33$. Currie et Rampersad ont ensuite raffiné la construction de Carpi pour l'étendre à $k \geq 27$. Plus récemment encore, ils ont utilisé les techniques de Moulin Ollagnier pour résoudre les cas restants, $15 \leq k \leq 26$. Par une autre technique, et presque simultanément, une preuve des cas restants a été donnée par Michaël Rao [3].

Les mots de Lyndon et la théorie des factorisations

Une des notions importantes sur les mots est celle de *mot de Lyndon*. La définition est très simple : c'est un mot x tel qu'en permutant circulairement ses lettres, on obtient toujours un mot strictement plus grand que x dans l'ordre alphabétique. Ainsi, 'abracadabra' n'est pas un mot de Lyndon parce qu'en faisant passer sa dernière lettre devant, on obtient 'aabracadabr', qui est moins agréable à prononcer, mais plus petit dans l'ordre alphabétique.

La notion a été introduite par R.C. Lyndon [20] et il suit de ses résultats [7] que tout mot peut se factoriser de façon unique en une suite décroissante de puissances de mots de Lyndon. Ainsi le mot 'abracadabra' se factorise en 'abracad abr a' où les trois mots 'abracad', 'abr' et 'a' sont des mots de Lyndon avec 'abracd'> 'abr' > 'a' dans l'ordre alphabétique.

Ce résultat a été à la base de la théorie, développée par MPS en 1965 [26], des *factorisations de monoïdes libres*. Une telle factorisation est donnée par un ensemble X totalement ordonné de mots sur un alphabet fixé A tel que tout mot w sur A s'écrive de façon unique

$$w = x_1 x_2 \cdots x_n$$

avec $x_1, \ldots, x_n$ dans X et $x_1 \geq x_2 \geq \ldots \geq x_n$. Ainsi les mots de Lyndon fournissent un exemple d'une telle factorisation. De nombreux autres exemples avaient été étudiés par ailleurs, en particulier dans la théorie des algèbres de Lie (voir la section Codes).

Le résultat principal de [26] dit en particulier que si X est une factorisation du monoïde libre sur A, tout mot sur A a un unique transformé par permutation circulaire de ses lettres égal à une puissance de l'un des éléments de X (comme c'est évidemment le cas pour les mots de Lyndon).

Les factorisations de monoïdes libres ont été étudiées par la suite par Viennot [29] (voir aussi [23] ou [1]). Une application intéressante des mots de Lyndon est la transformation de Burrows-Wheeler [4].

3. Le paragraphe ci-dessus est reproduit de `https://fr.wikipedia.org/wiki/Utilisateur:ManiacParisien/Brouillons/Math-12`. Voir aussi le blog de J. Shallitt `http://recursed.blogspot.fr/2009/05/dejeans-conjecture-solved.html`.

4. Voir `https://en.wikipedia.org/wiki/Burrows%E2%80%93Wheeler_transform`

Les algorithmes de la combinatoire

L'objet combinatoire *permutation* a beaucoup fasciné M.-P. Schützenberger, probablement parce qu'il est central en informatique et en mathématiques. En effet, dans les années entre 1960 et 1970, il était courant de dire que plus de la moitié du temps de calcul d'un ordinateur était occupé à trier des données. Il était donc important d'évaluer les performances des algorithmes de tri. On utilise pour cela des permutations tirées au hasard uniformément et on applique l'algorithme de tri à analyser à celles-ci, en évaluant le nombre moyen de comparaisons et d'échanges de valeurs effectués en fonction de la taille de la permutation. Dans ces analyses interviennent des paramètres classiques sur les permutations : par exemple le nombre d'inversions, de maxima partiels, de montées, les longueurs des sous-suites croissantes.

La publication principale de MPS dans ce domaine est l'opuscule, co-écrit avec Dominique Foata, intitulé *Traité Géométrique des Polynômes Eulériens* [16]. Il reprend les résultats classiques de détermination du nombre de permutations suivant les paramètres dont il est question plus haut. Plusieurs ont été obtenus par des mathématiciens célèbres : Euler, Bernouilli, Mac Mahon, Stirling. Les méthodes utilisées dans cet opuscule sont particulièrement élégantes et nouvelles. Elles font intervenir des arguments combinatoires (que les auteurs appellent géométriques) qui évitent les longs calculs analytiques des anciens. L'outil principal mis en avant est celui des bijections. Il consiste, pour obtenir une formule donnant le nombre d'objets d'un ensemble E donné, à trouver un autre ensemble F dont on soupçonne qu'il a le même nombre d'éléments que E et pour lequel on dispose d'une formule donnant son nombre d'éléments. Il s'agit ensuite de construire une bijection entre E et F. On voit que cette technique a une application en terme algorithmique, par exemple tirer au hasard un objet dans l'ensemble F peut se réaliser en tirant un objet dans l'ensemble E puis en appliquant la bijection entre les deux ensembles qui, si elle est simple, donne lieu à un algorithme efficace.

Décrivons un exemple de transformation qui est centrale dans le traité de Foata et Schützenberger. Elle permet par exemple de montrer que le nombre de permutations de taille n présentant k montées est égal au nombre de celles présentant k excédants. Ces nombres sont appelés nombres eulériens. Une permutation α se représente par une suite

$$a_1, a_2, \ldots, a_n$$

où les a_i sont tous les nombres de l'ensemble $E_n = \{1, 2, \ldots, n\}$. On peut aussi la considérer comme une application de cet ensemble dans lui-même en notant $\alpha(i) = a_i$. Une montée est un i tel que $a_i < a_{i+1}$ et un excédant est un i tel que $a_i > i$. La bijection fondamentale consiste à découper la suite des a_i en déterminant les minima partiels c'est-à-dire les $a_{i_1}, a_{i_2}, \ldots, a_{i_k}$ tels que $a_{i_j} < a_p$ pour tous les $p < i_j$. Les sous-suites qui commencent par un minimum partiel et se terminent juste avant le suivant forment alors les cycles d'une permutation β dont le nombre d'excédants est égal au nombre de montées de α. Plus précisément, on peut définir β en posant

$$\beta(a_i) = \begin{cases} a_{i+1} & \text{si } i < n \text{ et } a_{i+1} \text{ n'est pas un minimum partiel,} \\ a_j = \min_{p \leqslant i} a_p & \text{sinon.} \end{cases}$$

Par exemple, la permutation $\alpha = 5,2,7,4,1,8,6,3$ a deux montées $i = 2,5$. On a trois minima partiels pour $i = 1,2,5$ et on trouve sous forme de cycles $\beta = (5)(274)(1863)$, c'est-à-dire, sous forme de suite $\beta = 8,7,1,2,5,3,4,6$, qui a deux excédants pour $i = 1,2$.

Cette méthode, dite bijective, est devenue très à la mode en combinatoire et, à la suite de Schützenberger, de nombreux chercheurs ont tenté, certains avec grand succès, d'obtenir des preuves par bijection de formules fort complexes.

Les contributions de MPS dans le domaine de l'algorithmique et de la combinatoire des permutations sont loin de s'arrêter là. En effet, motivé par l'article de C. Schensted paru en 1961 [24], qui contient un algorithme de recherche de la plus longue sous-suite croissante d'une suite, il étudie en détail la construction proposée et obtient des résultats remarquables sur les tableaux de Young. Ceux-ci sont largement cités, en particulier par D. Knuth dans *The Art of Computer Programming (Vol. 3)* [18]. Il est remarquable de noter que cette construction avait aussi été obtenue par G. de B. Robinson [5] (20 ans plus tôt) en algèbre pour l'étude des représentations linéaires du groupe symétrique ; toutefois, ce dernier n'avait pas considéré l'aspect algorithmique de la recherche de sous-suites croissantes, dont les relations avec les considérations algébriques sur les groupes sont loin d'être évidentes.

Quelques années plus tard, Schützenberger revient sur les tableaux de Young en introduisant l'opération dite de *promotion*, qu'il illustre en utilisant le jeu de taquin et qui se révélera être un outil fécond en combinatoire algébrique. Ces développements sont décrits plus loin.

Les automates

On doit à Schützenberger l'idée de *multiplicité* dans les modèles formels comme les automates ou les grammaires [25]. L'idée de départ est de remplacer la notion booléenne de reconnaissance d'un objet (oui ou non suivant que l'objet est reconnu ou pas) par l'introduction du nombre de façons dont l'objet est reconnu.

Cette idée conduit naturellement à celle d'automate ou de grammaire avec multiplicités. On est ainsi conduit à remplacer la notion de langage formel par celle de multi-ensemble de mots, qui est une fonction associant à chaque mot une multiplicité, pouvant être un entier ou un nombre réel représentant une probabilité.

5. MPS soutenait que les initiales G. de B. étaient une abréviation pour « Gueule de Bois ». De même, R.P. (Richard Peter) Stanley était devenu pour lui le « Révérend Père » Stanley.

La beauté de la chose est que la notion de langage rationnel va ainsi être remplacée par celle de série rationnelle en variables non commutatives. On retrouvera les séries rationnelles classiques en une variable dans le cas d'un alphabet à une seule lettre. Les opérations booléennes sur les langages vont être remplacées par des opérations algébriques (la somme remplaçant l'union) et l'étoile X^* par l'opération de quasi-inverse

$$(1-X)^{-1} = 1 + X + X^2 + \dots$$

définie pourvu que la multiplicité du mot vide dans X soit 0.

De même, la notion de grammaire *context-free* sera remplacée par celle de système d'équations algébriques. Ainsi la grammaire définissant les mots bien parenthésés s'écrira

$$X = (X)X + 1$$

au lieu de $X \to (X)|\varepsilon$ (voir à ce sujet la section Langages Formels).

D'autre part, la notion d'automate fini est essentiellement équivalente à celle de représentation linéaire du monoïde libre, ce qui conduit à la généralisation du théorème de Kleene, souvent nommée théorème de Kleene-Schützenberger. Cette notion de rationalité non commutative conduit aussi à la notion d'inverse de matrices de séries formelles. Ceci conduit au principe d'algèbre non commutative selon lequel, pour construire les corps gauches, il faut inverser des matrices et pas seulement des éléments (comme, dans le cas commutatif, pour construire les rationnels à partir des entiers). Ce principe a été abondamment utilisé par P. M. Cohn [11] (pour la construction du corps libre) et par ses successeurs.

Ce point de vue a été considérablement développé depuis. La théorie des ensembles avec multiplicité dans un semi-anneau a été formulée en toute généralité par Eilenberg dans son livre *Automata, Languages, and Machines (Vol. A)* [13]. Sur les séries rationnelles non-commutatives, on peut consulter [2] et sur les séries algébriques [15].

Les semi-groupes (monoïdes)

Dans les années d'après-guerre, les mathématiciens algébristes influents étaient en France les membres du groupe Bourbaki. Les structures algébriques qui ne comportent qu'une seule opération (treillis, monoïdes, et même groupes) ne faisaient pas partie de leurs préoccupations. Le problème de la classification des groupes finis mobilisait des énergies surtout aux États-Unis. Ainsi la structure de semi-groupe (on disait aussi demi-groupe) — ensemble muni d'une opération qui ne vérifie que la loi d'associativité (ou celle de monoïde qui possède en plus un élément neutre) — était considérée comme particulièrement pauvre et ne devait pas mériter l'attention d'un mathématicien sérieux.

Allant à l'encontre de cette tendance, Schützenberger, qui avait le goût de l'originalité, a choisi de considérer pour ses premières contributions significatives en algèbre les « *block designs* » (publication en 1949), les treillis (en 1953) et les semi-groupes (en 1957).

En ce qui concerne les semi-groupes, il a cherché après d'autres à trouver comment des groupes pouvaient être construits afin de rendre compte des propriétés de cette structure. Un des premiers résultats notables en théorie des semi-groupes est la découverte par J. A. Green de plusieurs relations binaires qui permettent de diviser tout semi-groupe en classes d'équivalences ayant des propriétés algébriques intéressantes. Une des relations de Green, notée $\mathcal{H}$, a la propriété suivante : si une $\mathcal{H}$-classe contient un idempotent e (élément tel que $e.e = e$) alors elle constitue un sous-groupe du semi-groupe. Le premier résultat de Schützenberger est d'avoir généralisé ce résultat, en montrant que l'on pouvait toujours associer un groupe aux $\mathcal{H}$-classes d'un semi-groupe, même si elles ne contiennent pas d'idempotent. Ce résultat figure en bonne place dans l'ouvrage classique de Clifford et Preston [10], dans lequel une section s'intitule : « *The Schützenberger group* ». L'utilisation des automates en informatique et les développements récents de la modélisation des phénomènes physiques par des systèmes dynamiques discrets ont montré que la notion de semi-groupe était centrale dans les mathématiques permettant de modéliser ces phénomènes. Ainsi, l'article de Schützenberger paru en 1957 est cité par B. Bond et L. Levine [3] dans une série d'articles de référence (*Abelian Networks*) sur les modèles combinatoires des phénomènes physiques à paramètres discrets.

À peu près en même temps que Schützenberger découvrait ce groupe associé à un semi-groupe, les travaux de Kleene, Rabin et Scott créaient la théorie des automates : se trouvaient ainsi définis des semi-groupes correspondant aux transformations réalisées sur ces automates. L'introduction du semi-groupe (appelé *syntaxique*) associé à un automate est due à Schützenberger. Elle permet de donner des propriétés de l'automate et du langage qu'il reconnaît en étudiant celles de son semi-groupe syntaxique.

Un des plus importants résultats est le théorème qui affirme qu'un langage est exprimable par des opérations d'union, produit et complémentaire à partir des langages finis (les langages *sans étoile*) si et seulement si le semi-groupe associé à son automate minimal n'admet que des sous-groupes ne contenant qu'un seul élément (les langages *apériodiques*). Ce résultat, démontré en [27], est considéré par Eilenberg comme « *next to Kleene's Theorem, probably the most important result dealing with recognizable sets* » [14].

Il a été le point de départ d'un grand nombre de contributions récentes faisant intervenir de l'algèbre et de la logique. Le lien avec la logique est le point de départ d'un voyage passionnant. Les automates finis permettent de définir des propriétés des mots exprimables dans une logique qui s'appelle la logique monadique du second ordre. On admet dans cette logique de faire porter les quantificateurs sur des objets du second ordre à condition qu'ils désignent des sous-ensembles (et pas des relations binaires par exemple). On interprète les variables comme des entiers avec la relation

$<$ et les relations $a(n)$ pour exprimer que la lettre d'indice n est a. L'équivalence entre les deux notions a été établie par Büchi [5]. On doit à Robert McNaugthton d'avoir montré que les langages sans étoile sont ceux qui sont définissables dans la théorie correspondante du premier ordre. Ainsi, on pourra définir le langage des mots contenant au moins un b par la formule $\exists n b(n)$. Le théorème de MPS montre qu'on peut décider si un langage rationnel est définissable au premier ordre. Par exemple, le langage formé des mots de longueur paire, qui est bien entendu rationnel, n'est pas apériodique, donc pas sans étoile, et donc pas définissable au premier ordre. Chapeau bas !

Les codes

L'un des sujets favoris de MPS a été celui des codes à longueur variable. C'est un domaine né des travaux de Shannon qui soulève, sous la forme la plus élémentaire, le problème de l'ambiguïté, ou son contraire, la non-ambiguïté. En effet, un code est un ensemble X de mots tel qu'on ne peut pas trouver de produit $x_1 x_2 \cdots x_n$ de n mots de X égal à un autre produit $y_1 y_2 \cdots y_m$ de m mots, sans que $n = m$ et $x_1 = y_1$, $\ldots$, $x_n = y_n$. Il s'agit donc de la non-ambiguïté de l'expression X^*. Par exemple $\{a, ba\}$ est un code mais $\{a, ab, ba\}$ n'en est pas un puisque $a(ba) = (ab)a$. Un cas particulier très simple est celui des codes préfixes : aucun mot n'est le début d'un autre.

MPS a formulé une série de conjectures sur les codes et en particulier sur les codes maximaux qui sont ceux pour lesquels le rajout d'un mot détruit la propriété d'être un code. On peut facilement vérifier qu'un code sur un alphabet de k lettres est maximal en calculant sa somme de Kraft, $\sum_{n \geq 1} u_n k^{-n}$, où u_n est le nombre de mots du code de longueur n, et en vérifiant que celle-ci vaut bien 1. La principale conjecture de MPS sur les codes maximaux est toujours non résolue. Elle dit que tout code maximal fini peut, par réarrangement des lettres dans les mots du code, être transformé en un code préfixe. Ainsi $X = \{aa, ab, aab, abb, bb\}$ est un code maximal (on le vérifie en calculant la somme de Kraft $3 \times 1/4 + 2 \times 1/8 = 1$). On le réarrange en un code préfixe comme ci-dessous :

$$aa, ab, aab, abb, bb \;\rightarrow\; aa, ab, baa, bab, bb.$$

(Sur cette conjecture et ce qui est connu aujourd'hui, voir [1].)

La notion de code (et donc plus généralement de non-ambiguïté) est encore pleine de mystères. On sait facilement vérifier si un ensemble fini est un code mais on ne connaît aucun procédé direct permettant de les obtenir tous. Par exemple, on sait qu'il n'y a qu'un nombre fini de codes maximaux sur un alphabet donné ayant un nombre donné d'éléments. Mais on ne connaît pas de formule permettant d'exprimer ce nombre (par exemple comme solution d'une formule de récurrence).

La contribution essentielle de MPS à ce domaine est d'avoir montré que l'étude des codes était liée au domaine de l'algèbre non commutative. En particulier, il a montré que les propriétés de synchronisation intervenant dans des familles de codes nommés *comma-free*, ou plus généralement celle des codes circulaires, étaient liées à celles des algèbres de Lie libres. Les codes circulaires sont ceux pour lesquels le décodage peut s'effectuer de façon unique pour un mot lu sur un cercle. Ainsi $\{ab, ba\}$ est un code mais n'est pas un code circulaire, puisque le mot ab écrit circulairement se décompose de deux façons comme une suite de ab ou comme une suite de ba. Quant à l'algèbre de Lie libre $L(A)$ sur un alphabet A, elle est obtenue à partir des lettres en utilisant la somme ordinaire mais en remplaçant le produit par le *crochet de Lie* :

$$[xy] = xy - yx.$$

Le degré de $[xy]$ est la somme des degrés de x et y (les lettres ayant degré 1). On obtient ainsi une algèbre non commutative (on a même toujours $[xy] + [yx] = 0$) et non associative.

L'un des résultats obtenus par MPS [28], répondant à une conjecture de Golomb et Gordon [17], dit qu'il existe un code circulaire ayant pour tout $n \leq N$, u_n mots de longueur n si et seulement si l'algèbre de Lie libre $L(B)$ sur un alphabet B ayant $\sum_{n \leq N} u_n$ éléments peut être plongée dans $L(A)$ par un isomorphisme qui envoie u_n éléments de B sur des éléments de degré n de $L(A)$. Ainsi, il existe un code circulaire formé d'une lettre et de deux mots de longueur 3 (comme $\{b, aab, abb\}$) parce qu'avec $B = \{u, v, w\}$, on plonge $L(B)$ dans $L(A)$ par le morphisme $u \to a$, $v \to [a[ab]]$, $w \to [[ab]b]$.

Ce point de vue a été largement développé à sa suite par Viennot [29] et par Reutenauer [23].

Les langages formels

Dans ce domaine, c'est certainement l'article en commun avec Noam Chomsky, *The algebraic theory of context-free languages* [8], qui est le plus connu, en tout cas le plus cité. Il a fait l'objet d'une traduction en allemand et d'une traduction en français [9], peut-être dans d'autres langues. Cet article esquisse, et même décrit de manière assez détaillée, une théorie des langages dits *context-free*, qui serait algébrique plutôt que constructive ou combinatoire. Par exemple, une grammaire pour un langage est considérée comme un système d'équations, les variables de la grammaire jouant le rôle des inconnues. Les solutions de ce système d'équations sont calculées par approximations successives (c'est la méthode des points fixes) dans l'algèbre des séries formelles non commutatives, et prennent ainsi en compte les ambiguïtés éventuelles dans les dérivations. L'article contient de nombreux exemples et des propriétés de clôture. Il contient aussi la définition des langages de Dyck, des langages rationnels locaux, et le fameux théorème de Chomsky-Schützenberger selon lequel tout langage algébrique est l'image homomorphe de l'intersection d'un langage de Dyck et d'un langage rationnel. L'article contient aussi des résumés ou des citations de nombreux résultats démontrés par M.-P. Schützenberger dans d'autres articles.

La vision d'un langage comme une série formelle en variables non commutatives proposée par MPS a eu des répercussions dans le domaine de la combinatoire énumérative. En effet, MPS a remarqué qu'en rendant commutatives les variables intervenant dans les équations d'une série correspondant à un langage *context-free*, on obtient une série génératrice énumérant les mots du langage par longueur. De plus, cette série est le développement de Taylor d'une fonction algébrique (vérifiant une équation ne faisant intervenir que des polynômes). Une des premières applications est d'obtenir le nombre de mots de Dyck de longueur donnée en résolvant une équation du second degré.

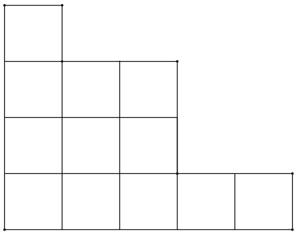
La technique prônée par MPS pour énumérer certains objets combinatoires consiste à trouver une bijection entre ces objets et les mots d'un langage *context-free*, à écrire l'équation correspondant à la grammaire, la résoudre et calculer les coefficients de Taylor au voisinage de l'origine pour la fonction algébrique obtenue. Cette méthode a connu un certain succès pour l'énumération de diverses familles de graphes planaires et permis de retrouver ainsi des formules obtenues par W. Tutte. MPS pensait que planarité et algébricité des séries énumératives étaient deux facettes d'une même propriété combinatoire. Des résultats relativement récents ont montré que cela n'était pas tout à fait le cas puisque des graphes sur le tore ou sur des surfaces de genre plus élevé sont aussi énumérées par des séries algébriques. De fait, l'intuition de Schützenberger est correcte si on remplace la contrainte de planarité sur les graphes par celle d'être plongés sur une surface de genre donné inférieur à une constante (voir [6]).

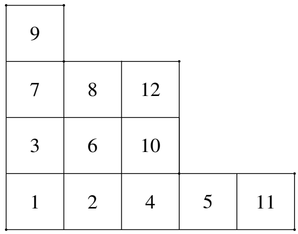
Cette méthode générale a été utilisée pour énumérer un grand nombre de familles d'objets. Elle est exposée dans un article de Mireille Bousquet-Mélou [4] ainsi que dans l'ouvrage de Philippe Flajolet et Robert Sedgwick [15].

La combinatoire algébrique

À partir de 1977, les travaux de recherche de Schützenberger sont principalement consacrés à des questions de combinatoire liées à la représentation des groupes. Alain Lascoux, géomètre algébriste qui avait travaillé dans la mouvance de Grothendieck, le rejoint pour participer avec plusieurs autres mathématiciens dont Gian Carlo Rota, Richard Stanley, Adriano Garsia, à la création de cette discipline maintenant en vogue : la Combinatoire Algébrique. L'objet initial dans le développement de ces travaux est le *tableau de Young*, dont il a été déjà question. Un tableau est une forme géométrique associée à une partition de l'entier n, c'est-à-dire une suite d'entiers

$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ faiblement décroissante (i. e. $\lambda_i \leqslant \lambda_{i-1}$ pour $1 < i \leqslant k$). On dessine k lignes composées chacune de λ_i cases les unes au dessus des autres, comme dans la figure ci-dessous représentant la partition $(5,3,3,1)$ de 12 :





Un tableau de Young (standard) de forme λ est ensuite construit en plaçant les entiers entre 1 et n dans les cases de façon telle que les entiers sont strictement croissants, dans une même ligne de gauche à droite, et dans une même colonne de bas en haut. Un exemple de tableau de Young est donné ci-dessus.

Un algorithme de construction d'un tableau de Young à partir d'une permutation a permis à Schensted de déterminer la longueur de sa plus longue sous-suite croissante et de sa plus longue sous-suite décroissante. Ce sont respectivement le nombre de cases de la première ligne et le nombre total de lignes. L'algorithme de Schensted utilise une procédure d'ajout successif d'un élément de la permutation qui peut constituer un exemple intéressant de programmation pour des étudiants en licence d'informatique. Schützenberger en a proposé un autre, qu'il a appelé *jeu de taquin,* et qui lui a permis de proposer une nouvelle approche à ces questions en introduisant un monoïde défini par générateurs et des relations qui peuvent être interprétées comme des commutations conditionnelles. Ce monoïde a été par la suite appelé *plactique*.

Une propriété remarquable des tableaux de Young, qui les a fait entrer dans le monde des algébristes, en est le lien avec une famille de polynômes symétriques à plusieurs variables. Il s'agit de ce qu'on appelle les *fonctions de Schur*. Pour les définir, il faut généraliser la définition de ces tableaux en considérant les tableaux non standards pour lesquels les entiers peuvent apparaître plusieurs fois et la condition sur les valeurs dans une ligne est la croissance au sens large. Un tableau non standard est donné ci-dessous. D. E. Littlewood a remarqué que les fonctions de Schur, définies généralement par des déterminants assez difficiles à manipuler, pouvaient être retrouvées en associant à tout tableau de Young T un monôme en les variables $x_1^{i_1}, x_2^{i_2}, , \ldots, x_n^{i_n}$, tel que i_j est le nombre de fois où le nombre j apparaît dans T. Par exemple, $x_1^2 x_2 x_3^3 x_4^2 x_5^3 x_7$ est le monôme associé au tableau suivant.

5				
4	5	7		
2	3	4		
1	1	3	3	5

La fonction de Schur associée à une partition λ est alors la somme de tous les monômes associés aux tableaux de Young de forme λ. Cette formulation a été utilisée par Schützenberger pour donner une preuve combinatoire illuminante de la formule de Littlewood-Richardson qui permet de retrouver les coefficients du produit de deux fonctions de Schur de formes λ et ν. Cette preuve fait intervenir le monoïde plactique, ce qui a valu à Schützenberger la qualification suivante dans la préface de l'ouvrage de référence sur les fonctions symétriques écrit par I. G. Macdonald [22] :

> « *In the past years the combinatorial structure based on the jeu de taquin has become much better understood. Schützenberger, the main architect of this theory, has recently published a complete exposition.* »

Il s'agit là d'une référence à un article de 1979. Entre 1979 et 1996, Lascoux et Schützenberger ont contribué à donner d'autres titres de noblesse à cette théorie basée sur le jeu de taquin. Ils ont proposé de nombreux résultats profonds sur des polynômes considérés en Géométrie Algébrique, en liaison par exemple avec les travaux de Grothendieck.

Remerciements. Les auteurs tiennent à remercier Christophe Reutenauer pour l'aide qu'il leur a apportée.

Références

[1] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and automata*, volume 129 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2010.

[2] Jean Berstel and Christophe Reutenauer. *Noncommutative rational series with applications*, volume 137 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2011.

[3] Benjamin Bond and Lionel Levine. Abelian networks II : halting on all inputs. *Selecta Math. (N.S.)*, 22(1) :319–340, 2016.

[4] Mireille Bousquet-Mélou. Rational and algebraic series in combinatorial enumeration. In *International Congress of Mathematicians. Vol. III*, pages 789–826. Eur. Math. Soc., Zürich, 2006. `https://arxiv.org/abs/0805.0588`.

[5] J. Richard Büchi. On a decision method in restricted second order arithmetic. In *Logic, Methodology and Philosophy of Science (Proc. 1960 Internat. Congr .)*, pages 1–11. Stanford Univ. Press, Stanford, Calif., 1962.

[6] Guillaume Chapuy, Michel Marcus, and Gilles Schaeffer. A bijection for rooted maps on orientable surfaces. *SIAM J. Discrete Math.*, 23(3) :1587–1611, 2009.

[7] K.-T. Chen, R. H. Fox, and R. C. Lyndon. Free differential calculus. IV. The quotient groups of the lower central series. *Ann. of Math. (2)*, 68 :81–95, 1958.

[8] Noam Chomsky and Marcel-Paul Schützenberger. The algebraic theory of context-free languages. In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 118–161. North-Holland, Amsterdam, 1963. Traduction : "Algebraische Theorie kontextfreier Sprachen", Kibern. Sb., Nov. Ser. 3, 195-242 (1966).

[9] Noam Chomsky and Marcel-Paul Schützenberger. Théorie algébrique des langages "context-free". In Maurice Gross, editor, *Les modèles en linguistique*, volume 9 of *Langages*, pages 77–118. Didier/Larousse, mars 1968. Traduit par G. Fauconnier.

[10] A. H. Clifford and G. B. Preston. *The Algebraic Theory of Semigroups. Vol. I.* Mathematical Surveys, No. 7. American Mathematical Society, Providence, R.I., 1961.

[11] P. M. Cohn. *Free ideal rings and localization in general rings*, volume 3 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2006.

[12] Françoise Dejean. Sur un théorème de Thue. *J. Combinatorial Theory Ser. A*, 13 :90–99, 1972.

[13] Samuel Eilenberg. *Automata, Languages, and Machines. Vol. A*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58.

[14] Samuel Eilenberg. *Automata, Languages, and Machines. Vol. B*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1976. With two chapters ("Depth decomposition theorem" and "Complexity of semigroups and morphisms") by Bret Tilson, Pure and Applied Mathematics, Vol. 59.

[15] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[16] Dominique Foata and Marcel-Paul Schützenberger. *Théorie géométrique des polynômes eulériens*. Lecture Notes in Mathematics, Vol. 138. Springer-Verlag, Berlin, 1970.

[17] Solomon W. Golomb and Basil Gordon. Codes with bounded synchronization delay. *Information and Control*, 8 :355–372, 1965.

[18] Donald E. Knuth. *The Art of Computer Programming. Vol. 3*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973. Sorting and searching, Addison-Wesley Series in Computer Science and Information Processing.

[19] M. Lothaire. *Combinatorics on Words*. Cambridge University Press, second edition, 1997. (First edition 1983).

[20] R. C. Lyndon. On Burnside's problem. *Trans. Amer. Math. Soc.*, 77 :202–215, 1954.

[21] Roger C. Lyndon and Marcel-Paul Schützenberger. The equation $a^m = b^n c^p$ in a free group. *Michigan Math. J.*, 9 :289–298, 1962.

[22] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Classic Texts in the Physical Sciences. The Clarendon Press, Oxford University Press, New York, second edition, 2015. With contribution by A. V. Zelevinsky and a foreword by Richard Stanley, Reprint of the 2008 paperback edition [MR1354144].

[23] Christophe Reutenauer. *Free Lie algebras*, volume 7 of *London Mathematical Society Monographs. New Series*. The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.

[24] C. Schensted. Longest increasing and decreasing subsequences. *Canad. J. Math.*, 13 :179–191, 1961.

[25] Marcel-Paul Schützenberger. On the definition of a family of automata. *Information and Control*, 4 :245–270, 1961.

[26] Marcel-Paul Schützenberger. On a factorisation of free monoids. *Proc. Amer. Math. Soc.*, 16 :21–24, 1965.

[27] Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8 :190–194, 1965.

[28] Marcel-Paul Schützenberger. Sur une question concernant certains sous-monoïdes libres. *C. R. Acad. Sci. Paris*, 261 :2419–2420, 1965.

[29] Gérard Viennot. *Algèbres de Lie libres et monoïdes libres*, volume 691 of *Lecture Notes in Mathematics*. Springer, Berlin, 1978. Bases des algèbres de Lie libres et factorisations des monoïdes libres.