

UNE GENERALISATION DES ENSEMBLES DE DYCK

PAR
Y. COCHET ET M. NIVAT*

ABSTRACT

A certain class of Thue congruence relations are studied, which abstract some of combinatorial properties of the special congruence relation defining Dyck sets. It is proved that the congruence classes of these "Dyck congruences" as well as their complements are, unambiguous context-free languages.

Introduction

On sait le rôle important que jouent, dans la théorie des langages algébriques, les langages de Dyck et de semi-Dyck (cf [2] ou [3]) définis comme les classes d'équivalence du mot vide pour les congruences de Thue engendrées sur un alphabet à $2n$ lettres $Z = \{x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n\}$ par les systèmes de relations

$$\{x_i \bar{x}_i = \bar{x}_i x_i = e \mid i = 1, \dots, n\} \text{ et}$$

$$\{x_i \bar{x}_i = e \mid i = 1, \dots, n\} \text{ respectivement —}$$

Le travail qui suit a pour but de décrire des ensembles, dits ensembles de Dyck, qui jouissent de propriétés très voisines de celles de ces deux langages. Définis comme classes d'équivalence du mot vide pour des congruences de Thue dont les générateurs vérifient certaines relations combinatoires, ce sont aussi des langages algébriques non ambigus dont les complémentaires sont encore algébriques et non ambigus.

1. Congruences parfaites

NOTATIONS. Soient $X = \{x_1, x_2, \dots, x_M\}$ un ensemble fini, appelé alphabet, X^* le monoïde libre engendré par X , (dont nous notons e l'élément neutre) et S une partie de $X^* \times X^*$ telle que, pour tout $(\phi, \gamma) \in S$, on ait $|\phi| \leq |\gamma|$. La congruence notée $\overset{*}{\leftrightarrow}$, la plus grossière telle que, pour tout $(\phi, \gamma) \in S$, on ait

* Cet auteur a préparé ce travail lors d'un séjour de quelques semaines à l'Université de Jerusalem en Avril, 1969. Que le département de mathématiques de cette université et tout particulièrement le professeur E. Shamir soit ici chaleureusement remerciés.

Received December 21, 1969 and in revised form May 26, 1970

$\phi \overset{*}{\leftrightarrow} \gamma$ n'est autre que la fermeture transitive de la relation notée \leftrightarrow , définie, pour tous $f, g \in X^*$ par $f \leftrightarrow g$ si et seulement si il existe $(\phi, \gamma) \in S$ et $f', f'' \in X^*$ qui vérifient soit $f = f'\phi f''$ et $g = f'\gamma f''$, soit $f = f'\gamma f''$ et $g = f'\phi f''$. Par définition, la congruence $\overset{*}{\leftrightarrow}$ est la congruence de Thue engendrée par le système de règles $(\phi \leftrightarrow \gamma), (\phi, \gamma) \in S$. Nous noterons \rightarrow (resp. \vdash) la relation définie, pour tous $f, g \in X^*$, par $f \rightarrow g$ (resp. $f \vdash g$) si $f \leftrightarrow g$ et $|f| \geq |g|$ (resp. $|f| = |g|$), et $\overset{*}{\rightarrow}$ (resp. $\overset{*}{\vdash}$) sa fermeture transitive.

DÉFINITION 1.1. Le système de règles S est dit *quasi-parfait* si, pour tous $f, g, g' \in X^*$ tels que $f \overset{*}{\rightarrow} g$ et $f \overset{*}{\rightarrow} g'$, il existe $h \in X^*$ qui vérifie $g \overset{*}{\rightarrow} h$ et $g' \overset{*}{\rightarrow} h$.

Si, en outre, pour tout $(\phi, \gamma) \in S$, on a $|\phi| < |\gamma|$, le système est dit *parfait*.

Une congruence de Thue est une congruence *quasi-parfaite* (resp. *parfaite*) si elle admet un système générateur quasi-parfait (resp. parfait).

LEMME 1.1. Une condition nécessaire et suffisante pour qu'un système S soit quasi-parfait est : Pour tous $f, g \in X^*$, on a $f \overset{*}{\leftrightarrow} g$ si et seulement si il existe $h \in X^*$ tel que $f \overset{*}{\rightarrow} h$ et $g \overset{*}{\rightarrow} h$.

DÉMONSTRATION. Cette condition est évidemment suffisante, prouvons qu'elle est nécessaire. Raisonnons par récurrence sur la longueur, c'est-à-dire le nombre de pas, d'une dérivation $f \overset{*}{\leftrightarrow} g$; si l'on a $f \leftrightarrow g$, le lemme est évident; supposons le résultat vrai lorsque f se dérive en g en moins de p pas. Si $f \overset{*}{\leftrightarrow} g$ est une dérivation de longueur $p + 1$, on peut l'écrire $f \leftrightarrow f_1 \overset{*}{\leftrightarrow} g$, où $f_1 \overset{*}{\leftrightarrow} g$ est de longueur p ; par récurrence, il existe $h_1 \in X^*$ qui vérifie $f_1 \overset{*}{\rightarrow} h_1$ et $g \overset{*}{\rightarrow} h_1$. Deux cas se présentent: Si le pas $f \leftrightarrow f_1$ est une réduction $f \rightarrow f_1$, on a $f \overset{*}{\rightarrow} h_1$ et $g \overset{*}{\rightarrow} h_1$; sinon la quasi-perfection du système implique l'existence d'un $h \in X^*$ tel que $f \overset{*}{\rightarrow} h$ et $h_1 \overset{*}{\rightarrow} h$, donc $g \overset{*}{\rightarrow} h_1 \overset{*}{\rightarrow} h$.

DÉFINITION 1.2. Un mot h de X^* est dit *minimal* pour $\overset{*}{\leftrightarrow}$ si, pour tout $h' \in X^*$, la relation $h \overset{*}{\rightarrow} h'$ entraîne $|h'| = |h|$.

LEMME 1.2. Une condition nécessaire et suffisante pour qu'un système S soit quasi-parfait est: Quels que soient les mots minimaux h, h' de X^* , on a $h \overset{*}{\leftrightarrow} h'$ si et seulement si $h \vdash h'$.

DÉMONSTRATION. Il est évident que cette condition est nécessaire, montrons qu'elle est suffisante. Prouvons d'abord que si un mot f de X^* n'est pas minimal, il existe un mot minimal h tel que $f \overset{*}{\rightarrow} h$. Si $|f| = 1$, c'est trivial; supposons ce résultat démontré pour tous les mots g de longueur inférieure à $|f|$; si le mot f n'est pas minimal, il existe un mot g tel que $f \overset{*}{\rightarrow} g$ et $|g| < |f|$; si g est minimal,

le résultat est immédiat, sinon, par hypothèse de récurrence, il existe un mot minimal h qui vérifie $g \xrightarrow{*} h$, donc $f \xrightarrow{*} g \xrightarrow{*} h$. Démontrons maintenant que le système S est quasi-parfait en utilisant la caractérisation du lemme 1.1. ; soient f et g deux mots de X^* tels que $f \leftrightarrow g$; on peut évidemment supposer que ni f ni g n'est minimal, sinon le résultat est évident ; il existe alors deux mots minimaux h et h' tels que $f \xrightarrow{*} h$ et $g \xrightarrow{*} h'$, on a donc $h \leftrightarrow h'$ et, par hypothèse, on en déduit $h \vdash h'$, par conséquent $f \xrightarrow{*} h$ et $g \xrightarrow{*} h$.

THÉOREME 1.1. *Si \leftrightarrow^* est une congruence quasi-parfaite engendrée par le système quasi-parfait fini S , il existe un algorithme pour décider si $f \leftrightarrow^* g$, pour tous $f, g \in X^*$.*

DÉMONSTRATION. En un nombre fini de pas on construit les deux ensembles finis de mots h (resp. h') de X^* tels que $f \xrightarrow{*} h$ (resp. $g \xrightarrow{*} h'$). D'après le lemme 1.1., on a $f \leftrightarrow^* g$ si et seulement si ces deux ensembles ont une intersection non-vide.

LEMME 1.3. *Si S est un système générateur fini de la congruence \leftrightarrow^* et si, en outre, on a $|\phi| < |\gamma|$, pour tout $(\phi, \gamma) \in S$, il existe un algorithme pour décider si S est un système générateur parfait.*

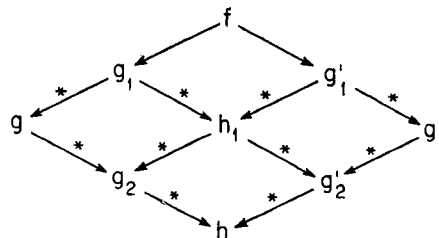
DÉMONSTRATION. Il suffit de démontrer que le système S est parfait si et seulement si les conditions suivantes sont satisfaites :

i) Pour tous $(\phi, \gamma), (\phi', \gamma') \in S$, et tous $u, v, w \in X^*$ tels que $\gamma = uv$ et $\gamma' = vw$, il existe $h \in X^*$ qui vérifie $u\phi' \xrightarrow{*} h$ et $\phi w \xrightarrow{*} h$;

ii) Pour tous $(\phi, \gamma), (\phi', \gamma') \in S$, et tous $u, v \in X^*$ tels que $\gamma = u\gamma'v$, il existe $h \in X^*$ qui vérifie $\phi \xrightarrow{*} h$ et $u\phi'v \xrightarrow{*} h$.

Il est clair que ces conditions en nombre fini peuvent être effectivement vérifiées. Ces conditions sont évidemment nécessaires, prouvons qu'elles sont suffisantes.

Raisonnons par récurrence sur la longueur d'un mot f . Si $|f| = 1$, c'est trivial ; supposons le lemme vrai pour tous les mots g tels que $|g| < |f|$ et considérons deux dérivations $f \xrightarrow{*} g$ et $f \xrightarrow{*} g'$, que l'on peut écrire $f \rightarrow g_1 \xrightarrow{*} g$ et $f \rightarrow g'_1 \xrightarrow{*} g'$; le diagramme



se lit : les conditions (i) et (ii) impliquent l'existence d'un $h_1 \in X^*$ tel que $g_1 \xrightarrow{*} h_1$ et $g'_1 \xrightarrow{*} h_1$, puis l'hypothèse de récurrence entraîne l'existence d'un $g_2 \in X^*$ qui vérifie $g \xrightarrow{*} g_2$ et $h_1 \xrightarrow{*} g_2$, et d'un $g'_2 \in X^*$ tel que $g' \xrightarrow{*} g'_2$ et $h_1 \xrightarrow{*} g'_2$, enfin d'un $h \in X^*$ qui vérifie $g_2 \xrightarrow{*} h$ et $g'_2 \xrightarrow{*} h$, donc $g \xrightarrow{*} h$ et $g' \xrightarrow{*} h$.

2. Congruences de Dyck

DÉFINITION 2.1. Un système S est dit *trivial* si et seulement si $S \subset \{e\}xX^*$.

Une congruence de Thue est dite congruence de Dyck si et seulement si elle est engendrée par un système trivial et parfait. On appelle ensemble de Dyck et l'on note D^* la classe d'équivalence du mot vide pour une telle congruence.

Soit S un système trivial fini que l'on notera $\{e \leftrightarrow \gamma \mid i = 1, \dots, N\}$; rappelons que d'après le lemme 1.3., il existe un algorithme pour décider si S est un système générateur parfait. On peut supposer qu'il n'existe aucun $i, j \in [N]$, aucun mots $f, f' \in X^*$ tels que $\gamma_i = f\gamma_j f'$; on en déduit que $\gamma_i = fg$, $\gamma_j = gh$ et $g \neq \rho$ impliquent $f = h$.

DÉFINITION 2.2. Un mot $f \in D^*$ est *premier* (relativement à D^*) s'il ne possède aucun facteur gauche autre que e qui soit dans D^* .

On note D l'ensemble des mots premiers.

LEMME 2.1. D est un code bipréfixe sur X^* , i.e. tout mot $f \in D^*$ admet une factorisation unique $f = f_1 f_2 \dots f_k$, où, pour tout $i \in [k]$, f_i appartient à D .

DÉMONSTRATION. L'existence d'une telle factorisation est évidente; l'unicité se démontre par l'absurde: Supposons que f admette deux factorisations $f = f_1 \dots f_k$ et $f = f'_1 \dots f'_m$, telles que $f_1 \neq f'_1$ et posons, par exemple, $f_1 = f'_1 g$, où $g \in XX^*$; alors, contrairement à l'hypothèse, f_1 n'est pas premier.

LEMME 2.2. Si f est un mot premier, il n'existe qu'un indice $i \in [N]$ tel que l'on ait $f \xrightarrow{*} \gamma_i \rightarrow \rho$.

DÉMONSTRATION. Soit f le mot premier le plus court tel qu'il existe $i, j \in [N]$ qui vérifient $f \xrightarrow{*} \gamma_i$ et $f \xrightarrow{*} \gamma_j$; ces réductions s'écrivent $f \rightarrow g \xrightarrow{*} \gamma_i$ et $f \rightarrow g' \xrightarrow{*} \gamma_j$, et l'on passe de f à g et de f à g' par application des règles $\gamma \rightarrow e$ et $\gamma' \rightarrow e$ respectivement. Deux cas peuvent se présenter: (i) Il existe $u, v \in XX^*$ et $w \in X^*$ tels que $f = u\gamma w\gamma'v$ et, par conséquent, le mot $f' = uwv$ est premier, plus court que f et tel que $f' \xrightarrow{*} \gamma_i$, $f' \xrightarrow{*} \gamma_j$, d'où la contradiction; (ii) Il existe $u, v, w, s \in XX^*$ tels que $f = uws wv$, $\gamma = ws$ et $\gamma' = sw$; mais alors le mot premier $f' = g = g' = uwv$ contredit encore la minimalité de f .

Pour tout $i \in [N]$ et tout f , facteur gauche propre de γ_i , on note D_i l'ensemble, des $g \in D$ tels que $g \xrightarrow{*} \gamma_i$, Δ_f^- l'ensemble des γ_j , $j \in [N]$, tels qu'il existe f' facteur droit propre de f , et γ'_j , facteur droit propre de γ_j , qui vérifient $\gamma_j = f'\gamma'_j$, et D_f la réunion des D_m , $m \in [N]$, tels qu'il existe $\gamma_n \in \Delta_f^-$, $\gamma_n = f'\gamma'_n$, qui vérifie $\gamma'_n f' = \gamma_m$.

LEMME 2.3. *Tout élément f de D_i admet une factorisation unique*

$$f = x_{i_1} f_1 x_{i_2} f_2 \cdots x_{i_{k-1}} f_{k-1} x_{i_k},$$

où, pour tout $m \in [k-1]$, f_m appartient à $(D \setminus \overline{D_{x_{i_1} \cdots x_{i_m}}})^*$ et $x_{i_1} x_{i_2} \cdots x_{i_k} = \gamma_i$ pour quelque $i \in [N]$.

DÉMONSTRATION Supposons qu'une telle factorisation existe et montrons qu'elle est unique; si f admettait deux factorisations, il existerait $m, n \in [k-1]$, et $f_n \in D \setminus \overline{D_{x_{i_1} \cdots x_{i_n}}}$ tels que $x_{i_m} x_{i_{m+1}} \cdots x_{i_n} f'_n$ appartienne à D^* , où f'_n est facteur gauche propre de f_n ; donc, il existerait un $m' \in [m, n]$, un facteur gauche propre f' de f_n et un mot f'' tels que $f' \xrightarrow{*} f''$ et $x_{i_m} \cdots x_{i_n} f''$ soit membre droit d'une règle du système S ; on en déduit, si $f_n = f' f'''$:

$$x_{i_m} \cdots x_{i_n} f' f''' \xrightarrow{*} f''' \text{ et } x_{i_m} \cdots x_{i_n} f' f''' \xrightarrow{*} x_{i_m} \cdots x_{i_n}$$

d'où $f''' \xrightarrow{*} x_{i_m} \cdots x_{i_n}$; si $f_n \in D_j$, on a donc $f' f''' \xrightarrow{*} f'' x_{i_m} \cdots x_{i_n} = \gamma_j$, où $x_{i_m} \cdots x_{i_n} f''$ est membre droit d'une règle, par conséquent f_n appartient à $\overline{D_{x_{i_1} \cdots x_{i_n}}}$, ce qui est exclu. On montre de la même manière que f est premier.

Pour établir l'existence d'une telle factorisation, il suffit de remarquer que, pour tout $m \in [k-1]$, $x_{i_1} f_1 \cdots x_{i_m} f_m$ est le plus long facteur gauche propre de f qui se réduise en $x_{i_1} \cdots x_{i_m}$.

On en déduit que la grammaire G_{D^*} suivante, d'axiome ξ_0 , engendre l'ensemble de Dyck de façon non-ambigüe

$$G_{D^*} = \left\{ \begin{array}{l} \xi_i \mapsto x_{i_1} \hat{\xi}'_{i_1} x_{i_2} \hat{\xi}'_{i_2} \cdots x_{i_{k-1}} \hat{\xi}'_{i_{k-1}} x_{i_k}, \\ \qquad \qquad \qquad i = 1, \dots, N \\ \xi'_{i_m} \mapsto \sum_{\gamma_j \notin \overline{D_{x_{i_1} \cdots x_{i_m}}}} \xi_j + \left(\sum_{\gamma_j \notin \overline{D_{x_{i_1} \cdots x_{i_m}}}} \xi_j \right) \xi'_{i_m} \\ \xi_0 \mapsto \sum_{i=1}^N \xi_i + \left(\sum_{i=1}^N \xi_i \right) \xi_0 \end{array} \right.$$

où le symbole $\hat{\xi}'_{i_m}$ indique la présence ou l'absence de l'occurrence de la variable ξ'_{i_m} ($\hat{\xi}'_{i_m}(\xi'_{i_m} = e + \xi_{i_m})$)

THÉORÈME 2.1. *Tout ensemble de Dyck est un langage algébrique nonambigu.*

Soit maintenant $u = x_{j_1}x_{j_2} \cdots x_{j_k}$ un mot minimal de X^* , soit $\lambda = \sup_{i \in [1, N]} |\gamma_i|$, et, si $f \in X^*$, convenons de poser, pour tout facteur gauche f' de f ,

$$\begin{aligned} D_{f'}^{\leftarrow} &= D_{f'}^-, \text{ si } |f'| \leq \lambda, \\ &\text{et } f' = f''f''', \text{ avec } |f''| = \lambda. \\ D_{f'}^{\leftarrow} &= D_{f'}^-, \text{ si } |f'| > \lambda \end{aligned}$$

LEMME 2.4. *Tout mot f de X^* tel que $f \xrightarrow{*} u$ se factorise d'une manière et d'une seule en*

$$f = f_0 x_{j_1} f_1 \cdots x_{j_k} f_k,$$

où $f_0 \in D^*$ et, pour tout $m \in [1, k]$, $f_m \in (D \setminus D_{x_{j_1} \cdots x_{j_m}}^{\leftarrow})^*$.

DÉMONSTRATION Supposons d'abord que $k = 1$. Montrons l'unicité d'une factorisation $f = f_0 x_{j_1} f_1$, $f_1 \in (D \setminus D_{x_{j_1}}^{\leftarrow})^*$, en prouvant que f_0 est le plus long facteur gauche propre de f qui appartienne à l'ensemble de Dyck D^* . Supposons que f admette deux factorisations $f = f_0 x_{j_1} f_1$ et $f = g_0 x_{j_1} g_1$ telles que, par exemple, $|f_0| < |g_0|$; alors on a $f_1 = f' x_{j_1} g_1$, ce qui implique $f' x_{j_1} \xrightarrow{*} e$, et, d'après le lemme 2.2., $f' x_{j_1}$ admet une factorisation unique en mots premiers $f' x_{j_1} = h_1 h_2 \cdots h_m$, où $h_m = h'_m x_{j_1}$; de même g_1 admet l'unique factorisation $g_1 = g'_1 g'_2 \cdots g'_n$, donc $f_1 = h_1 h_2 \cdots h'_m x_{j_1} g'_1 g'_2 \cdots g'_n$; mais on a aussi $x_{j_1} f' \xrightarrow{*} e$, par conséquent $x_{j_1} h'_m \xrightarrow{*} e$, et $h_m \in D_{x_{j_1}}$, ce qui contredit l'hypothèse.

Montrons maintenant l'existence d'une telle factorisation. Soit $f = f_0 f'$, où f_0 est le plus long facteur gauche propre de f qui soit dans D^* ; montrons qu'alors $f' = x_{j_1} f_1$, $f_1 \in (D \setminus D_{x_{j_1}}^{\leftarrow})^*$;

si l'n'en était pas ainsi, on aurait

— Soit $f' = x_{j_m} f''$, $x_{j_m} \neq x_{j_1}$ et la condition $f' \xrightarrow{*} x_{j_1}$ entraîne l'existence d'un f''' et d'un $f''v$ tels que $f' = x_{j_m} f''' x_{j_1} f''v$ et $x_{j_m} f''' \xrightarrow{*} e$, ce qui contredit la maximalité de f_0 ;

— Soit $f' = x_{j_1} f''$, $f'' \in D^*$; donc $f'' = f'_1 \cdots f'_m$ et l'existence d'un $n \in [1, m]$ tel que $f''_n \in D_{x_{j_1}}^{\leftarrow}$ contredirait encore la maximalité de f_0 .

Supposons le lemme vrai pour $m < k$ et considérons f tel que $f \xrightarrow{*} x_{j_1} \cdots x_{j_k}$. Comme précédemment, on remarque que si f se factorise en $f_0 x_{j_1} f_1 \cdots x_{j_k} f_k$, $f' = f_0 x_{j_1} f_1 \cdots x_{j_{k-1}} f_{k-1} f$ est le plus long facteur gauche de f tel que $f' \rightarrow x_{j_1} \cdots x_{j_{k-1}}$, d'où l'unicité.

Réciproquement, supposons $f = f' f''$, où f' est le facteur gauche le plus long de f tel que $f' \xrightarrow{*} x_{j_1} \cdots x_{j_{k-1}}$; le même raisonnement que ci-dessus montre que $f'' = x_{j_k} f_k$, $f_k \in (D \setminus D_{x_{j_1} \cdots x_{j_k}}^{\leftarrow})^*$, et l'hypothèse de récurrence entraîne que f' se factorise de façon unique en $f' = f_0 x_{j_1} f_1 \cdots x_{j_{k-1}} f_{k-1}$, d'où le lemme.

D'après le lemme précédent, la grammaire G_u suivante, d'axiome ξ_u , engendre la classe du mot minimal u de manière non-ambigüe

$$G_u = \begin{cases} G_{D^*} \\ \xi'_{j_m} \mapsto \sum_{\gamma_i \notin D_{x_{j_1} \dots x_{j_m}}} \xi_i + \left(\sum_{\gamma_i \notin D_{x_{j_1} \dots x_{j_m}}} \right) \xi'_{j_m} \\ \xi_u \mapsto \xi_0 x_{j_1} \xi'_{j_1} \dots x_{j_k} \xi'_{j_k} \end{cases}$$

THÉORÈME 2.2. *Toute classe d'équivalence d'une congruence de Dyck est un langage algébrique non-ambigü.*

Nous savons par ailleurs que l'ensemble des mots minimaux de X^* , qui n'est autre que le complémentaire dans X^* de $\bigcup \{X^* \gamma_i X^* \mid i \in [N]\}$, est un langage rationnel, que l'on peut engendrer de façon non-ambigüe au moyen d'une grammaire linéaire à gauche. (Une telle construction se trouve dans [1]). La même construction que celle de [2], corollaire IV.4.2. page 425, permet alors d'affirmer que le complémentaire de l'ensemble de Dyck est encore un langage algébrique non-ambigü. D'où le

THÉORÈME 2.3. *Le complémentaire d'un ensemble de Dyck est un langage algébrique non-ambigü.*

REMARQUES. Il est assez évident que D^* est un langage algébrique déterministe, c'est-à-dire susceptible d'être reconnu par un automate à mémoire pile déterministe (cf. [3] et [2]). Le théorème 2.3. résulte alors du théorème 8 de [3], ainsi d'ailleurs que la non-ambigüité de D^* . La construction effective des grammaires non-ambigües de ces deux langages, reposant sur les propriétés de factorisation unique qu'énoncent nos lemmes 2.2., 2.3. et 2.4., et qui généralisent la propriété 4.4.1. de [2] ne résulte pas cependant des propriétés énoncées en [3]. C'est à cette construction que va tout notre intérêt dans la mesure où nous poursuivons ce travail en cherchant à étendre nos résultats à des classes de congruence parfaite mais non-triviales. [4]

BIBLIOGRAPHIE

1. P. Jullien, *Contribution à l'étude des types d'ordres dispersés*, Thèse, Faculté des Sciences de Marseille, 1969.
2. M. Nivat, *Transduction des langages de Chomsky*, Ann. Inst. Fourier **18** (1968), 339-456.
3. S. Ginsburg and S. Greibach, *Deterministic context-free languages*, Information and Control **9** (1966), 620-648.
4. M. Nivat, *On some families of languages related to the Dyck set*, 2nd Annual ACM Symposium on Theory of Computing, 1970.