LOGIC, METHODOLOGY AND PHILOSOPHY OF SCIENCE IV

PROCEEDINGS OF THE FOURTH INTERNATIONAL CONGRESS FOR LOGIC, METHODOLOGY AND PHILOSOPHY OF SCIENCE, BUCHAREST, 1971

Edited by

PATRICK SUPPES
Stanford University, Stanford, USA
LEON HENKIN
University of California, Berkeley, USA
ATHANASE JOJA
Académie Roumaine, Bucarest, Roumaine
GR. C. MOISIL
Université de Bucarest, Bucarest, Roumaine

1973

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • LONDON

AMERICAN ELSEVIER PUBLISHING COMPANY, INC.
NEW YORK

SUR UN LANGAGE EQUIVALENT AU LANGAGE DE DYCK

M. P. SCHÜTZENBERGER

Faculté des Sciences, Paris, France

1. Introduction

Le langage de Dyck D_2 sur *deux* paires de lettres $\{\bar{y}, y, \bar{y}', y'\}$ est comme on sait, défini de façon équivalente, comme la solution de l'équation

$$\xi = 1+\bar{y}\xi y\xi+\bar{y}'\xi y'\xi$$

ou comme la classe de 1 pour la congruence

$$1 \equiv \bar{y}y \equiv \bar{y}'y'$$

sur le monoïde libre $\{\bar{y}, y, \bar{y}', y'\}^*$.

Appelant suivant Eilenberg *cône* $\mathscr{C}(L)$ d'un langage L, la famille des langages qui peuvent être déduits de L par une relation rationnelle, on sait que le cône $\mathscr{C}(D_2)$ est l'ensemble des langages algébriques.

Cette propriété n'est pas partagée par le langage de Dyck $D_1 = D_2 \cap \{\bar{y}, y\}^*$ sur une seule paire de lettres.

On se propose de montrer qu'au contraire la même propriété est possédée par le langage $L \subset \{\bar{y}, y\}^*$ défini de façon équivalente comme:

— le quotient de $\{\bar{y}, y\}^* = Y^*$ par la congruence définie par:

$$\bar{y}y \equiv \bar{y}\bar{y}y\bar{y}yy;$$

— la solution de l'équation

$$\xi = \bar{y}y+\bar{y}\xi\xi y;$$

— la partie L de $D_1\backslash 1$ formée des mots $d \in D_1\backslash 1$ tels que pour chaque factorisation $d = a\bar{y}b$ $(a, b \in Y^*)$ le mot $\bar{y}b$ ait exactement un facteur gauche dans $D_1\backslash 1$ ssi $a \in 1 \cup Y^*y$.

On observera enfin que *l'équivalence rationnelle* de D_2 et L $\big(\mathscr{C}(D_2) = \mathscr{C}(L)\big)$ entraîne celle de D_2 et de tout langage défini par une équation de la forme $\xi = a+b\xi c\xi d$ où cette fois a, b, c et d sont quatre mots vérifiant des conditions assez peu restrictives (par exemple que $\{a, b, c\}$ engendre librement le monoïde $\{a, b, c\}^*$). En raison de sa simplicité cette dernière équation peut présenter des avantages techniques dans certains problèmes d'équivalence rationnelle avec D_2.

2. Les différentes définitions de L

Nous considérons $L \subset Y^*$ comme étant défini par l'équation

$$\xi = \bar{y}y + \bar{y}\xi\xi y.$$

Posant $X = \{\xi, \bar{y}, y\}$ ceci équivaut à $L = M \cap Y^*$, où M est la plus petite partie de X^* contenant ξ qui soit telle que

$$f, g \in X^*, \quad f\xi g \in M \Rightarrow f\bar{y}yg, \quad f\bar{y}\xi\xi yg \in M.$$

De façon équivalente, M peut être considéré comme le langage obtenu en remplaçant x par ξ dans la solution de l'équation

$$\xi = x + \bar{y}y + \bar{y}\xi\xi y.$$

Notant comme d'usage $|h|_z$ le nombre d'occurrences de la lettre z dans le mot h, nous déduisons immédiatement de la définition de M les deux propriétés suivantes :

2.1. $e \in M \Rightarrow e \in \{\xi\} \cup X^*(\bar{y}y + \bar{y}\xi\xi y)X^*$

2.2. $e \in M \Rightarrow |e|_{\bar{y}} = |e|_y$

et, si $e = fg$ où $f, g \in X^* \backslash \xi^*$, alors

$$|f|_{\bar{y}} > |f|_y.$$

Preuve: En effet ces assertions sont trivialement vraies pour $e = \xi$ et si elles sont vraies pour un mot $e \in X^*$ elles le demeurent pour tout mot e' obtenu en remplaçant dans e une occurrence de ξ par $\bar{y}y$ ou par $\bar{y}\xi\xi y$.

Ecrivons $h' \in h\delta_i^{-1}$ $(i = 0, 1)$ ssi réciproquement il existe des mots $f, g \in X^*$ tels que $h = fu_ig$, $h' = f\xi g$ où $u_0 = \bar{y}y$, $u_1 = \bar{y}\xi\xi y$.

Rappelons que deux mots $a, b \in X^*$ *ne chevauchent pas* ssi il n'y a aucun mot $c \in XX^*$ tel que :

$$a \in X^*Xc, \quad \text{et} \quad b \in cXX^* \quad \text{ou}$$
$$a \in cXX^*, \quad \text{et} \quad b \in X^*Xc.$$

Cette condition est satisfaite ssi pour toute relation $fag = f'bg'$ $(f, f', g, g' \in X^*)$ l'on a l'une des quatre éventualités suivantes :

—$f'b$ un facteur gauche de f;

— $f'b$ un facteur gauche de fa et f un facteur gauche f' ;
— fa un facteur gauche de $f'b$ et f' un facteur gauche de f;
— fa un facteur gauche f'.

2.3. $h \in X^* \Rightarrow h\delta_0^{-1}\delta_1^{-1} = h\delta_1^{-1}\delta_0^{-1}$.

PREUVE: Ceci résulte immédiatement de ce que deux mots (éventuellement égaux) de $\{u_0, u_1\}$ ne chevauchent pas et que, de plus, aucun d'eux n'est facteur de l'autre quand ils sont différents.

Soit maintenant $\equiv$ la plus petite congruence sur X^* telle que

$$\xi \equiv \bar{y}y \equiv \bar{y}\xi\xi y.$$

2.4. *M est la classe de ξ pour $\equiv$ et L est la classe de $\bar{y}y$ pour la congruence $(\equiv)$ sur Y^* définie par $\bar{y}y \; (\equiv) \; \bar{y}\bar{y}y\bar{y}yy$.*

PREUVE: Supposant $e \in M$ tel que $e\delta_0^{-1} \cup e\delta_1^{-1} \subset M$, et $e = f\xi g$ $(f, g \in Y^*)$, il résulte de 2.3 que l'on a encore $e'\delta_0^{-1} \cup e'\delta_1^{-1} \subset M$ pour $e' = fu_i g$ $(i = 0, 1)$. Par induction ceci implique $M = \{h \in X^* : h(\delta_0^{-1} \cup \delta_1^{-1})^* = \xi\}$, donc

$$M = \{h \in X^* : h \equiv \xi\}.$$

Comme $\bar{y}\xi\xi y \in \bar{y}\bar{y}y\bar{y}yy\delta_0^{-1}\delta_0^{-1}$, la deuxième partie de l'énoncé résulte de la première.

COROLLAIRE 2.5: *Le langage L est rationnellement équivalent au langage N solution de l'équation*

$$\xi = x + \bar{y}\xi\xi y.$$

PREUVE: Soit φ le morphisme de $\{x, \bar{y}, y\}^*$ sur Y^* tel que $x\varphi = \bar{y}y$, $\bar{y}\varphi = \bar{y}$, $y\varphi = y$.

Il est clair que la restriction $\varphi|N$ est une bijection sur L.

Réciproquement, si $e = f\bar{y}zg \in L$ $(z = \bar{y}$ ou $y)$ il résulte de 2.5 que l'on a $f\xi g \in M$ ssi $z = y$. De même si $fzyg \in L$ $z = \bar{y}$ ou y on a $f\xi g \in M$ ssi $z = \bar{y}$. Il en résulte que $N = L\bar{\varphi}$ (avec $\varphi\bar{\varphi} = 1$) où $\bar{\varphi} : Y^* \to \{\bar{y}, y, x\}^*$ est la relation rationnelle remplaçant par x tout $\bar{y}$ suivi d'un y, et par 1 tout y précédé d'un $\bar{y}$ et laissant inchangées les autres lettres.

2.6. *Soit $e = f\bar{y}g \in M$. On a $\bar{y}g = mg'$ où $g' \in yY^*$ et où $m \in M$ ou $\in M^2$ selon que $f \in Y^*(y+\xi)$ ou $\in 1 \cup Y^*\bar{y}$.*

PREUVE: L'énoncé est vrai pour $e = \xi$ et s'il est vrai pour $e = f'\xi g'$, il est encore vrai pour $f'u_i g'$.

Rappelons maintenant que le langage de Dyck D_1 est le sous-monoïde D^* de Y^* librement engendré par l'ensemble D des mots $d \in YY^*$ tels que :

— $|d|_{\bar{y}} = |d|_y$;

— $d = fg, \quad f, g \in YY^* \Rightarrow |f|_{\bar{y}} > |f|_y$.

Cette condition implique que deux mots de D ne se chevauchent pas et que tout mot de Y^* ait au plus un facteur gauche (et droit) dans D. On sait aussi que :

Pour toute factorisation $f\bar{y}g$ d'un mot de D_1 $(f, g \in Y^*)$, on a $\bar{y}g = d'g'$ où $g' \in 1 \cup yY^*$ et $d' \in D_1 \backslash 1$, cette factorisation étant unique. On a alors $fg' \in D_1$.

2.7. *Les deux conditions suivantes* (D) *et* $(D)'$ *sur un mot* $d \in D_1 \backslash 1$ *sont équivalentes* :

Pour toute factorisation $d = f\bar{y}g$ $(f, g \in Y^*)$,

(D) $\bar{y}g$ *a exactement un facteur gauche dans* $D_1 \backslash 1$ *ssi* $f \in 1 \cup Y^*y$;

$(D)'$ $\bar{y}g = d'g'$ *où* $g' \in 1 \cup yY^*$ *et où* $d' \in D$ *ou* $\in D^2$ *selon que* $f \in Y^*\bar{y}$ *ou* $\in 1 \cup Y^*y$.

Preuve: En vertu de $D_1 = D^*$ on peut écrire $\bar{y}g = d_1 d_2 \dots d_p g'$ où $p \geqslant 1$, $d_1, d_2, \dots, d_p \in D$, $g' \notin DY^*$, donc $g' \in 1 \cup yY^*$. Comme $D \subset D_1 \backslash 1$ p est le nombre de facteurs gauche dans $D_1 \backslash 1$ de $\bar{y}g$.

Maintenant, comme $D \subset Y^*\bar{y}$, la condition (D) équivaut à la condition que $p \leqslant 2$ avec égalité ssi $f \in Y^*\bar{y}$, c'est-à-dire a $(D)'$.

Propriété 2.8.: *L est la partie de $D_1 \backslash 1$ définie par* (D).

Preuve: La remarque 2.2 montre que $L \subset D \subset D_1 \backslash 1$ et le fait que L satisfait (D) résulte de 2.6 et 2.7.

Réciproquement, on a $\bar{y}y \in D \cap L$. Supposons que l'implication $(D)' \Rightarrow L$ soit établie pour tous les mots plus courts que le mot $d \in D_1 \backslash (1 + \bar{y}y)$ satisfaisant $(D)'$.

On peut écrire $d = f\bar{y}g$ avec $f = 1$. Donc, $d \in D$, d'après $(D)'$ et plus précisément, $d = \bar{y}d_1 d_2 y$ où $d_1, d_2 \in D$.

Comme la condition (D) sur un mot, implique la même condition sur tous les facteurs dans $D_1 \backslash 1$ de ce mot, l'hypothèse d'induction donne $d_1, d_2 \in L$, d'où évidemment,

$$d = \bar{y}d_1 d_2 y \in L.$$

3. Equivalence rationnelle de D_2 et de L

Comme L est algébrique et que par conséquent $L \in \mathscr{C}(D_2)$, il suffit de montrer que réciproquement $D_2 \in \mathscr{C}(L)$. Pour simplifier les calculs on établira plutôt $D'_2 \in \mathscr{C}(N)$ où $D'_2 \subset Y'^*$ $(Y' = \{\bar{y}, y, \bar{y}', y', y'', x\})$ est défini par l'équation

$$\xi = x+\bar{y}\xi y\xi y''+\bar{y}'\xi y'\xi y''$$

et M par l'équation

$$\xi = x+\bar{y}\xi\xi y.$$

Comme $\mathscr{C}(L) = \mathscr{C}(N)$ ainsi qu'on l'a vu en 2.6 et comme $D'_2 \in \mathscr{C}(D_2)$ puisque D'_2 est algébrique, l'équivalence des inclusions $D_2 \in \mathscr{C}(L)$ et $D'_2 \in \mathscr{C}(N)$ résulte immédiatement de la relation $D_2 = D'_2\varphi$ où φ est le morphisme de Y'^* dans lui-même envoyant y'' et x sur 1 et laissant inchangées les autres lettres.

De façon inverse, l'inclusion $D'_2 \in \mathscr{C}(N)$ sera vérifiée en prouvant l'existence d'un morphisme $\psi: (\xi \cup Y')^* \to (\xi \cup X')^*$ $(X' = \{x, \bar{y}, y\})$ tel qu'en posant $a = \bar{y}\psi$, $a' = \bar{y}'\psi$, $b = y\psi$, $b' = y'\psi$, $c = y''\psi$, $d = x\psi$ les deux conditions suivantes soient satisfaites:

(1) $\xi\psi = \xi$ et ψ est injectif.

(2) Il existe un langage rationnel $R \subset Y'^*$ tel que $D'_2 \subset R$ et que $R\psi \cap N$ soit la solution P de l'équation

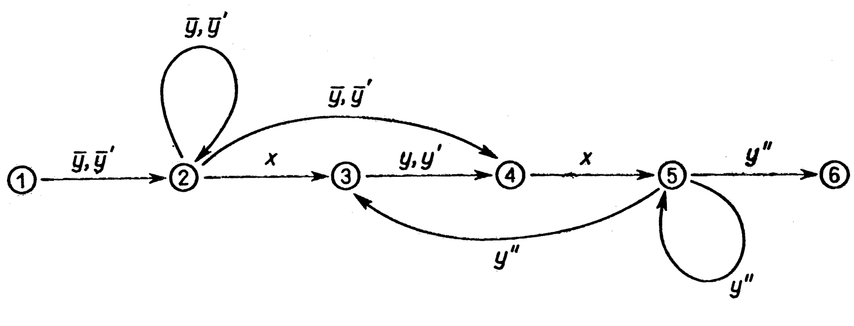
$$\xi = d+a\xi b\xi c+a'\xi b'\xi c.$$

En effet sous ces conditions on aura:

$$D'_2 = (N \cap R\psi)^{-1}\psi, \text{ puisque } D'_2\psi \subset P, \text{ par construction.}$$

Nous commençons par la construction de R.

3.1. *$D'_1\backslash x$ est contenu dans le langage R' accepté par l'automate à 6 états avec 1 et 6 comme état initial et final respectivement.*



PREUVE: Soit $R_0 = (\bar{y}+\bar{y}')x(y+y')xy''$. On a $R_0 \subset R'$ et on voit sur le graphe que $fxg \in R' \Rightarrow fR_0g \subset R'$. Donc, par induction $\Rightarrow fR'g \subset R'$ ce qui établit $D_2'\backslash x \subset R'$.

Soit maintenant $X'' = \{\xi, x, \bar{y}, y\}$, et soit $\equiv$ la congruence sur X''^* définie par :

$$\xi \equiv x \equiv \bar{y}y \equiv \bar{y}\xi\xi y.$$

Sa restriction à X est la congruence étudiée plus haut et notant M' la classe de ξ on voit d'après 2.4 que $N = M'\backslash X''^*(\xi+\bar{y}y)X''^*$.

Nous choisissons maintenant le morphisme ψ et nous posons :

$$a = \bar{y}\psi = \bar{y}\bar{y}x\bar{y}\bar{y}\,; \quad b = y\psi = wyxyy$$
$$a' = \bar{y}'\psi = \bar{y}\bar{y}\bar{y}x\bar{y}\,; \quad b' = y'\psi = wyyxy$$
$$c = y, \quad d = x$$

où $w = \bar{y}xxy$.

Ce morphisme est injectif puisqu'aucun mot de $A = Y'\psi$ n'est facteur gauche d'un autre mot de A.

3.2. *P est contenu dans N et axb' et a'xb appartiennent à* $0 = \{h \in M' : X''^*hX''^* \cap M' = \emptyset\}$.

PREUVE: La première assertion résulte de $a\xi b\xi c$, $a'\xi b'\xi c \in M'$ qui découle elle-même du parenthésage :

$$a\xi b\xi c = \bar{y}\Big(\bar{y}x\big(\bar{y}(\bar{y}\xi(w)y)(xy)y\big)\xi y\Big) \in M'$$
$$a'\xi b'\xi c = \bar{y}\Big(\bar{y}\big(\bar{y}x(\bar{y}\xi(w)y)y\big)xy\Big)\,\xi y \in M'.$$

Pour vérifier la seconde assertion, nous notons que d'après 2.6, on a $\bar{y}\xi y$, $\xi\xi\xi \in 0$. Par conséquent,

$$axb' \equiv a\xi b' = \bar{y}\bar{y}x\bar{y}\big(\bar{y}\xi(w)y\big)yxy \equiv \bar{y}\bar{y}x\bar{y}\xi yxy \in 0$$

et

$$a'xb \equiv a'\xi b = \bar{y}\bar{y}\bar{y}x(\bar{y}\xi wy)xyy \equiv \bar{y}\bar{y}\bar{y}\xi\xi\xi yy \in 0.$$

Fin de la preuve de $D_2 \in \mathscr{C}(L)$.

Il ne nous reste plus qu'à vérifier :

$$sx \in (x \cup S) \cap N \Rightarrow s \in P$$

où $S = R'\psi$. Comme le résultat est vrai pour $s = x$, $axbc$ et $a'xbx'c$ où les deux derniers sont les mots les plus courts de S, nous pouvons, par induction sur la longueur, nous borner à vérifier que tout mot $s \in S \cap N \backslash\{axbxc+a'xb'xc\}$ admet une factorisation $s = s_1vs'$ telle que $v = axbxc$ ou $a'xb'xc$ et que $s_1xs' \in S \cap N_1+x$.

Soit donc un tel mot s. Considérant l'automate qui définit R', l'hypothèse $s \in S$ implique l'existence d'une factorisation $s = s_1 a s_2$ où $s = s_1 a' s_2$ avec $s_1 \in A^*$ et $s_2 \in A^* \backslash A^*(a+a')A^*$. De plus, la même hypothèse entraîne $s_2 \in x(b+b')x \subset s'$ où $s_1 x s' \in x \cup S$.

On a donc $s = s_1 r s'$ où $r \in (a+a')x(b+b')xc$. Comme $r \in P$ si $r = axbxc$ ou $r = a'xb'xc$ il n'y a donc qu'à vérifier $r \neq axb'xc$, $a'xbxc$, ce qui résulte immédiatement de 3.2 puisque $s \in N$.

Note bibliographique. Le langage défini par l'équation $\xi = d + a\xi b\xi c$ a été introduit par M. Nivat (*Thèse*, Paris, 1967) qui l'appele „langage de expressions arithmétiques". Un élève de M. Nivat, Y. Cochet a traité de façon approfondie dans sa *Thèse* (Rennes, 1971) les langages qui sont classe d'une congurence. Enfin une théorie complète des langages rationellement équivalents à D_2 est en cours de publication sous les signatures conjoints de L. Boasson et M. Nivat.

Nous avons fait dans le présent travail de nombreux emprunts aux recherches de ces auteurs.