M.-P. Schützenberger et la théorie des langages algébriques

Jean Berstel et Luc Boasson

le 29 novembre 2005

Lorsque l'on parcourt la très longue liste des publications de M.-P. Schützenberger, on constate que les articles portant sur la théorie des langages algébriques (appelés "context-free" à l'époque) sont somme toute assez peu nombreux. Pourtant, ces publications ont une importance qu'il ne faut pas sous-estimer tant leurs contenus sont denses et tant les idées développées se sont révélées essentielles dans de nombreux travaux ultérieurs.

C'est à la rencontre de M.-P. Schützenberger et Noam Chomsky que l'on doit quelques résultats parmi les plus importants et dont les conséquences ont été de complètement modifier l'approche de la théorie des langages formels.

Les grammaires formelles étaient alors en usage pour la formulation de propriétés syntaxiques de langues naturelles. Une théorie mathématique est issue de cette coopération, dont la portée dépasse de loin la linguistique.

Ces grammaires sont utilisées encore de nos jours pour la description de la syntaxe des langages de programmation. À cette époque, les techniques d'évaluation et de compilation des expressions arithmétiques complexes ont fait l'objet d'études nombreuses motivées par les premiers compilateurs. Les langages ont d'ailleurs été longtemps qualifiés de *Algol-like*, en référence au premier langage de programmation (Algol 60) dont la syntaxe a été conçue sur ce modèle. Le célèbre théorème dit de Chomsky–Schützenberger a mis en lumière le rôle remarquable du langage de Dyck. De façon un peu grossière, ce théorème montre que le langage des expressions bien parenthésées (sur deux types de paires de parenthèses) capture l'essence des langages algébriques.

Mais surtout, ce théorème a fait passer cette famille de langages du statut d'un modèle adapté à la description formelle de la syntaxe des langages de programmation, à celui d'une classe d'objets susceptibles d'un traitement mathématique. Cette entrée dans le monde mathématique débute par la définition des langages comme points fixes de systèmes d'équations polynomiales en variables non commutatives. Cette caractérisation équationnelle peut remplacer la dérivation séquentielle. En outre, elle permet l'interprétation algébrique de l'ambiguïté des dérivations au moyen de coefficients dans les séries associées, dont on parlera plus loin.

Plus tard, M.-P. Schützenberger a montré que le rôle du langage de Dyck pouvait être joué par un autre langage qui représente de façon symbolique les expressions arithmétiques bien construites. La preuve est à nouveau algébrico–combinatoire. Elle a ouvert le champs à la recherche de langages “générateurs” pour de nombreuses (sous)familles de langages. Il a initié lui-même l’étude des hiérarchies de familles de langages algébriques en s’intéressant en particulier aux langages linéaires, aux langages à un compteur, et aux langages obtenus par leur substitution itérée.

Cette description montre que M.-P. Schützenberger, lors de ces travaux, a constamment poursuivi ses objectifs initiaux : relier le monde des langages et automates aux mathématiques et, plus généralement, tenter d’étendre au monde non commutatif les résultats des mathématiques commutatives classiques. Ce travail sur les langages algébriques l’a conduit à poser les premiers éléments de la théorie des séries formelles. Les langages réguliers (et les séries rationnelles en variables non commutatives) correspondent aux nombres ou fonctions rationnels, et les langages et séries algébriques aux nombres ou fonctions algébriques. C’est ici qu’aparaissent les notions d’ambiguïté, voire d’ambiguïté non bornée. D’autres ont exploité cette correspondance pour des problèmes d’énumération ou des preuves de transcendance.

Outre les développements propres qui les ont suivi, les résultats de M.-P. Schützenberger ont une nouvelle fois été à l’origine de nombreux travaux. Les automates à pile ont, quant à eux, fait émerger la notion de langage déterministe. La place respective des langages déterministes et des langages inambigus n’est d’ailleurs toujours pas entièrement claire.

“L’école française” d’informatique théorique doit à ce précurseur la place et la renommée qu’elle a acquies sous l’impulsion des idées toujours intéressantes de celui que ceux qui ont travaillé avec lui appellaient parfois “le bon maître”.