ALGÈBRE. — *Une théorie algébrique du codage.*

Note de M. **Marcel Paul Schützenberger**, présentée par M. Georges Darmois.

Étant donné un ensemble A_0 de « messages élémentaires » et un ensemble A_0 de « lettres » on peut définir un code comme une correspondance $\mathcal{C}$ entre les éléments de A_0 et certaines séquences de lettres (les « mots ») telle qu'inversement à toute semblable séquence corresponde au plus une suite unique de messages. Ceci revient à dire que le décodage doit être sans ambiguïté quand il est possible ou encore que $\mathcal{C}^{-1}$ est quasi fonctionnelle (1). Ces structures jouent un certain rôle dans la théorie des algorithmes et, en calcul des probabilités, dans celles des événements récurrents (2) puisque cette dernière étudie les processus stochastiques sur les suites de lettres associés par $\mathcal{C}$ à d'autres définis sur A_0.

Le but de la présente Note est d'indiquer comment cette théorie peut être replacée dans ce qui semble être son domaine naturel : la théorie des demi-groupes et l'on observera que les notions utilisées ici, qui sont classiques (3) dans ce domaine ont une interprétation très immédiate et souvent importante sur le plan de la réalisation physique des machines codeuses et transcodeuses.

Définition. — 1° Soient A et A respectivement les demi-groupes libres engendrés par A_0 et A_0. On dira que la structure $(A_0, A_0, \mathcal{C})$ est un code (4) si l'extension de $\mathcal{C}$ à A est un isomorphisme entre A et un sous-demi-groupe P de A.

2° Un code sera dit fini si la longueur de tous ses mots (c'est-à-dire des éléments de $P_0 = \mathcal{C}A_0$) est finie; *unitaire* (3) à gauche ou à droite, *net* (3) à gauche ou à droite s'il en est de même du sous-demi-groupe $P \subset A$.

3° K étant une partie quelconque d'un demi-groupe quelconque aussi D, on appellera « *équivalence syntactique* » $[\equiv (D; K)]$ la relation entre éléments de D : $a \equiv a'$ (D; K) si et seulement si pour tout $x, y \in D$ $xay \in K \Rightarrow xa'y \in K$.

(1) J. Riguet, *Bull. Soc. Math. France*, 76, 1948, p. 129.

(2) *Cf.* B. Mandelbrot, *Contr. théorie math. des jeux de communications*, Paris, 1953, p. 124, pour une étude des relations entre codage et événements récurrents.

(3) *Cf.* P. Dubreil, *Mém. Acad. Sc.*, 63, 1941, p. 16, auquel sont empruntés les éléments de la théorie des demi-groupes utilisés ici.

On démontre :

$\equiv (D; K)$ est une relation d'équivalence régulière ([3]). (En effet la définition s'écrit aussi $K \cdot .ay = K \cdot .a'y$ pour tout y. Donc $\equiv$ qui est manifestement une équivalence, est régulière à droite et à gauche.)

Si K est un sous-demi-groupe Q de D, $\equiv (D; Q)$ est la moins fine des relations d'équivalence régulières pour lesquelles Q soit saturé ([3]). (En effet si ρ est une telle relation $a \rho a'$ entraîne $xay \rho xa'y$ et, par conséquent, $xay \in Q$ entraîne $xa'y \in Q$.)

Evidemment si D était un groupe $\equiv (D; K)$ serait l'équivalence normale associée au plus grand sous-groupe invariant contenu dans K. On appellera φ_K l'homomorphisme attaché à $\equiv (D; K)$. La relation $\equiv (\varphi_K D; \varphi_K K)$ est réduite à l'égalité et l'on dira que $\varphi_K K$ est *syntactiquement simple* dans $\varphi_K D$. Si $A \supset P$ est un code $\varphi_P A$ et $\varphi_P P$ en seront les *demi-groupes fondamentaux*. On démontre : Si le code est fini, $\varphi_P A$ est fini (la réciproque n'est pas vraie : en particulier si $\varphi_P A$ est un groupe, le code n'est fini que s'il est cyclique et $\varphi_P P$ réduit à l'élément unité). P est unitaire ou net en même temps que $\varphi_P P$. Réciproquement il est important de savoir si un couple de demi-groupes $A' \supset P'$ syntactiquement simples peuvent être les demi-groupes fondamentaux d'un code pour un certain choix de générateurs :

Une condition nécessaire et suffisante est que pour tout $p, q \in P'$ $ps \in P'$ et $sq \subset P'$ *entraîne* $s \in P'$.

La propriété est donc indépendante du choix des générateurs. La démonstration que nous ne pouvons donner ici repose sur le fait que si $s \notin P'$ la suite psq correspondrait à deux suites de messages irréductiblement distinctes. Il en résulte la possibilité de construire explicitement tous les codes (qui sont en nombre infini) correspondant à des demi-groupes fondamentaux donnés.

Enfin des méthodes d'énumération permettent d'énoncer :

Si $(\Lambda_0, A_0, \mathcal{C})$ est un code fini et si P est un demi-groupe net, P est unitaire. (La réciproque n'est pas vraie.)

En effet le fait que P soit unitaire correspond à la possibilité d'ordonner ses mots de façon lexicographique sans qu'aucun ne soit le commencement d'un autre (ou encore qu'il existe un « arbre de codage ») et le fait qu'il soit net à la possibilité qu'une séquence quelconque de lettre puisse être complétée en une suite de mots.

([4]) J. Riguet emploie le mot code dans une acception plus générale. La définition adoptée ici est celle implicitement utilisée dans la théorie des communications.

Ceci permet de répondre à un problème pratique d'optimalité ([5]) en montrant que la classe des codes unitaires est admissible. En outre ceci indique l'existence de codes finis nets et unitaires à gauche et à droite à la fois qui ne correspondent à aucun groupe. L'inventaire de ces demi-groupes particuliers reste encore à faire.

([5]) Posé par : A. A. SARDINAS et G. W. PATTERSON, *Res. Div. Reports*, 1950, p. 50-27. University of Pennsylvania.

(Extrait des *Comptes rendus des séances de l'Académie des Sciences*, t. 242, p. 862-864, séance du 13 février 1956.)

GAUTHIER-VILLARS,
ÉDITEUR-IMPRIMEUR-LIBRAIRE DES COMPTES RENDUS DES SÉANCES DE L'ACADÉMIE DES SCIENCES
149553-56 Paris. — Quai des Grands-Augustins, 55.