Groupe d'étude d'algèbre

YVES COCHET

Langages définis par des congruences

Groupe d'étude d'algèbre, tome 1 (1975-1976), exp. nº 16, p. 1-7
<http://www.numdam.org/item?id=GEA_1975-1976__1__A16_0>

Groupe d'étude d'ALGÈBRE
(Marie-Paule MALLIAVIN)
1re année, 1975/76, n° 16, 7 p.

LANGAGES DÉFINIS PAR DES CONGRUENCES

par Yves COCHET

Les congruences et langages quasi parfaits définis ci-dessous ont été introduits par M. NIVAT en 1970 (cf. [9]) afin d'étendre à une vaste famille de langages les agréables propriétés algébriques dont jouissent les langages de Dyck sur le monoïde libre.

L'aspect algébrique de la théorie des congruences quasi parfaites est développé par M. BENOIS (cf. [1]) au cours de ces mêmes journées sur les demi-groupes. Nous en exposons ici l'aspect combinatoire. Les propriétés fondamentales, rappelées au paragraphe 1, sont démontrées en [5], [8] et [9], et dans l'exposé précédent de M. NIVAT (cf. [10]). Les autres assertions, énoncées sans démonstration (sauf au paragraphe 2), sont issues d'articles parus ou à paraître.

1. Systèmes et congruences quasi parfaites.

Soit X un alphabet fini, X^* le monoïde libre sur X et 1 son élément neutre. Si f est un mot de X^*, $|f|$ est sa longueur. Appelons système une relation finie Σ sur X^* et, pour tous $f , g \in X^*$, notons :

$f \longleftrightarrow_{\Sigma} g$ si, et seulement si, il existe $\alpha , \beta \in X^*$, $(u , v) \in \Sigma$ tels que $f = \alpha u \beta$ et $g = \alpha v \beta$, ou $f = \alpha v \beta$ et $g = \alpha u \beta$,

$f \longrightarrow_{\Sigma} g$ si, et seulement si, $f \longleftrightarrow_{\Sigma} g$ et $|f| > |g|$,

$f \longmapsto_{\Sigma} g$ si, et seulement si, $f \longleftrightarrow_{\Sigma} g$ et $|f| = |g|$.

Désignons par $\longleftrightarrow_{\Sigma}^*$, $\longrightarrow_{\Sigma}^*$, $\longmapsto_{\Sigma}^*$ respectivement les fermetures réflexives et transitives de $\longleftrightarrow_{\Sigma}$, $\longrightarrow_{\Sigma}$ et $\longmapsto_{\Sigma}$. La relation $\longleftrightarrow_{\Sigma}^*$ est la congruence de Thue engendrée par le système Σ. Les relations $\longrightarrow_{\Sigma}^*$ et $\longmapsto_{\Sigma}^*$ sont respectivement baptisées réduction et équipollence. Un mot $f \in X^*$ est irréductible modulo Σ si, et seulement si, l'ensemble $\{g ; f \longrightarrow_{\Sigma} g\}$ est vide.

On suppose, ce qui n'entrainera aucune restriction, que l'on a $|u| \leqslant |v|$, pour tout $(u , v) \in \Sigma$. On note souvent $\longleftrightarrow^*$, $\longrightarrow^*$, $\longmapsto^*$ sans " Σ " en indice inférieur si aucune confusion n'est à craindre. Un couple $(u , v) \in \Sigma$ est noté $u \longleftrightarrow v$, et s'appelle d'ordinaire une règle.

DÉFINITION et PROPOSITION 1. - Un système Σ est quasi parfait si, et seulement si, il satisfait à l'une des conditions équivalentes suivantes :

(QP1) Pour tous mots $h , h' \in X^*$, si h et h' sont irréductibles et congrus, alors ils sont équipollents,

(QP2) Pour tous mots $f , f' \in X^*$, si f et f' sont congrus, alors il existe deux mots h , h' irréductibles tels que l'on ait $f \longrightarrow^* h$, $h \vdash\!\!\dashv^* h'$ et $f' \longrightarrow^* h'$,

(QP3) Pour tous mots f , g , $g' \in X^*$, si l'on a $f(\longrightarrow \cup \vdash\!\!\dashv)g$ et $f(\longrightarrow \cup \vdash\!\!\dashv)g'$, alors il existe deux mots h , h' tels que l'on ait $g \longrightarrow^* h$, $h \vdash\!\!\dashv^* h'$ et $g' \longrightarrow^* h'$,

(QP4) 1° Pour tous couples (u , v) , (u' , v') dans Σ , et tous mots α , $\beta \in X^*$, si $v = \alpha v' \beta$, alors existent h , $h' \in X^*$ tels que l'on ait $u \longrightarrow^* h$, $h \vdash\!\!\dashv^* h'$ et $\alpha u' \beta \longrightarrow^* h'$,

2° Pour tous couples (u , v) , (u' , v') dans Σ , et tous mots α , β , $\gamma \in X^*$, si $v = \alpha\beta$ et $v' = \beta\gamma$, alors il existe h , $h' \in X^*$ tels que l'on ait $u\gamma \longrightarrow^* h$, $h \vdash\!\!\dashv^* h'$ et $\alpha u' \longrightarrow^* h'$.

De nombreuses propriétés des systèmes quasi parfaits sont décidables.

COROLLAIRE 1. - Si Σ est un système quasi parfait, il existe un algorithme pour décider si $f \longleftrightarrow^* g$, pour tous mots f , $g \in X^*$.

Autrement dit, le problème des mots est décidable dans le monoïde quotient $X^*/\longleftrightarrow^*$, finiment présenté par X et Σ .

COROLLAIRE 2. - Il existe un algorithme pour décider, étant donné un système Σ , s'il est quasi parfait ou non.

COROLLAIRE 3. - Il existe un algorithme pour décider, étant donnés deux systèmes quasi parfaits Σ et Σ' , s'ils engendrent la même congruence ou non (si oui, on dit que les systèmes Σ et Σ' sont équivalents).

DÉFINITION et PROPOSITION 2. - Un système Σ est parfait si, et seulement si, il satisfait à l'une des conditions équivalentes suivantes :

(P1) Pour tous mots h , $h' \in X^*$, si h et h' sont irréductibles et congrus, alors ils sont égaux,

(P2) Le système Σ est quasi parfait, et il ne possède pas de couple (u , v) de mots irréductibles et de même longueur.

Une congruence θ sur X^* est quasi parfaite (resp. parfaite) si, et seulement si, il existe un système quasi parfait (resp. parfait) Σ tel que $\theta = \longleftrightarrow^*_{\Sigma}$.

2. Langages quasi parfaits.

Une partie L de X^* est un langage quasi parfait si, et seulement si, L est réunion finie de classes d'une congruence quasi parfaite.

Quelques exemples (on consultera aussi [7]) :

1° Le langage rationnel $L = X^* abX^*$ (avec $X = \{a , b\}$) n'est autre que la classe d'équivalence du mot $ab \in X$ pour la congruence quasi parfaite $\longleftrightarrow^*$ engendrée par le système $\Sigma = \{a \longleftrightarrow a^2 , b \longleftrightarrow b^2 , ab \longleftrightarrow aba + bab\}$. En fait, tous les langages rationnels sont quasi parfaits (cf. § 3).

2° Le langage algébrique $L = \{a^n b^n ; n \geqslant 1\} \subseteq \{a , b\}^*$ est engendré à partir du mot ab par le système quasi parfait $\Sigma = \{ab \longleftrightarrow a^2 b^2\}$.

3° Les langages de Dyck sont des langages quasi parfaits.

LEMME 1. - <u>Un langage</u> L <u>sur une lettre est quasi parfait si, et seulement si, il est rationnel.</u>

<u>Preuve</u>. - On peut considérer, sur une lettre, que l'application d'une règle $u \longleftrightarrow v$ d'un système quasi parfait Σ s'effectue sur un segment gauche d'un mot. Le système Σ est alors un "système pur régulier" (cf. [2], p. 100) qui engendre L . Le théorème 2 de [2] implique que L est rationnel. La réciproque vient de la propositon 4, §3

LEMME 2. - <u>La famille des langages quasi parfaits est strictement contenue dans la famille des langages context-sensitive.</u>

<u>Preuve</u>. - Soit L un langage quasi parfait défini par un système quasi parfait Σ et un ensemble fini H de mots irréductibles. Notons $\Sigma_<$ l'ensemble des règles (u , v) du système Σ telles que $|u| < |v|$, et munissons-nous d'un alphabet auxiliaire $\bar{X}$ en correspondance biunivoque avec X et d'un nouveau symbole d'axiome S . Nous supposons que le mot vide 1 n'est pas membre gauche d'une règle de Σ . Une grammaire context-sensitive engendrant L est alors définie par les productions :

$$S \longmapsto \bar{h} , \quad \text{pour tout} \quad h \in H ,$$
$$\bar{u} \longmapsto \bar{v} , \quad \text{pour tout} \quad (u , v) \in \Sigma_< ,$$
$$\bar{x} \longmapsto x , \quad \text{pour tout} \quad x \in X .$$

Si 1 est membre d'une règle du système Σ , la construction d'une grammaire context-sensitive engendrant L est aussi facile à concevoir que longue à décrire.

D'autre part un langage sur une lettre tel que $\{a^{2^n} ; n \in \underset{\sim}{N}\}$ est context-sensitive non rationnel donc non quasi parfait, d'où l'inclusion stricte annoncée.

LEMME 3. - <u>Il existe un langage quasi parfait non algébrique.</u>

<u>Preuve</u>. - Sur l'alphabet $x = \{a, b, c\}$, la classe L du mot abc modulo le système quasi parfait $\Sigma = \{ab \longleftrightarrow aabv , vb \longleftrightarrow bbv , vc \longleftrightarrow bcc\}$ est telle que

$$L \cap a^* b^* c^* = \{a^n b^{2^{n-1}} c^n ; n \geqslant 1\} ,$$

par conséquent elle ne saurait être un langage algébrique.

LEMME 4. - Les langages $\{ca^m b^m ; m \geqslant 1\} \cup \{a^n b^{2n} ; n \geqslant 1\}$, $\{a^m b^m a^n b^{2n} ; m , n \geqslant 1\}$ <u>et</u> $\{a^n b^{2n-1} c^n ; n \geqslant 1\}$ <u>ne sont pas quasi parfaits.</u>

<u>Preuve</u>. - S'il existait un système quasi parfait Σ pour l'un des deux premiers langages, il contiendrait nécessairement des règles de la forme $a^i b^j \longleftrightarrow a^{i+k} b^{j+k}$ et $a^p b^q \longleftrightarrow q^{p+r} b^{q+2r}$ avec $i , k , p , q , r \quad \underset{\sim}{N}$. En prenant m , n assez

grands, la première de ces règles s'applique aussi bien à un segment $a^m b^m$ qu'à un segment $a^n b^{2n}$ et engendre ainsi des mots n'appartenant pas au langage. Un argument similaire s'applique à $\{a^n b^{2n-1} c^n \ ; \ n \geq 1\}$.

COROLLAIRE 4. - Il existe un langage simple deterministe qui n'est pas quasi parfait.

Preuve. - Il s'agit du langage $\{ca^m b^m \ ; \ m > 1\} \cup \{a^n b^{2n} \ ; \ n \geq 1\}$ engendré par la grammaire simple

$S \longmapsto cA_1 + aC$

$A_1 \longmapsto aA_2$

$A_2 \longmapsto b + aA_2 B$

$B \longmapsto b$

$C \longmapsto bB + aCBB$.

PROPOSITION 3. - La famille des langages quasi parfaits n'est fermée par aucune des opérations de reunion, intersection, complémentation, produit, étoile, homomorphisme propre et intersection avec les langages rationnels.

Preuve.

Réunion. - les langages $\{a^m b^m \ ; \ m \geq 1\}$ et $\{a^n b^{2n} \ ; \ n \geq 1\}$ sont quasi parfaits, mais leur réunion ne l'est pas, d'après le lemme 4.

Intersection. - Voir la preuve du lemme 3 et l'énoncé du lemme 4.

Complémentation. - Le complémentaire dans $\{a , b\}^*$ du langage quasi parfait $\{a^n b^n \ ; \ n \geq 1\}$ est réunion infinie de classes pour sa congruence syntactique, la plus grossière qui le sature. A fortiori, ne peut-il être réunion finie de classes pour une autre congruence.

Produit. - Le langage non quasi parfait $\{a^m b^m a^n b^{2n} \ ; \ m , n \geq 1\}$ est le produit de deux langages quasi parfaits.

Etoile. - Considérons le langage quasi parfait $L = \{a^n b^n \ ; \ n \geq 1\} \cup \{a\}$. Le mot a^n est dans L^* , pour tout $n \in \underset{\sim}{N}$. Si L^* était quasi parfait, il existerait dans le système quasi parfait Σ engendrant L^* une règle de la forme $u \longleftrightarrow a^n$, avec $u \in L^*$ et $|u| < n$. Mais alors le mot $a^n b^n$ se réduit en ub^n qui ne peut être dans L^* .

Homomorphisme propre. - Le langage $\{a^m b^m \ ; \ m \geq 1\} \cup \{c^n d^{2n} \ ; \ n \geq 1\}$ est un langage quasi parfait du monoïde libre $\{a, b, c, d\}^*$. Son image par l'homomorphisme $\varphi : \{a, b, c, d\}^* \longrightarrow \{a, b\}^*$ défini par $a\varphi = a$, $b\varphi = b$, $c\varphi = a$, $d\varphi = b$, est le langage non quasi parfait

$$\{a^m b^m \ ; \ m > 1\} \cup \{a^n b^{2n} \ ; \ n \geq 1\} .$$

Intersection avec les langages rationnels. - Voir la preuve du lemme 3 et l'énoncé du lemme 4.

Enfin, il est assez évident que la famille des langages quasi parfaits est fermée pour l'opération miroir.

3. Applications aux langages rationnels.

Les relations entre langages quasi parfaits et langages rationnels reposent sur les deux remarques suivantes.

D'une part, il est facile de constater que l'ensemble des mots irréductibles modulo une congruence quasi parfaite est un langage rationnel. On trouvera quelques conséquences de ce fait en [5] et [6]. Incidemment, émettons la conjecture suivante. Si L est un langage algébrique déterministe sur X^*, l'ensemble des mots de longueur minimale dans leur classe modulo la congruence syntactique de L, n'est-il pas un langage rationnel ?

D'autre part, les langages rationnels sont des langages quasi parfaits d'un type particulier. Plus précisément, étant donné un système quasi parfait Σ, disons qu'une réduction $f \longleftrightarrow g$ est <u>à gauche</u> si f et g se factorisent en vf' et uf' respectivement, avec $(u , v) \in \Sigma$, $|u| < |v|$. Un système quasi parfait Σ est <u>à gauche</u>, si tout mot réductible modulo Σ admet une réduction à gauche. Du même mouvement, on définit les congruences quasi parfaites à gauche et les langages quasi parfaits à gauche.

PROPOSITION 4. - <u>Un langage est rationnel, si et seulement si, il est quasi parfait à gauche.</u>

Cette proposition est un corollaire de l'énoncé plus général ci-dessous.

LEMME 5. - <u>Toute congruence d'indice fini sur</u> X^* <u>est une congruence quasi parfaite à gauche.</u>

Question de décidabilité, on a les résultats suivants.

LEMME 6.

1° <u>Il existe un algorithme pour décider, étant donné un système quasi parfait</u> Σ, <u>si le quotient</u> $X^*/\overset{*}{\underset{\Sigma}{\longleftrightarrow}}$ <u>est fini ou non.</u>

2° <u>Soit</u> Σ <u>un système quasi parfait tel que le quotient</u> $X^*/\overset{*}{\underset{\Sigma}{\longleftrightarrow}}$ <u>soit fini. On peut effectivement construire un système quasi parfait à gauche équivalent à</u> Σ.

3° <u>Il existe un algorithme pour décider, étant donné un système quasi parfait</u> Σ, <u>s'il est à gauche ou non.</u>

Terminons ce paragraphe par une question. Si l'on qualifie de <u>rationnel</u> un système quasi parfait Σ tel que toutes ses classes soient des langages rationnels (autrement dit que le monoïde quotient $X^*/\overset{*}{\underset{\Sigma}{\longleftrightarrow}}$ soit ponctuellement reconnaissable), existe-t-il un algorithme pour décider si un système quasi parfait donné Σ est rationnel ou non ?

4. Applications aux langages algébriques.

Contrairement aux langages rationnels, les langages algébriques ne sont pas tous quasi parfaits (cf. corollaire 4). La question se pose donc de caractériser ceux qui le sont. D'un autre côté les langages quasi parfaits ne sont pas tous algébriques (cf. lemme 3). Si l'on appelle algébrique un système quasi parfait tel que toutes ses classes soient des langages algébriques, une autre question est de caractériser les systèmes quasi parfaits algébriques. A ces deux questions n'ont été apportées que des réponses partielles dont nous allons donner quelques exemples en ce qui concerne la seconde.

4.1. - Si l'on impose que les règles du système quasi parfait soient toutes de la forme $(1 , v)$, $v \in X^*$, la classe d'équivalence du mot vide, que l'on appelle langage de Dyck généralisé, est un langage algébrique non ambigu dont le complémentaire est encore algébrique non ambigu comme il a été montré en [6].

4.2. - Soient U , V deux ensembles de mots sur X^* . On dit que U est une base de V si, et seulement si, pour tous mots $u \in U$, $v \in V$, et $f_1 , f_2 , f_3 \in X^*$,

$$v = f_1 f_2 \quad \text{et} \quad u = f_2 f_3 \quad \text{impliquent} \quad f_2 = 1 \quad \text{ou} \quad f_3 = 1 ,$$

$$v = f_2 f_3 \quad \text{et} \quad u = f_1 f_2 \quad \text{impliquent} \quad f_1 = 1 \quad \text{ou} \quad f_2 = 1 .$$

Maurice NIVAT a énoncé en [8] la condition suivante, suffisante pour qu'un système quasi parfait Σ soit algébrique.

PROPOSITION 5. - *Soit Σ un système quasi parfait. Posons*

$$U = \{u \in X^* \; ; \; \exists\, v \in X^* , \; |u| \leq |v| \; , \; (u , v) \in \Sigma\}$$

et

$$V = \{v \in X^* \; ; \; \exists\, u \in X^* , \; |u| \leq |v| \; , \; (u , v) \in \Sigma\} .$$

Si U est une base de V , Σ est un système quasi parfait algébrique.

4.3. - Les deux exemples précédents sont des cas particuliers, mais remarquables, d'une classe récursive de systèmes quasi parfaits, dits minces, dont nous avons montré qu'ils sont algébriques (cf. [5]).

Enfin, dans le but de résoudre le problème de l'équivalence des grammaires séparées, puis des grammaires simples, P. BUTZBACH a montré en [3] et [4] que les langages séparés sont des langages quasi parfaits et que les langages simples sont classes, non plus d'une congruence, mais d'une "substitution quasi parfaite".

BIBLIOGRAPHIE

[1] BENOIS (M.). - Présentation et simplifiabilité de certains monoïdes, Séminaire Dubreil, 29e année, 1975/76, Journées sur les demi-groupes : Algèbre et combinatoire [1976. Paris], n° 13, p.

[2] BUCHI (J. R.). - Regular canonical systems, Arch. math. Logik und Grundlagenforsch., t. 6, 1964, p. 91-111.

[3] BUTZBACH (P.). - Une famille de congruences de Thue pour lesquelles le problème de l'équivalence est décidable. Application à l'équivalence des grammaires séparées, "Automata, langages and programming", p. 3-12. - Amsterdam, North-Holland publishing Company, 1973.

[4] BUTZBACH (P.). - Equivalence des grammaires simples, "Actes de l'école de printemps sur les langages algébriques [1973. Bonascre]" (à paraître).

[5] COCHET (Yves). - Sur l'algébricité des classes de certaines congruences définies sur le monoïde libre, thèse 3e cycle, Mathématiques, Rennes, 1971.

[6] COCHET (Y.) et NIVAT (M.). - Une généralisation des ensembles de Dyck, Israël J. Math., t. 9, 1971, p. 389-395.

[7] GABRIELIAN (A.). - Pure grammars and pure languages, University of Waterloo, Dept of appl. Anal. and Comput. Sc., Res. Rep. CSRR 2027, 1970.

[8] NIVAT (M.). - On some families of languages related to the Dyck language, "Second Annual symposium on theory of computing" [1970. Northampton], p. 221-225. - New York, Association for Computing Machinery, 1970.

[9] NIVAT(M.) et (BENOIS (M.)). - Congruences parfaites et quasi parfaites, Séminaire Dubreil : Algèbre, 29e année, 1971/72, n° 7, 9 p.

[10] NIVAT (M.). - Congruences et théorème de Church-Rosser, Journées sur les demi-groupes ; Algèbre et combinatoire [1976. Paris] (Communication orale).

Yves COCHET
I. N. S. A.
20 avenue des Buttes de Coësmes
35031 RENNES CEDEX