Marcel-Paul Schützenberger

ŒUVRES COMPLÈTES

éditées par
Jean Berstel, Alain Lascoux et Dominique Perrin

*

Tome 5 : 1961–1963

Institut Gaspard-Monge, Université Paris-Est
2009

Introduction
Tome V : 1961–1963

C'est une période extrêmement féconde, qui voit la parution de nombreux articles dans plusieurs de ses domaines principaux de recherche :

- les langages algébriques (ou context-free) avec [1961-1], [1962-6], [1963-1], [1963-5] et [1963-7],
- les séries rationnelles : notamment [1961-4], [1962-1], [1962-3] et [1963-4]
- la combinatoire, avec [1962-5] et [1963-7]
- la théorie des codes, avec [1961-2] et [1961-5]
- les automates et les transductions, avec [1961-3].

C'est certainement l'article en commun avec Noam Chomsky, *The algebraic theory of context-free languages* [1963-7] qui est le plus connu, en tout cas le plus cité. Il a fait l'objet d'une traduction en allemand, et d'une traduction en français ([1968-4]), peut-être dans d'autres langues. Cet article esquisse et même décrit de manière assez détaillée une théorie des langages dits context-free qui serait algébrique plutôt que constructive ou combinatoire. Par exemple, une grammaire pour un langage est considérée comme un système d'équations, les variables de la grammaire jouant le rôle des inconnues. Les solutions de ce système d'équation sont calculées par approximations successives (c'est la méthode des points fixes) dans l'algèbre des séries formelles non commutatives, et prennent ainsi en compte les ambiguïtés éventuelles dans les dérivations. L'article contient de nombreux exemples et des propriétés de clôture. On y trouve aussi la définition des langages de Dyck, des langages rationnels locaux, et le fameux théorème de Chomsky-Schützenberger selon lequel tout langage algébrique est l'image homomorphe de l'intersection d'un langage de Dyck et d'un langage rationnel local. L'article contient aussi des résumés ou des citations de nombreux résultats démontrés par M.-P. Schützenberger dans d'autres publications.

Ce volume contient d'autres travaux sur les langages *context-free*. Publié avant son travail en commun avec Chomsky, l'article *Some remarks on Chomsky's context-free languages* [1961-1] est une étude sur les liens entre langages et séries formelles. En particulier, il y est établi que le support des séries $\mathbb{N}$-rationnelle (resp. $\mathbb{N}$-algébriques) est un langage rationnel (resp. algébrique). C'est ici que se trouve l'origine de la terminologie « moderne » de langage *rationnel* ou *algébrique* pour *regular* ou *context-free*.

L'article *On probabilistic push-down storage* [1962-6] décrit une variante du fonctionnement usuel des automates à pile : ces automates peuvent effacer des « segments rationnels » dans leur pile ; ceci veut dire que l'automate peut, en

un seul mouvement, effacer un mot de pile pourvu qu'il soit dans un langage rationnel donné. Cette extension donnera naissance beaucoup plus tard au « automates à taquets » de Sheila Greibach [4] qui permettra une classification assez fine des langages algébriques déterministes. Dans *On context-free languages and push-down automata* [1963-5], M.-P. Schützenberger montre que, dans le théorème de Chomsky-Schützenberger, le langage de Dyck peut être remplacé par le noyau d'un groupe libre (aussi appelé « langage de Dyck bilatère »), tentant ainsi de relier la théorie des langages algébriques à l'algèbre plus classique. Cette tentative donnera des résultats beaucoup plus tard, comme par exemple les travaux de Muller et Schupp [6] sur le lien entre les langages algébriques et la théorie des bouts.

La formulation algébrique d'une propriété – que nous venons de mentionner à propos des grammaires – est l'un des objectifs récurrents de M.-P. Schützenberger qui considère les objects non commutatifs, les mots, codes, grammaires ou séries, comme relevant d'un traitement algébrique.

Les travaux de M.-P. Schützenberger ont montré que les grammaires des langages algébriques se traduisent très simplement en des équations pour les séries formelles génératrices énumérant les mots du langage correspondant. Sans que Schützenberger lui-même ait publié des écrits sur ce sujet, ce constat et son enseignement ont permis à certains de ses élèves d'élaborer, en liaison aussi avec la méthode bijective, une technique très puissante pour obtenir des formules d'énumération. Cette technique, dont les premières prémices remontent peut-être à Maurice Gross [5], a été particulièrement féconde pour énumérer plusieurs famille d'objets proches des mots de Dyck, comme par exemple les cartes et triangulations planaires qui constituent un des chapitres les plus actuels des interactions entre physique statistique et combinatoire.

Les articles *On a special class of recurrent events* [1961-2] et *On a family of submonoids* [1961-5], contiennent les principaux résultats de la théorie des codes bifixes. Le premier est publié dans une revue de théorie des probabilités (Annals of Mathematical Statistics). Comme beaucoup d'articles de M.-P. Schützenberger, il met en valeur un énoncé présenté comme le résultat principal de l'article, alors que celui-ci en contient beaucoup d'autres. Le résultat en question (Property 1) est le fait remarquable que les codes bifixes maximaux coupants ont une longueur moyenne qui est un entier quelle que soit la distribution de Bernoulli sur l'alphabet. L'énoncé est exprimé dans la terminologie de Feller des évènements récurrents. Il avait été formulé dans un cas particulier par Gilbert et Moore dans leur célèbre article de 1959. L'article présente, sous forme d'énoncés numérotés comme des remarques, de très nombreuses propriétés des codes bifixes, parmi lesquelles le fait qu'il n'existe qu'un nombre fini de codes bifixes maximaux finis de degré donné et un grand nombre de constructions, comme la transformation interne dont Cesari devait prouver en 1972 ([2]) qu'elle permet d'obtenir tous les codes bifixes maximaux finis.

Le deuxième article, *On a family of submonoids*, est publié dans une revue à moins grande diffusion (la revue de l'Académie des Sciences Hongroise). Il est destiné à un public d'algèbristes et présente les sous-monoïdes engendrés par les codes bifixes comme une généralisation intéressante des sous-groupes d'un groupe. L'article développe les constructions utilisant la théorie des semigroupes et en particulier la structure de l'idéal minimal et du groupe de Suschkevitch d'un semigroupe ayant des idéaux minimaux.

Introduction

L'article *A remark of finite transducers* [1961-3] est le premier dans lequel M.-P. Schützenberger traite des transducteurs non ambigus qui apparaissent en particulier pour le décodage des codes à longueur variable. Il définit les bimachines, reprises plus tard dans le livre d'Eilenberg en 1974, comme un modèle permettant de représenter toutes les fonctions rationnelles et en particulier la composition d'une transduction séquentielle gauche et d'une transduction séquentielle droite.

L'article *On the definition of a family of automata* [1961-4] est le document fondateur pour la théorie des séries rationnelles en variables non commutatives, vues comme l'extension des langages rationnels. Il contient ce que l'on appelle maintenant le théorème de Kleene–Schützenberger, et aussi l'algorithme de construction de la représentation minimale. Depuis, le théorème de Kleene–Schützenberger a été redémontré et étendu à des situations plus générales, comme les séries en variables partiellement commutatives. L'article *Finite counting automata* [1962-1] contient l'un des théorèmes les plus difficiles de la théorie, à savoir la caractérisation des séries rationnelles à croissance polynomiale. Ce théorème a, comme cas particulier, le théorème de Burnside pour les semigroupes de matrices. Une démonstration retravaillée de ce résultat se trouve dans le livre [1] qui contient aussi les résultats sur les séries rationnelles non commutatives postérieurs aux travaux de Schützenberger, notamment ceux de Fliess, Jacob, Soittola, Berstel, Reutenauer.

L'article [1962-2] *Certain infinite formal products and their combinatorial applications*, qui fait partie des actes d'un colloque, est la première contribution à la théorie des factorisations des monoïdes libres qui donnera lieu en particulier à l'article [1965-6].

L'article *The equation* $a^m = b^n c^p$ *in a free group* [1962-5] contient ce que l'on appelle le théorème de Lyndon et Schützenberger selon lequel l'équation dans le groupe libre $x^n y^m = z^p$ n'a, pour $n, m, p \geq 2$, que des solutions triviales, c'est-à-dire que x, y, z sont tous puissances d'un même élément. L'un des lemmes utilisés (Lemma 4) est le fait que si les mots x^n et y^m ont un préfixe commun de longueur $|x| + |y|$, alors x et y sont puissances d'un même élément. La borne peut en fait être remplacée par $|x| + |y| - 1$ mais cela n'apparaît qu'un peu plus tard, dans l'article de Fine et Wilf de 1965 ([3]).

L'article *On the minimum number of elements in a cutting set of words* [1962-8] est un rapport de recherche d'IBM sur les ensembles coupants. Il donne une borne asymptotique sur la taille des ensembles coupants minimaux de mots de même longueur. Le résultat sera repris dans l'article de [1964-1].

L'article *Quelques remarques sur une construction de Schensted* [1963-6], repris dans l'exposé [1963-3], est le début d'une longue série d'articles concernant les *tableaux de Young*. Schensted a défini une bijection entre permutations w et paires de tableaux $(P(w), Q(w))$ de même forme (construction liée en fait à un travail plus ancien de Robinson). M.-P. Schützenberger montre que la paire correspondant à l'inverse de w est $(Q(w), P(w))$, et que par ailleurs la conjugaison par rapport à la permutation maximale induit une involution sur les tableaux, qui peut se réaliser par une procédure d'*évacuation*. Cette involution apparaît fréquemment dans la littérature sous le nom d'*involution de Schützenberger*, terme désavoué par M.-P. Schützenberger qui lui préférait le nom d'*involution naturelle*.

[1] Jean Berstel and Christophe Reutenauer. *Noncommutative Rational Series with Applications*. 2009. to appear.

[2] Yves Césari. Sur un algorithme donnant les codes bipréfixes finis. *Math. Systems Theory*, 6 :221–225, 1972.

[3] N. J. Fine and H. S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.*, 16 :109–114, 1965.

[4] Sheila A. Greibach. Jump PDA's and hierarchies of deterministic context-free languages. *SIAM J. Comput.*, 3 :111–127, 1974.

[5] Maurice Gross. Applications géométriques des langages formels. *International Computer Centre Bulletin*, 5 :141–167, 1966.

[6] David E. Muller and Paul E. Schupp. The theory of ends, pushdown automata, and second-order logic. *Theoret. Comput. Sci.*, 37(1) :51–75, 1985.

Année 1961

Bibliographie

[1] Marcel-Paul Schützenberger. Some remarks on Chomsky's context-free languages. *Quarterly Progress Report of the Research Lab. of Electronics, MIT*, 68 :155–170, 1961.

[2] Marcel-Paul Schützenberger. On a special class of recurrent events. *Ann. Math. Statist.*, 32 :1201–1213, 1961.

[3] Marcel-Paul Schützenberger. A remark on finite transducers. *Information and Control*, 4 :185–196, 1961.

[4] Marcel-Paul Schützenberger. On the definition of a family of automata. *Information and Control*, 4 :245–270, 1961.

[5] Marcel-Paul Schützenberger. On a family of submonoids. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 6 :381–391, 1961.

[6] David D. Rutstein, Murray Eden, and Marcel-Paul Schützenberger. Report on mathematics in the medical sciences. *New England J. Medecine*, 265 :172–176, 1961.

Reprinted from

QUARTERLY PROGRESS REPORT No. 63, October 15, 1961

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Cambridge, Massachusetts

References

1. R. Jakobson, Word 4, 155-167 (1960).

C. SOME REMARKS ON CHOMSKY'S CONTEXT-FREE LANGUAGES

1. Introduction

This report is devoted to the examination of several families of subsets of a free monoid that arise rather naturally when generalizing some definitions of classical analysis to the noncommutative case. These families contain, in particular, the regular events of Kleene and the context-free languages of Chomsky.

The main tool is the so-called formal power series with integral coefficients in the noncommutative variates $x \in X$.

By definition, such a formal power series, r, is a mapping that assigns to every word $f \in F(X)$, (where F(X) is the free monoid generated by X) a certain positive or negative integral "weight" $\langle r, f \rangle$, the coefficient of f in r. Thus, in fact, a formal power series is just an element of the free module with basis F(X).

In fact, if instead of considering only a subset F' of F(X) we specify a process producing its words, it seems natural to count how many times each of them is obtained and the formal power series is the tool needed for handling this more detailed information.

Of course, with this interpretation we only get positive power series, i.e., power

(XII. LINGUISTICS)

series in which every coefficient $\langle r, f\rangle$ is non-negative. The general case may be thought of as being associated with two processes and then the coefficient of f is the difference between the number of times f is obtained by each of these processes.

In any case, we shall define the support of r as the subset $F'_r = \{f \in F : \langle r, f\rangle \neq 0\}$.

The power series form a ring $\bar{A}(X)$ with respect to the following operation:

multiplication by an integer: the coefficient of f in nr is simply $n\langle r, f\rangle$

addition: $\langle r+r', f\rangle = \langle r, f\rangle + \langle r', f\rangle$, for all f

multiplication: $\langle rr', f\rangle = \Sigma\langle r, f\rangle\langle r', f''\rangle$, where the sum is extended to all factorizations $f = f'f''$.

It is clear that when r and r' are positive, the support $F'_{r+r'}$ of r+r' is just the union of the supports of r and of r'; similarly, the support of rr' is the set product $F_r F_{r'}$. For arbitrary r the interpretation is more complicated.

It is convenient to introduce a topology in $\bar{A}(X)$ in order to be able to define the limit of a sequence. Among the many possibilities that are available the simplest one is based upon the following definition of the distance: $\|r-r'\| = 1/n$ if and only if $\langle r, f\rangle = \langle r', f\rangle$ for every word $f \in F$ of degree ("length") strictly, less than n and $\langle r, f\rangle \neq \langle r', f\rangle$ for at least one $f \in F$ of degree n.

Thus, $\|r-r'\| = 0$ if $\langle r, e\rangle \neq \langle r', e\rangle$, where e is the empty word and $\|r-r'\| = 0$ if $r = r'$.

It is easily checked that $\|r-r'\| \leq \sup(\|r-r''\|, \|r'-r''\|)$ for any $r, r', r'' \in \bar{A}(X)$, and that the addition and multiplication are continuous. The norm $\|r\|$ of r is just $\|r-0\|$. Clearly, $\|r\| = 1/n$, where n is the smallest integer such that $\langle r, f\rangle \neq 0$ for some f of degree (= length) n. Thus r has a finite norm if and only if $\langle r, e\rangle \neq 0$.

We now introduce the important notion of an inverse.

By definition $r \in \bar{A}(X)$ is invertible if $r' = e-r$ has a finite norm, i.e., if $\langle r, e\rangle = 1$.

If this is so, the infinite sum $e + \sum_{n>0} r'^n = r''$ satisfies the identity $r'' - r''r' = r'' - r'r'' = e$, i.e., $r''r = rr'' = e$.

This suggests the notation $r'' = r^{-1}$ and, since r'' is invertible, one can also construct $(r'')^{-1}$.

It is easily verified that $(r'')^{-1} = r$, and thus there is no inconvenience in considering the infinite sum r'' as the inverse r^{-1} of r. It is worth noting that if r_1 is a positive element with finite norm, then $(e-r_1)^{-1}$ is positive and has as its support the subset $F^*_{r_1} = \bigcup_{n>0} (F_{r_1})^n$ in Kleene's notation.

Thus we are able to interpret all of the usual set theoretic operations except for complementation and intersection.

With respect to the first, we can observe that by construction the formal power series $\left(e - \sum_{x \in X} x\right)^{-1}$ is equal to $\Sigma\{f : f \in F(X)\}$.

Consequently, if we associate with the subset F' of F the formal power series

(XII. LINGUISTICS)

$r_{F'} = \sum_{f \in F'} f$ (i.e., the power series with $\langle r_{F'}, f \rangle = 1$ if $f \in F' = 0$, otherwise) the support of $\left(e - \sum_{x \in X} x\right)^{-1} - r_{F'}$ is precisely the complement of F' in F.

With respect to the intersection, we can define a Hadamard product which associates with any $r, f \in \bar{A}(X)$ the new power series $r \otimes r'$, defined by $\langle r \otimes r', f \rangle = \langle r, f \rangle \langle r', f \rangle$ for all f. Clearly, the support of $r \otimes r'$ is the intersection of the supports of r and r'.

However, the Hadamard product is no longer an elementary operation and this may explain why some otherwise reasonable families of subsets are not closed under intersection (cf. below).

2. Relation with Ordinary Power Series

This can be expressed in a loose way by saying that ordinary power series are obtained from the elements of $\bar{A}(X)$ by disregarding the order of the letters in the words $f \in F$. Formally, let α be a bijection (one-to-one mapping onto) $X \to \bar{X}$. An ordinary power series $\bar{r}$ in the variates $\bar{x}_i \in \bar{X}$ is an infinite sum $\bar{r} = \sum a_{n_1 n_2 \ldots n_m} \bar{x}_1^{n_1} \bar{x}_2^{n_2} \ldots \bar{x}_m^{n_m}$ extended to all the monomials $\bar{x}_1^{n_1} \bar{x}_2^{n_2} \ldots \bar{x}_m^{n_m}$.

We can consider that any such r (with integral coefficients $a_{n_1 n_2 \ldots n_m}$) is the image by the homomorphism α of at least one $r \in \bar{A}(X)$ by defining $a_{n_1 n_2 \ldots n_m}$ as the sum of $\langle r, f \rangle$ extended to all of the words $f \in F(X)$ containing the letters x_2 n_1 times; the letters x_2 n_2 times ... etc.; i.e., to all words f such that $\alpha f = \bar{x}_1^{n_1} \bar{x}_2^{n_2} \ldots \bar{n}_m^{n_m}$, where α is the homomorphism sending $F(X)$ onto the free commutative monoid generated by $\bar{X}$. It is trivial that $\alpha(r_1 \pm r_2) = \alpha r_1 \pm \alpha r_2$; $\alpha r_1 r_2 = \alpha r_1 \alpha r_2 = \alpha r_2 \alpha r_1$; $\alpha\left(r_1^{-1}\right) = (\alpha r_1)^{-1}$ identically.

Also, when X contains a single letter no difference need be made between formal (noncommutative) and ordinary (commutative) power series.

Since the theory of ordinary power series is an extremely well-developed chapter of mathematics, the existence of the homomorphism α may at times be used for the study of the formal power series and of their support. The discussion of some elementary examples of this approach is, in fact, the main content of this report.

3. The Algebraic Elements of $\bar{A}(X)$

In ordinary calculus, one usually considers as especially elementary functions the polynomials, the rational functions, and the algebraic functions.

By definition, a polynomial is the function represented by an ordinary power series with only finitely many nonzero coefficients; a rational function is the quotient of two polynomials; an algebraic function is a function of the variates with the property that

(XII. LINGUISTICS)

it satisfies identically some algebraic relation expressed by a polynomial.

For example, $\frac{\bar{x}_1\bar{x}_2^2}{1-\bar{x}_2^2}$ is a rational function of $\bar{x}_1$ and $\bar{x}_2$, and the function r of the commutative variates $\bar{x}_1$ and $\bar{x}_2$ that is such that $\bar{x}_1\bar{x}_2\bar{r}^2 - \bar{r}+1 = 0$ identically is an algebraic function.

We can imitate this hierarchy by introducing the following definitions: an element $r \in \bar{A}(X)$ is a polynomial if its support is a finite set; an element $r \in \bar{A}(X)$ is rational if it can be obtained from the generators $x \in X$ as a finite expression using only the sum, the product, and the inversion (of invertible elements).

It is clear that the polynomials form a subring of $A(X)$. Indeed, this ring is what is usually called the free ring generated by X.

In a similar manner, the set $\bar{R}(X)$ of the rational elements is a ring, i.e., it is closed under addition, subtraction, and multiplication. Furthermore, it is closed under inversion (of invertible elements). In fact, $\bar{R}(X)$ is the smallest subring of $\bar{A}(X)$ closed under this last operation and containing X.

It is easily verified that for any $r \in \bar{R}(X)$ the "Abelianized" ordinary power series $\vec{r} = \alpha r$ represents a rational function.

Consider, for instance, the formal power series r with $\langle r, f\rangle = 1$ if and only if $f = x_2^7 x_1^{3+2n_1} x_2 x_1^{3+2n_2} x_2 \cdots x_2 x_1^{3+2n_m} x_2^5$, and $\langle r, f\rangle = 0$, otherwise. This series r belongs to $\vec{R}(X)$ because r is equal to $x_2^6\left(e - x_2 x_1^3\left(e - x_1^2\right)^{-1}\right)^{-1} x_2^5$, and αr can be reduced to the quotient of two polynomials by writing

$$\begin{aligned}
\alpha r &= \bar{x}_2^6\left(e - \bar{x}_2\bar{x}_1^3\left(e - \bar{x}_1^2\right)^{-1}\right)^{-1}\bar{x}_2^5 \\
&= \bar{x}_2^6\left(\left(e - \bar{x}_1^2\right)^{-1}\left(e - \bar{x}_1^2 - \bar{x}_2\bar{x}_1^3\right)\right)^{-1}\bar{x}_2^5 \\
&= \left(e - x_1^2\right)\left(e - x_1^2 - x_2 x_1^3\right)^{-1} x_2^{11} \\
&= \frac{x_2^{11}\left(1 - x_1^2\right)}{1 - x_1^2 - x_1^2 x_2}
\end{aligned}$$

The family of all subsets of F that can be the support of a rational element of $\bar{A}(X)$ has been defined elsewhere.[7] It is not difficult to verify that it is closed under union, intersection, set product, and Kleene's star operation.

Having recalled these facts, we proceed to the definition of an algebraic element of $\bar{A}(X)$.

For this purpose, we consider a finite set $\dot{\square}$ of m new elements ξ_i, and we denote by $\bar{\sigma}$ an m-tuple of polynomials σ_ξ in the (noncommutative) variates $y \in Y = X \cup \dot{\square}$ that satisfy the condition that $\langle \sigma_\xi, e\rangle = \langle \sigma_\xi, \xi'\rangle = 0$ for all $\xi, \xi' \in \dot{\square}$.

(XII. LINGUISTICS)

Now let W denote the set of all m-tuples $w = (w_1, w_2, \ldots w_m)$ of elements of $\bar{A}(X)$. We consider $\bar{\sigma}$ as a mapping of W into itself by defining the coordinate $\bar{\sigma}w_\xi$ of the transformed vector σw as the element of $\bar{A}(X)$ obtained by replacing in the polynomial $\bar{\sigma}_\xi$ every symbol ξ' by the corresponding coordinate $w_{\xi'}$ of w_ξ.

For instance, if $\sigma_{\xi_1} = x_1\xi_2x_2$; $\sigma_{\xi_2} = x_1x_2 + x_1\xi_1x_2\xi_2$; and if w is the vector $\left(3x_1-x_2x_1, x_2^2+2x_3^4\right)$, the w coordinates of $\bar{\sigma}w$ are

$$\bar{\sigma}w_{\xi_1} = x_1\left(x_2^2+2x_3^4\right) x_2 = x_1x_2^3 + 2x_1x_3^4x_2$$

$$\bar{\sigma}w_{\xi_2} = x_1x_2 + x_1(3x_1-x_2x_1)\, x_2\left(x_2^2+2x_3^4\right)$$

$$= x_1x_2 + 3x_1^2x_2^3 + 6x_1^2x_2x_3^4 - x_1x_2x_1x_2^3 - 2x_1x_2x_1x_2x_3^4.$$

It is clear that $\bar{\sigma}$ is a continuous mapping in the sense that if $w, w' \in W$ are such that $\|w-w'_\xi\| \leq 1/n$ for each $\xi \in \dot{\square}$ (i.e., for short, if $\|w-w'\| \leq 1/n$, then $\|\bar{\sigma}w-\bar{\sigma}w'\| \leq 1/n$.

Indeed, the relation $\|w-w'\| \leq 1/n$ expressed the fact that the coefficients $\langle w_\xi, f\rangle$ and $\langle w'_\xi, f\rangle$ are equal for every coordinate $\xi \in \dot{\square}$ and for every word f of degree $\leq n$. Since the coefficient of every word of degree n in the polynomial in the letters $x \in X$ obtained by the substitution $\xi' \to w_{\xi'}$ or $\xi' \to w'_{\xi'}$ in σ_ξ depends only on the terms of lower degree, the result is a simple consequence of the definition.

In fact, because of our hypothesis on $\bar{\sigma}$, a stronger result can be proved when w and w' satisfy the supplementary condition that $\langle w_\xi, e\rangle = \langle w'_\xi, e\rangle = 0$ for all ξ. Then, obviously, this last condition is still verified for $\bar{\sigma}w$ and $\bar{\sigma}w'$ (because $\langle \sigma_\xi, e\rangle = 0$). Furthermore (because $\langle \sigma_\xi, \xi'\rangle = 0$), we can conclude from $\|w-w'\| \leq 1/n$ that $\|\bar{\sigma}w-\bar{\sigma}w'\| \leq 1/n+1$. This, again, is a direct consequence of the fact that the coefficients of the terms of degree n+1 of σw are determined univocally by the coefficients of the terms of degree $\leq n$ of w.

Let us now consider the infinite sequence $w, w_1, \ldots, w_n, \ldots$, where $w_o = (0, 0, \ldots, 0)$ and $w_{n+1} = \bar{\sigma}w_n$. By applying our previous remarks and using induction, we can easily show that for all n and $n' > 0$ we have $\|w_n-w_{n+n'}\| \leq 1/n$. Consequently, we have proved that $w = \lim_{n\to\infty} w_n$ is a well-defined element of W and that $\lim_{n\to\infty} \bar{\sigma}w_n - w_n = 0$. This suggests that we speak of w as of a solution of the system of equations $\xi = \sigma_\xi$ (i.e., $w = \bar{\sigma}w$), since, in fact, for each ξ, w_ξ is equal to the formal power series in the $x \in X$ obtained by replacing in σ_ξ each ξ' by the coordinate $w_{\xi'}$.

We shall say, accordingly, that w_ξ is an algebraic element of $\bar{A}(X)$. Because of our definition of $\bar{\sigma}$, any w has a finite norm (i.e., $\langle w_\xi, e\rangle = 0$). This restriction would be artificial; we shall denote by $\bar{S}(X)$ the set of all formal power series that is the sum of a polynomial and of a coordinate w_ξ, defined above, for some suitable finite set of polynomials $\bar{\sigma}$, or, as we prefer to say, by a set of "equations" $\xi = \sigma_\xi$.

(XII. LINGUISTICS)

It is not difficult to verify the fact that $\bar{S}(X)$ is a ring closed under the formation of inverses (of invertible elements). Indeed, let r and r' be obtained as the coordinates w_ξ and w'_ξ of the solutions w and w' of the equations $w = \bar{\sigma}w$ and $w' = \bar{\sigma}'w'$. For the sake of clarity, we assume that $\bar{\sigma}$ and $\bar{\sigma}'$ are defined by two disjoint sets $\boxed{\cdot}$ and $\boxed{\cdot}'$ of m and m' elements, and we consider $\boxed{\cdot}''$ the union of $\boxed{\cdot}$, $\boxed{\cdot}'$ and of a new letter ξ''. Then, if we denote by σ'' the direct sum of $\bar{\sigma}$ and $\bar{\sigma}'$, it is clear that the new equation $\xi'' = \sigma_\xi + \sigma'_\xi$, determines $w''_{\xi''} = r+r'$. Similarly, the equation $\xi'' = \xi\xi'$ determines $w''_\xi = rr'$.

In order to get $(e-r)^{-1} - e \left(= \sum_{n>0} r^n\right)$ it is enough, for instance, to add the new equation $\xi'' = \xi\xi'' - \sigma$.

As a final remark it may be pointed out that (as for rational elements) the homomorphism α sends the algebraic elements of $\bar{A}(X)$ onto the Taylor series of the ordinary algebraic functions. These last series are easily proved to converge in some small enough domain around 0. Let us also mention that $\bar{S}(X)$, as defined constructively here, can also be shown to be identical to the set of all formal power series with integral coefficients that satisfy a set of equations of the type $w = \bar{\sigma}w$, described above, provided, of course, that such solutions exist.

Example 1.

Let $\sigma_{\xi_1} = x_1\xi_1x_2 + x_1x_2$

$$\sigma_{\xi_2} = \xi_1\xi_2 + x_1\xi_1x_2 + x_1x_2.$$

Since the first equation involves only ξ_1, it can be solved for its own sake, and one easily obtains $r = w_{\xi_1} = \sum_{n>0} x_1^n x_2^n$. Then the second equation gives

$w_{\xi_2} = rw_{\xi_2} + r$, that is, $w_{\xi_2} = r(e-r)^{-1}$.

Thus, by definition, a word f belongs to the support of w_{ξ_2} if and only if it can be factorized as a product $\left(x_1^{n_1}x_2^{n_1}\right)\left(x_1^{n_2}x_2^{n_2}\right)\ldots\left(x_1^{n_m}x_1^{n_m}\right)$ of words belonging to the support of r.

Since, trivially, this factorization is unique, we always have $\langle w_{\xi_2}, f\rangle = 0$ or 1.

Example 2.

Let $\bar{\sigma}_{\xi_1} = x_1\xi_1x_2\xi_1 + x_1x_2\xi_1 + x_1\xi_1x_2 + x_1x_2$.

After setting $r = e + \xi_1$ we get the simpler form $r = x_1rx_2r + e$, instead of the equation $\xi_1 = \sigma_{\xi_1}$. Again, $\langle r, f\rangle = 0$, or 1; with $\langle r, f\rangle = 1$ if and only if

1°. f contains as many x_1 as x_2.

2°. any left factor f' of f contains at least as many x_1 as x_2.

(XII. LINGUISTICS)

Since the equation can also be written in the form $r = (e - x_1 r x_2)^{-1}$, it follows that every $f \in F_r$ has one and only one factorization as a product of words belonging to the support F' of $x_1 r x_2$.

F' is closely related to the well-formed formulas in Lukasiewicz' notation because f belongs to F' if and only if it satisfies 1^o and, instead of 2^o, condition 3^o. Any factor f' of f contains, strictly, more x_1 than x_2, unless $f' = e$ or $f' = f$.

Let us now observe that $x_1 r x_2$ satisfies the equation $x_1 (e - x_1 r x_2)^{-1} x_2 = x_1 r x_2$. Taking the homomorphic image as α and writing $\bar{r}' = \alpha(x_1 r x_2)$, we get the ordinary equation $\bar{x}_1 \bar{x}_2 (1 - \bar{r}')^{-1} = \bar{r}'$; i.e., $\bar{r}'^2 - \bar{r} + \bar{x}_1 \bar{x}_2 = 0$.

By construction, the ordinary power series r' takes the value 0 for $x_1 x_2 = 0$ and thus, as is well known,

$$\bar{r}' = \frac{1 - \sqrt{1 - 4\bar{x}_1 \bar{x}_2}}{2} = \frac{1}{2} \sum_{n>0} (-\bar{x}_1 \bar{x}_2)^n \begin{bmatrix} 1/2 \\ n \end{bmatrix}$$

where $\begin{bmatrix} 1/2 \\ n \end{bmatrix}$ is the binomial coefficient.

Because $\langle x_1 r x_2, f \rangle = 0$ or 1, we can conclude that $(-1)^n \begin{bmatrix} 1/2 \\ n \end{bmatrix}$ is the number of distinct words of degree $2n$ in the support of $x_1 r x_2$.

The reader may notice that our present computation is exactly the one used in the classical problem of the return to equilibrium in coin-tossing games.

Example 3.

Let $\dot{\cup}$ be the union of ζ, η and of ξ_i $(i = 1, \ldots 2m)$ and agree that $\xi_{i+m} = \xi_{i'}$, when $i = i' + m$. Let $X = \{x_i\}$ $i = 1, 2, \ldots, 2m$, and consider the $2m$ equations

$$\xi_i = x_i x_{i+m} + x_i \left(\zeta + \zeta^2 + \zeta\eta\zeta - \sum_{j=1}^{2n} x_j (e+\eta)\, x_{j+m} \right) x_{i+n}$$

$$\zeta = \sum_{1=1}^{2m} \xi_i; \quad \eta = \zeta + \zeta\eta.$$

Simple transformations reduce these to standard form, and it can be proved that $\langle e+\eta, f \rangle = 0$, or 1 with $\langle e+\eta, f \rangle = 1$ if and only if f belongs to the kernel K of the homomorphism ϕ, which sends $F(X)$ onto the corresponding free group (with $(\phi x_i)^{-1} = \phi x_{i+n}$).

After performing the homomorphism α, we compute the value of $\alpha\eta = u(t)$ for $\bar{x}_1 = \bar{x}_2 = \ldots = \bar{x}_{2m} = \frac{t}{2m}$. By construction, $u(t)$ is the generating function of the recurrent event consisting in the return to K, and $u(1)$ is the probability that a random word ever belongs to K when the letters $x_i \in X$ are produced independently with constant probability $1/2m$.

We find that $u = u(t)$ is defined by the quadratic equation $(4m^2 - t^2) u^2 - 4m^2 u + 2mt^2 = 0$, which is in agreement with similar results of Kesten[3] to which we refer for a more

explicit interpretation of u(t).

4. Some Subfamilies of $\overline{S}(X)$

It seems natural to distinguish in $\overline{S}(X)$ the subset $\overline{S}_1(X)$ of those elements that are obtained when each σ_ξ of $\overline{\sigma}$ has the special form

$\sigma_\xi = f + \Sigma\, f'\xi'f''$, where $f, f'f'' \in F(X)$, and the summation is over any finite set of triples (f', ξ', f'') (with, eventually, the same ξ' occurring several times; i.e., when each σ is linear in the variates $\xi \in$ [illegible].

Within $\overline{S}_1(X)$ itself we shall distinguish the special case $S_o(X)$ for which $\sigma = f + \Sigma\, \xi' f''$; i.e., only one-sided linear equations are considered.

Clearly, after taking the homomorphic image as a, both $\overline{S}_1(X)$ and $\overline{S}_o(X)$ collapse onto the ring of the ordinary rational functions but, at the level of $\overline{A}(X)$, the sets from $\overline{S}_o(X)$ form only a very restricted subset of $\overline{S}_1(X)$, as we shall see.

A second principle of classification is provided by the restriction that every coefficient in the polynomials σ_ξ is non-negative.

Under this hypothesis, the same is true of the power series w_ξ, and, correspondingly, we obtain three subsets (in fact, three semirings) which we denote $\overline{S}^+(X)$, $\overline{S}_1^+(X)$, and $\overline{S}_o^+(X)$. It is to be stressed that the converse is not true. Indeed, it is quite easy to display examples of formal power series having only non-negative coefficients that belong to $\overline{S}_o(X)$, but not even to $\overline{S}^+(X)$.

A priori the inclusion relations shown in Fig. XII-1 hold. Here, $P_o(X)$ and $P_o^+(X)$

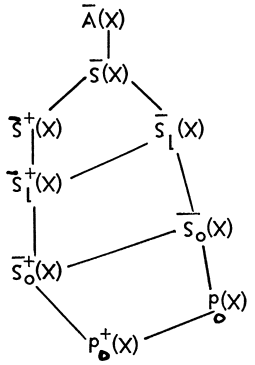


Fig. XII-1.

denote the polynomials and the positive polynomials, respectively. Insofar as the corresponding supports are concerned, three theorems summarize the results.

THEOREM I. (Ginsburg-Rice). The family of the supports of the elements of $\overline{S}^+(X)$ is identical to the family $\mathscr{C}$ of Chomsky's context-free languages.

(XII. LINGUISTICS)

THEOREM II. (Chomsky). The family of the supports of the elements of $\overline{S}_o^+(X)$ is identical to the family $\mathscr{R}_o$ of Kleene's regular events.

THEOREM III. The family of the supports of the elements of $\overline{S}_o(X)$ is identical to the family $\mathscr{R}$ of the sets of words accepted by an automaton of the type $\mathscr{A}$ (i. e., it is identical to the family of the supports of the rational elements of $\overline{A}(X)$).

In order to prove Theorem I we need to alter slightly Chomsky's definition and we propose

DEFINITION. A context-free grammar is given by

i. Two disjoint finite sets $\boxdot$ and X;

ii. A finite set G of pairs (ξ, g), where $\xi \in \boxdot$, $g \in F(X \cup \boxdot)$, $g \neq e$, $g \notin \boxdot$.

iii. A distinguished element $\xi_o \in \boxdot$.

The language $D_X(\xi_o, G)$ produced by G is the intersection $F(X) \cap D(\xi_o, G)$, where $D(\xi_o, G)$ is the smallest subset of $F(X \cup \boxdot)$ which is such that $\xi_o \in D(\xi_o, G)$ and $g_1\xi' g_2 \in D(\xi_o, G)$, and $(\xi', g) \in G$ implies $g_1 g g_2 \in D(\xi, G)$. In the usual terminology, $\boxdot$ (resp. X) is the nonterminal (resp. terminal) vocabulary, and G is the grammar; our definition departs from Chomsky's by the easily met restriction $g \notin \boxdot$ for each rule (ξ, g) of G.

With this notation the equivalence of $\mathscr{C}$ with the set of all supports F'_r: $r \in \overline{S}^+(X)$ is trivial.

Let G be given, and define for each $\xi \in \boxdot$ the polynomial σ_ξ as the sum $\Sigma\, g$ extended to all g so that $(\xi, g) \in G$.

If we interpret the support of w_ξ as the set $D_X(\xi, G)$, it is clear that any equation $w_\xi = \overline{\sigma} w$ can be interpreted as describing $D_X(\xi, G)$ as the union of the sets $D_X(g, G)$ $((\xi, g) \in G)$ obtained by replacing in g every letter ξ' by a terminal word $f \in D_X(\xi', G)$.

Conversely, let us assume that $\overline{\sigma}$ is such that $\langle \sigma_\xi, g \rangle \geqslant 0$ for all $\xi \in \boxdot$ and $g \in F(X, \boxdot)$.

By introducing enough new variates ξ', we can find $\overline{\sigma}'$ which is such that $\langle \sigma'_\xi, g \rangle = 0$ or 1, and the new polynomials σ'_ξ reduce to old polynomials σ_ξ when the new variates ξ' are identified with the old ones in a suitable manner. Furthermore, for every new ξ' (corresponding to the old variate ξ) we add an equation $\sigma_{\xi'}$ identical to σ_ξ.

Thus the original w_ξ is equal to a sum $\Sigma\, w'_\xi$, (with $w' = \overline{\sigma}' w'$) and $\overline{\sigma}'$ can be associated with a grammar in a unique fashion, since $\langle \sigma'_{\xi'}, g \rangle = 0$ or 1.

This interpretation throws some light on the other families. Thus, $\overline{S}_1^+(X)$ corresponds to the family $\mathscr{C}_1$ of the context-free languages in which every rule has the form $(\xi; f'\xi' f'')$ or (ξ, f) with $f, f', f'' \in F(X)$.

In turn, $\overline{S}_o^+(X)$ is obtained by restricting the rules to have the form $(\xi, \xi' f)$ or (ξ, f) with $f \in F(X)$.

Observe now that, in any case, the coefficient $\langle w_\xi, f \rangle$ of the word f expresses the number of distinct factorizations of f according to the rule of grammar. Thus, for

(XII. LINGUISTICS)

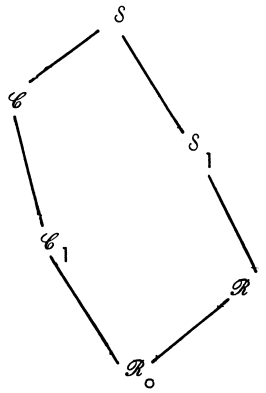


Fig. XII-2.

any two r, r' $\in \overline{S}^{+}(X)$, the support of r-r' consists precisely of those words that have a different number of factorizations in the two grammars associated with r and r', respectively.

Reciprocally, given any r" $\in \overline{S}(X)$, it is easy to prove that r" = r-r' for at least one pair r, r' $\in \overline{S}^{+}(X)$, and the same is true for $\overline{S}_1$ and $\overline{S}_1^{+}$ or for $\overline{S}_o$ and $\overline{S}_o^{+}$.

Summarizing our remarks, we obtain (on top of the family of the finite subsets) the six families illustrated in Fig. XII-2. Here, $\mathcal{S}$ and $\mathcal{S}_1$ correspond to $\overline{S}(X)$ and $\overline{S}_1(X)$, respectively. In order to prove that these six families are all different and do not enjoy further inclusion relations, it would be enough to build three subsets, say F_1, F_2, F_3 of F(X) having the following properties:

$R_1 \in \mathcal{R}$, $F_1 \notin \mathcal{C}$

$F_2 \in \mathcal{C}_1$, $F_2 \notin \mathcal{R}$

$F_3 \in \mathcal{C}$, $F_3 \notin \mathcal{S}_1$.

I am not able to construct a set such as F_3,[10] but there exists an F_4 which is such that $F_4 \in \mathcal{C}$ and $F_4 \notin \mathcal{C}_1$. Thus the only possible diagrams apart from Fig. XII-2 are Fig. XII-3a and 3b. Again, there is no further inclusion relation. In fact, it seems

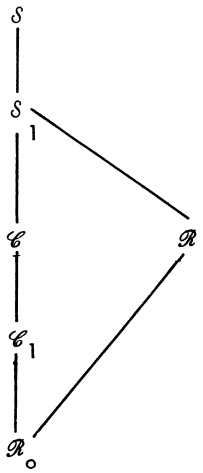


Fig. XII-3a.

or

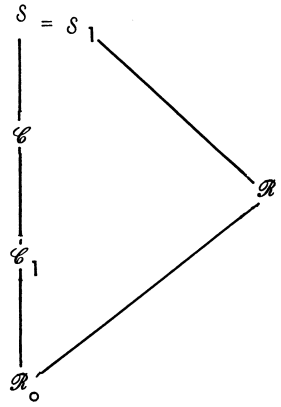


Fig. XII-3b.

(XII. LINGUISTICS)

most unlikely that $\mathscr{C} \subset \mathcal{S}_1$, and the original scheme probably represents the true situation.

The counterexamples F_2 and F_3 are very simple:

Let $F_2 = \{x_1^n x_2 x_1^n : n \geq 0\}$. This set is produced by $G = \{(\xi, x_1 \xi x_1), (\xi, x_2)\}$, and thus $F_2 \in \mathscr{C}_1 \subset \mathscr{C}$. On the other hand, it is known that F_2 does not belong to $\mathscr{R}$.[8]

Let $F_3 = \{x_1^n x_2 x_2^{n'} : n, n' \geq 0\ n \neq n'\}$. It is known[8] that $F_3 \in \mathscr{R}$, and it is not difficult to show that F_3 does not belong to $\mathscr{C}$ because of the relatively simple structure of a grammar G which produces infinite sets of the form $\{x_1^n x_2 x_1^{n'} : n, n'$ linked by a certain relation$\}$.

Indeed, as the reader can verify, any set of this type is a finite union of finite sets and of sets having the form:

$\{x_1^{n+N} x_2 x_1^{n'+N'} : N, N', n, n' \geq 0;\ n \equiv 0 \pmod{p},\ n' \equiv 0 \pmod{p'}\}$ for some integers N, N', p, p'.

(The proof is based upon the fact that, when X has a single letter, $\mathscr{C}$ reduces to the family of regular events.)

For the construction of F_4 we need a more explicit description of $\mathscr{C}_1$:

F' belongs to $\mathscr{C}_1$ if and only if there exist

(a) A finite set Y;

(b) Two mappings ϕ and Φ from F(Y) to F(X) that are a homomorphism and an anti-isomorphism (i. e., $\Phi\, g\, g' = \Phi\, g'\, \Phi\, g$);

(c) A regular event $G' \subset F(Y)$ that is such that $F' = \{\phi\, g\, \Phi\, g : g \in G'\}$.

The proof of this statement follows the same lines as Chomsky's proof[1] of the fact that the support of any $r \in \overline{S}_o^+(X)$ is a regular event.

The same technique, of course, is valid for the more general case of $\overline{S}_1(X)$ (with the obvious modifications) and it displays every element of $\mathscr{C}_1$ obtained by the three following steps:

1. Taking the words g from some regular event on $F(\boxdot)$;

2. Forming the products $g \xi^* \tilde{g}$, where ξ^* is a new symbol, and $\tilde{g}$ is the "mirror image" of g;

3. Making a transduction θ of $\tilde{g}$ and of g into F(X), and erasing ξ^*.[7]

Let us now return to our problem. For any $f \in F(X)$ $(X = \{x_1, x_2\})$ let λf denote the difference between the number of times x_1 and x_2 appear in f.

We claim that F_4 belongs to $\mathscr{C}$ and not to $\mathscr{C}_1$, where $F_4 = \{f : \lambda f = 0;\ \lambda f' > 0$ for all proper left factors of $f\}$.

The first part of the claim has already been verified (Example 2).

Let us now observe that if $F' \in \mathscr{C}_1$ is such that for all integers $n < 0$ there exists a $g \in F(\boxdot)$ which is such that $g_1 g\, g_2 \in G'$ for some g_1, g_2, and that $\lambda \phi\, g < n$, then $F' \neq F_4$. Indeed, since G' is a regular event, there exists a finite set of pairs (g_1, g_2') which are such that for any $g \in F(\boxdot)$ either $F(\boxdot)\, g\, F(\boxdot) \cap G'$ is empty, or else

$g_1 g\, g'_2 \in G'$ for some of these pairs. Thus, under this hypothesis, we can find that $f = \phi\, g_i g\, g'_1 \Phi\, g_i g\, g'_i \in F'$ which is such that its left factor $f' = \phi\, g_i g$ satisfies $\lambda f' < 0$, and thus $f \in F'$ and $f \notin F_4$.

It follows that if $F'' \in \mathscr{C}_1$ is contained in F_4, we can find a large enough integer n' which is such that no $f \in F''$ has a factorization $f = f_1 f_2 f_3 f_4$ with $\lambda f_1 > n$; $\lambda f_1 f = 1$; $\lambda f_1 f_2 f_3 > n$ (and $\lambda f_1 f_2 f_3 f_4 = 0$ because by hypothesis $f \in F_4$). Since clearly F_4 contains such words, we have proved that $F'' \in \mathscr{C}_1$ and $F'' \subset F_4$ implies $F'' \neq F_4$; that is, $F_4 \notin \mathscr{C}_1$.

These remarks can be pictorially expressed by saying that the words of F'' have, at most, one arbitrarily high peak. It follows from the definition of F_4 that this last set contains words having an arbitrary number of arbitrarily high peaks. Thus, incidentally, we have proved the stronger result that $\mathscr{C}$ is different from the family of subsets obtained from $\mathscr{C}_1$ by closure under a finite number of set products or set unions.

5. Some Miscellaneous Remarks

a. As an easy source of counterexamples we could consider the special case of X reduced to a single element because then no difference exists between commutative and noncommutative power series.

The results known thus far contribute to the statement that in this case $\mathscr{C}$ and $\mathcal{S}_1$ are equivalent to $\mathscr{R}_o$. No result is known for $\mathcal{S}$.

However, although the proofs that $\mathscr{C} = \mathscr{R}_o$ and that $\mathcal{S}_1 = \mathscr{R}$ are quite easy, the proof that $\mathscr{R} = \mathscr{R}_o$ is a rather deep theorem of Skolem.[8] Nonetheless, the fact that when $X = \{x\}$ any $r \in \bar{S}(X)$ is the Taylor series of some ordinary algebraic function of n allows us to construct simple families of sets that cannot belong to $\mathcal{S}$.

A rather general instance is the family of the infinite sets $\left\{ x^{N_1}, x^{N_2}, \ldots, x^{N_m}, \ldots \right\}$ which have the property that $\lim \frac{N_{m+1}}{N_m}$ is infinite (i.e., which have the property that the ratio N_{m+1}/N_m exceeds for some finite m any prescribed finite value).

In order to prove that no set of this type belongs to $\mathcal{S}$ we consider any $r \in \bar{S}(X)$ $(X = x)$. Without loss of generality we may assume that $\langle r, e \rangle = 0$. By definition, $r = a_o + \sum_{i\,2}^{m} a_i r^i$, where m is finite and the a_i's are polynomial in x. By comparing the two members of the equation, we see that for each n $\langle r, x^n \rangle$ must be equal to a linear combination with fixed coefficients of sums of the type $\Sigma \left\langle r, x^{n_1} \right\rangle \left\langle r, x^{n_2} \right\rangle \ldots \left\langle r, x^{n_{m'}} \right\rangle$ extended to all representations of n–h as a sum $n_1 + n_2 \ldots n_{m'}$ where $h \geqslant 1$ is bounded by the degrees of the a_i's, and m' are bounded by the degree m of the equation. It follows that if N is such that $\langle r, x^{N+k} \rangle = 0$ for $0 \leqslant k < mN$, then $\langle r, x^{n'} \rangle = 0$ for all $n' \geqslant N$; i.e., r is a polynomial.

Since the condition imposed on the set $\left\{ x^{N_1}, x^{N_2}, \ldots \right\}$ amounts to the existence of at least one such N for every finite m, our contention is proved.

(XII. LINGUISTICS)

A similar method could be applied to show that $\{x^{n^2} : n > 0\}$ does not belong to $\mathcal{S}$.

b. Our next example shows that the intersection problem even for so restricted a family as $\mathscr{C}_1$ is an undecidable one.

Let $\overbrace{\cdot}$ = $\{\xi\}$, X = {a, b, c} and consider the two grammars:

G = $\{(\xi, a\xi a), (\xi, b\xi b), (\xi, c)\}$

G' = $\{(\xi, c), (\xi, f_i \xi f_i'): i \in I\}$, where $(f_i, f_i')(i \in I)$ is an arbitrary set of pairs of elements from F = F(a, b).

The language $D_X(\xi, G)$ is a special instance of Chomsky's mirror-image languages and there exists an $f \in D_X(\xi, G) \cap D_X(\xi, G')$ if and only if one can find a finite sequence $i_1, i_2, i_3 \ldots, i_n$ of indices such that the word $f_{i_1} f_{i_2} \ldots f_{i_n}$ is equal to the mirror image of $f'_{i_n} f'_{i_{n-1}} \ldots f'_{i_2} f'_{i_1}$. Thus clearly the intersection problem for G and G' is equivalent to the classical correspondence problem of Post[5] and since this last one is undecidable, our contention is proved.

c. It may be mentioned that other principles could be used for distinguishing interesting subsets of words. For example, Ginsburg and Rice[2] have shown that $\mathscr{C}$ contains as a proper subset the family $\mathscr{C}'$ corresponding to the case in which the set of equations $w = \tilde{\sigma} w$ has the following property which these authors call the "= sequential property": There exists an indexing $\xi_1, \xi_2 \ldots \xi_m$ of the variates $\xi \in \overbrace{\cdot}$ which is such that for all j the polynomial $\sigma\xi_j$ does not involve the variates $\xi_{j'}$, with $j' > j$.

In Chomsky's terminology this means that no $\xi_{j'}$, $(j' > j)$ appears in a word g that is such that $(\xi_j, g) \in G$. (Then, clearly the rewriting process must be started from $\xi_o = \xi_m$).

Another possibility is to consider the subset $\tilde{S}^1(X)$ of those $s \in \tilde{S}(X)$ that are such that $\langle s, f \rangle = 0$, or 1, for all f.

It has been shown by Parikh[4] that there exist sets of words in $\mathscr{C}$ (in fact, in the closure of $\mathscr{C}_1$ by finite union and set product) which cannot be the support of an $s \in \tilde{S}^1(X)$ having this property.

In our notation, Parikh's example is described as follows:

$$\sigma_{\xi_o} = \xi_1\xi_2 + \xi_4\xi_3; \quad \sigma_{\xi_1} = x_1\xi_1x_1 + x_1\xi_2x_1; \quad \sigma_{\xi_2} = x_2 + x_2\xi_2;$$

$$\sigma_{\xi_3} = x_2\xi_3x_2 + x_2\xi_4x_2; \quad \sigma_{\xi_4} = x_1 + x_1\xi_4.$$

From this reasoning we deduce the following equations in which, for short, w_i denotes the coordinate of w whose index is ξ_i:

$$w_o = w_1w_2 + w_4w_3;$$
$$w_1 = x_1(w_1+w_2)\, x_1$$
$$w_3 = x_2(w_3+w_4)\, x_2$$
$$w_2 = x_2 + x_2w_2$$
$$w_4 = x_1 + x_1w_4.$$

These equations can easily be solved because they are "sequential" in the sense of

(XII. LINGUISTICS)

Ginsburg and Rice. Indeed, the last equation can be written $w_4 - x_1 w_4 = x_1$. That is, $(e-x_1)\, w_4 = x_1$ and we have $w_4 = (e-x_1)^{-1} x_1 = \sum_{n>0} x_1^n$.

Similarly, $w_2 = (e-x_2)^{-1} x_2 = \sum_{n>0} x_2^n$. Thus $w_3 = x_2 w_3 x_2 + \sum_{n>0} x_2 x_1^n x_2$, and, consequently, $w_3 = \sum_{n>0} \sum_{m>0} x_2^m x_1^n x_2^m$; $w_1 = \sum_{n>0} \sum_{m>0} x_1^m x_2^n x_1^m$. Thus we finally obtain

$$w_o = \left(\sum_{n>0} \sum_{m>0} x_1^m x_2^n x_1^m \right) \left(\sum_{n'>0} x_1^{n'} \right) + \left(\sum_{m'>0} x_1^{m''} \right) \left(\sum_{n>0} \sum_{m>0} x_2^n x_1^m x_2^n \right)$$

$$= \sum x_1^m x_2^n x_1^{m'} x_1^{n'} \, r(m, n, m'n')$$

The last summation is, after all, quadruples (m, n, m', n') of positive integers, and the coefficient $r(m, n, m', n')$ has the following values:

$r(m, n, m', n') = 0$ if $m \neq m'$ and $n \neq n'$
$= 1$ if $m \neq m'$ and $n = n'$
$= 1$ if $m = m'$ and $n \neq n'$
$= 2$ if $m = m'$ and $n = n'$.

The fact that this coefficient is equal to 2 for certain words exactly measures the "ambiguity" of the grammar. It would be interesting to give examples in which this grammatical ambiguity is unbounded.

I mention that conversely the following process gives elements $s \in \bar{S}^+(X)$ with $\langle s, f \rangle = 0$, or 1.[9]

Let ϕ be a homomorphism of $F(X)$ into a finite monoid H (i.e., let us consider a finite automaton), and β a mapping that assigns to every pair $(h, h') \in (H, H)$ an integer $\beta(h, h')$. For any word $f \in F(X)$ let $\beta^* f$ be the sum $\Sigma \beta(\phi f_1, \phi f_2)$ extended to all factorizations $f = f_1 f_2$ of f, and say that f is accepted if and only if $\beta^* f$ does not belong to a prescribed finite set Z' of integers.

Then the formal sum $s = \Sigma f'$ extended to all f' which are not accepted (i.e., $s = \Sigma \{f' : \beta^* f' \in Z'\}$ belongs to $\bar{S}^+(X)$.

An equivalent definition[8] is: Let u be a representation of $F(X)$ by finite integral matrices uf and assume that there exists a constant K which is such that for all words f the value $(uf)_{1,N}$ of the $(1, N)$ entry of uf is, at most, equal to K times the degree (length) of f.

Then the set of all f with the property that $uf_{1,N} \neq 0$ is the set of the words accepted by an algorithm of the type described above (and reciprocally). As an auxiliary result, we have shown that the complement of a set F' belonging to the simplest subfamily of $\mathscr{R}$ which is different from $\mathscr{R}_o$ belongs itself to the far higher family $\mathscr{C}$. In general, the complement of a set from F' does not.

Trivially, this construction applies to sets of words defined by the condition that some linear function of the number of times each letter $x \in X$ appears in them has a

given value. It is quite remarkable that the sets defined by two or more such constraints (for instance, the sets of words which contain the same number of times x_1, x_2 and x_3 or the set $\{x_1^n x_2^n x_1^n = n \geqslant 0\}$) do not seem to have any relation to $\mathscr{C}$.

I conclude these rather disconnected remarks by an interesting construction of $\mathscr{C}$ which is due to Parikh and which can also be applied to S(X).

6. Parikh's Construction[4]

Let us consider a grammar G satisfying the usual conditions and extend to a homomorphism $A \to A$ the mapping $j : \square \to A$ defined by $j\xi = \Sigma\{g: (\xi, g) \in G\}$.

For any $g \in F(X \cup \square)$, the support of jg is the set of all words which can be derived from g by the application of one rule of G to each of the occurrences of a symbol $\xi' \in \square$. Every element of this set has either a strictly larger total degree (length) than f or the same total degree but a strictly larger partial degree in the variates $x \in X$. Thus the supports of the elements f, jf, j^2f, ..., j^n_f, ... are all disjoint. Their union, say F', is a subset of the set $D^*(f, G)$ of all words derivable from f.

Of course, F' is, in general, different from $D^*(f, G)$ because of the extra condition that every $\xi' \in \square$ is rewritten at each step. However, when considering only the intersection $D^*(f, G) \cap F(X) = D_X(f, G)$ we have $F' \cap F(X) = D^*(f, G) \cap F(X)$, since in order to get an element $f \in F(X)$ we have to rewrite each $\xi \in \square$ at least once at one time or another.

Let us now denote by u the sum $\Sigma\{\xi: \xi \in \square'\}$ for any subset $\square'$ of $\square$. The element $t = u + \sum_{n>0} j^n u$ belongs to A, as we have seen, and it satisfies the Schröder-like equation $u + jt = t$.

Conversely, we can write $t = (\epsilon - j)^{-1} u$, where ϵ is the identity mapping $A \to A$. Let δ_o denote the retraction $A(X \cup \square) \to A(X)$ induced by $\delta_o \xi = 0$ for each $\xi \in \square$; (a retraction is a homomorphism that allows a subset invariant and sends everything else into this subset; here the subset is that of the words not containing a single $\xi \in \square$.)

Example. $\square = \{\alpha, \beta\}$; $X = \{a, b\}$; $\square' = \{\alpha\}$

$G = \{(\alpha, \alpha\alpha\beta), (\alpha, a), (\beta, b)\}$.

We have

$j\alpha = \alpha\alpha\beta + a$

$j\beta = b$.

Thus $u = \alpha$; $ju = a + \alpha\alpha\beta$;

$j^2u = (a+\alpha\alpha\beta)(a+\alpha\alpha\beta)\, b = a^2b + a\alpha\alpha\beta b + \alpha\alpha\beta a\alpha\beta b$

$j^3u = a(a+\alpha\alpha\beta)(a+\alpha\alpha\beta)\, bb + (a+\alpha\alpha\beta)(a+\alpha\alpha\beta)\, bab + ((a+\alpha\alpha\beta)^2 b)^2\, b$

$= a^3b^2 + a^2bab + a^2ba^2b^2$ + terms of degree $\geqslant 1$ in the $\xi \in \square$, etc.

The support F'_t of $\delta_o t$ is the set of the well-formed formulas in Lukaciewicz notation.

M. P. Schützenberger

(XII. LINGUISTICS)

References

1. N. Chomsky, On certain formal properties of grammars, Information and Control 2, 137-167 (1959).

2. S. Ginsburg and H. G. Rice, Two Families of Languages Related to Algol, Technical Memorandum, System Development Corporation, Santa Monica, California, January 1961.

3. H. Kesten, Trans. Am. Math. Soc. 92, 336-354 (1959).

4. R. J. Parikh, Language-generating devices, Quarterly Progress Report No. 60, Research Laboratory of Electronics, M.I.T., January 15, 1961, pp. 199-212.

5. E. L. Post, A variant of a recursively unsolvable problem, Bull. Am. Math. Soc. 52, 264-268 (1946).

6. M. O. Rabin and D. Scott, Finite automata and their decision problems, IBM J. Res. Develop. 3, 114-125 (1959).

7. M. P. Schützenberger, A remark on finite transducers, Information and Control 4, 185-196 (1961).

8. M. P. Schützenberger, On the definition of a class of automata, Information and Control 4, 245-270 (1961).

9. M. P. Schützenberger, Un problème de la théorie des automates (Séminaire Dubreil Pisot, Paris, 1959-1960).

10. (Added in Proof): Since this paper was written I have obtained a set F_3 and, consequently, the true diagram is Fig. XII-2.

ON A SPECIAL CLASS OF RECURRENT EVENTS

By M. P. Schützenberger

Université de Poitiers and University of North Carolina

I. Introduction. Let F be the set of all finite sequences (*words*) in the symbols $x \varepsilon X$. According to W. Feller ([2], Chap. VIII), a recurrent event $\mathfrak{E}$ is a pair (A, μ) where A is a subset of F and μ a probability measure fulfilling the conditions recalled below; one says that the event $\mathfrak{E} = (A, \mu)$ occurs at the last letter x_{i_n} of a word $f = x_{i_1}x_{i_2} \cdots x_{i_n}$ if and only if f belongs to the set A; we shall call A the *support* of $\mathfrak{E}$ and denote by $T(A, \mu)$ the mean recurrence time of the event $\mathfrak{E}$.

If the pair (B, μ') defines another recurrent event on F, the pair $(A \cap B, \mu')$ defines also a recurrent event. It results from the general theory of Feller ([2], Chap. VIII) that, when $T(B, \mu')$ is finite, the ratio $\pi = T(B, \mu')/T(A \cap B, \mu')$ is, in a certain sense, the limit of the conditional probability that a random word $f \varepsilon F$ belongs to A when it is known to belong to B. For given arbitrary A, it is in general possible to find infinitely many (B, μ') having finite $T(B, \mu')$ which are such that $\pi = 0$.

The main point of this note is to verify several statements which, together, imply the following property:

Property 1. *If the support A is such that $T(A \cap B, \mu')$ is finite for* every recurrent event (B, μ') having *finite $T(B, \mu')$, then, for every such (B, μ'), π^{-1} is an integer at most equal to a certain finite number δ^* which depends only upon A.*

Classical examples of this occurrence are the return to the origin in random walks over a finite group [3] and, in particular, the recurrent event which occurs at the end of every word whose length is an integral multiple of a particular integer.

In Section II, we discuss some properties of a class of recurrent events which we shall call *birecurrent*; in Section III, we verify the statements mentioned above, and in Section IV we describe examples of birecurrent supports.

II. Preliminary remarks. We consider F as the free monoid ([1], Chap. 1) generated by X; the empty word e is the neutral element of F and the product ff' of the words f and f' is the word f'' made up of f followed by f'; $f(f')$ is called a *left (right) factor* of f''; a word is *proper* if it is different from e.

Feller's condition ([2], Chap. VIII) that the non empty subset A of F is the support of a recurrent event can be expressed as follows: U_r : *if $a \varepsilon A$ and $f \varepsilon F$, then, $af \varepsilon A$ if and only if $f \varepsilon A$*. This condition implies that A is a submonoid of F (i.e., that $e \varepsilon A$ and $A^2 \subset A$). We shall say that A is *birecurrent* if it satisfies U_r and the symmetric condition U_l , U_l : *if $a \varepsilon A$ and $f \varepsilon F$, then, $fa \varepsilon A$ if and only if $f \varepsilon A$*.

It follows immediately that, if $\{A_i\}$ is any collection of supports of recurrent

Received May 26, 1960; revised February 14, 1961.

(birecurrent) events, the same is true of the intersection C of the sets A_i ; indeed, C is a submonoid because every A_i is a submonoid and, if, e.g., a, $af \varepsilon C$, the word f belongs to all the sets A_i (because of U_r) and consequently it belongs also to C.

Throughout this paper, A will denote a recurrent (or, eventually, birecurrent) support and we shall use the following notations:

A^* = the set of all the proper words at the end of which the event whose support is A occurs for the first time; for any recurrent support B, B^* is defined similarly.

$S = F - A^*F$ (= the complement in F of the right ideal A^*F);

$R = F - FA^*$.

We state explicitly the following well known facts:

II.1. Every $f \varepsilon F$ admits one and only one factorization $f = as$ with $a \varepsilon A$ and $s \varepsilon S$ and at least one factorization $f = ra'$ with $a' \varepsilon A$ and $r \varepsilon R$. If and only if A is birecurrent the second factorization is unique for all $f \varepsilon F$.

II.1′. Every proper a of A admits a unique factorization as a product of elements of A^*.

The two statements are quite intuitive but a formal proof of them has been given in ([5]); II.1′ shows that any bijection (i.e., one to one mapping onto) of A^* onto a set Y can be extended to an isomorphism of A onto the free monoid generated by Y.

The following remark will be used repeatedly in the course of this paper:

II.1″. When A is birecurrent, if s, $s' \varepsilon S$ (r, $r' \varepsilon R$) are such that s is a right factor of s' (r is a left factor of r') and that sf, $s'f \varepsilon A$ (fr, $fr' \varepsilon A$) for some $f \varepsilon F$, then $s = s'$ ($r = r'$). If, furthermore, $f \varepsilon R$ ($f \varepsilon S$), then $sf \varepsilon A^* \cup \{e\}$.

Proof. Because of the perfect symmetry of U_r and U_l we can limit ourselves to the proof of the statement concerning s and s'. By hypothesis, $s' = f's$ for some $f' \varepsilon S$ and sf, $f'sf \varepsilon A$; because of U_l, this implies $f' \varepsilon A$. Because of $s' \varepsilon S = F - A^*F$ and II.1′, this, in turn, implies $f' = e$, and we have proved that $s' = es = s$. Let us assume now that $sr \varepsilon A$ with $s \varepsilon S$ and $r \varepsilon R$. If, in addition, $sr = e$, the result is proved. If $sr \varepsilon A - \{e\}$, II.1′ shows that $sr = aa'$ with $a \varepsilon A^*$ and $a' \varepsilon A$; as above, a cannot be a left factor of s and, consequently, a' is a right factor of r; but, by a symmetrical argument, this shows that $a' = e$ and that consequently $sr = a \varepsilon A^*$. This concludes the proof of II.1″.

Let us assume now that A is birecurrent; we denote by $\Delta Sf(\Delta Rf)$ the set of the right (left) factors of f that belong to $S(R)$ and by Δf the set of the triples (r, a, s) such that $f = ras$ and that $r \varepsilon R$, $a \varepsilon A$, $s \varepsilon S$; such a triple will be called an *A-factorization* of f and δf will denote the number of distinct triples in the set of the A-factorizations of f.

II.2. For any $f, f' \varepsilon F$, $\delta ff' \geqq \max(\delta f, \delta f')$ and $\delta ff' = \delta f$ ($= \delta f'$) if and only if for every left (right) factor f'' of f' (of f) the product ff'' ($f''f'$) has a factorization $ff'' = sa$ ($f''f' = ar'$) where $a \varepsilon A$ and where f'' is a right (left) factor of a.

PROOF. Let us consider any element $g \varepsilon F$ and prove that there exists a bijection $\sigma_g : \Delta Rg \to \Delta Sg$. Indeed, by II.1, to any $r \varepsilon \Delta Rg$ (i.e., to any $r \varepsilon R$ which is such that $g = rg'$ for some $g' \varepsilon F$) there corresponds a unique $s \varepsilon \Delta Sg$ (determined by the conditions $g' = as$, $a \varepsilon A$, $s \varepsilon S$) which we call $\sigma_g r$; because of the symmetry implied by the hypothesis that A is birecurrent we can construct in a similar manner a mapping $\Delta Sg \to \Delta Rg$ which we call σ_g^{-1}. Since, clearly, for any $r \varepsilon \Delta Rg$ we have $(\sigma_g^{-1} \circ \sigma_g)r = r$ and similarly for any $s \varepsilon Sg$, this shows that σ_g is a bijection and also that the A-factorizations of g are in a one-to-one correspondence with the elements of ΔRg.

We now revert to the proof of II.2. By the above construction we know that $\delta ff'$ is equal to δf (i.e., to the number of elements in ΔRf) plus the number of proper $r' \varepsilon \Delta Rf'$ such that $fr' \varepsilon R$. Thus, $\delta ff' \geqq \delta f$ with the equality sign if and only if we do not have $ff'' \varepsilon R - \Delta Rf$ for some left factor f'' of f', i.e., if and only if every such ff'' satisfies the condition stated in II.2. Because of the symmetry this concludes the proof.

For any $f \varepsilon F$, let us denote by αf the smallest positive integer for which $f^{\alpha f} \varepsilon A$; αf is infinite if the only finite power of f that belongs to A is f^0 $(= e$, by definition).

II.3. A sufficient condition that the recurrent support A is birecurrent is that αf is finite for all $f \varepsilon F$; reciprocally if A is a birecurrent support, then, for any $f \varepsilon F$, αf is at most equal to the supremum $\delta' f$ of δf^m over all the positive powers of f.

PROOF. By hypothesis, A satisfies U_r and, in order to show that it is birecurrent, it will be enough to show that if a and fa belong to A then f also belongs to A. Let us assume that $(af)^m \varepsilon A$ for some positive finite m; we have $(af)^m = a(fa)^{m-1} f \varepsilon A$ and, because of the fact that a, $(fa)^{m-1} \varepsilon A$ and U_r, this implies $f \varepsilon A$. This proves the first part of II.3.

Now let A be birecurrent and f such that $\delta' f$ is finite; by II.1, any $f^n (0 \leqq n \leqq \delta' f)$ admits an A-factorization (e, a_n, s_n) and, by II.2, to each such s_n there corresponds one A-factorization of $f^{\delta' f}$. Since, by definition, $\delta f^{\delta' f} \leqq \delta' f$, we must have $s_n = s_m$ $(= s$, say) with $0 \leqq m, n \leqq \delta' f$ and, e.g., $m < n$. Thus, $f^n = as$ and $f^m = a's$ with $a, a' \varepsilon A$ and, after cancelling s, we obtain $f^{n-m} a' = a$. Because of U_l, this last relation shows that f^{n-m} belongs to A and, since $0 < n - m \leqq \delta' f$, by construction, the result is entirely proved.

Let us assume now that A is birecurrent and that f is such that $\delta f = \delta f^2 < \infty$. We consider the set K (containing at least f^2) defined by $K = \{f' \varepsilon fFf : \delta f' = \delta f\}$.

II.4. There exists a group G, a subgroup H of G and a mapping $\sigma : K \to G$ that have the following properties: σ is an epimorphism (i.e., homomorphism onto) and G is finite; $\sigma^{-1} H = K \cap A$ and the index of H in G is at most δf.

PROOF. According to II.2, the hypothesis $\delta f = \delta f^2$ implies the existence of a bijection $\sigma^* : \Delta Sf \to \Delta Rf$ defined for each $s \varepsilon \Delta Sf$ by $\sigma^* s$, the unique $r \varepsilon \Delta Rf$ which is such that $sr \varepsilon A$; trivially, $\sigma^* e = e$. Also, by II.2 and the very definition of K, we have $\Delta Rk = \Delta Rf$ and $\Delta Sk = \Delta Sf$ for any $k \varepsilon K$; consequently, $K^2 \subset K$.

Thus, recalling the definition of σ_f given in the proof of II.2, we can associate to any $k \varepsilon K$ a bijection $\sigma_k^* : \Delta Rf = \Delta Rf$ defined by $\sigma_k^* = \sigma^* \circ \sigma_k$.

Let us now verify that for any $k, k' \varepsilon K$ we have $\sigma_{kk'}^* = \sigma_k^* \circ \sigma_{k'}^*$. Indeed, if $(r, a, s) \varepsilon \Delta k$ and $(r', a', s') \varepsilon \Delta k'$ we shall have $(r, a'', s') \varepsilon \Delta kk'$ for some $a'' \varepsilon A$ if and only if $sr' \varepsilon A$ and the identity is verified. Because of the hypothesis that δf is finite, this construction shows that the set $\{\sigma_k^*\}$ $(k \varepsilon K)$ is a group G and that the mapping σ which sends every $k \varepsilon K$ onto σ_k^* is an epimorphism.

Observe now that k belongs to A if and only if $(e, k, e) \varepsilon \Delta k$, that is, if and only if σ_k^* keeps e invariant. Again, because G is finite, the elements $k \varepsilon K$ which have this last property map onto a subgroup H of G and, clearly, $\sigma^{-1}H$ is contained in A. The fact that the index of H in G is at most equal to the number of elements in ΔRf (i.e., to the number δf) is a standard result from group theory. As a corollary of II.4 we state II.4′.

II.4′. If A is such that the supremum δ^* of $\delta f'$ over all $f' \varepsilon F$ is finite and if $\delta f = \delta^*$, then the representation $\{\sigma_k^*\}$ described in II.4 is isomorphic to the representation of G over the cosets of H.

Proof. The property stated amounts to the statement that the group $G = \{\sigma_k^*\}$ is transitive or, in an equivalent fashion, to the fact that for every $s \varepsilon \Delta Sf$ there exists at least one $k \varepsilon K$ such that $\sigma_k e = s$, i.e., such that $k = as$ with $a \varepsilon A$.

In order to prove this, let $(r, a', s) \varepsilon \Delta f$. By II.3 we know that there exist finite positive integers m and m' such that $f^m \varepsilon A$ and $r^{m'} \varepsilon A$. Thus the product $f^m f^{m'-1} f = f^m f^{m'} as$ admits the factorization $a''s$ with $a'' = f^m f^{m'} a' \varepsilon A$ and it belongs to K since, under the hypothesis that δf is maximal, K is identical to fFf.

The next statement is not needed for the verification of property 1. Its aim is to show that the representation described in Section IV below covers all the birecurrent supports with finite $\delta^* = \sup \delta f$.

II.5. If A is a birecurrent support with finite δ^* there exists a monoid M and an epimorphism (homomorphism onto) $\gamma : F \to M$ such that $\gamma^{-1}\gamma A = A$, and that M admits minimal ideals.

Proof. Let us consider any $f \varepsilon F$ and denote by $\{\gamma f\}$ the set of all $f' \varepsilon F$ which satisfy the following condition: for any $f_1, f_2 \varepsilon F$, $f_1 f f_2 \varepsilon A$ if and only if $f_1 f' f_2 \varepsilon A$. The relation $f' \varepsilon \{\gamma f\}$ is reflexive and transitive and it is well known that it is compatible with the multiplicative structure of F (i.e., it is a congruence); thus we can identify each set $\{\gamma f\}$ with an element γf of a certain quotient monoid M of F. Since $f \varepsilon A$ if and only if $f_1 f f_2 \varepsilon A$ with $f_1 = f_2 = e$, A is the union of the sets $\{\gamma a\}$ $(a \varepsilon A)$ and, trivially, $\gamma^{-1}\gamma A = A$.

Let us now take an element f such that $\delta f = \delta^*$, a finite quantity; according to II.2, the maximal character of δf implies that for every f_1 the product $f_1 f$ has a left factor $f_1 r \varepsilon A$ for some $r \varepsilon \Delta Rf$. Thus, because of the symmetry, any relation $f_1 f f_2 \varepsilon A$ implies $f_1 r, s f_2 \varepsilon A$ with $(r, a, s) \varepsilon \Delta f$.

It follows immediately that for any two $k, k' \varepsilon K (= fFf)$, the relation $\gamma k = \gamma k'$ is equivalent to the relation $\sigma k = \sigma k'$ in the notations of II.4. Thus, σK is

isomorphic to a group and since K is the intersection of a right and of a left ideal of F, this shows that M admits minimal ideals.

We now revert to the preparation of the proof of the main property and we consider A, a birecurrent support, B a recurrent support and $C = A \cap B$; we assume that C does not reduce to $\{e\}$ and that consequently C^* (the set of the proper words at the end of which the events whose supports are A and B respectively occur together for the first time) is not empty.

II.6. Any element f from $F - C^*F$ has a unique factorization $f = f_1 f_2$ with $f_1 \varepsilon B - C^*B$ and $f_2 \varepsilon F - B^*F$; conversely any such product $f_1 f_2$ belongs to $F - C^*F$.

PROOF. Because of II.1 any f has a unique factorization $f = f_1 f_2$ with $f_1 \varepsilon B$ and $f_2 \varepsilon F - B^*F$. Since C is a recurrent support contained in B, any product $f_1' f_2'$ with $f_1' \varepsilon B$ and $f_2' \varepsilon F - B^*F$ belongs to $F - C^*F$ if and only if f_1' belongs to $B - C^*B$ and this concludes the proof.

As mentioned in II.1′, there exists an isomorphism $\beta : B \to Q$ where Q is the free monoid generated by $Q^* = \beta B^*$ and it is easily verified that the image P of C by β satisfies U_r and U_l when, according to our hypothesis, A is birecurrent. Indeed, P is surely a submonoid of Q and it is enough to verify that the relations $p, p', pqp' \varepsilon P$ imply $q \varepsilon Q$ (because $\beta^{-1}p, \beta^{-1}p', \beta^{-1}pqp' \varepsilon A$ imply, e.g., $\beta^{-1}qp \varepsilon A$, by U_r, then $\beta^{-1}q \varepsilon A$, by U_l and, finally $q \varepsilon P = \beta(A \cap B)$).

As before, we define a P-factorization of an element $q \varepsilon Q$ as a triple $(\bar{r}, p, \bar{s})$ such that $q = \bar{r}p\bar{s}$ and that $\bar{r} \varepsilon \bar{R} = Q - QP^*$, $p \varepsilon P$, $\bar{s} \varepsilon \bar{S} = Q - P^*Q$ with $P^* = \beta C^*$. All the remarks made in II.2 apply here since P is a birecurrent support in Q, and we define $\bar{\delta}q$ as the number of P-factorizations of q.

II.7. For any $b \varepsilon B$, $\bar{\delta}\beta b \leqq \delta b$.

PROOF. Let $\bar{r}$ be any element of $\bar{R}$ and define $\beta^*\bar{r}$ as the (uniquely determined) element $r \varepsilon R$ such that $(r, a, e) \varepsilon \Delta b$ for some $a \varepsilon A$. We show that the restriction of the mapping β^* to any set $\Delta\bar{R}q$ ($q \varepsilon Q$) is an injection (i.e., is one to one into). Indeed, if $\bar{r}, \bar{r}' \varepsilon \Delta\bar{R}q$ we have, e.g., $\bar{r}' = \bar{r}q'$ for some $q' \varepsilon Q$; thus, if $\beta^*\bar{r} = \beta^*\bar{r}'$ ($= r$, say), we have the following relations: $\beta^{-1}\bar{r} = ra \varepsilon B$ with $a \varepsilon A$; $\beta^{-1}\bar{r}' = ra' \varepsilon B$ with $a' \varepsilon A$; $ra' = rab'$ with $b' = \beta^{-1}q\beta \varepsilon B$. Consequently, $a' = ab'$ and, because of U_r, $b' \varepsilon A$. This shows that $q' = \beta b'$ belongs to P and that finally, $q' = e$ because of the relation $\bar{r}' = \bar{r}q' \varepsilon \bar{R}$. Thus, $\bar{r}' = \bar{r}$ and our contention is proved.

The remark II.7 is also proved since we have shown that for any $b \varepsilon B$ there exists an injection of $\Delta\bar{R}\beta b$ into ΔRb.

II.8. If δ^* ($= \sup \delta f$) is finite and if $\delta b = \delta^*$ for at least one $b \varepsilon B$, then $\bar{\delta}^*$ ($= \sup \bar{\delta}q$) is a divisor of δ^*.

PROOF. Under these hypotheses, we may assume without loss of generality that B contains an element f such that $\delta f = \delta^*$ and $\bar{\delta}\beta f = \bar{\delta}^*$. We use the notations of II.4 and II.4′. By construction, the image G' by σ of $B \cap K$ is a subgroup of G and we have $B \cap \sigma^{-1}(H \cap G') = A \cap B \cap K$. Thus, by a standard result of

group theory the index δ'^* of $H \cap G'$ in G' is a divisor of the index of H in G (i.e., of δ^*). We prove now that δ'^* is in fact equal to $\bar{\delta}^*$; for this we repeat the construction of II.4 and II.4′ with $\beta(B \cap K)$ in the role of K and we obtain an epimorphism $\bar{\alpha}: \beta(B \cap K) \to \bar{G}$ such that $\bar{\delta}^*$ is the index of the subgroup $\bar{H}$ of $\bar{G}$. We recall the definition of the mapping β^* used in II.7 and we observe that we can define a bijection $\beta^{*-1}: \Delta Rf \cap \beta^* \Delta \bar{R} \beta f \to \Delta \bar{R} \beta f$ such that $\beta^{*-1} \circ \beta^*$ is the identity mapping of $\Delta \bar{R} \beta f$ onto itself; β^{*-1} induces in a natural fashion an epimorphism $\beta^{**}: G' \to \bar{G}$ and, trivially, $H \cap G'$ is the inverse image of $\bar{H}$ by β^{**}. Thus $\bar{\delta}^*$ is equal to δ'^* and II.8 is proved.

III. Verification of property 1. We keep the notations already introduced and we assume that (A, μ) is a recurrent event. According to Feller, μ satisfies the two conditions:

M_0 : $\mu e = 1$ and for any $f \varepsilon F$, $\mu f = \sum (\mu fx: x \varepsilon X)$,

M_r : if $a \varepsilon A$ and $f \varepsilon F$ then $\mu af = \mu a \mu f$.

We shall say that μ is a *positive product measure* if $\mu ff' = \mu f \mu f' > 0$ for any f, $f' \varepsilon F$, and, in this case, M_r is trivially satisfied.

We denote by $|f|$ the length of the element f and for any subset F' of F we use the following notations: $F'_n = \{f \varepsilon F': |f| \leqq n\}$; $\mu F' = \lim_{n \to \infty} \sum \{\mu f: f \varepsilon F'_n\}$. It follows that $\mu F' \leqq 1$ if F' is such that any $f \varepsilon F$ has at most one left factor which belongs to F'; this condition is satisfied in particular by any subset of A^* and, according to Feller's definition, we shall say that (A, μ) is *persistent* if and only if $\mu A^* = 1$. The next two statements are verified by an imitation of Feller's proof procedure.

III.1. For any recurrent event (A, μ) we have $T(A, \mu) = \mu S$.

Proof. Let us introduce for any $s \varepsilon S$ the notation $S(s) = S \cap sF$. We verify the identities

(III.1). for all $m \geqq |s|$: $0 \leqq \mu s - \mu A^*_{m+1}(s) = \mu S_{m+1}(s) - \mu S_m(s)$;

(III.1′). for all $m \geqq 1$: $(1 - \mu A^*) + (\mu A^* - \mu A^*_m) = \mu S_m - \mu S_{m-1}$

Indeed, (III.1) is an immediate consequence of M_0 and of the fact that the sets $\{s\} \cup S_m(s)X$ and $S_{m+1}(s) \cup A^*_{m+1}(s)$ are identical for any $m \geqq |s|$. (III.1′) is the special case of (III.1) for $s = e$.

From this second identity we deduce that if $\mu A^* = 1$ we have $\lim_{m \to \infty} (\mu S_m - \mu S_{m-1}) = 0$. Thus, *a fortiori* (from the first identity) $\mu A^* = 1$ implies $\mu s = \mu A^*(s)$. We now sum the second identity from $m = 1$ to $m = n$. After rearranging terms, we obtain:

(III.1″). $\mu S_n = (n + 1)(1 - \mu A^*_n) + \sum \{|a| \mu a: a \varepsilon A^*_n\}$.

This shows that if (A, μ) is not persistent, μS is infinite and we assume now that $\mu A^* = 1$. Under this hypothesis, $T(A, \mu)$ is defined as $\lim_{n \to \infty} \sum \{|a| \mu a: a \varepsilon A^*_n\}$, and since $\mu A^* = 1$ implies that

$$(n + 1)(1 - \mu A^*_n) = \sum \{(n + 1)\mu a: a \varepsilon A^* - A^*_n\},$$

we can write for all n

$$\sum\{|a|\ \mu a : a \in A_n^*\} \leqq \mu S_n \leqq \sum\{|a|\ \mu a : a \in A^* - A_n^*\} + \sum\{|a|\ \mu a : a \in A_n^*\}.$$

This concludes the proof since it shows that $\mu S = T(A, \mu)$ if this last quantity is finite and that μS is infinite if $T(A, \mu)$ is so.

For any $s \in S$ let us define $R^*(s)$ as $\{e\}$ if $s = e$ and, as the set of those $f \in F$ such that $sf \in A^*$, if $s \neq e$.

III.2. If A is birecurrent, μ a product measure and (A, μ) persistent, we have $T(A, \mu) = \mu R$ and, for all $s \in S$, $1 = \mu R^*(s)$.

PROOF. Under these hypotheses all the notions are perfectly symmetrical. Thus, the identity (III.1″) shows that $\mu R_n = \mu S_n$ and, as a special case, that $\mu R = T(A, \mu)$. Since any $a \in A^*(s)$ has a unique factorization $a = sf$ with $f \in R^*(s)$, and since μ is a product measure, we have for all $m \geqq |s|$ the identity

$$\mu A_m^*(s) = \mu s \mu R_{m-|s|}^*(s). \tag{III.2}$$

Thus, we have in any case $\mu R(s) = \mu A^*(s)/\mu s \leqq 1$ because of the formula (III.1); with the equality sign when (A, μ) is persistent because as seen above $\mu s = \mu A^*(s)$.

III.3. If A is birecurrent and μ a product measure, $T(A, \mu) = \delta^*$.

PROOF. We use the notations of Section II and we recall the following facts:

(1) According to II.1″, $R^*(s)$ is a subset of R;

(2) for the same reason, if $s, s' \in \Delta Sf$ for some $f \in F$, the sets $R^*(s)$ and $R^*(s')$ are disjoint.

(3) if δ^* is finite and $\delta f = \delta^*$ then, by II.2, to every $r \in R$ there corresponds one $s \in \Delta Sf$ such that $sr \in A^*$. Thus, in this case, the union of the sets $R^*(s)$ over all $s \in \Delta Sf$ is equal to R. Now to the proof! We shall show that if $\delta f = \delta^*$ we have the inequalities $\mu R \leqq \delta f \leqq \mu R$ and, trivially, the result will follow by III.2.

The second inequality is vacuously true when (A, μ) is not persistent since, then, μR is infinite. When (A, μ) is persistent we have for any $f' \in F$ the inequality $\delta f' = \sum\{\mu R^*(s) : s \in \Delta Sf'\} \leqq \mu R$ since, then, $\mu R^*(s) = 1$ and since the sets $R^*(s)$ are pairwise disjoint. Thus the second inequality is always true. If now $\delta f = \delta^*$, we know by 3 above that $\sum\{\mu R^*(s) : s \in \Delta Sf\} = \mu R$. Since in any case, as we have seen in the proof of III.2, we have $\mu R^*(s) \leqq 1$, it follows that $\mu R \leqq \delta^*$ and the result is proved.

III.4. If (B', μ) is a recurrent event and if A is birecurrent we have

$$T(A \cap B', \mu) = \bar{\delta}^* T(B', \mu)$$

where $\bar{\delta}^*$ is defined below.

PROOF. Let $B = \{b \in B' : \mu b > 0\}$ and $C = A \cap B$; it is easily verified that (B, μ) is again a recurrent event and that, according to III.1. we have

$$T(A \cap B', \mu) = T(A \cap B, \mu) = \mu(F - C^*F)$$
$$T(B', \mu) = T(B, \mu) = \mu(F - B^*F).$$

We keep the notations used in the proofs of II.6 and II.7 and we observe that, by taking into account II.6 and the condition M_r on μ, the remark III.4 is equivalent to the relation $\mu(B - C^*B) = \bar{\delta}^*$. In order to prove this identity we define a measure ν on Q by the relation $\nu\beta b = \mu b$, for all $b \varepsilon B$; because of M_r and of the definition of B, ν is a positive product measure and, since we know that $P = \beta C$ is birecurrent, (P, ν) is a recurrent event on Q. Because of III.1 and III.3 $T(P, \nu) = \nu(Q - P^*Q) = \bar{\delta}^*$. But, by definition, $\nu(Q - P^*Q) = \nu\beta(B - C^*B) = \mu(B - C^*B)$ and the result is proved.

III.5. If δ^* is finite, and (B', μ) persistent for some measure μ which satisfies the condition that for every $f \varepsilon F$ at least one element from FfF has positive measure, then $\bar{\delta}^*$ is a divisor of δ^*.

Proof. Because of the conditions satisfied by μ and δ^* we can find an element f such that $\delta f = \delta^*$ and that $\mu f > 0$; we have $f = b's'$ with $b' \varepsilon B$ and $s' \varepsilon F - B^*F$. Because (B, μ) is persistent, it follows from III.1 that $\mu(B^* - s'F) = \mu s'$. Since this last quantity is positive, there exists at least one element $b \varepsilon B^* \cap s'F$. Finally, because of II.2 we have $\delta b'b = \delta^*$ with $b'b \varepsilon B$. Thus, we can apply II.8 and the result is proved.

The next statement is intended to give a characterization of the birecurrent supports in terms of their intersection with other recurrent events; by E we mean any fixed birecurrent support such that $T(E, \mu)$ is finite for some positive product measure μ; E^* is defined as usual and we say that (E', μ') belongs to the family $((E))$ if the two following conditions are met:

(i). (E', μ') is a recurrent event on F;

(ii). there exists a finite integer m such that any element from E'^* is the product of m words from E^*. It is trivial that under these hypotheses E' is birecurrent. Since F itself is a birecurrent support (with $F^* = X$) a simple example of a family $((E))$ is the family of the birecurrent events $(F_{(m)}, \mu_m)$ where $F_{(m)}$ is the set of all words whose length is a multiple of m and where μ_m is a suitable measure.

III.6. If the recurrent support A is such that $(A \cap E', \mu')$ is persistent for every $(E', \mu') \varepsilon ((E))$, then, A is a birecurrent support.

Proof. This is a simple application of II.3 and we use the notations of this remark. If αf is finite for all f, then we know by II.3 that A is birecurrent. Thus we may suppose that A and f are such that αf is infinite and we show that $(A \cap E', \mu')$ is not persistent for some suitable (E', μ'). Indeed, by the second part of II.3 we know that $f^m \varepsilon E$ for some finite positive m. Thus f^m admits a factorization as a product of m' elements from E^*. We take E' defined by the condition $E'^* = E^{*m'}$ and μ' defined by the condition that $\mu' f^m = 1$ and $\mu' f' = 0$ for any other $f' \varepsilon E'^*$. The conditions M_0 and M_r recalled at the beginning of this section are obviously satisfied and $T(E', \mu')$ is finite. Finally, $(A \cap E', \mu')$ cannot be persistent since $A \cap E'$ reduces to $\{e\}$ and this ends the proof.

Clearly, the conditions of III.6 are satisfied if A is such that $T(A \cap B, \mu) < \infty$ for any (B, μ) with finite $T(B, \mu)$.

The next statement is a simple application of II.2.

III.7. If A is birecurrent and if δ^* is finite, then, for any product measure, μ, the distribution of the recurrence time of (A, μ) has moments of every order.

PROOF. Let $A' = \{a \varepsilon A : \mu a > 0\}$. Trivially, A' is birecurrent and, by II.7 we know that every $f \varepsilon F$ has at most δ^* A'-factorizations. Since the distribution of the recurrence times of (A, μ) and (A', μ) are the same, there is no loss of generality in assuming that $A = A'$, i.e., that μ is positive.

Since δ^* is finite there exists an element $f \varepsilon F$ which, because of II.2, has the property that for any proper $s \varepsilon S$ the product sf has a factorization $sf = ar$ with $a \varepsilon A^*A$. Thus, for any integer n, the definition $S = F - A^*F$ allows us to write the inequality

$$\Sigma\{\mu f' : f' \varepsilon A^*, n|f| < |f'| \leqq (n+1)|f|\} = \mu A^*_{(n+1)}|f| - \mu A^*_n|f| \leqq (1 - \mu f)^{n+1}.$$

Consequently the distribution of the $|a|$ for $a \varepsilon A^*$, i.e., of the recurrence time of A^*, is dominated by an exponential distribution and this proves the result.

IV. Examples. We want to describe a class of monoids, V, which allows the construction of birecurrent supports. For this purpose, we consider a group G' (whose elements are identified with the corresponding elements of its Frobenius algebra) and a subgroup H' which contains no proper normal subgroup of G'; $I = \{i\}$ and $J = \{j\}$ are two sets of indices and w is a $I \times J$ matrix with entries w_{ij} in H'. Without loss of generality we can assume that there exists no pair of indices $j, j' \varepsilon J$ $(i, i' \varepsilon I)$ and no element $h \varepsilon H'$ such that $w_{ij}h = w_{ij'}$ $(hw_{ij} = w_{i'j})$ identically for all $i \varepsilon I$ $(j \varepsilon J)$.

We shall denote by V the set of all $I \times I$ matrices v with entries in $G' \cup \{0\}$ that have the following property: for each $j \varepsilon J$ there exists an index $j' \varepsilon J$ and an element $g_{jj'} \varepsilon G'$ which are such that the product $vw_{.j}$ (with $w_{.j}$ = the jth column vector of w) is equal to $w_{.j'}g_{jj'}$ (i.e., to the vector whose ith entry is equal to $w_{ij'}g_{jj'}$). Trivially, this condition implies that v has one and only one non zero entry in each line; it also implies the existence of an isomorphism $v \to \bar{v}$ which sends V onto the monoid $\bar{V}$ of the $J \times J$ matrices defined by the symmetric condition and which is such that $vw = w\bar{v}$, identically; V is a monoid and it contains as minimal ideal the set V_0 of all matrices whose ith column vector is equal to $w_{.j}g$ (with any $i \varepsilon I$, $j \varepsilon J$, $g \varepsilon G'$) and whose i'th column vector is zero for $i' \neq i$.

IV.1. The subset $L \subset V$ of the matrices of V which have at least one entry in H' satisfies U_r and U_l.

PROOF. L is not empty since it contains at least the neutral element of V. Let us assume that $v \varepsilon L$ and that $v_{ii'} \varepsilon H'$. Because of the hypothesis that all the entries of w belong to H', the ith coordinate of $vw_{.j}$ for any $j \varepsilon J$, (that is, $v_{ii'}w_{i'j}$) belongs to H'. Thus, $vw_{.j} = w_{.j'}h$ for some $j' \varepsilon J$ and $h \varepsilon H$; it follows that all the non zero entries of v belong to H'. This shows that L is a monoid and, trivially, that it satisfies U_r and U_l.

IV.1′. If F is a free monoid and $\gamma' : F \to V$ an homomorphism, then the subset $A = \gamma'^{-1}(L \cap \gamma' F)$ is a birecurrent support and the corresponding parameter, δ^*, is at most equal to the index of H' in G'.

PROOF. The first part of the statement does not need a proof; we verify the second part by showing that for any $f \varepsilon F$ (with the notations of II.2) there exists an injection of ΔRf into the set of the left H'-cosets. Let r, $r' \varepsilon \Delta Rf$ with, e.g. $r' = rf'$; for any $i \varepsilon I$, the condition $(\gamma' r)_{ii'} \neq 0$ defines in a unique manner $i' \varepsilon I$ and $g = (\gamma' r)_{ii'} \varepsilon G'$. In a similar way, we define $i'' \varepsilon I$ and $g' \varepsilon G'$ by the condition $0 \neq (\gamma' f')_{i'i''} (= g')$. Since $(\gamma' r')_{ii''} = (\gamma' r f')_{ii''} = gg'$ we see that g and g' belong to the same H'-coset if and only if $g \varepsilon H'$, that is, if and only if $f' \varepsilon A$, that is, finally, if and only if $r = r'$ and this ends the proof.

Reciprocally, if A is a birecurrent support with finite δ^* we can take (with the notations of II.5) $G' = G$ and $H' = H$ and find, I, J and w such that $\gamma F = M$ is a submonoid of V. Then $V_0 \subset \gamma F$ and a sufficient condition that $\gamma f \varepsilon V_0$ is $\delta f = \delta^*$. We shall not prove these results here since they are a straightforward application of Clifford's theory [4].

IV.1″. If δ^* is finite and if for each $f \varepsilon F$ there exists a finite positive m such that $\gamma f^m \varepsilon V_0$, then the parameter $\bar{\delta}^*$ defined in II.7 is always a divisor of δ^*.

PROOF. We consider the group G' defined in II.8. According to the general theory of monoids [4] the only groups contained in γF under the hypothesis of IV.1″ are in fact contained in V_0. Consequently, they are isomorphic to subgroups of G and this concludes the proof.

IV.2. If A is a birecurrent support such that A^* is a finite set then either there exists an $s \varepsilon S$ for which $sF \cap A = \phi$ (and then (A, μ) is not persistent for any positive product measure μ) or else, the conditions of IV.1″ are satisfied by A. In this second case, γF is a group if and only if A^* reduces to the set of all the words having some fixed finite length. [5].

PROOF. We assume that A^* is finite and that $A \cap sF \neq \phi$ for all $s \varepsilon S$; then, by the very definition of γ the monoid γF is finite. By II.2 we see that if r, $r' \varepsilon \Delta Rf$ for some $f \varepsilon F$, then the equation $\gamma r = \gamma r'$ implies $r = r'$. Thus, the parameter δ^* is finite. Let us take any element $f \varepsilon F$; the hypothesis that $\delta f < \delta^*$ implies that for some pair (f', f'') one has $f'ff'' \varepsilon A^*$. Thus for all $f \varepsilon F$, $\delta f^m = \delta^*$ for large enough m since, otherwise, A^* would not be finite. This proves that A satisfies the conditions of IV.1″.

We now make the supplementary assumption that γF is a group G with $\gamma A = H$, and we consider a, an element of maximal length of A^*. If $|a| = 1$ the result is vacuously true since, then, $A = F$. If $|a| \geqq 2$ we write $a = sxx'$ with x, $x' \varepsilon X$. Because of U_r, no left factor of a belongs to A^* and because of the maximality of $|a|$, we have $sxx'' \varepsilon A$ for all $x'' \varepsilon X$. Thus, all the generators of F belong to the same left H-coset. For this reason, we cannot have $sx'' \varepsilon A^*$ for any $x'' \varepsilon X$ and, because again of the maximal character of $|a|$ this implies that $sx''x''' \varepsilon A^*$ for any two x'', $x''' \varepsilon X$. Thus, for any two elements x, $x' \varepsilon X$, the left coset $xx'H$ does not depend upon the choice of x and x'. If $|a| = 2$, this proves the result. If $|a| \geqq 3$ we can write $s = s'y$ with $y \varepsilon X$ and by the same argument we prove that for any x, x', $x'' \varepsilon X$ the coset $xx'x''H$ does not depend

upon the choice of these three elements. Since $|a|$ is finite, by hypothesis, a simple induction gives the result.

The next statements discuss the existence of birecurrent supports with finite δ^*. Without loss of generality, we shall assume from now on that X contains a *finite* number $\geqq 2$ of elements.

IV.3. For any finite $n \geqq 3$ there exist infinitely many different birecurrent supports with this value of δ^*.

PROOF. In the next section we shall show the existence of at least one birecurrent support with $\delta^* = 2$ and A^* infinite. In this section we show that to every birecurrent support A and element $u \varepsilon A^*$ we can associate one other birecurrent support B with $\delta_B^* = \delta^* + 1$ and B^* infinite and that, for the same A^* and different choice of $u \varepsilon A^*$, the two corresponding new supports are different. Thus IV.3. will be entirely proved with the help of IV.4.. Let us now take $u \varepsilon A^*$, a fixed element, and define: $J = (uF \cap Fu) - \{u\}$; $J^* = J - J^2$ (i.e., = the subset of those elements of J that cannot be written as the product of two elements of J). With the help of II.1″, it is easily verified that there exists a birecurrent support B which is such that $B^* = J^* \cup (A^* - \{u\})$ and we prove that for all $f \varepsilon F$ the number (say, $\delta(B, f)$) of its B-factorizations is at most equal to $\delta f + 1$. In order to do this, we slightly extend the notations of II.2, and for any subset F' of F we say that the triple (f'', f', f''') is a F'-factorization of f if $f' \varepsilon F'$ and $f''f'f''' = f$; also, we denote by $\delta(F', f)$ the number of distinct F'-factorizations of f and we observe that by induction on the length of f, the result of II.3 can be summarized by the identity $|f| + 1 = \delta(A, f) + \delta(A^*, f)$.

Here, we have

$$\delta(A^*, f) = \delta(A^* - \{u\}, f) + \delta(\{u\}, f),$$

$$\delta(B^*, f) = \delta(A^* - \{u\}, f) + \delta(J^*, f),$$

We want to show that $\delta(B^*, f) \leqq \delta(A^*, f) + 1$. If $\delta(\{u\}, f) = 0$ or 1, we have $\delta(J^*, f) = 0$ and the result is proved; consequently, we assume now that $\delta(\{u\}, f) \geqq 2$ and we consider two $\{u\}$-factorizations (f_1, u, f_1') and (f_2, u, f_2') with, e.g. $|f_1| \leqq |f_2|$. The element w determined by the equation $f = f_1 w f_2'$ belongs to J; it belongs to J^* if and only if there is no $\{u\}$-factorization (f_3, u, f_3') for which $|f_1| < |f_3| < |f_2|$; it follows instantly that $\delta(J^*, f) = \delta(\{u\}, f) - 1$ and the result is proved.

IV.4. For each finite $n \geqq 3$ there exist at least two different birecurrent supports with A^* finite and $\delta^* = n$.

PROOF. One of these supports has been described in IV.2; in order to produce the other one, we take a birecurrent support A, a fixed element $u \varepsilon (F - A^*F) \cap (F - FA^*)$ and we construct another birecurrent support B with $\delta_B^* = \delta^*$; in the last part of the proof we verify that by a proper choice of u and A^* we can make B^* finite.

Let the following sets be defined:

$$C^* = A^* \dot{-} A^* \cap (uF \cup Fu),$$
$$Z = \{f: uf \varepsilon A^* - A^* \cap Fu\},$$
$$Z' = \{f: fu \varepsilon A^* - A^* \cap uF\},$$
$$J^* = A^* \cap uF \cap Fu,$$
$$P^* = \{f: fu \varepsilon A^* \cap uF\}.$$

Thus, A^* admits a partition into the sets C^*, uZ, $Z'u$ and J^*; by construction, there exists a recurrent support P such that $P^* = P^2 - P$ (with $P = \{e\}$ if P^* is empty) and one can verify that there exists a birecurrent support B such that B^* admits a partition into the sets C^*, $\{u\}$ and $Z'PuZ$.

In order to verify that $\delta_B^* = \delta^*$ we take an arbitrary positive product measure μ and, for any $F' \subset F$, we write $T(F')$ as an abbreviation for $\sum (|f| \mu f: f \varepsilon F')$. Thus, by III.3, we have, e.g., $\delta^* = T(A, \mu) = T(A^*)$.

By a simple computation, we obtain when δ^* is finite: $\delta^* = T(A^*) = T(C^*) + T(P^*) + |u|(\mu Z + \mu Z' + \mu P^*)\mu u + (T(Z) + T(Z') + T(P^*))\mu u$. Also, $\mu Z = \mu Z' = 1 - \mu P^*$; $\mu P = (1 - \mu P^*)^{-1}$; $T(P) = (1 - \mu P^*)^{-2} T(P^*)$. Now, $T(B^*)$ $(= \delta_B^*)$ is equal to the sum $T(C^*) + |u|\mu u + T(Z'PuZ)$; because of the above relations, we have $T(Z'PuZ) = |u|\mu u \mu Z + (T(Z) + T(Z') + T(P^*))\mu u$ and this concludes the second part of the proof.

Let us now observe that B^* is finite if and only if C^* is finite and $P = \{e\}$. The first condition is surely satisfied when A^* is finite and the second one is equivalent to $P^* = \phi$, that is, to $A^* \cap uF \cap Fu = \phi$.

Thus, if A^* is the set of all words of length $n > 2$ and if $x_1, x_2 \varepsilon X$, the word $u = x_1^{n-2} x_2$ belongs to $F - A^*F$ and to $F - FA^*$ and it satisfies our last condition; this ends the proof of IV.4.

If we take $n = 2$ and $u = x_1$ we find that $P^* = x_1$ and the corresponding B^* is infinite; this is the example needed for IV.3.

IV.5. For each finite n there exists only a finite number of birecurrent supports A with $\delta^* = n$ which satisfy one or the other of the two following supplementary conditions: that γF is a group or that A^* is finite.

Proof. This is obvious for the first condition since, because of II.4′, it amounts to the fact that for any finite n there exist only finitely many groups of permutation on n symbols.

With respect to the second condition we first verify the following elementary remark: let $K_0 = F - \{e\}$, $K_1, K_2, \cdots$ be a decreasing sequence of subsets of F defined inductively by the relation $K_{i+1} = \{fFf: f \varepsilon K_i\}$. If X is finite there exists for every finite i a finite value $d(i)$ which is such that every word of length at least $d(i)$ has at least one factor belonging to K_i. Indeed, if $d(i)$ has already been defined, we take $d(i+1)$ as $d(i)$ $(1 + |X|^{d(i)})$ where $|X|$ denotes the number of elements of X. Then, every word of length $d(i+1)$ contains at

least two disjoint identical factors of length $d(i)$ and the result follows by induction.

We now observe that if $f \neq e$ the hypothesis that A^* is a finite set (with finite δ^*) implies that $\delta ff'f \geqq \inf(\delta^*, \delta f + 1)$. Indeed, this is surely true if $\delta f' > \delta f$ or if $\delta ff'f = \delta^*$; in the remaining case, i.e., in the case that $\delta f = \delta ff'f < \delta^*$, we would have according to II.2, for all finite m, $\delta(ff')^m f = \delta f < \delta^*$ and, according to the same remark, there would exist for all finite m at least one $a \varepsilon A^*$ admitting $(ff')^m f$ as a factor, which is impossible since A^* is assumed to be finite.

Thus, by induction, every word f of length $\geqq d(\delta^*)$ is such that $\delta f = \delta^*$ and, consequently, it cannot be a factor of a word $a \varepsilon A^*$. This proves that for given δ^* the hypothesis that A^* is finite imposes that the lengths of the words from A^* is bounded and it concludes the proof (cf.[6]).

REFERENCES

[1] Chevalley, C., *Fundamental Concepts of Algebra*, Academic Press, New York, 1956.

[2] Feller, William, *An Introduction to Probability Theory and its Applications*, Vol. 1, 2nd ed., John Wiley and Sons, New York, 1957.

[3] Kesten, H., "Symmetric random walks on groups," *Trans. Amer. Math. Soc.*, Vol. 92 (1959), pp. 336–354.

[4] Miller, D. D., and Clifford, A. H., "Regular D-classes in semigroups," *Trans. Amer. Math. Soc.*, Vol. 82 (1956), pp. 270–280.

[5] Schützenberger, M. P., "Une théorie algébrique du codage," *Séminaire Dubreil-Pisot*, exp. no 15, Paris, 1955–1956.

[6] Schützenberger, M. P., "On a family of submonoids." To appear in *Pub. Math. Inst. Hungarian Academy of Science*, December, 1961.

Reprinted from Information and Control, Volume 4, Nos. 2–3, September 1961

INFORMATION AND CONTROL **4**, 185–196 (1961)

A Remark on Finite Transducers

M. P. Schützenberger

Faculte des Sciences, Rue de la Trinite, Poitiers, France

I. INTRODUCTION

In this note we consider a very restricted class of transducers, i.e., of automata which transform finite input words into finite output words (cf. Moore, 1956). The simplest case is the transformation consisting in the replacement of every input letter x by an output word $\eta(x)$ which is eventually the empty word e_Y. Algebraically, since the set $F_X(F_Y)$ of all finite input (output) words is the free monoid (Chevalley, 1956) generated by the input alphabet $X = \{x\}$ (the output alphabet $Y = \{y\}$), this transformation is simply an homomorphism $\eta: F_X \to F_Y$.

If η is such that $\eta(f) = \eta(f')$ only if $f = f'$, it is called an *encoding* (with unique decipherability) and then η is an isomorphism.

Next in simplicity are the transformations realized by a conventional [one way, one tape (Rabin and Scott, 1959)] automaton supplemented by a printing device (Huffman, 1959). Upon reading x on the input tape and, accordingly, going from the state s to the state $s' = sx$, a word $\eta(s; x)$ function of s and x only is printed on the output tape which is moved the corresponding length. Trivially, any mapping from F_X to F_Y can be performed by a transformation of this type if no restriction is imposed on the number of states. We shall always assume here that $S = \{s\}$ is a finite set. This forces drastic limitations on η and, in particular, it introduces a difference between the *right* transformations (where reading and printing are done from left to right) and the *left* transformations (where both operations are done in the opposite direction). For example no (finite) right automaton can perform the task of reproducing the input word when it ends with a given letter and of printing nothing when it does not.

Consequently the composite operation which consists of transforming first the input word by a right automaton, and then of transforming again the output word by a left automaton cannot as a rule be carried

out in a single pass; we shall call it a *transduction* and we shall describe some of its elementary properties:

1. The transductions form a set closed by finite composition and also by inversion when this last operation has a meaning (Huffman, 1959).

2. The transductions transform regular events (Kleene, 1956) on the input words into regular events on the output words and any regular event can be obtained in this manner.

These two properties indicate that there is no difference between the languages which can be accepted by finite automata and the languages which can be produced by any *bounded* number of finite automata; here, the boundedness condition cannot be omitted as is easily shown by Chomsky's counter examples (cf. Chomsky, 1959).

For notational reasons it is more convenient to define a transduction η with sets of states (S, S') as the transformation from an input word $f = x_1 x_2 \cdots x_n$ and a pair of states $s_1 \varepsilon S$, $s_1' \varepsilon S'$ to an output word that is obtained by replacing every letter x_i by a fixed output word $\eta(s_i ; x_i ; s'_{n-i+1})$ where the states are given inductively by the equations $s_{j+1} = s_j x_j$ and $s'_{n-j+2} = x_j s'_{n-j+1}$. With this definition, right (left) transductions correspond to the special case where $\eta(s_i ; x; s'_{i'})$ does not depend effectively upon its right (left) argument and where, consequently, S' (S) can be taken as reduced to a single state and, finally, omitted.

The finite closure property 1 shows that this new construct is equivalent to the composition of a right and of a left transduction; encodings correspond to the case where S and S' reduce to a single state and, then, the property 1 shows that the deciphering can always be performed by a transduction.

Example. Let $X = \{x_1, x_2\}$ and $Y = \{x_1, x_2, y_3\}$. Every input word has a unique factorization $f = x_i^{n_1} x_{i'}^{n_2} \cdots (i \neq i')$ into *runs* $x_j^{n_k}$ consisting of the same letter x_j repeated n_k times, and we suppose that we want to perform the transformation η which lets invariant the runs of *even* length and replaces every run of *odd* length by y_3.

Thus, for example,

$$\eta x_1^3 x_2^2 x_1 x_2^3 x_1^4 = y_3 x_2^2 y_3^2 x_1^4.$$

This can be realized if for any factorization $f = f' x f''$ we follow the two instructions: (1) Print out x if it belongs to a run of even length. (2) Print out y_3 or nothing when x belongs to a run of odd length according to whether x is or is not the last letter of this run.

In order to carry them out it is sufficient to know that f' and f'' respectively end and begin by runs of length $n' > 0$ and $n'' > 0$ in the letters x' and x'' because: (1) x belongs to a run of even length if $x' = x = x''$ and n' and n'' have different parity or if $x' = x \neq x''$ and n' is odd or if $x' \neq x = x''$ and n'' is odd; (2) x is the last letter of a run of odd length if $x \neq x''$ and n' is even, or if $x' \neq x \neq x''$.

Consequently, all that is needed is the parity of n' and n'' and the last and first letter respectively of f' and f''. As we shall see below this information can be supplied by two finite state automata, one having read f' from left to right and the other one having read f'' in the opposite direction.

Let us now consider how this transformation could be achieved in *two* passes.

The first one is performed by a right transduction with states $\{s_i\}$ $(0 \leqq i \leqq 4)$, initial state s_0 and transitions:

$$s_0x_1 = s_2x_1 = s_3x_1 = s_4x_1 = s_1 ,$$

$$s_1x_1 = s_2 ,$$

$$s_0x_2 = s_1x_2 = s_2x_2 = s_4x_2 = s_3 ,$$

$$s_3x_2 = s_4 .$$

Thus for any input word f' the last state reached, s_j, has index of the same parity as the last run of f', and $j \leqq 2$ if and only if the last letter of f' is x_1. The machine has an output alphabet $Z = \{z_i\}$ $(1 \leqq i \leqq 4)$ with the printing rule $\eta'(s_i ; x_j) = z_{i\cdot}$ when $s_ix_j = s_{i\cdot}$. For example,

$$\eta'(x_1^3x_2^2x_1x_2^3x_1^4) = z_1z_2z_1z_3z_4z_1z_3z_4z_3z_1z_2z_1z_2 = \eta'f.$$

The second pass is performed by a left transducer with states $\{s_i'\}$ $(0 \leqq i \leqq 4)$, initial state s_0' and transitions:

$z_1s_i' = s_1'$ if $i \neq 2$ and $z_1s_2' = s_2'$; $z_2s_i' = s_2'$ if $i \neq 1$ and $z_2s_1' = s_1'$;

$z_3s_i' = s_3'$ if $i \neq 4$ and $z_3s_4' = s_4'$; $z_4s_i' = s_4'$ if $i \neq 3$ and $z_4s_3' = s_3'$.

The printing rule is given by $\eta''(z_i ; s_j') = x_1x_1$ when $i = 2$ and $j \neq 1$; $= x_2x_2$ when $i = 4$ and $j \neq 3$; $= y_3$ when $i = 1$ and $j = 0, 2, 4$ or $i = 3$

and $j = 0, 1, 2$; $= e_Y$ (nothing) in all other cases. For example,

$$\eta''(\eta' f) = e_Y e_Y y_3 e_Y x_2 x_2 y_3 e_Y e_Y y_3 e_Y x_1 x_1 e_Y x_1 x_1 ,$$

that is, $\eta(s_0 ; f; s_0')$.

II. FORMAL DEFINITION AND NERODE'S THEOREM

A transduction η is given by the following structures:

1. A finite input alphabet $X = \{x\}$ and an output alphabet $Y = \{y\}$.
2. Two finite sets of states $S = \{s\}$ and $S' = \{s'\}$.
3. Two mappings $(S, X) \to S$ and $(X, S') \to S'$ written respectively sx and xs'.
4. A mapping η: $(S, X, S') \to F_Y$ written $\eta(s; x; s')$. These mappings are extended in a natural fashion to any $f \varepsilon F_X$ by the following inductive rules:

$se_X = s$ and $e_X s' = s'$, $\eta(s; e_X ; s') = e_Y$ for any $(s, s') \varepsilon (S, S')$.

For any $f \varepsilon F_X$, $x \varepsilon X$, $(s, s') \varepsilon (S, S')$: $s(fx) = (sf)x$, $(fx)s' = f(xs')$, $\eta(s; fx; s') = \eta(s; f; xs')\eta(sf; x; s')$.

It is easily checked that these rules are equivalent to the ones given in the introduction. By induction the last rules gives the following identity which could be taken as a definition and which displays η as a two-sided coset mapping $F_X \to F_Y$: for any $f_1 , f_2 , f_3 \varepsilon F_X$

$$\eta(s; f_1 f_2 f_3 ; s') = \eta(s; f_1 ; f_2 f_3 s')\eta(sf_1 ; f_2 ; f_3 s')\eta(sf_1 f_2 ; f_3 ; s').$$

In a more concrete manner η can be realized by finite matrices whose entries belong to the union of F_Y and of a zero, 0. Indeed for any $x \varepsilon X$ let μx be a square matrix whose rows and columns are indexed by the pairs $(s_i , s_{i'}') \varepsilon (S, S')$ and whose entries are

$$\mu x((s_i , s_{i'}'), (s_j , s_{j'}')) = \eta(s_i ; x; s_{j'}') \quad \text{if} \quad s_i x = s_j \quad \text{and} \quad s_{i'}' = x s_{j'}' ,$$
$$= 0, \quad \text{otherwise.}$$

Then if $f = x_1 x_2 \cdots x_n$ the corresponding output word $\eta(s; f; s')$ is equal to the entry $\mu f((s, fs'), (sf, s'))$ of $\mu f = \mu x_1 \mu x_2 \cdots \mu x_n$.

PROOF. For any $f \varepsilon F_X$ and $x \varepsilon X$ we have

$$\mu f x((s_i , s_{i'}'), (s_j , s_{j'}')) = \Sigma[\mu f((s_i , s_{i'}'), (s_k , s_{k'}'))][\mu x((s_k , s_{k'}'), (s_j , s_{j'}'))]$$

where the summation is over all the pairs $(s_k , s_{k'}') \varepsilon (S, S')$. The only

nonzero term in the sum is the one corresponding to the pair defined by the equations $s_k = s_i f$ and $s'_{k'} = x s'_{j'}$; we have then $s_j = s_k x$ and $s'_{i'} = f s'_{k'}$, that is, $s_j = s_i f x$ and $s'_{i'} = f x s'_{j'}$. Thus the entry under consideration is equal to $\eta(s_i\,;\, f;\, x s'_{j'})\,\eta(s_i f;\, x;\, s'_{j'})$, that is, to $\eta(s_i\,;\, fx;\, s'_{j'})$ and the result follows by induction.

Example. Let $X = \{a, b\}$; $Y = \{c, d\}$; $S = \{s_1, s_2\}$; $S' = \{t_1, t_2\}$; $s_1 a = s_2 a = s_2 b = s_1$; $s_1 b = s_2$; $b t_1 = b t_2 = a t_2 = t_1$; $a t_1 = t_2$. $\eta(s_i\,;\, x;\, t_j = cc$ if $x = a$ and $i = j$; $= d$ if $x = a$ and $i \neq j$ or if $x = b$ and $1 = i \neq j$; $= c$ if $x = b$ and $i = j = 2$; $= e_Y$ in all other cases.

Then, for instance, $\eta(s_1\,;\, bbab;\, t_1) = ccc$ according to the following self-explanatory scheme

$$\begin{matrix} s_1 & & s_2 & & s_1 & & s_1 & & s_2 \\ & b(e_Y) & & b(c) & & a(cc) & & b(e_Y) & \\ t_1 & & t_1 & & t_2 & & t_1 & & t_1 \end{matrix} .$$

Also we have

$$\mu a = \begin{pmatrix} 0 & d & 0 & 0 \\ cc & 0 & 0 & 0 \\ 0 & cc & 0 & 0 \\ d & 0 & 0 & 0 \end{pmatrix}; \quad \mu b = \begin{pmatrix} 0 & 0 & e_Y & d \\ 0 & 0 & 0 & 0 \\ e_Y & c & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad \mu bbab = \begin{pmatrix} 0 & 0 & ccc & cccd \\ 0 & 0 & 0 & 0 \\ 0 & 0 & dd & ddd \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and $\eta(s_1\,;\, bbab;\, t_1)$ is equal to $\mu bbab((s_1, t_1), (s_2, t_1))$.

As an immediate consequence of the definitions we derive the following weak form of Nerode's *ultimate periodicity theorem* (Nerode, 1958). *There exist finite integers m and n which are such that for any $f, f', f'' \in F_X$, $(s, s') \in (S, S')$, $p, r \geqq 0$, and $r \leqq n$ one has $\eta(s;\, f' f^{2m+pn+r} f'';\, s') = g' g^p g''$ where $g, g', g'' \in F_Y$ do not depend on p.*

Proof. Since S and S' are finite we can find integers m and n such that for all $(s, s') \in (S, S'), f \in F_X$, $p \geqq 0, 0 \leqq r \leqq n$, we have: $s f^{m+pn+r} = s f^{m+r}$, $f^{m+pn+r} s' = f^{m+r} s'$. Thus

$$\eta(s;\, f' f^{2m+pn+r} f'';\, s') = \eta(s;\, f' f^{m+r};\, f^{pn+m} f'' s')\,\eta(s f' f^{m+r};\, f^{pn};\, f^m f'' s')\,\eta(s f' f^{m+pn+r};\, f^m f'';\, s').$$

Because of our choice of m and n the second factor is equal to p times the word $g = \eta(s f' f^{m+r};\, f^n;\, f^m f'' s')$ and the result is proved.

III. FINITE CLOSURE PROPERTIES

A. To any two transductions $\eta\colon (S; F_X\,;\, S') \to F_Y$ and $\xi\colon (T; F_Y\,;\, T') \to F_Z$ there corresponds a transduction $\pi\colon (R;\, F_X\,;\, R') \to F_Z$ which is

such that for any $f \varepsilon F_X$, $(s, s') \varepsilon (S, S')$, $(t, t') \varepsilon (T, T')$ one has identically $\xi(t; \eta(s; f; s'); t') = \pi(r; f; r')$ where the states $r \varepsilon R$ and $r' \varepsilon R'$ are functions of s and t and of s' and t' respectively.

PROOF. We define an equivalence relation σ on F_X by the following rules: $\sigma f = \sigma f'$ (to be read: the σ-class of f is the same as that of f') if and only if (1) for all $s \varepsilon S$, $sf = sf'$; (2) for all $(s, s') \varepsilon (S, S')$ and $t \varepsilon T$, $t\eta(s; f; s') = t\eta(s; f'; s')$.

The relation σ has at most $|S|^{|S|} \times |T|^{|T|\times|S|\times|S'|}$ ($|S|$ the number of states in S) distinct classes. Furthermore it is right regular (i.e., $\sigma f = \sigma f'$ implies $\sigma ff'' = \sigma f'f''$ for all f'') since when $\sigma f = \sigma f'$ we have (1) $sff'' = sf'f''$ for any $s \varepsilon S$ (because $sf = sf'$); (2) for any $(s, s') \varepsilon (S, S')$ and $t \varepsilon T$,

$$t\eta(s; ff''; s') = t\eta(s; f; f''s')\eta(sf; f''; s')$$
$$= t\eta(s; f'; f''s')\eta(sf'; f''; s') = t\eta(s; f'f''; s').$$

We now define R as the set of all triplets $r = (s, t, \sigma f)$ and the mapping $(R, X) \to R$ by $(s, t, \sigma f)x = (s, t, \sigma fx)$. In a perfectly symmetric manner we construct a left regular equivalence σ', a set of states $R' = \{r'\} = \{(s', t', \sigma' f)\}$, and a mapping $(X, R') \to R'$. Finally, we put

$$\pi((s, t, \sigma f); x; (s'; t'; \sigma' f')) = \xi(t\eta(s; f; xf's'); \eta(sf; x; f's'); \eta(sfx; f'; s')t').$$

This definition is free from ambiguity because the three expressions $\eta(\ ;\ ;\ )$ entering in it depend only upon the classes σf and $\sigma' f'$; this is a direct consequence of the definition of σ and σ' and it concludes the proof since it is sufficient now to check by developing the expressions that if $f = f'xf''$ we have $\xi(t; \eta(s; f; s'); t') = \pi(r; f; r')$ where $r = (s, t, \sigma e_X)$ and $r' = (s', t', \sigma' e_X)$. Before verifying the second closure properties we recall the following facts:

1. Let R_Z denote the family of the subsets $F' \subset F_Z$ that are regular events in the sense of Kleene (1956). The specification of an $F' \varepsilon R_Z$ is equivalent (cf. Shepherdson, 1959) to that of an homomorphism γ: $F_Z \to P$ where P is a finite monoid together with the subset P' of P associated to F' by the relations $\gamma F' = P'$; $F' = \gamma^{-1}P'$ $(= \{f : \gamma f \varepsilon P'\})$. The equivalence on F_Z defined by $\gamma f = \gamma f'$ is at the same time left and right regular and it has only finitely many classes.

2. According to D. Huffman's theory (1959) the transformation η can

be said to be *information lossless* on the subset F' of F_X if the equations $\eta(s; f; s') = \eta(s; f'; s'); sf = sf'; fs' = f's'; f, f' \varepsilon F'$ imply $f = f'$.

B. If η is information lossless on the subset $F' \varepsilon R_X$ there exists a transduction $\xi(= \eta^{-1})$ which is such that for any $f \varepsilon F'$, $(s, s') \varepsilon (S, S')$ we have $\xi(t; \eta(s; f; s'); t') = f$ where the states $t \varepsilon T$ and $t' \varepsilon T'$ are functions of s and fs' and of s' and sf respectively.

PROOF. Let H be the set of all the words $\eta(s; x; s')$ with $x \varepsilon X$ and $K(K')$ the set of all proper right (left) factors of the words of H (i.e., $k \varepsilon K$ if and only if $kf \varepsilon H$ for some $f \neq e_Y$). If σ is a right regular equivalence on F_X with $|\sigma|$ classes, we say that $g \varepsilon F_Y$ admits a factorization of type $(\sigma f, s_i, s'_{j'}, s_j, s'_{i'}, k)$ (where σf is any σ-class; $s_i, s_j \varepsilon S$; $s'_{i'}, s'_{j'} \varepsilon S'$; $k \varepsilon K$) if there exist $f' \varepsilon F_X$ and $g' \varepsilon F_Y$ such that the following relations are satisfied:

$$\sigma f = \sigma f'; \qquad g = g'k; \qquad g' = \eta(s_i; f'; s'_{i'}); \qquad s_i f' = s_j; \qquad f' s'_{i'} = s'_{j'}.$$

Clearly, if $|h|$ is the maximal length of an element of H, there exist at most $|\sigma| \times (|S| \times |S'|)^2 \times |h|$ different types of factorization. Thus if we write $\lambda g_1 = \lambda g_2$ when the elements $g_1, g_2 \varepsilon F_Y$ admit exactly the same set of types of factorization, the relation λ has only a finite number of classes when the same is true of σ and, by construction, λ is right regular. In perfectly symmetric manner we associate a left regular equivalence λ' on F_Y to any left regular σ' on F_X.

We now come to the construction of ξ. As indicated above we construct the relations λ and λ' on F_Y associated with the (left and right regular) relation γ on F_X used for the definition of F', and we define T as the set of all triples $(s, s', \lambda g)$ with $(s, s') \varepsilon (S, S')$ and λg a λ-class; the mapping $(T, Y) \to T$ is given by $(s, s', \lambda g)y = (s, s', \lambda gy)$. The set of states T' and the mapping $(Y, T') \to T'$ are defined in symmetric manner with the help of the relation λ'.

For each triple $(s, s', x \varepsilon X)$ such that $\eta(s; x; s') \neq e_Y$ we select arbitrarily one factorization kyk' of $\eta(s; x; s')$ and we define ξ by the following rules: $\xi((s_1, s_4', \lambda g); y; (s_4, s_1', \lambda' g')) = x$ if there exists $\bar{g}, \bar{g}' \varepsilon F_Y$; $k \varepsilon K$; $k' \varepsilon K'$; $f, f' \varepsilon F_X$; $s_2, s_3 \varepsilon S$; $s_2', s_3' \varepsilon S'$ satisfying the following relations:

$$\lambda g = \lambda \bar{g} k; \lambda' g' = \lambda' k' \bar{g}';$$

$$\bar{g} = \eta(s_1; f; s_3'); s_1 f = s_2; f s_3' = s_4'; s_2 x = s_3;$$

$$\bar{g}' = \eta(s_3; f'; s_1'); s_3 f = s_4; f' s_1' = s_2'; x s_2' = s_3';$$

kyk' is the selected factorization of $\eta(s_2\ ;\ x;\ s_3')$;

fxf' belongs to F'.

In all other cases the value of $\xi(\quad)$ above is e_Y.

The possibility of solving all except the last of the above equations for given $y \varepsilon Y$, λg, $\lambda' g'$, s_1 , s_4 , s_1', s_4' is a direct consequence of the definition of λ and λ'. Taken together these equations imply that there exists at least one triple f, x, $f' \varepsilon F_X$ for which $\eta(s_1\ ;\ fxf';\ s_1') = \bar{g}kyk'\bar{g}' = g''$ with a selected factorization; $s_1 fxf' = s_4$ and $fxf's_1' = s_4'$; $fxf' \varepsilon F'$.

Thus, if the word $g'' = \bar{\bar{g}}kyk'\bar{\bar{g}}'$ has been obtained from a word f'' in F' by a transduction with the indicated initial and final states, it follows from Section III, A that $\xi(t;\ g'';\ t')$ will be identical to f'' and, because of the hypothesis that η is information lossless on F', this proves a posteriori that the above equations have a unique solution.

REMARK.

Because of the assumption that S' is finite it is always possible to realize in a single pass any arbitrary transduction if one is allowed to use a bounded number of output tapes and if one has the possibility of erasing on them.

Since the general case is rather cumbersome it may be sufficient to restrict ourselves to the detailed examination of the procedure needed for the deciphering of an encoding. Thus, let us assume now that S and S' are reduced to a single element and that consequently η is an isomorphism $F_X \to F_Y$. The sets H and K have the same meaning as in Section III, B and $P = \eta F_X$ is the submonoid of F_Y generated by H.

To any $g \varepsilon F_Y$ we associate the set λg of those $k \varepsilon K$ which are such that $g = pk$ for some $p \varepsilon P$; λg contains at most $|\ h\ |$ elements and, consequently, the equivalence relation on F_Y defined by $\lambda g = \lambda g'$ has only finitely many classes; since, furthermore, it is right regular we can construct a conventional automaton whose states are identified with the various possible λg's and whose transitions are given by $(\lambda g)y = \lambda(gy)$. We still observe that for any $g \varepsilon F_Y$ either λg is empty (and in this case g cannot be a left factor of a word in P) or, if $k \varepsilon \lambda g$ there exists a uniquely determined element $f = \xi p \varepsilon F_X$ such that $g = (\eta f)k = pk$.

Let us now consider a word $g'' \varepsilon P$ and any factorization $g'' = gyg'$ of it; let us assume also that we have been able to record on $|\ \lambda g\ |$ tapes the words ξp_i corresponding to the $|\ \lambda g\ |$ elements $k_i \varepsilon \lambda g$. The automaton is in state λg and upon reading the letter y it will go to the state $\lambda(gy)$.

For each $k_i \varepsilon \lambda g$ four cases are possible and we list below the printing instructions to be followed in each of them:

1. $k_i y$ does not belong to H nor to K (i.e., $k_i y$ cannot be a left factor of a word of P); then the machine erases the corresponding word ξp_i.

2. $k_i y$ belongs to H and not to K; the machine writes on the corresponding tape the letter $x \varepsilon X$ such that $\eta x = k_i y$; thus, on this tape we now have $(\xi p_i)x$.

3. $k_i y$ belongs to K and not to H; the machine does nothing on the corresponding tape.

4. $k_i y$ belongs to H and to K; the machine does as in 2 above but also it takes a new tape and it reproduces on it the word ξp_i. This new tape corresponds to the element $k_i y \varepsilon \lambda(gy)$ and the old tape corresponds to the element $e_Y \varepsilon \lambda(gy)$.

At the end of the reading of g'', $\lambda g''$ contains e_Y because, by hypothesis, $g'' \varepsilon P$ and the corresponding tape carries the word $\xi g''$ such that $\eta(\xi g'') = g''$. It is clear that, at any given stage of the procedure $| h |$ tapes, at most, are needed since we can use the tapes made free by the operation 1 above.

The proof of the validity of the algorithm is left to the reader and in Tables I and II we give a complete account of the construction of the state diagram and of the deciphering of the word $a^4ba^4b^2a^2$ for the following example:

$$X = \{x_i\}(1 \leqq i \leqq 5); \qquad Y = \{a, b\};$$

$$\eta x_1 = aa; \quad \eta x_2 = baa; \quad \eta x_3 = bb; \qquad \eta x_4 = ba; \quad \eta x_5 = bb;$$

$$K = \{e_Y, a, b, ba, bb\}.$$

(This is an encoding because it is a left prefix code in the three words: $u = a; v = ba; w = bb$) (Schützenberger, 1956). We find $\xi(a^4ba^4b^2a^4) = x_1^2x_2x_1x_3x_1$.

IV. RELATIONSHIP WITH REGULAR EVENTS

As we shall deal here with fixed initial states, we write for any subset $F'(G')$ of $F_X(F_Y)$:

$$\eta F' = \{g \varepsilon F_Y : g = \eta(s_1 ; f; s_1'), f \varepsilon F'\};$$

$$\eta^{-1}G' = \{f \varepsilon F_X : \eta(s_1 ; f; s_1') \varepsilon G'\}.$$

A. The subset G' of ηF_X belongs to R_Y if and only if $\eta^{-1}G'$ belongs to R_X.

TABLE I

K	e_Y	a	b	ba	bb	Corresponding states and transitions
e_Y	+					t_1
a		+				$t_2 = t_1a$
b			+			$t_3 = t_1b$
aa	+					$t_1 = t_2a$
ab						$t_0 = t_1b = t_0a = t_0b$
ba				+		$t_4 = t_3a$
bb	+				+	$t_5 = t_3b$
baa	+	+				$t_6 = t_4a$
bab			+			$t_3 = t_4b$
bba	+	+				$t_6 = t_5a$
bbb			+			$t_3 = t_5b$
$baaa$	+	+				$t_6 = t_6a$
$baab$			+			$t_3 = t_6b$

Proof. By definition there corresponds to every $F' \varepsilon R_X$ a right regular equivalence γ with finitely many classes such that F' is a union of γ-classes; in the proof of Section III, B we have seen how to construct λ associated to γ and such that $\eta F'$ is a union of λ-classes; since λ is right regular and has only finitely many classes, this proves the forward implication. In particular, since F_X belongs to R_X, this shows that the total output ηF_X is a regular event.

Now let G' be a subset of ηF_X that belongs to R_Y; G' is defined by a certain right regular relation λ with finitely many classes and we construct the relation σ on F_X by the following conditions:

$\sigma f = \sigma f'$ if and only if (1) $s_1 f = s_1 f'$; (2) for any $s' \varepsilon S'$, $\lambda\eta(s_1 ; f; s') = \lambda\eta(s_1 ; f'; s')$. σ is right regular because, if $\sigma f = \sigma f'$, we have $s_1 f f'' = s_1 f' f''$ and $\lambda\eta(s_1 ; ff''; s') = (\lambda\eta(s_1 ; f; f''s'))\eta(s_1 f; f''; s') = (\lambda\eta(s_1 ; f'; f''s'))\eta(s_1 f'; f''; s') = \lambda\eta(s_1 ; f'f''; s')$ where the second and third equality result from the right regularity of λ and where the second equality is a consequence of $\sigma f = \sigma f'$. Also, σ has at most $|S| \times |\lambda|^{|s'|}$ classes and $\eta^{-1}G'$ is a union of σ-classes. This concludes the proof.

B. Provided that X contains two letters or more, there corresponds to each $G'' \varepsilon R_Y$ a right transduction η such that $G'' = \eta F_X$.

Proof. Because of our hypothesis on X it is sufficient to prove the same statement for an arbitrarily large (finite) input alphabet and then to perform a preliminary encoding.

The result is trivial if G'' is finite and, by Kleene's theory, it is sufficient to show that if G and G' are the total outputs respectively of the right transductions η and η' (with the disjoint input alphabets X and X') we can construct right transductions η_1, η_2, η_3 (with input alphabet $X \cup X'$) such that their total output is respectively $G \cup G'$, GG', and G^* in Kleene's notation. The construction given below is the simplest to describe.

Let S and S' be the set of states of the right transducers η and η'; we can assume that S and S' are disjoint and we define S'' as the union of S, S' and of two new states s_1^* and s_0^* for which we have:

TABLE II

Input word	States	Tapes T1	Tapes T2	Tapes T3	Instructions
	t_1				
a	t_2				
a	t_1				$x_1 \to T1$
a	t_2				
a	t_1				$x_1 \to T1$
b	t_3				
a	t_4				$T1 \to T2$; $x_4 \to T1$
a	t_6				$x_2 \to T2$
a	t_6				$x_1 \to T1$
a	t_6				$x_1 \to T2$
b	t_3				$0 \to T1$
b	t_5				$T2 \to T3$; $x_5 \to T2$
a	t_6				$x_3 \to T3$
a	t_6				$x_1 \to T2$

1. $s_1{}^*x'' = s_1x''$ or $= s_1'x''$ and $\eta_i(s_1{}^*; x'') = \eta(s_1\,; x'')$ or $= \eta'(s_1'; x'')$ according to $x'' \varepsilon X$ or $\varepsilon X'$.

2. $s_0{}^*x'' = s_0{}^*$ for all x'' and $\eta_i(s''; x'') = e_Y$ for all s'', x'' such that $s''x'' = s_0{}^*$.

3. $s''x''$ and $\eta_i(s''; x'')$ are the same as in the original transducers when $s'' \varepsilon S$ and $x'' \varepsilon X$ or when $s'' \varepsilon S'$ and $x'' \varepsilon X'$.

4. η_1: $s''x'' = s_0{}^*$ when $s'' \varepsilon S$ and $x'' \varepsilon X'$ or when $s'' \varepsilon S$ and $x'' \varepsilon X$. For η_2: $s''x'' = s_1x''$ and $\eta_2(s''; x'') = \eta_2(s_1{}^*; x'')$ when $s'' \varepsilon S'$ and $x'' \varepsilon X'$ and $s''x'' = s_0{}^*$ when $s'' \varepsilon S'$ and $x'' \varepsilon X$. For η_3, we take $G = G'$ (and S identical to S') and we define $s''x'' = s_1{}^*$ and $\eta_3(s''; x'') = \eta_3(s_1{}^*; x'')$ when $s'' \varepsilon S$ and $x \varepsilon X'$ or when $s'' \varepsilon S'$ and $x'' \varepsilon X$.

The verification is left to the reader.

RECEIVED April 3, 1960

REFERENCES

CHEVALLEY, C. "Fundamental Concepts of Algebra." Academic Press, New York, 1956.

CHOMSKY, N. (1959). *Information and Control* **2,** 137.

GILBERT, E. N., AND MOORE, E. F. (1959). *Bell System Tech. J.* **38,** 933.

HUFFMAN, D. (1959). *IRE Trans. on Inform. Theory* **IT-5,** 41.

KLEENE, S. C. (1956). "Automata Studies." Princeton Univ. Press, Princeton, New Jersey.

MOORE, E. F. (1956). "Automata Studies." Princeton Univ. Press, Princeton, New Jersey.

NERODE, A. (1958). *Proc. Am. Math. Soc.* **9,** 541.

RABIN, M., AND SCOTT, D. (1959). *I.B.M. Research J.* **3,** 114.

SCHÜTZENBERGER, M. P. (1956). *IRE Trans. on Inform. Theory* **IT-2,** 47.

SHEPHERDSON, J. C. (1959). *I.B.M. Research J.* **3,** 198.

Reprinted from INFORMATION AND CONTROL, Volume 4, Nos. 2-3, September 1961

INFORMATION AND CONTROL **4**, 245–270 (1961)

On the Definition of a Family of Automata*

M. P. SCHÜTZENBERGER

Faculté des Sciences (Poitiers)
and
University of North Carolina, Chapel Hill, N.C.

I. INTRODUCTION

In this note we discuss the definition of a family $\mathfrak{A}$ of automata derived from the family $\mathfrak{A}_0$ of the finite one-way one-tape automata (Rabin and Scott, 1959).

In loose terms, the automata from $\mathfrak{A}$ are among the machines characterized by the following restrictions:

(a) Their output consists in the acceptance (or rejection) of input words belonging to the set F of all words in the letters of a finite alphabet X.

(b) The automaton operates sequentially on the sucessive letters of the input word without the possibility of coming back on the previously read letters and, thus, all the information to be used in the further computations has to be stored in the internal memory.

(c) The unbounded part of the memory, V_N, is the finite dimensional vector space of the vectors with N integral coordinates; this part of the memory plays only a passive role and all the control of the automaton is performed by the finite part.

(d) Only elementary arithmetic operations are used and the amount of computation allowed for each input letter is bounded in terms of the total number of additions and subtractions.

(e) The rule by which it is decided to accept or reject a given input word is submitted to the same type of requirements and it involves only the storage of a finite amount of information.

Thus the family $\mathfrak{A}$ is a very elementary modification of $\mathfrak{A}_0$ and it is not

* This work has been done in part at the Department of Statistics of the University of North Carolina under contract number AF 49 (638)-213 of the United States Air Force.

claimed that it relates usefully to the Turing machines or to the algorithms used in actual computing practice. In a more formal manner we have

DEFINITION 1. An automaton $\alpha \in \mathfrak{A}$ is given by the following structures:

(1) An automaton $\alpha_0 \in \mathfrak{A}_0$ (the *finite part of* α), that is, a finite set of states Σ, a mapping $(\Sigma, X) \to \Sigma$, an initial state $\sigma_1 \in \Sigma$, a distinguished subset of Σ' of Σ.

(2) A finite integer N, an initial vector v_1 from V_N and for each state σ in Σ' a distinguished finite union V_σ' of homogeneous linear subspaces of V_N.

(3) For each pair (σ, x) in (Σ, X) a mapping $\eta: V_N \to V_N$ which is such that each of the coordinates v_j' of $\eta(v, \sigma, x)$ can be computed by a finite computing program independent of the vector v and involving only the following operations: reduction of an integer modulo a positive integer at most equal to a finite bound $K_1(\sigma, x, j)$, multiplication of an integer by an integer of absolute value at most equal to a finite bound $K_2(\sigma, x, j)$, addition and subtraction of two integers.

(4) For each input word $f = x_{i_1} x_{i_2} \cdots x_{i_n}$ the automaton computes recursively the sequence of states $\sigma_{i_0}, \sigma_{i_1}, \sigma_{i_2}, \cdots, \sigma_{i_n} = \sigma_1 f$ and the sequence of vectors $v_{i_0}, v_{i_1}, v_{i_2}, \cdots, v_{i_n} = v(f)$ by the rules

$$\sigma_{i_0} = \sigma_1 \quad \text{and} \quad \sigma_{i_m} = (\sigma_{i_{m-1}}, x_{i_m})$$

$$v_{i_0} = v_1 \quad \text{and} \quad v_{i_m} = \eta(v_{i_{m-1}}, \sigma_{i_{m-1}}, x_{i_m}).$$

(5) The input word f belongs to the set F_α of the words accepted by α if and only if $\sigma_1 f \in \Sigma'$ and, then, if the vector $v(f) = v_{i_n}$ *does not* belong to $V'_{\sigma_1 f}$.

As expected, this definition can be considerably simplified and in Section I we verify that it is equivalent to the following one:

DEFINITION 1′. An automaton $\alpha \in \mathfrak{A}$ is given by a (homomorphic) representation μ of the monoid F in the ring Z_N of the integral $N \times N$ matrices (N, finite) together with the rule

$$F\alpha = \{f \in F : \mu f_{1,N} \neq 0\}$$

where $\mu f_{1,N}$ denotes the $(1, N)$ entry of the matrix μf.

It follows that the theory of Kleene (1956) can be applied and in Section III we verify that the family R of all the sets F_α with $\alpha \in \mathfrak{A}$ has the following property

If F' and F'' belong to R the same is true of their intersection, union, set product $F'F''$ (i.e., of the set of all words $f = f'f''$ with $f' \in F'$ and $f'' \in F''$), and formal inverse F'^0 (i.e., of the infinite union of all the set products $F', F'F', F'F'F', \cdots, F'F' \cdots F', \cdots$).

However, because of the arbitrariness implied in the conditions (2) and (5) of Definition 1, it is not necessarily true that the complement $F - F'$ of an F' from R also belongs to R. This, together with some miscellaneous remarks of a negative character, is verified in Section II by way of counterexamples.

Furthermore, I am unable to formulate for the family $\mathfrak{A}$ the deep part of Kleene's theory, namely to characterize R starting from a reasonably simple subfamily of sets in terms of meaningful set theoretical operations.

In Section IV, the family $\mathsf{R}_0 = \{F_\alpha : \alpha \in \mathfrak{A}_0\}$ of the regular events is characterized in terms of our present notations and in the same section we apply some elementary remarks from the theory of *sequential machines* (Moore, 1956) or *transducers* (Huffman, 1959) in order to obtain a third definition of $\mathfrak{A}$.

I am most indebted to Professor D. Arden from M.I.T. for many discussions of the content matter of this paper which have greatly contributed to the development or to the clarification of several points.

A. Preliminary Reduction

We shall say that the automaton α from $\mathfrak{A}$ is *semi-reduced* if there exists a collection of finite integral matrices $\mu(\sigma, x)$ which are such that the vector $\eta(v, \sigma, x)$ is simply the product $v\mu(\sigma, x)$.

I.A.1. To any $\alpha \in \mathfrak{A}$ there corresponds one equivalent semireduced $\alpha' \in \mathfrak{A}$ which is such that for every input word f the vector $v(f)$ is a projection on a subspace of the vector $v'(f)$ of the automaton α'.

Proof. Let us consider a fixed triple (σ, x, j) and write in explicit form the computing program giving the jth coordinate of the vector $\eta(v, \sigma, x)$.

Since this program is assumed to be finite there exists a finite natural number $M[=M(\sigma, x, j)]$ and a set of M quadruples of integers $(i, i_1(i), i_2(i), o(i))$ satisfying the conditions: (a) $1 \leqq i \leqq M$; (b) for all values of i, $i_1(i)$ and $i_2(i)$ are nonegative numbers at most equal to $N + i$; (c) $o(i) = 1, 2, 3$ or 4.

We now define a sequence of $1 + N + M$ numbers $a(i, v)$ by the following conditions: (a) $a(0, v) = 1$; (b) $a(i, v)$ is the ith coordinate

of v when $1 \leqq i \leqq N$; (c) $a(N + M, v)$ is the jth coordinate of $\eta(v, \sigma, x)$ which we try to compute; (d) for each $i(1 \leqq i \leqq M)$, $a(N + i, v)$ is the value of $a(i_1(i), v)$, reduced modulo $a(i_2(i), v)$, when $o(i) = 1$ and, when $o(i) = 2$, 3 or 4, it is respectively the value of the product, the sum, or the difference of $a(i_1(i), v)$ and $a(i_2(i), v)$. Here, as usual, by the value of a reduced modulo b we mean the smallest nonnegative integer which is congruent to a modulo b.

At the cost of some increase in the length of this program we can assume that only multiplications by bounded nonegative factors are used. Indeed, since by hypothesis $| a(i_2(i), v) | \leqq K_2(\sigma, x, j)$ when $o(i) = 2$, we can always replace this line of the program by a subroutine which consists in a multiplication by the nonnegative number $K_2(\sigma, x, j) + a(i_2(i), v)$ followed by $K_2(i, x, j)$ subtractions of the multiplicand.

Also, since there exist only finitely many triples (σ, x, j) we can take a fixed finite constant K which is larger than 2 and larger than any of the numbers $2K_1(\sigma, x, j)$ and $2K_2(\sigma, x, j)$. We shall always denote by $\bar{a}(i, v)$ the value of $a(i, v)$ reduced modulo $K!$.

Let W_σ be the set of the vectors v which can occur when the finite part of the automaton is in state σ, i.e., the set of all v which are such that $v = v(f)$ for at least one input word f satisfying $\sigma_1 f = \sigma$. Let I' be the set of the addresses $i(0 \leqq i \leqq N + M)$ which are such that for every v in W_σ, the value of $a(i, v)$ is nonnegative and at most equal to K.

Because of the hypothesis and of our convention that only multiplications by nonnegative factors are allowed we have:

If $o(i) = 1$, then $i \in I'$.

If $o(i) = 1$ or 2, then $i_2(i) \in I'$.

Let us denote by $\bar{v}$ the vector whose coordinates are those of v reduced modulo $K!$ and verify the following statement: If v, $v' \in W_\sigma$ and $\bar{v} = \bar{v}'$ then, for all i, $\bar{a}(i, v) = \bar{a}(i, v')$, and for all $i \in I'$, $a(i, v) = a(i, v')$.

Indeed, because of our choice of K and I', $a(i, v)$ is always equal to $\bar{a}(i, v)$ when $v \in W$ and $i \in I'$. Thus, since the statement is true by hypothesis when $i \leqq N$, we can apply induction and it is an elementary consequence of the properties of the congruences that when $\bar{a}(i_1(i), v) = \bar{a}(i_1(i), v')$ and $\bar{a}(i_2(i), v) = \bar{a}(i_2(i), v')$ we also have $\bar{a}(i, v) = \bar{a}(i, v')$ for each of the four possible cases $o(i) = 1, 2, 3$, or 4.

Consequently, any $\bar{a}(i, v)$, and in particular $\bar{a}(N + M, v)$, depends

only upon $\bar{v}$. But, the set $\bar{V}_N$ of these reduced vectors contains only $(K!)^N$ distinct elements and for all practical purposes here it may be considered as an abstract finite set of states. Thus, we can replace α by an automaton α' for which the finite part α_0' has the union of $\bar{V}_N$ and Σ as a set of states. Then, in α', the computing program for any triple (σ', x, j) admits the following simplifications: $o(i)$ is never 1; when $o(i) = 2$, the instruction consists in the multiplication of $a(i_1(i), v)$ by a factor which does not depend upon the vector v but only upon i, j, x and the state σ' in which is the finite part of α'.

By a simple induction, it follows that we can find integers $c_{j'}(\sigma, x, j)$ $(0 \leqq j' \leqq N)$ which are such that the jth coordinate of $\eta(v, \sigma, x)$ is the linear function

$$c_0(\sigma, x, j) + \sum_{1 \leqq j' \leqq N} v_{j'} c_{j'}(\sigma, x, j)$$

of the coordinates $v_{j'}$ of v; since, at the cost of increasing N by one unit we can always have a coordinate v_0 which is identically 1 for all f and, since, consequently we can make the above relations homogeneous the result is entirely proved.

Let us recall that a representation of the monoid F in the ring Z_N of the integral $N \times N$ matrices is a mapping $\mu: F \to Z_N$ which is such that $\mu ff' = \mu f \, \mu f'$ for all $f, f' \in F$. For any matrix m, $\mathrm{Tr}(m)$ denotes the sum of the elements lying in the main diagonal of m.

I.A.2. To any $\alpha \in \mathfrak{A}$ there corresponds one representation μ' in $Z_{N'}$ and a finite set P of matrices from the same ring that are such that

$$F_\alpha = \{f \in F : \mathrm{Tr}(p\mu' f) \neq 0 \quad \textit{for all} \quad p \in P\}.$$

Proof. Let us assume that α is a semireduced automaton with M states in its finite part and, for each $x \in X$, let $\mu' x$ be the $(M \times N) \times (M \times N)$ matrix defined by

$$\mu' x_{ij, i'j'} = (\mu(\sigma, x))_{j, j'} \quad \text{if} \quad \sigma_i x = \sigma_{i'} \; ; = 0, \quad \text{otherwise.}$$

Assuming that σ_1 and v are respectively the initial state and the initial vector of α, we define the $M \times N$ vector v' by

$$v'_{ij} = v_j \quad \text{if} \quad i = 1; = 0 \quad \text{if} \quad i \neq 1.$$

It is easily verified that μ' is a representation of F in $Z_{N'}$ $(N' = M \times N)$ and that for any input word f one has $(v' \mu' f)_{ij} = (v(f))_j$ if $\sigma_1 f = \sigma_i$; $= 0$, otherwise.

Let us now revert to the condition (5) of Definition 1 and observe that it implies that for each $\sigma \in \Sigma'$ a finite collection $W(\sigma)$ of N-vectors is given together with the rule that $f \in F_\alpha$ if and only if $\sigma_1 f \in \Sigma'$ and $v(f)w \neq 0$ for all w in $W(\sigma_1 f)$. For each state $\sigma_{i'} \in \Sigma'$ and, then, for each vector w in $W(\sigma_{i'})$ let w' be the $M \times N$ vector defined by $w'_{ij} = w_j$ if $i = i'$; $= 0$ otherwise.

Because of the relations established above we have $(v'\mu' f)w' = 0$ when $\sigma_1 f \neq \sigma_{i'}$. Thus if W' denotes the set of all the vectors such as w' we have $f \in F_\alpha$ if and only if not all the products $(v'\mu' f)w'$ $(w' \in W')$ are zero. This practically ends the proof because if p is the $M \times N$ matrix defined by

$$p_{ij,i'j'} = (w'_{ij}) \times (v'_{i'j'})$$

the relation $(v'\mu' f)w' \neq 0$ is equivalent to $\mathrm{Tr}(p\mu' f) \neq 0$.

B. Equivalence of Definitions 1 and 1′

We recall that if $m \in Z_N$ and $m' \in Z_{N'}$, the kroneckerian product $m'' = m \otimes m'$ of m and m' is a matrix from $Z_{NN'}$ whose entries are defined by

$$m_{ii',jj'} = (m_{i,j}) \times (m'_{i',j'}).$$

Then, identically, for any $a, b \in Z_N$ and $a', b' \in Z_{N'}$ one has

$$(a \otimes a')(b \otimes b') = (ab) \otimes (a'b')$$

and

$$\mathrm{Tr}(a \otimes a') = \mathrm{Tr}(a)\ \mathrm{Tr}(a').$$

I.B.1. The Definitions 1 *and* 1′ *of the family* $\mathfrak{A}$ *are equivalent.*

Proof. On the one hand the statement is trivial because an automaton as defined by 1′ is a special case of an automaton as defined by 1; indeed, given a representation μ of F, we take as initial vector v_1 the first row of μe. For any input word f the vector $v_1\mu f$ is obtained by performing for each input letter a bounded number of additions and multiplications by bounded factors. Finally, f is accepted if and only if $v_1\mu f$ does not belong to the linear subspace of the vectors whose last coordinate is zero.

On the other hand the statement is also trivial. Because of I.A.1 and I.A.2. we may assume that α is in reduced form, i.e., that α is given by a representation $\mu : F \to Z_N$ together with a finite subset P of Z_N. For

each $f \in F$, let $\mu' f = (\mu f) \otimes (\mu f)$ $(\in Z_{N^2})$ and let $\bar{p}$ be the sum of the kroneckerian squares $p \otimes p$ over all $p \in P$. Because of the identities recalled above, μ' is a representation and for any $f \in F$, $\mathrm{Tr}(\bar{p}\mu' f)$ is the sum over all $p \in P$ of the square of $\mathrm{Tr}(p\mu f)$; thus, $\mathrm{Tr}(\bar{p}\mu' f) \neq 0$ if and only if $f \in F_\alpha$. It follows that without loss of generality we may reduce the verification of the statement to that of the following:

If μ is a representation of F in Z_N and p a matrix from the same ring there exists a representation μ' of F in Z_{N^2+2} which is such that for all f, $\mathrm{Tr}(p\mu f) = \mu' f_{1,N^2+2}$. Indeed, for each $f \in F$ let $\mu' f$ be the following $N' \times N'$ matrix $(N' = N^2 + 2)$:

(i) $\mu' f_{N',j} = \mu' f_{j,1} = 0$ for all $1 \leqq j \leqq N'$; (i.e., the last row and the first column of every $\mu' f$ are identically zero).

(ii) $\mu' f_{1,1+j+(k-1)N}$ for each pair (j, k) $(1 \leqq j, k \leqq N)$ is equal to the (j, k) entry of the matrix $p\mu f$; (i.e., for each k the subvector $\mu' f_{1,1+j+(k-1)N}$ (when $1 \leqq j \leqq N$) of the first row of $\mu' f$ is equal to the kth row vector of the matrix $p\mu f$).

(ii′) $\mu' f_{1+j+(k-1)N,N'}$ for each pair (j, k) $(1 \leqq j, k \leqq N)$ is equal to the (j, k) entry of μf.

(iii) $\mu' f_{1,N'}$ is equal to $\mathrm{Tr}(p\mu f)$.

(iv) The restriction of $\mu' f$ to the set of indices (i, j) strictly larger than 1 and strictly less than N' is the direct sum of N matrices identical to the matrix μf.

The verification that μ' is a representation is a straightforward computation and the result is proved because of the condition (iii).

As a simple consequence of these constructions we have:

I.B.2. The family R *of all* F_α $(\alpha \in \mathfrak{A})$ *is closed under finite intersections and unions.*

Proof. Let F' and F'' be defined respectively by $\mu' : F \to Z_{N'}$ and $\mu'' : F \to Z_{N''}$. If for every f we define $\mu_i f$ as the kroneckerian product $(\mu' f) \otimes (\mu'' f)$ we have $\mu_i f_{1,N'N''} \neq 0$ if and only if both $\mu' f_{1,N'}$ and $\mu'' f_{1,N''}$ are different from zero; thus μ_i defines the intersection of F' and F''.

If for every f we define $\mu_u f$ as the direct sum of the kroneckerian squares of $\mu' f$ and $\mu'' f$, μ_u is still a representation and we can easily find a $(N'^2 + N''^2) \times (N'^2 + N''^2)$ matrix p which is such that for all f, $\mathrm{Tr}(p\mu_u f)$ is the sum of the square of $\mu' f_{1,N'}$ and $\mu'' f_{1,N''}$; thus, by our last reduction μ_u can be used for defining the union of F' and F''. D. Arden has pointed out to me that by using the kroneckerian product of $\mu' f$ and $\mu'' f$ one can obtain more economically the same result.

II. COUNTER EXAMPLES

II.1. If X has a single letter, R *reduces to* R_0 (= *the set of all regular events*).

PROOF. Let α be defined by the representation $\mu : F \to Z_N$ and consider the following integral power series in the variate t:

$$a(t) = a_0 + \sum_{n>0} t^n (\mu x^n)_{1,N} .$$

By definition, $F\alpha$ is the set of those words x^n which are such that $(\mu x^n)_{1,N} \neq 0$; however, as a function of t, $a(t)$ is the Taylor series of a rational function whose denominator is a factor of $\det(1 - t\,\mu x)$. Thus according to the theorem of Skolem (1934), there exists a finite set of finite integers m, p, d_1, d_2, $\cdots$, d_k which have the property that for any n larger than m, the coefficient of t^n in $a(t)$ (i.e., $(\mu x^n)_{1,N}$) is zero if and only if n is congruent modulo p to one of the d_j's. Consequently F_α reduces to a regular event when X has a single letter and there exist quite simple sets (as, e.g., the set of the words x^{n^2} where n runs over all integers) which do not belong to R. It can be observed that Skolem's theorem shows that for any $F_\alpha \in \mathsf{R}$ and $f \in F$, the intersection of F_α with the infinite set $f, f^2, f^3, \cdots, f^n, \cdots$ also reduces to a regular event.

II.2. When X has two letters or more there exists at least one $F_\alpha \in \mathsf{R}$ *which has the following properties:* F_α *does not belong to* R_0 ; *the complement* $F - F_\alpha$ *of* F_α *does not belong to* R.

PROOF. Let $X = \{x, y\}$; $\Sigma = \{\sigma_i\}$, $i = 1, 2, 3, 4, 5$, and $(\Sigma, X) \to \Sigma$ defined by

$$\sigma_1 x = \sigma_2 x = \sigma_2 ; \qquad \sigma_3 x = \sigma_4 x = \sigma_4 ; \qquad \sigma_5 x = \sigma_5$$

$$\sigma_1 y = \sigma_3 y = \sigma_4 y = \sigma_5 y = \sigma_5 ; \qquad \sigma_2 y = \sigma_3 .$$

Let $\mu(\sigma_i , x)$ be the following matrices

$$\mu(\sigma_1 , x) = \mu(\sigma_2 , x) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix};$$

$$\mu(\sigma_3 , x) = \mu(\sigma_4 , x) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix};$$

$\mu(\sigma_2 , y)$ = the identity matrix; $\mu(\sigma, x)$ = the zero matrix in all other cases.

The initial state is σ_1 and the initial vector $v = (1, 0)$. Σ' is Σ itself and f is accepted in all cases except if $\sigma_1 f = \sigma_4$ and then if the second coordinate of $v(f)$ is zero.

By looking at the diagram (Fig. 1) one easily sees that $\sigma_1 f = \sigma_4$ if and only if $f = x^{1+n}yx^{1+n'}$ and that, then, $v(f) = (1, n - n')$.

Thus F_α consists of all words except those which have the form $x^{1+n}yx^{1+n}$; and, according to a metamathematical proof of Calvin Elgot (1960), F_α is not a regular event.

Let us verify that $F' = F - F\alpha$ does not belong to R; indeed, let us assume that there exists a representation $\mu' : F \to Z_N$ which has the property that $\mu' f_{1,N} \neq 0$ if and only if f does not belong to F_α.

Since $\mu' x$ is a $N \times N$ matrix it satisfies an equation of degree at most N and for every pair (i, j) and matrix m from Z_N there exists a linear relationship between the (i, j) entries of the $N + 1$ matrices m, $m\mu x$, $m\mu x^2$, $\cdots$, $m\mu x^N$. Since, by hypothesis, for every finite n the $(1, N)$ entry of the matrix $\mu' x^{1+n} y x^{n'}$ $(= \mu' x^{1+n} y \mu' x^{n'})$ is zero for $0 \leqq n' \leqq n$ and different from zero for $n' = n + 1$, we have shown that N' must be at least equal to every finite integer n. Consequently, the representation μ' is an infinite representation, i.e., $F - F_\alpha$ does not belong to R. Some elementary properties of this type of sets have been described in Schützenberger (1959).

II.3. The family $\bar{\mathsf{R}} = \{F - F_\alpha : \alpha \in \mathfrak{A}\}$ *is closed under (finite) union and intersection but not under set multiplication.*

Proof. By definition, $\bar{F}_\alpha \in \bar{\mathsf{R}}$ if and only if there exists a representa-

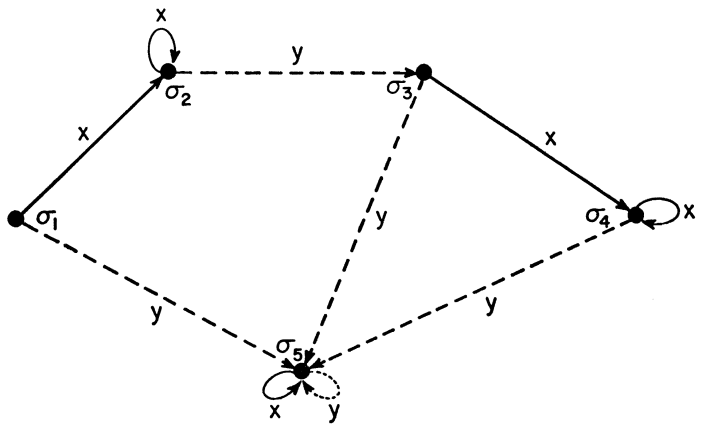


Fig. 1

tion $\mu: F \to Z_N$ (with N finite), that is, such that $\bar{F}_\alpha = \{f \in F : \mu f_{1,N} = 0\}$. Thus the closure properties of $\bar{\mathsf{R}}$ for the union and the intersection are a simple consequence of I.3.2.

Let $X = \{x, y\}$;

$$\mu x = \mu' x = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; \qquad \mu y = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}; \qquad \mu' y = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

The sets

$$\bar{F}_\alpha = \{f \in F : \mu f_{1,2} = 0\} = \{f \in F : |f|_x = |f|_y\}$$

and

$$\bar{F}_{\alpha'} = \{f \in F : \mu' f_{1,2} = 0\} = \{f \in F : |f|_x = 2\,|f|_y\}$$

(where $|f|_z$ denotes the number of times the letter z appears in f) both belong to $\bar{\mathsf{R}}$ and we shall verify that $F'' = \bar{F}_\alpha \bar{F}_{\alpha'}$ does not belong to $\bar{\mathsf{R}}$.

Let us define $W(f)$ as the set of the words f' which are such that $ff' \in F''$. We have (i) For all $f \in F$ and $f'' \in \bar{F}_\alpha$, $W(f)$ is contained in $W(f''f)$. Indeed, $ff' \in F''$ means that $ff' = f_1 f_2$ where $f_1 \in \bar{F}_\alpha$ and $f_2 \in \bar{F}_{\alpha'}$. Thus $f''ff' = f''f_1 f_2$ belongs to F'' since by hypothesis $f''f_1 \in \bar{F}_\alpha$. (ii) If both f and f'' belong to $\bar{F}_\alpha$, $W(f''f) = W(f)$ implies that $f'' = e$ (the empty word of F). Indeed, let $f''f \in \bar{F}_\alpha$, that is $|f''f|_x = |f''f|_y = k$, say. The product $f''fx^k$ satisfies the relations $|f''fx^k|_x = 2k$ and $|f''fx^k|_y = k$ and, consequently, it belongs to $\bar{F}_{\alpha'}$; thus, since $\bar{F}_{\alpha'}$ is a subset of F'' (because $e \in \bar{F}_\alpha$) the word x^k belongs to $W(f''f)$ and we shall show that it does not belong to $W(f)$.

Assume for the sake of contradiction that $fx^k = f_1 f_2$ with $f_1 \in \bar{F}_\alpha$ and $f_2 \in \bar{F}_{\alpha'}$; this implies that $f = f_1 f_3$ with $f_3 \in \bar{F}_\alpha$ and, consequently, $|f_3 x^k|_x = |f_3|_x + k = 2\,|f_3|_y$; $|f_3|_x = |f_3|_y$. It follows that $|f_3|_x = k$ and finally that $|f''f|_x = |f''f_1 f_3|_x = k$, i.e., $f'' = e$.

Thus, using (i) and (ii), we can find at least one strictly increasing infinite sequence of sets W, viz., $W(f)$, $W(f''f)$, $W(f''^2 f)$, $\cdots$, $W(f''^m f)$, $\cdots$.

Let us assume now that $F'' = \{f \in F : \mu'' f_{1,N''} = 0\}$; we observe that for given f the set $\bar{W}(f)$ of the vectors consisting of the N''th row of the matrices $\mu'' f'$ where $f' \in W(f)$ form a linear space whose dimension is at most N''. Since we can build an infinite strictly increasing sequence $\bar{W}(f''^m f)$ of such spaces it follows that N'' is infinite, i.e., that $\bar{F}_\alpha \bar{F}_{\alpha'} = F''$ does not belong to $\bar{\mathsf{R}}$.

II.4. If X contains two letters or more, there corresponds to any subset F' of F one automaton satisfying the conditions (1), (3), *and* (4) *of Definition* 1 *and having F' as its set of accepted words.*

PROOF. This is a trivial consequence of the existence of isomorphic, integral, finite dimensional representations of the monoid F.

Let first $X = \{x, y\}$;

$$\mu x = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; \qquad \mu y = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix};$$

and take $v = (1, 1)$ as initial vector. According to a theorem of Harrington (1951), the relation $v\mu f = v\mu f'$ implies $f = f'$; thus for any subset F' of F if the subset V' of V_2 is defined by $V' = \{v' = v\mu f : f \in F'\}$, we have reciprocally $F' = \{f \in F : v\mu f \in V'\}$. Clearly, this algorithm satisfies the conditions (1), (3), and (4) but not necessarily (2) and (5).

When X contains $n \geqq 3$ letters z_i, the same result subsists because we can associate to each z_i the matrix μf_i where $f_1 = x$, $f_n = y^{n-1}$ and $f_i = y^{i-1}x$ when $2 \leqq i \leqq n - 1$.

III. KLEENE'S THEOREM

Although this part could be written without explicitly using the notion of the ring $\bar{A}$ of the formal integral power series in the noncommutative variates $x \in X$, it seems more natural to do so and we recall here, without proofs, a few definitions and results on $\bar{A}$. These are very special and shallow cases of theorems used by many authors in the study of other problems. An especially valuable reference is Lazard (1955).

DEFINITION 2. $\bar{A}$ is the ring of all formal infinite sums

$$a = \sum_{f \in F} f(a, f)$$

with integral coefficients (a, f).

The addition and multiplication in $\bar{A}$ are defined respectively by:

$$a + a' = \sum_{f \in F} f((a, f) + (a'f)); \qquad aa' = \sum_{f \in F} f\Big(\sum_{f'f''=f} (a, f')(a', f'')\Big)$$

where, as always in this section, $\sum_{f'f''=f}$ means a summation over all factorizations $f = f'f''$ of f.

It may be easier to visualize any element of $\bar{A}$ as a generating function in which every word f has a (positive or negative) integral coefficient

(a, f). Thus, in particular, to each subset F' of F there corresponds the formal sum (its *characteristic function*)

$$\sum_{f\in F} f = \sum_{f\in F} f\chi_{F'}(f)$$

with $\chi_{F'}(f) = 1$ or 0 according to $f \in F'$ or not.

The multiplication is simply the ordinary multiplication of series using infinite distributivity, that is aa' can also be formally expressed as

$$\sum_{f\in F} f(a,f)a' = \sum_{f\in F} af(a',f) = \sum_{f,f'\in F} ff'(a,f)(a',f').$$

It may not be unnecessary to stress that this product is *not* the Hadamard product $\Sigma f(a,f)(a',f)$ to which we were led by the construction of the kroneckerian product of matrices in Section I.

We shall always denote the empty word by e and by $\bar{A}^*$ the subset of all $a \in \bar{A}$ in which (a, e), the coefficient of e, is zero. The elements of $\bar{A}^*$ are usually called quasi regular and we denote by a^* the mapping $\bar{A} \to \bar{A}^*$ defined by $a - (a, e)e$. If and only if a is quasi regular (i.e., $a = a^*$), it has a *quasi inverse* $a^0 = \sum_{n\geq 1} a^n$ which satisfies $aa^0 + a = a^0a + a = a^0$. In a perfectly equivalent manner an element $a \in \bar{A}$ has an inverse $a^{-1}(a^{-1}a = a^{-1}a = e)$ if and only if it belongs to the group $G \subset \bar{A}$ of the elements a' which are the sum of e and of the quasi regular element a'^*; then $a'^{-1} = e + (-a'^*)^0$ because

$$\begin{aligned} a'a'^{-1} &= (e + a'^*)(e + (-a'^*)^0) \\ &= e + a'^* + (-a'^*)^0 + a'^*(-a'^*)^0 \\ &= e + a'^* - a'^* = e. \end{aligned}$$

We shall find it more convenient to deal with the *quasi* notions because if a has nonnegative coefficients the same is true of $(a^*)^0$ but not necessarily of $(e + a^*)^{-1}$.

All the above operations are legitimate because $\bar{A}^*$ is a continuous topological algebra with continuous inverse when the distance between a and $a'(a \neq a')$ is defined as the supremum of the inverse of the length of those f for which $(a, f) \neq (a', f)$.

A. The Subring $\bar{R}$ of the Rational Elements

Definition 3. The subset $\bar{R}$ of $\bar{A}$ is the subset of the formal power series r which have the form $r = \sum_{f\in F} f\mu f_{1,N}$ for some representation μ of F in Z_N (N finite) which is said to *produce* r.

We shall verify that $\bar{R}$ is in fact the smallest subring of $\bar{A}$ that contains all the generators $x \in X$ of $\bar{A}$ and is such that its intersection with G is a subgroup; this last restriction is not trivial because, for an arbitrary subring $\bar{A}'$ of $\bar{A}$ it may well happen that some a belongs to the intersection of G and $\bar{A}'$ but that a^{-1} does not belong to $\bar{A}'$. If the variates x were commutative, $\bar{R}'$ would be the ring of the ordinary rational functions with integral coefficients and it seems natural to extend this terminology to the noncommutative case. A slightly different definition of $\bar{R}$ is given below.

III.A.1. $\bar{R}$ is a submodule of $\bar{A}$.

Proof. If

$$a = \sum_{f \in F} f(\mu f)_{1,N}, \qquad a' = \sum_{f \in F} f(\mu' f)_{1,N'}$$

we take the direct sum $\mu'' : F \to Z_{N+N'}$ of μ and μ' and we apply the remarks of I.3.1. for reducing to the desired form.

III.A.2. $\bar{R}$ is a subring.

Proof. We have to prove that if a is produced by μ and a' by μ' we can construct some μ'' which produces aa'. It will be simpler to prove the result under the additional assumption that $a, a' \in \bar{A}^*$ and to observe that the general case follows from III.A.1 because $aa' = a^*a'^* + (a, e)a'^* + a^*(a', e) + (a, e)(a', e)e$. We can also assume that μe and $\mu' e$ are the identity matrices of Z_N and $Z_{N'}$ respectively. After these preliminaries we proceed to the actual construction.

For each $x \in X$ we define $\mu'' x \in Z_{N+N'}$ as the matrix

$$\begin{pmatrix} \mu x & (\mu x)u \\ 0 & \mu' x \end{pmatrix}$$

where by $(\mu x)u$ we mean the $N \times N'$ matrix in which all columns are zero except for the first one which is equal to the Nth one of μx. Then, after taking $\mu'' e$ as the identity matrix of $Z_{N+N'}$, we extend μ'' to a representation $\mu'' : F \to Z_{N+N'}$ in the usual manner.

Because of our assumptions the following relations are surely true if $f = e$ or x:

$$\mu'' f_{1,i} = \mu f_{1,i} \text{ when } 1 \leqq i \leqq N;$$

$$\mu'' f_{1,N+i} = \sum_{f'f''=f} \mu f'_{1,N} \mu' f''_{1,i} \text{ when } 1 \leqq i \leqq N'$$

Let us verify now that if they hold for f they also hold for fx. Indeed we

have for $1 \leqq i \leqq N$:

$$\mu''fx_{1,i} = \sum_{1 \leqq j \leqq N+N'} \mu''f_{1,j}\mu''x_{j,i} = \sum_{1 \leqq j \leqq N} \mu f_{1,j}\mu x_{j,i} = \mu f x_{1,i}$$

and for $1 \leqq i' \leqq N'$:

$$\mu''fx_{1,N+i'} = \sum_{1 \leqq j \leqq N} \mu f_{1,j}((\mu x)u)_{j,i'} + \sum_{1 \leqq j \leqq N'} \mu''f_{1,N+j}\mu' x_{j,i'} .$$

The first sum is just $\mu f x_{1,N}$ when $i = 1$ and zero otherwise. By the induction hypothesis the second sum is

$$\sum_{f'f''=f} \sum_{1 \leqq j \leqq N} \mu f'_{1,N}\mu' f''_{1,j}\mu' x_{j,i'} = \sum_{f'f''=f} \mu f'_{1,N}\mu' f'' x_{1,i'} .$$

When $i' \neq 1$ this can also be written as

$$\sum_{g'g''=fx} \mu g'_{1,N}\mu' g''_{1,i'}$$

since $\mu' e_{1,i'} = 0$. On the contrary when $i' = 1$ we have

$$\mu''fx_{1,N+1} = \mu f x_{1,N} + \sum_{\substack{g'g''=fx \\ g'' \neq e}} \mu g'_{1,N}\mu' g''_{1,1} = \sum_{g'g''=fx} \mu g'_{1,N}\mu' g''_{1,1}$$

and the above relations are true for all cases. Since they imply that

$$\sum_{f \in F} f(\mu''f)_{1,N+N'} = \sum_{f \in F} f\Big(\sum_{f'f''=f} (\mu f')_{1,N}(\mu' f'')_{1,N'}\Big) = aa'$$

the result is proved.

III.A.3. $\bar{R}$ contains the quasi inverse of each of its quasi-regular elements.

Proof. As above we assume that μe is the identity matrix and we define $\bar{\mu} e$ as μe. For each $x \in X$, we take $\bar{\mu} x$ equal to the sum of μx and of a matrix $(\mu x)u \in Z_N$ which has all columns zero except for the first one which is equal to the Nth column of μx. For $f = e$ or x we have

$$\bar{\mu} f_{1,i} = \mu f_{1,i} + \sum_{f'f''=f} \bar{\mu} f'_{1,N}\mu f''_{1,i} .$$

As in the last proof above:

$$\mu f x_{1,i} = \sum_{1 \leqq i' \leqq N} \bar{\mu} f x_{1,i'}\bar{\mu} x_{i',i} + \sum_{f'f''=f} \sum_{1 \leqq i' \leqq N} \bar{\mu} f'_{1,N}\mu f''_{1,i'}\bar{\mu} x_{i',i} .$$

Thus: if $i = 1$,

$$\bar{\mu} f x_{1,1} = \mu f x_{1,1} + \mu f x_{1,N} + \sum_{f'f''=f} \bar{\mu} f'_{1,N}\mu f'' x_{1,1} + \sum_{f'f''=f} \bar{\mu} f'_{1,N}\mu f'' x_{1,N} .$$

If $i \neq 1$,

$$\mu f x_{1,i} = \mu f x_{1,i} + \sum_{g'g''=g} \bar{\mu} g'_{1,N}\mu g''_{1,i} .$$

Consequently, the initial relation is valid in all cases. Let us now compute $\bar{a}_i$. We have

$$\bar{a}_i = \sum_{f\in F} f\bar{\mu}f_{1,i} = \sum_{f\in F} f\mu f_{1,i} + \sum_{f',f''\in F} (f'\bar{\mu}f'_{1,N})(f''\mu f''_{1,i}).$$

In particular, for $i = N$, we have $\bar{a}_N = a + \bar{a}_N a$, that is, $e = (e + \bar{a}_N)(e - a)$ and, since a is assumed to be quasi regular, $a_N = (e - a)^{-1} - e = a^0$.

III.A.4. Reciprocally, any element $a = \sum_{f\in F} f(\mu f)_{1,N}$, *of* $\bar{R}$, *can be obtained from the generators* $x \in X$ *by a finite number of ring operations and formation of the quasi inverse (of quasi-regular elements).*

Proof. It is convenient to verify first the following statement: If s is a $N \times N$ matrix whose entries s_{ij} are N^2 distinct noncommutative variates, any entry of the quasi inverse u of s is a rational element with integral coefficients in the ring of the formal power series in the variates s_{ij}.

When $N = 1$, the statement is trivial because, then, u reduces to u_{11} which is equal to the quasi inverse of s_{11} ; when $N \geqq 2$ we shall use induction and base the verification upon the popular fact that any entry u_{ij} can be interpreted as the sum of all paths from i to j on the complete graph with vertices $1, 2, \cdots, N$. Since any such path can be decomposed in a unique manner described below with respect to the return of the vertex 1, the verification is a straightforward clerical operation.

Let us assume that the result is already proved for $N - 1$ and consider the $(N - 1) \times (N - 1)$ matrix t obtained from s by replacing by zero all the entries s_{1j} and $s_{j1} (1 \leqq j \leqq N)$ of s; by the induction hypothesis the quasi inverse v of t does exist and its entries $v_{ij} (2 \leqq i, j \leqq N)$ have the desired properties. We define a $N \times N$ matrix u by the following relations below and we shall verify later that it is the quasi inverse of s by showing that $us = u - s$.

$$u_{11} = (s_{11} + \sum_{2\leqq i\leqq N} s_{1i}s_{is} + \sum_{2\leqq i,j\leqq N} s_{1i}v_{ij}s_{j1})^0$$

If $i \neq 1$,

$$u_{1i} = \bar{u}_{1i} + u_{11}\bar{u}_{1i} ; \qquad u_{i1} = \bar{u}_{i1} + \bar{u}_{i1} + u_{i1}u_{11}$$

where, as an abbreviation,

$$\bar{u}_{1i} = s_{1i} + \sum_{2\leqq j\leqq N} s_{1j}v_{ji} ; \qquad \bar{u}_{i1} = s_{i1} + \sum_{2\leqq j\leqq N} v_{ij}s_{j1} .$$

If $i, j \neq 1$,

$$u_{ij} = v_{ij} + \bar{u}_{i1}(e + u_{11})\bar{u}_{1j} .$$

By the induction hypothesis, all the u_{ij}'s can be obtained from generators by the specified operations and we verify that $us = u - s$. We have to examine four cases:

Case 1.

$$(us)_{11} = u_{11}s_{11} + \sum_{2 \leqq j \leqq N} u_{1j}s_{j1} = u_{11}s_{11} + \sum_{2 \leqq j \leqq N} \bar{u}_{1j}s_{j1} + u_{11} \sum_{2 \leqq j \leqq N} \bar{u}_{1j}s_{j1}$$

$$= u_{11}s_{11} + (e + u_{11})\Big(\sum_{2 \leqq j \leqq N} s_{1j}s_{j1} + \sum_{2 \leqq j,j \leqq N} s_{1j'}v_{j'j}s_{j'1}\Big)$$

$$= u_{11}s_{11} + (e + u_{11})(e - s_{11} - (e + u_{i1})^{-1}) = u_{11} - s_{11}\,.$$

Case 2. If $2 \leqq i \leqq N$

$$(us)_{1i} = u_{11}s_{1i} + (e + u_{11}) \sum_{2 \leqq j \leqq N} \bar{u}_{1j}s_{ji}$$

$$= u_{11}s_{1i} + (e + u_{11})\Big(\sum_{2 \leqq j \leqq N} s_{1j}s_{ji} + \sum_{2 \leqq j,j' \leqq N} s_{1j'}v_{j'j}s_{ji}\Big)$$

$$= u_{11}s_{1i} + (e + u_{11})\Big(\sum_{2 \leqq j \leqq N} s_{ij}s_{ji} + \sum_{2 \leqq j' \leqq N} s_{1j'}(v_{j'i} - s_{j'i})\Big)$$

$$= u_{11}s_{1i} + (e + u_{11}(\bar{u}_{1i} - s_{1i})$$

$$= u_{1i} - s_{1i}\,.$$

Case 3. If $2 \leq i \leq N$

$$(us)_{i1} = u_{i1}s_{11} + \sum_{2 \leqq j \leqq N} u_{ij}s_{j1}$$

$$= \bar{u}_{i1}s_{11} + \bar{u}_{i1}u_{11}s_{11} + \sum_{2 \leqq j \leqq N} v_{ij}s_{j1} + \sum_{2 \leqq j \leqq N} \bar{u}_{i1}(e + u_{11})\bar{u}_{1j}s_{j1}$$

$$= \bar{u}_{i1}[s_{11} + u_{11}s_{11} + e + \sum_{2 \leqq j \leqq N} (e + u_{11})\bar{u}_{1j}s_{j1}] - s_{i1}$$

$$= \bar{u}_{i1}[e + (e + u_{11})(s_{11} + \sum_{2 \leqq j \leqq N} s_{1j}s_{j1} + \sum_{2 \leqq j,j \leqq N} s_{1j'}v_{j'j}s_{j1})] - s_{i1}$$

$$\bar{u}_{i1}[e + (e + u_{11})(e - (e + u_{11})^{-1})] - s_{i1}$$

$$= \bar{u}_{i1}(e + u_{11}) - s_{i1} = u_{i1} - s_{i1}\,.$$

Case 4. Finally, if $i, j \neq 1$

$$(us)_{ij} = u_{is}s_{ij} + \sum_{2 \leqq j' \leqq N} u_{ij'}s_{j'j}$$

$$= \bar{u}_{i1}s_{1j} + \bar{u}_{i1}u_{11}s_{ij} + v_{ij} - s_{ij} + \sum_{2 \leqq j' \leqq N} \bar{u}_{i1}(e + u_{11})u_{1j'}s_{j'j}$$

$$= v_{ij} - s_{ij} + \bar{u}_{1i}(e + u_{11})[s_{1j} + \sum_{2 \leqq j' \leqq N} \bar{u}_{1j'}s_{j'j}] = u_{ij} - s_{ij}$$

and the statement is verified.

We now revert to the proof of III.A.4 and we consider an element $r \in \bar{R}$ produced by the representation $\mu : F \to Z_N$. Without loss of generality we may assume that r is quasi regular and we consider the formal sum $s = \sum_{x \in X} x\mu x$, that is a $N \times N$ matrix whose entries s_{ij} are the elements $\sum_{x \in X} x\mu x_{ij}$ from $\bar{R}$. The sum s can be interpreted as a quasi-regular element of the ring of the power series in the variates $x \in X$ with coefficients in Z_N. Let us observe that for any two elements of this ring having the form $f\mu f$ and $f'\mu f'$ the product $(f\mu f)(f'\mu f')$ is equal to $ff'\mu ff'$. Consequently, the quasi inverse u of s is equal to the sum of $f\mu f$ extended of all the elements of F except e and the entry u_{ij} is the sum of $f\mu f_{ij}$ extended to the same set. Since we have seen previously that u_{ij} is a rational element in the entries of s, the same is true in particular of $u_{1N} = a$ and the result is verified.

As a point of marginal interest in the applications of probabilities to regular events, we consider the homomorphism (of ring) λ which sends every $r \in \bar{R}$ with finitely many nonzero coefficients onto the corresponding ordinary polynomial in the commutative variates $\bar{x} = \lambda x$; λ extends in a natural fashion to $\bar{R}$ and we have

III.A.5. For each $r \in \bar{R}$, λr is a power series, converging in some open domain around zero and representing there an ordinary rational function in the commutative variates $\bar{x} = \lambda x$.

PROOF. Let $r = \sum_{f \in F} f\mu f_{1,N}$. We consider the matrix $\sum_{x \in X} \bar{x}\mu x = s$ and the ordinary polynomial equal to $\det(I - s)$ in the commutative variates $\bar{x} = \lambda x$. For small enough ϵ, $\det(I - \lambda s)$ has its value arbitrarily close to 1 when all the $\bar{x}$ are less than ϵ. Under this condition the matrix $I + \sum_{n \geq 1} s^n = (I - s)^{-1} = [\det(I - s)]^{-1} \operatorname{Adj}(I - s)$ exists and its $(1, N)$ entry, that is, λr, is a rational function of the ordinary variates $\bar{x}$.

B. Reduction to Standard Form

In this section, we apply classical algebraic techniques to obtain a minimal representation producing a given element r of $\bar{R}$; at variance with the other parts of this paper we deal here with arbitrary (not necessarily integral) numbers.

III.B.1. To any element $r \in \bar{R}$, there corresponds a unique integer $\bar{N}$ and a representation $\bar{\mu}$ of F by $\bar{N} \times \bar{N}$ rational matrices that has the following properties:

(i) *There exists two sets T and S of $\bar{N}$ words each, a finite integer K and $\bar{N}^2$ matrices $\pi(t, s)$ from Z_N which are such that for all $f \in F$ the matrix $\bar{\mu}f$*

is identically equal to the sum $\Sigma K^{-1}(r, tfs)\pi(t, s)$ *extended to all the pairs* $(t, s) \in T \times S$.

(ii) *The representation* $\bar{\mu}$ *produces* r *in that sense that* $\bar{\mu}f_{1N} = (r, f)$ *if* r *is quasi regular and that* $\bar{\mu}f_{11} = (r, f)(r, e)^{-1}$ *if* $(r, e) \neq 0$ *(that is,* $\pi(e, e)$ *has a single nonzero entry).*

(iii) *If* μ *is any representation of* F *by* $N \times N$ *matrices that produces* r, *then* $N \geqq \bar{N}$ *and there exists a pair of matrices* $(\bar{u}, \bar{u}')$ *which is such that* $(\bar{u}\mu f\bar{u}')_{ij} = \bar{\mu}f_{ij}$, *if* $i \leqq i, j, \leqq N$ *and* $= 0$, *otherwise (i.e.* $\bar{\mu}$ *is a projection of* μ).

Proof. In the first steps of the proof we start from a given N-dimensional representation μ of F that produces r and we construct by iteration of the procedure described in 1 below the $\bar{N}$-dimensional representation $\bar{\mu}$ which has the properties (i), (ii), and (iii) with respect to μ; in the last step we verify that this representation does not depend upon μ but only upon r.

1. Let I be a fixed one to one mapping of F onto the natural numbers that satisfies the inequality $If \leqq Iff'$, for all words f and f', and let v be the vector equal to the first row of μe. We construct a set of words T' by the two following rules: (a) $I^{-1}1$ (that is, e) belongs to T'; (b) inductively, $I^{-1}j = f$ belongs to T' if and only if $f = f'x$ where $f' \in T'$ and $x \in X$ and if $v\mu f$ is (linearily) independent of the vectors $v\mu f''$ where $f'' \in T'$ and $If'' < If$.

By construction, T' contains $N' \leqq N$ elements and, without loss of generality, it may be assumed that $T' = \{f : If \leqq N'\}$ since $ff' \in T'$ implies $f \in T'$.

Let χf be the $N' \times N$ matrix whose jth row is the vector $v\mu f'f$ where $If' = j$; by construction $\chi f = \chi e\mu f$ identically.

Observe that for any $t \in T'$ and $x \in X$ either $tx \in T'$ or, else, the vector $v\mu tx$ is a linear combination of the vectors $v\mu t'$ $(t' \in T')$, that is, of the rows of χe; in other words, the matrix χx is equal to the product $\mu' x \chi e$ where $\mu' x$ is a certain $N' \times N'$ matrix. For any word $f = x_{i_1}x_{i_2} \cdots x_{i_n}$ we define $\mu' f$ as $\mu' x_{i_1}\mu' x_{i_2} \cdots \mu' x_{i_n}$ and we verify by induction that the representation μ, the associated representation μ', and the interwinning matrix χe are linked by the identity

$$\chi f = \chi e\mu f = \mu' f \chi e.$$

In fact, since the rank of χe is by construction equal to the number N' of its rows, there exists a pair (a, b) of nonsingular matrices which is such that $(a\chi e\, b)_{ij} = 1$ if $1 \leqq i = j \leqq N'$; $= 0$, otherwise.

The identity $(a\chi e\, b)(b^{-1}\mu f\, b) = (a\mu' f a^{-1})(a\chi e\, b)$ shows that

$$(b^{-1}\mu f\, b)_{ij} = (a\mu' f\, a^{-1})ij \quad \text{if} \quad 1 \leqq i, j \leqq N'$$
$$= 0 \quad \text{if} \quad 1 \leqq i \leqq N' \quad \text{and} \quad N' < j \leqq N.$$

Thus, there exists a pair (u, u') of matrices which are such that the restriction of $u\mu f u'$ to the indices less than N' is equal to $\mu' f$ and that any other entry of $u\mu f u'$ is zero, that is, μ' is a projection of μ.

Finally, we point out that the construction of μ' implies that any vector $v\mu f (f \in F)$ is a linear combination of the N' independent vectors $v\mu t (t \in T')$ and that consequently N' can be defined, without reference to I, as the rank of the vector space spanned by the vectors $v\mu f (f \in F)$.

2. Let $\tilde{I}$ be a one to one mapping of F onto the natural numbers that satisfies $\tilde{I}f \leqq \tilde{I}f'f$ for all words f and f', and let v' be the (column) vector equal to the Nth column of χe. It is clear that by replacing I by $\tilde{I}$ and by exchanging everywhere left and right multiplications we obtain a set S analogous to T' and that we can associate to the representation μ' a third representation $\tilde{\mu}$ of dimension $\tilde{N} \leqq N'$ and an interwinning matrix $\tilde{\chi}e$ that satisfies the identity:

$$\tilde{\chi}f = \mu' f \tilde{\chi}e = \tilde{\chi}e\tilde{\mu}f.$$

Again, reverting to I and taking a basic vector $\tilde{v}$ equal to the first row of $\tilde{\chi}e$, we can apply the same construction once more and obtain a set T, a representation $\bar{\tilde{\mu}}$ of dimension $\bar{N} \leqq \tilde{N}$ associated to $\tilde{\mu}$, and an intertwinning matrix $\bar{\chi}e$.

However, by definition, $(\tilde{\chi}f)_{ij} = (r, tfs)$ where $It = i$ and $\tilde{I}s = j$. Consequently T is a subset of T' and $\bar{\chi}f$ is obtained from $\tilde{\chi}f$ by deleting a certain subset of $N' - \bar{N}$ rows. Let us observe that the rank of $\tilde{\chi}e$ is equal to $\tilde{N}$, its number of columns and that, by construction, T is a set of words corresponding to a maximal set of independent rows of $\tilde{\chi}e$. Thus, $\bar{N} = \tilde{N}$ and we conclude that $\bar{\chi}e$ is a nonsingular matrix.

3. Let us consider the intertwinning identity

$$\bar{\chi}f = \bar{\chi}e\tilde{\mu}f = \bar{\tilde{\mu}}f\bar{\chi}e.$$

Since $\bar{\chi}e$ is nonsingular, we have identically $\tilde{\mu}f = (\bar{\chi}e)^{-1}\bar{\chi}f$ and $\tilde{\mu}$ has the property i of the statement.

Since the $(1, 1)$ entry of $\bar{\chi}f$ is exactly (r, f) we have $\mathrm{Tr}(q\tilde{\mu}f) = (r, f)$ where q is obtained from $\bar{\chi}e$ by replacing by zero every row except the first one; thus, depending upon $(r, e) = 0$ or not we can find a nonsingular matrix m which is such that mqm^{-1} is a matrix in which all the entries

are zero except for the $(1,1)$ entry or for the $(N,1)$ entry, and the representation $\bar{\mu}f = m\tilde{\mu}fm^{-1}$ has the properties (i) and (ii).

We have already seen that μ' is a projection of μ and, by the same argument, it is easily verified that $\bar{\mu}$ is a projection of μ', that is, finally, of μ.

4. Let us say that the set of words F' is (right) independent if the only linear relation

$$\sum_{f' \in F'} c_{f'}(r, ff') = 0$$

which is valid for all $f \in F$ is the trivial one in which all the coefficients $c_{f'}$ are zero.

It results instantly from the construction of $\tilde{\mu}$ and $\bar{\mu}$ that for given $r \in \bar{R}$ and $\tilde{I}$, the set S can be defined intrinsically as the maximal (right) independent set which is such that $f \in S$ implies that the set union of f and of the words $s \in S$ with $\tilde{I}s < \tilde{I}f$ is not (right) independent. In similar manner T can be defined intrinsically in terms of r and I only.

Consequently, if ν is any N''-dimensional representation of F, that is, such that $(r,f) = \nu f_{1N''}$ identically, we can apply to ν the construction described in 1 and 2 for μ and, although the first set T'' may be different from T', we are sure to obtain at the second and third steps the same sets S and T. Thus, $\tilde{\tilde{\mu}}$ is also a projection of ν and, consequently, $N'' \geqq \bar{N}$; in particular, if $N'' = \bar{N}$, this implies that $a\nu fa^{-1} = \bar{\mu}f$ identically for some nonsingular matrix a and this concludes the proof.

This, of course, does not preclude the possibility that $(r,f) = \mathrm{Tr}(p\nu' f)$ for some representation ν' of dimension less than $\bar{N}$ and matrix p or sufficient rank.

However, the complete discussion of this case, i.e., of the algebra associated to r would take us too far away from the strictly linear techniques used in this note and it will be given elsewhere.

Since we have not proved that the matrices $\bar{\mu}x (x \in X)$ are integral matrices, it may be worthwhile verifying that the following definition of $\bar{R}$ is equivalent to our previous one (cf. Fatou, 1904).

DEFINITION 3′. *An element $a \in \bar{A}$ (i.e., a formal sum with integral coefficients) belongs to $\bar{R}$ if and only if there exists a representation ν of F by arbitrary finite dimensional matrices which is such that $(a,f) = Tr(p\nu f)$ for some fixed matrix p.*

PROOF. By using the construction given in I.B.1. and then the construction of III.B.1, we may assume without loss of generality that, in fact, (a,f) is equal to the $(1,N)$ entry of the matrix $\bar{\mu}f$ divided by a constant factor. Consequently, because of the intertwining identity $\bar{\chi}f =$

$\bar{\chi}em^{-1}\bar{\mu}fm$, i.e., $\bar{\mu}f = m(\bar{\chi}e)^{-1}\bar{\chi}fm^{-1}$, each entry of μf is a rational fraction in which the denominator K is an integer independent of f.

Thus, if $v(f)$ is the vector equal to the first row of $\bar{\mu}f$, we can write $v(f) = v'(f) + K^{-1}v''(f)$ where $v'(f)$ has integral coordinates and where the bounded vector $v''(f)$ is equal to $Kv(f)$ reduced modulo K. It follows that for any $x \in X$ the $2N$-dimensional vector $(v'(fx),v''(fx))$ is entirely determined by x and the $2N$-dimensional vector $(v'(f),v''(f))$ in the sense of Definition 1 and this concludes the proof via the reduction procedure of Section I.

Incidentally, it shows that if all the coefficients (r, f) of some $r \in \bar{R}$ are divisible by K, the element $K^{-1}r$ also belongs to $\bar{R}$.

C. Applications to the Theory of Kleene

III.C.1. If F_α, $F_{\alpha'} \in \mathsf{R}$, *then* $F_\alpha F_{\alpha'} \in \mathsf{R}$.

Proof. Let $F_\alpha = \{f:\mu f_{1N} \neq 0\}$; $F_{\alpha'} = \{f:\mu' f_{1,N'} \neq 0\}$. We can assume that for all $f,\mu f_{1,N}$ and $\mu' f_{1N'}$ are both nonnegative (cf. Section I.B). Then, if

$$r = \sum_{f \in F} f(\mu f)_{1N} \quad \text{and} \quad r' = \sum_{f \in F} f\mu' f_{1N'},$$

we have

$$(rr', f) = \sum_{f'f'' \in F} (\mu f')_{1,N}(\mu' f'')_{1,N'}$$

that is, $(rr', f) \neq 0$ if and only if there exists at least one factorization $f = f'f''$ for which $\mu f'_{1,N} \neq 0$ and $\mu' f''_{1,N'} \neq 0$. Thus $F_\alpha F_{\alpha'} = \{f:(rr',f) \neq 0\}$. Since we know by III.A.2 how to construct $\mu'':F \to Z_{N+N'}$ such that $(rr',f) = \mu'' f_{1,N+N'}$ the result is proved.

III.C.2. If $F_\alpha \in \mathsf{R}$ *then* $F_\alpha{}^0 \in \mathsf{R}$.

Proof. Let $F_\alpha = f:(z, f) = \mu f_{1,N} \neq 0$ with $(z, f) \geqq 0$ for all f, as above. We can write $F_\alpha{}^0 = \{e\} \cup (F_\alpha - \{e\})^0$ and, consequently, we can assume that F_α does not contain e. Then, as in III.B.3, it is easily checked that $F_\alpha{}^0 = \{f:(r^0, f) \neq 0\}$ and the result follows from III.A.3.

There exists another type of invariance of the family of regular events which carries over to the family R; in order to describe it we still need the following definition.

Definition 4. A restricted right transducer η is given by the following structure:

(1) A finite automaton with a (finite) input alphabet Y, a (finite) set of states Σ and a mapping $(\Sigma,Y) \to \Sigma$.

(2) A mapping $\eta:(\Sigma, Y) \to F$ where F is the monoid generated by

the (finite) alphabet X; η is extended in a natural fashion to a mapping $(\Sigma, F_Y) \to F$ (where F_Y is the monoid generated by Y) by the following rules:

For any state $\sigma_j \in \Sigma$, $\eta_j e = e$ (e: the empty word).

For any state $\sigma_j \in \Sigma$ and input word $g = y_{i_1} y_{i_2} \cdots y_{i_n}$,

$$\eta_j g = \eta(\sigma_{i_0}, y_{i_1})\ \eta(\sigma_{i_1}, y_{i_2}) \cdots \eta(\sigma_{i_{n-1}}, y_{i_n})$$

where $\sigma_{i_0} = \sigma_j$ and, inductively, $\sigma_{i_m} = (\sigma_{i_{m-1}}, y_{i_m})$.

(3) The two mappings $(\Sigma, Y) \to \Sigma$ and η satisfy the condition that if the state σ_j is such that $\sigma_j g = \sigma_j$ and $\eta_j g = e$ for some $g \neq e$, then $\eta_j g' = e$ for all input words $g' \in F_Y$ (we say then that σ_j is a *sink*).

Given a restricted right transducer η and an initial state σ_1, we shall define an element $a \in \bar{A}$ (the sum *produced* by η) according to the following rule: For each $f \in F$, (a, f) is equal to the number of distinct words $g \in F_Y$ which are such that $\sigma_1 g$ is not a sink and that $\eta_1 g = f$.

III.C.3. If the subset F' of F_Y belongs to R_Y (defined for Y as R was defined for X) then, for any state σ_j of Σ, the set $F_\alpha = \eta_j F' = \{\eta_j g : g \in F'$; $\sigma_j g$ is not a sink$\}$ belongs to R.

Proof. The result is true if every state of Σ is a sink; let Σ' be the set of the states of Σ which are not a sink and assume that Σ' contains $M \geqq 1$ elements.

For each finite N we shall consider the ring $\bar{B}_N$ of the $N \times N$ matrices whose entries belong to the ring $\bar{A}$ of the formal power series in the letters from X. To any $y \in Y$ we associate the matrix νy from $\bar{B}_M$ with entries

$$\nu y_{jj'} = \eta(\sigma_j, y) \quad \text{if} \quad \sigma_j y = \sigma_{j'}\ ; = 0, \quad \text{otherwise.}$$

The matrices νg form a representation of F_Y in $\bar{B}_N$ and for any $g \in F_Y$ and $\sigma_j, \sigma_{j'} \in \Sigma'$ we have

$$\nu g_{jj'} = \eta_j g \quad \text{if} \quad \sigma_j g = \sigma_{j'}\ ; = 0, \quad \text{otherwise.}$$

Let us now assume that $F' = \{g \in F_Y : \mu g_{1N} \neq 0\}$ where μ is a representation of F_Y in Z_N; for the sake of simplicity we assume that $r = \sum_{g \in F_Y} g \mu g_{1N}$ is quasi regular. By applying the construction described in I.B.2. we can also assume that $\mu g_{1N} \geqq 0$ for all $g \in F_Y$. Finally, for any $y \in Y$, let $\bar{\mu} y$ denote the matrix from $\bar{B}_{N \times M}$ obtained by replacing in μy each entry $\mu y_{ii'}$ by a submatrix identical to $(\mu y_{ii'}) \nu y$. Again this gives us a representation of F_Y which has the property that for any $g \in F_Y$ and $\sigma_j, \sigma_{j'} \in \Sigma$ the $(1j, Nj')$ entry of $\bar{\mu} g$ is equal to $(\mu g_{1N})\ \eta_j g$ if $\sigma_j g = \sigma_{j'}$ and to 0 otherwise.

Because of the condition (3) and of the hypothesis that r is quasi regular, the matrix $s = \sum_{y \in Y} \bar{\mu} y$ is also quasi regular and the $(1j, Nj')$ entry of the quasi inverse u of s is equal to the sum $b_{jj'}$ of $(\mu g_{1N})\ \eta_j g$ extended to all the words g from F_Y which send σ_j onto $\sigma_{j'}$. According to the remark III.A.4 and to the fact that every entry of s belongs to $\bar{R}$, this sum is also an element of $\bar{R}$. Consequently the sum b_j of all $b_{jj'}$ (where $\sigma_{j'} \in \Sigma'$) also belongs to $\bar{R}$ and this proves the statement since $F_\alpha = \{f \in F : (b_j, f) \neq 0\}$ because of our hypothesis that $\mu g_{1,N}$ is always nonnegative.

IV. AN ELEMENTARY CHARACTERIZATION OF REGULAR EVENTS

We begin by verifying two remarks that are needed later.

DEFINITION 5. Let $\bar{R}^{\text{pos}}$ be the smallest subset (in fact, the smallest *semiring*) of $\bar{A}$ which satisfies the following conditions:

(i) $x \in \bar{R}^{\text{pos}}$ for any $x \in X$ and $e \in \bar{R}^{\text{pos}}$.

(ii) If $a, a' \in \bar{R}^{\text{pos}}$ then $a + a'$ and aa' also belong to $\bar{R}^{\text{pos}}$.

(iii) If $a \in \bar{R}^{\text{pos}}$, then $a^{*0} \in \bar{R}^{\text{pos}}$.

IV.1. A necessary and sufficient condition that $a \in \bar{R}^{\text{pos}}$ is that $a = \sum_{f \in F} f \mu f_{1,N}$ where $\mu : F \to Z_N^{\text{pos}}$ and where Z_N^{pos} denotes the subset (in fact, the semiring) of the integral $N \times N$ matrices with nonnegative entries.

PROOF. It is sufficient to revert to I.B, III.A.2, and III.A.3 and to observe that if a, a' are produced by representations into Z_N^{pos} the same is true of $a + a'$, aa' and a^0; also, trivially, μe and all the matrices μx belong to Z_N^{pos}. The construction performed in III.A.4. does not use subtraction either, and consequently $\Sigma f \mu f_{1,N} \in \bar{R}^{\text{pos}}$.

IV.2. $\bar{R}$ is the smallest submodule of $\bar{A}$ that contains $\bar{R}^{\text{pos}}$ and any $r \in \Re$ can be written under the form $r = r' - r''$ with $r', r'' \in \bar{R}^{\text{pos}}$.

PROOF. Since every $r \in \bar{R}$ can be obtained from the generators x by a finite number of additions, subtractions, multiplications, and formation of inverses it is sufficient to prove that if the result is true for $r_1, r_2 \in \bar{R}$ it is still true for $r_3 = r_1 + r_2$; $r_4 = r_1 - r_2$; $r_5 = r_1 r_2$ and $r_6 = (e - r_1^*)^{-1}$. Let us assume that $r_1 = r_1' - r_1''$; $r_2 = r_2' - r_2''$ where r_1', $r_2' r_1''$ and r_2'' belongs to $\bar{R}^{\text{reg}}$. We have:

$$r_3 = (r_1' + r_2') - (r_1'' + r_2''); \qquad r_4 = (r_1' + r_2'') - (r_1'' + r_2');$$

$$r_5 = (r_1' r_2' + r_1'' r_2'') - (r_1' r_2'' + r_1'' r_2')$$

where again all the elements between brackets belong to $\bar{R}^{\text{pos}}$ since this set is a semiring.

With respect to r_6 we observe first that $r_1^* = r_1'^* - r_1''^*$ and that $r \in \bar{R}^{\text{pos}}$ implies $(e - r^*)^{-1} - e = s \in \bar{R}^{\text{pos}}$ and, then, $r^* \in \bar{R}^{\text{pos}}$ since $r^* = (e - s)^{-1} - e$.

Now, for any $a, b \in \bar{R}$ with $a = a^*$, $b = b^*$ we have

$$\begin{aligned} e - a + b &= (e - a)(e + (e - a)^{-1}b) \\ &= (e - a)(e + (e - a)^{-1}b)(e - (e - a)^{-1}b) \\ &\quad (e - (e - a)^{-1}b)^{-1} \\ &= (e - a)(e - ((e - a)^{-1}b)^2)(e - (e - a)^{-1}b)^{-1}. \end{aligned}$$

From this we get the identity

$$\begin{aligned} (e - a + b)^{-1} &= (e - (e - a)^{-1}b) \\ &\quad (e - (e - a)^{-1}b(e - a)^{-1}b)^{-1}(e - a)^{-1} \\ &= [(e - (e - a)^{-1}b(e - a)^{-1}b)^{-1}(e - a)^{-1}] \\ &\quad - [(e - a)^{-1}b(e - (e - a)^{-1}b(e - a)^{-1}b)^{-1}(e - a)^{-1}]. \end{aligned}$$

Thus, taking, $a = r_1'^*$ and $b = r_1''^*$, we can display $(e - r_1'^* + r_1''^*)^{-1}$ as the difference of two elements from $\bar{R}^{\text{pos}}$ and the result is proved.

IV.3. A necessary and sufficient condition that $F_\alpha \in \mathsf{R}_0$ is that there exists some $r \in \bar{R}^{\text{pos}}$ which is such that

$$F_\alpha = \{f \in F : (r, f) \neq 0\}.$$

Proof. The condition is necessary because, if $\alpha \in \mathfrak{A}_0$ is defined by a set Σ of N states, a mapping $(\Sigma, X) \to \Sigma$, an initial state σ_1, a distinguished subset Σ' of Σ we can associate to every f the $N \times N$ matrix $\mu f_{ii'} = 1$ if $\sigma_i f = \sigma_{i'}$; $= 0$, otherwise, which gives a representation of F in Z_N^{pos}. Trivially, if p is defined by $p_{ii'} = 1$ if $i' = 1$ and $\sigma_i \notin \Sigma'$; $= 0$, otherwise, we have $\operatorname{Tr}(p\mu f) = 1$ or 0 according to $f \in F_\alpha$ or not.

Thus, using the construction described in I.B.1 we can find a representation μ' of F in $Z_{N'}^{\text{pos}}$ which is such that the sum

$$r = \sum_{f \in F} f\mu' f_{1N'} = \sum_{f \in F} f \operatorname{Tr}(p\mu f)$$

has the desired properties.

For proving the sufficiency we start with any $\mu : F \to Z_N^{\text{pos}}$ and we consider the mapping β which sends 0 onto $\mathbf{0}$ and every positive integer onto $\mathbf{1}$ where $\mathbf{0}$ and $\mathbf{1}$ are boolean elements. (i.e., $\mathbf{0}\,\mathbf{0} = \mathbf{0}\,\mathbf{1} = \mathbf{1}\,\mathbf{0} = \mathbf{0} = \mathbf{0} + \mathbf{0}$ and $\mathbf{1}\,\mathbf{1} = \mathbf{1} = \mathbf{1} + \mathbf{0} = \mathbf{0} + \mathbf{1} = \mathbf{1} + \mathbf{1}$); β is an homomorphism of semiring and it can be naturally extended to Z_N^{pos} by defining βm when

$m \in Z_N^{pos}$ as the matrix whose entries are $\beta(m_{ij}) \in \mathbf{0,1}$. Trivially, for any $m,m' \in Z_N^{pos}$ we have $\beta mm' = \beta m \beta m'$ and βZ_N^{pos} has at most $2^{N^2} < \infty$ distinct elements. Thus, the set $\beta mf : f \in F$ is a finite monoid M and $F_\alpha = \{f : \mu f_N \neq 0\} = \{f : \beta\mu f_{1.N} \neq \mathbf{0}\}$ is the inverse image by the homomorphism $\beta\mu : F \to M$ of a subset of M. In other words, F_α satisfies the condition that $F_\alpha = \beta^{-1}\beta F_\alpha$ where β is a homomorphism of the free monoid F into a finite monoid and, according to the theorem 6 of Bar-Hillel and Shamir, this is a necessary and sufficient condition that F_α belongs to R_0.

IV.4. A necessary and sufficient condition that $F_\alpha \in \mathsf{R}_0$ is that there exists an element $r \in \bar{R}$ which is such that $|(r,f)|$ is bounded for all $f \in F$ and that $F_\alpha = \{f \in F : (r,f) \neq 0\}$.

Proof. The construction indicated in the proof of IV.3 shows that the condition is necessary. In order to prove that it is sufficient, it is enough to take any prime number p at least equal to twice the upper bound of $|(r,f)|$ and to observe that the homomorphism γ which sends every integer upon its residue modulo p extends naturally to an homomorphism of Z_N onto the finite algebra of the $N \times N$ matrices over the Galois field of characteristic p; thus $F_\alpha = \{f \in F : \mu f_{1N} \neq 0\} = \{f \in F : \gamma\mu f_{1N} \neq 0\}$ and our remark is again a simple consequence of the theorem of Bar-Hillel and Shamir.

A. An Intuitive Description of $\mathfrak{A}$

IV.A.1. A necessary and sufficient condition that the element a from $\bar{A}$ belongs to $\bar{R}^{pos}$ is that it be produced by a restricted right transducer.

Proof. It is trivial that $0, e$ and each letter x from X can be produced by a (restricted, right) transducer; let us assume that the elements r and r' of $\bar{R}^{pos}$ are produced by the transducers (η, Y, Σ) and (η', Y', Σ') respectively where, without loss of generality, we may assume that the two input alphabets Y and Y' and the two sets of states Σ and Σ' are disjoint. We consider new transducers η'' whose input alphabet Y'' is the union of Y and Y' and whose set of states Σ'' is the union of Σ, Σ', a new initial state σ_1'' and a new sink σ_0''; for any such η'' we shall have the following rules:

(i) $\sigma_1'' y'' = \sigma_1 y$ and $\eta''(\sigma_1'', y'') = \eta(\sigma_1, y)$ if $y'' = y \in Y$; $\sigma_1'' y'' = \sigma_1' y'$ and $\eta''(\sigma_1'', y'') = \eta'(\sigma_1', y')$ if $y'' = y' \in Y'$.

(ii) $\sigma'' y'' = \sigma y$ and $\eta''(\sigma'', y'') = \eta(\sigma, y)$ if $\sigma'' = \sigma \in \Sigma$ and $y'' = y \in Y$ and, similarly, $\sigma'' y'' = \sigma' y'$ and $\eta''(\sigma'', y'') = \eta'(\sigma' y')$ if $\sigma'' = \sigma' \in \Sigma'$ and if $y'' = y' \in Y'$.

1. Let now η_a'' be defined by the supplementary rule

(iii) $\sigma''y'' = \sigma_0''$ when $\sigma'' = \sigma' \in \Sigma'$ and $y'' = y \in Y$ when $\sigma'' = \sigma \in \Sigma$ and $y'' = y' \in Y'$.

By construction η_a'' produces the sum $r + r'$.

2. Let η_m'' be defined by the rule (iii) when $\sigma'' = \sigma' \in \Sigma'$ and $y'' = y \in Y$ and the rule

(iv) $\sigma''y'' = \sigma_1'y'$ and $\eta''(\sigma'', y'') = \eta'(\sigma_1', y')$ when $\sigma'' = \sigma \in \Sigma$ and $y'' = y' \in Y'$.

By construction, η_m'' produces rr'.

3. Let us assume that r is quasi regular and take for (η', Y', Σ') a copy of (η, Y, Σ); if η_g'' is defined by (iv) and the rule (iv′) obtained by exchanging in (iv) the alphabets Y and Y' and the sets Σ and Σ' we obtain a transducer which produces the quasi inverse of r. According to Definition 5 this proves the necessity of the condition IV.A.1; that this condition is sufficient is a simple consequence of the construction indicated in the verification of III.B.3.

Since it has been remarked in IV.2 that any element of $\bar{R}$ can be expressed as the difference of two elements of $\bar{R}^{\text{pos}}$ we have at the same time verified that the definition 1 of $\mathfrak{A}$ is equivalent with the following:

DEFINITION 1″. *An automaton α of $\mathfrak{A}$ consists of a pair of restricted right transducers together with the rule that a word $f \in F$ is accepted if and only if $(r, f) \not\equiv (r', f)$ where r and r' are the formal sums produced by the two transducers.*

RECEIVED: April 3, 1961
REVISED: June 7, 1961

REFERENCES

BAR-HILLEL, Y. AND SHAMIR. E. (1960). *Bull. Res. Council Israel.* **8F,** 155.
ELGOT, C. C. (1960). *Trans. Am. Math. Soc.* **98,** 21.
FATOU, P. (1904). *Compt. Rend. Acad. Sci. Paris* **138,** 342.
HARRINGTON, W. J. (1951). *Am. Math. Monthly* **58,** 114.
HUFFMAN, D. (1954). *Proc. Symp. Inf. Networks*, p. 291. Polytechnic. Inst., Brooklyn, New York.
KLEENE, S. C. (1956). "Automata Studies." Princeton Univ. Press, Princeton, New Jersey.
LAZARD, M. (1955). *Ann. Sci. Ecole Normale Sup.* (3) **72,** 299.
MOORE, E. F. (1956). In "Automata Studies." Princeton Univ. Press, Princeton, New Jersey.
RABIN, M. AND SCOTT, D. (1958). *I.B.M. Res. J.* **3,** 115.
SCHÜTZENBERGER, M. P. (1959). *Seminaire Dubreil-Pisot*, 1959–1960, Inst. H. Poincaré', Paris.
SKOLEM, T. (1934). *Compt. Rend.* 8 *éme Congres Math. Scandinaves, Stockholm*, p. 163.

ON A FAMILY OF SUBMONOIDS

by
M. P. SCHÜTZENBERGER[1]

§ 1. Introduction

As it is well known, only few of the properties of the subgroups of a group are still enjoyed by all the submonoids of a monoid [1] and in the applications it is sometimes useful to consider more restricted families of stable subsets (i. e. of subsets A which are such that $A^2 \subset A$).

In remote connection with a problem in communication theory (Cf. [12]) one encounters a family $\mathfrak{R}(F)$ of submonoids of a monoid F that is characterized by extremal properties and that, consequently, admits several slightly different definitions. When F is a group, $\mathfrak{R}(F)$ reduces to the lattice of the subgroups of F ; in the general case, it is not necessarily a lattice and its simplest definition is the following one.

Definition. The submonoid A of a monoid F belongs to $\mathfrak{R}(F)$ if and only if it satisfies the following three conditions :

1. There exists at least one homomorphism γ of F, compatible with A (i. e. $\gamma^{-1}\gamma A = A$) which is such that γA is isomorphic to a monoid admitting minimal left and right ideals ;

2. $(N_k) : A$ intersects every right and every left ideal of F ;

3. A is maximal among the submonoids of F that have the same intersectoin with an arbitrarily small two-sided ideal of F.

Let us abbreviate by $N_d(N_r, N_l, N_k)$ the condition that A intersects every two sided (right, left, right and left) ideal of F (i. e. that A is "net" in P. Dubreil's theory [5]), by $M_d(M_r, M_l, M_k)$ the condition that γF admits minimal two-sided (right, left, right and left) ideals for some homomorphism γ compatible with A.

We shall verify that $\mathfrak{R}(F)$ can also be defined by the following set of three conditions : A satisfies

1′. M_r ;

2′. N_l;

3′. There exists some right representation of F by mappings of a set into itself that is such that A is submonoid which lets invariant one element from the set.

[1] Cambridge (Mass).

Let us recall that LEVY's condition [8] that a stable subset A of a free monoid F is isomorphic to a free monoid can be expressed in the form (Cf. [12]).

$$(U_d): \quad fA \cap Af \cap A \neq \varnothing \quad \text{only if} \quad f \in A$$

which remains meaningful even when F is not a free monoid.

We shall also verify that $\mathfrak{R}(F)$ is characterized by the following set of conditions on A :

1″. M_k ;
2″. N_k ;
3″. U_d .

When F is finite, the conditions 1, 1′ or 1″ become vacuous. Then $\mathfrak{R}(F)$ can be characterized by 3, 3′ or 3″ and the requirement that A contains at least one positive power of each element from F.

In § 2, as a preliminary step, we apply the classical theory of SUSCHKEWITSCH [18] and REES [11] for obtaining a direct characterization of $\mathfrak{R}(F)$ when F admits minimal left and right ideals. In §§ 3 and 4 respectively we discuss the sets of conditions (1″, 2″, 3″) and (1′, 2′, 3′). In order to make the paper self contained several results which are special cases of theorems due to other authors are given complete proofs.

Applications of the remarks developed here to the less restricted family of the submonoids which satisfy U_d only will be considered in another paper.

§ 2. A direct definition of $\mathfrak{R}(F)$

Let us verify first the following

Remark 2.1. If the stable subset A of a monoid F satisfies N_r and admits minimal right ideals, then, F also admits minimal right ideals.

Proof. Let us consider any $a \in A$ such that aA is a minimal right ideal of A; by definition this is equivalent to the statement that, for any $a' \in A$, there exists at least one $a'' \in A$ which is such that $aa'a'' = a$ since, unless, the right ideal $aa'A$ would be a proper subset of aA.

Trivially, if aA is minimal, the same is true of any $a'''A$ where $a''' \in A\, a\, A$.

Let us show that if A satisfies N_r, a^2F is a minimal right ideal of F. Indeed, for any $f \in F$, N_r implies that $A \cap afF \neq \varnothing$, i. e. that $aff' = a_1 \in A$ for some $f' \in F$; multiplying on the left by a, we obtain $a^2ff' = aa_1$. By our previous remark, there exists at least one $a_1' \in A$ which satisfies $aa_1a_1' = a$. Thus, $a^2ff'a_1'a = a^2$ and the result is verified.

We observe that when the homomorphism γ is compatible with A any of the conditions M_x, N_x or $U_x(x = d, r, l, k)$ defined in the introduction (or later) is true for γA in γF if and only if it is true for A in F. Since we have seen that when A satisfies N_k the condition 1 implies M_k, there will be no loss in generality for the description of a given A from $\mathfrak{R}(F)$ in assuming that F itself admits minimal ideals.

This convention will be kept in the §§ 2 and 3 and we shall use the following standing notations :

The monoid F admits the minimal right ideals R_i ($i \in I$) and the minimal left ideals L_j ($j \in J$). The minimal two-sided ideal of F is denoted by D and the following facts are classical (Cf. [18], [3], [16])

1. $$D = \bigcup_{i \in I} R_i = \bigcup_{j \in J} L_j$$

2. every quasi ideal $K_{i,j} = R_i \cap L_j$ is isomorphic to a certain group G, the SUSCHKEWITSCH *group* of F. (A *quasi ideal* is the intersection of a left and of a right ideal [16]).

3. The idempotent $e_{i,j}$ of $K_{i,j}$ is such that $de_{i,j} = d$ and $e_{i,j} d' = d'$ for any $d \in L_j$ and $d' \in R_i$; thus, identically, $K_{i,j} = e_{i,j} F e_{i,j}$.

We select a fixed arbitrary quasi ideal $K_{1,1}$ and isomorphism $\sigma : K_{1,1} \to G$ and we introduce the following standing notations : $g_{j,i} = \sigma(e_{1,j} e_{i,1})$ $(= e_G$, the neutral element of G when i or j is equal to 1 since $e_{1,j} e_{1,1} e_{i,1} = e_{1,1}$ identically).

$G_0 =$ the subgroup of G generated by the elements $g_{j,i}$.

$\sigma' =$ the mapping $D \to G$ which is defined by $\sigma' d = \sigma(e_{1,j} de_{1,1})$ where j is the index of the left ideal L_j containing d.

$\tau_{i,j} =$ the mapping $G \to K_{i,j}$ which is defined by $\tau_{i,j} g = e_{i,1} \cdot \sigma^{-1}(g_{j,i}^{-1} g) \cdot e_{1,j}$.

It is classical that $\tau_{i,j}$ and the restriction of σ' to $K_{i,j}$ are mutually inverse isomorphisms (onto) (Cf. [11], [2], [10]). Indeed, $\tau_{i,j}$ is a homomorphism because of the following more general formula valid for any $g, g' \in G$

$$(\tau_{i,j} g)\,(\tau_{i',j'} g') = e_{i,1} \cdot \sigma^{-1}(g_{j,i}^{-1}\ g) \cdot e_{1,j} \cdot e_{i',1} \cdot\ \sigma^{-1}(g_{j',i'}^{-1}\ g') \cdot e_{1,j'} =$$
$$= e_{i,1} \cdot \sigma^{-1}(g_{j',i}^{-1} g'') \cdot e_{1,j'} = \tau_{i,j'}\, g''$$

where

$g'' = g_{j',i}\, g_{j,i}^{-1}\, g\, g_{j,i'}\, g_{j',i'}^{-1}\, g'$; thus, when $i = i'$ and $j = j'$, we have simply

$$(\tau_{i,j} g)(\tau_{i',j} g') = \tau_{i,j}(gg').$$

Because of the formula

$$\sigma'\, \tau_{i,j}\, g = \sigma\left(e_{1,j}\left(e_{i,1}\, \sigma^{-1}\,(g_{j,i}^{-1}\, g)\, e_{1,j}\right) e_{1,1}\right) =$$
$$= \sigma(e_{1,j}\, e_{i,1}) \cdot g_{j,i}^{-1}\, g \cdot \sigma(e_{1,j}\, e_{1,1}) = g,$$

we see that $\tau_{i,j}$ is a monomorphism (i. e. isomorphism into). Finally, it is proved that $\tau_{i,j}$ (and consequently the restriction of σ') is an isomorphism (onto) by the formula valid for any $d \in K_{i,j}$

$$\tau_{i,j}\, \sigma'\, d = e_{i,1} \cdot \sigma^{-1}\left(g_{j,i}^{-1}\, \sigma\,(e_{1,j}\, d\, e_{1,1})\right) \cdot e_{1,j} =$$
$$= \left(e_{i,1} \cdot \sigma^{-1}\,(g_{j,i}^{-1}\, e_G) .\ e_{1,j}\right) \cdot d \cdot e_{1,j} = e_{i,j}\, de_{1,j} = d.$$

We still need to recall the following simple statement. (Cf. [15], [17]).

Theorem 2.2. *For any non empty stable subset B of D the three following conditions are equivalent*

(i) *For at least one $K_{i,j}$ having a non empty intersection Q with B the subset $\sigma' Q$ of G contains the inverse of each of its elements;*

(ii) *There exist nonempty subsets* I_B *of* I *and* J_B *of* J *and a subgroup* G' *of* G *that have the following properties:* G' *contains every* $g_{j,i}\,\{(i,j) \in I_B \times J_B\}$, $B = \{d \in D : \sigma' d \in G' \text{ and } d \in K_{i,j}\,[(i,j) \in I_B \times J_B]\}$

(iii) B *admits minimal right* (*and left*) *ideals.*

Proof. (i) *implies* (ii). Because of the fact that the restriction of σ' to $K_{i,j}$ is an isomorphism, there is no loss in generality in taking $(i,j) = (1,1)$ in the condition (i) which then, (because B is stable) becomes equivalent to the condition that $G' = \sigma Q$ is a subgroup of G. Thus $e_{1,1} = \sigma^{-1} e_G$ belongs to B.

Trivially, if $b \in R_i \cap B$ and $b' \in L_j \cap B$, we have $bb' \in K_{i,j} \cap B$ and, thus, $K_{i,j} \cap B \neq \varnothing$ if and only if $(i,j) \in I_B \times J_B$ where I_B and J_B are subsets of I and J respectively.

Let b be any element from $K_{i,j} \cap B$; we have $\sigma(e_{1,1} b^3 e_{1\,1}) = g' \in G'$ and, since G' is a group, $b' = b\sigma^{-1}(g'^{-1})\, b$ belongs to $K_{i,j} \cap B$. A straightforward computation shows that $bb' = e_{i,j}$ and, thus, we have $e_{i,j} \in B$ and $g_{j,i} \in G'$ for all $(i,j) \in I_B \times J_B$. Consequently, for any such pair (i, j), the mappings $\tau_{i,j}$ and σ' can be carried out by using multiplications by elements from B only. It follows instantly that for any such (i, j) and $g \in G$ (rspectively, $d \in K_{i,j}$) one has $\tau_{i,j} g \in B$ (resp. $\sigma' d \in G'$) if and only if $g \in G'$ (resp. $d \in B$) and this is precisely the formula given in (ii).

(ii) *implies* (iii). Let I_B and J_B be any non empty subsets of I and J and G' any subgroup of G containing all the elements $g_{j,i}\,(i, j) \in I_B \times J_B$. In order to prove that B as defined in (ii) admits right and left ideals it is enough to show that for any $(i, j) \in (I_B \times J_B)$ one has

$$(\tau_{i,j}\, G')\, B(\tau_{i,j}\, G') = \tau_{i,j}\, G'.$$

This again is a straightforward computation, which also shows that B^2 is contained in B, i. e. that B is stable.

(iii) *implies* (i). Let us assume only that the stable subset B admits minimal right ideals and, for simplicity, that $b \in K_{1,1} \cap B$ is such that bB is minimal. This implies in particular that, to any $b' \in K_{1,1} \cap B$, there corresponds at least one b'' in some suitable $K_{i,1}$ that is such that $bb'b'' = b$; writing $g = \sigma b$, $g' = \sigma b'$, $g'' = \sigma' b''$, it follows that $gg'g'' = g$, i. e. that $g'' = g'^{-1}$. Thus, since $b'b''b'' \in K_{1,1}$, the set $G' = \sigma(K_{1,1} \cap B)$ contains $\sigma(b'b''b'') = g'^{-1}$ whenever it contains g'. Consequently, G' is a subgroup of G and the proof is concluded.

It is useful to observe that the apparently weaker conditions (iii)′ below is in fact equivalent to (iii).

(iii)′. *There exists a homomorphism* γ *of* F *which is such that* $D \not\supset \gamma^{-1}\gamma B = B$ *and that* γB *admits minimal right ideals.*

Indeed, since $K_{1,1} F K_{1,1} = K_{1,1}$, any homomorphism γ of F sends $K_{1,1}$ onto a minimal quasi-ideal of $\gamma F = \overline{F}$ and, consequently, γ induces an epimorphism γ' (homomorphism onto) of G onto the SUSCHKEWITSCH group $\overline{G}$ of $\overline{F}$.

Let us assume now that γB admits minimal right ideals ; because of theorem 2.2, γB admits minimal quasi-ideals and, since $(K_{1,1} \cap B)\, B(K_{1,1} \cap B)$ is contained in $K_{1,1} \cap B$, at least one of these minimal quasi ideals, $Q_{1,1}$ say, is contained in $\overline{K}_{1,1} = \gamma K_{1,1}$. Thus $\gamma'\, Q_{1,1}$ is a subgroup $\overline{G}'$ of $\overline{G}$ and the stable

subset $\bar{G}'' = \gamma' \gamma(K_{1,1} \cap B)$ of $\bar{G}$ satisfies the conditions $\bar{G}' \bar{G}'' \bar{G}' = \bar{G}'$ and $\bar{G}' \subset \bar{G}''$.

From this we conclude that $\bar{G}' = \bar{G}''$, that is, $Q_{1,1} = \gamma(K_{1,1} \cap B)$.

This ends the proof because, when $D \cap \gamma^{-1} \gamma B = B$, it shows that the stable subset $K_{1,1} \cap B = D \cap \gamma^{-1} Q_{1,1}$ is equal to $\sigma^{-1} \gamma'^{-1} G'$ where $\gamma'^{-1} \bar{G}'$ is a subgroup of G, and that, consequently, the condition (i) is satisfied.

Let us define a mapping χ from F to the set of right cosets of G over G_0 by the rule

$$\chi f = G_0 \sigma(e_{1,i} f e_{1,1}).$$

We have

Remark 2.3. (i) If $f \in D$, $\chi f = G_0 \sigma' d$;

(ii) for any f, $f' \in F$, $\chi(ff') \subset (\chi f)(\chi f')$.

Proof. We verify first that for any $f \in F$ and $j \in J$, $\sigma(e_{1,j} f e_{1,1})$ belongs to χf. Indeed, $f e_{1,1}$ belongs to a well defined $K_{i,1}$ and, using $\tau_{i,1}$ we obtain

$$f e_{1,1} = e_{i,j} \cdot \sigma^{-1}\left(g_{1,i}^{-1} \sigma'(f e_{1,1})\right) \cdot e_{1,1} = e_{i,1} \cdot \sigma^{-1}\left(\sigma'(f e_{1,1})\right).$$

Thus, for any $e_{1,j}$,

$$\sigma(e_{1,j} f e_{1,1}) = \sigma(e_{1,j} e_{i,1}) \sigma'(f e_{1,1}) \in G_0 \chi f.$$

This proves the statement (i).

Let now $f, f' \in F$. The product $e_{1,1} f$ belongs to a well defined $K_{1,j}$ and we have

$$\sigma(e_{1,1} f f' e_{1,1}) = \sigma(e_{1,1} f e_{1,j} f' e_{1,1}) = \sigma(e_{1,1} f e_{1,1}) \sigma(e_{1,j} f' e_{1,1}),$$

that is,

$\chi(ff') = (\chi f) \sigma(e_{1,j} f' e_{1,1})$ and the statement (ii) follows from our initial remark.

Theorem 2.4. *A necessary and sufficient condition that A belongs to $\mathfrak{R}(F)$ is that*

$$A = \{f \in F : \quad \chi f \subset G'\}$$

where G' is any subgroup of G that contains G_0.

Proof. The condition is necessary because, if A belongs to $\mathfrak{R}(F)$, its intersection B with any $K_{i,j}$ is not empty (condition 2) and, according to the condition 1, it satisfies the condition (iii)′ of theorem 2.2. Thus, by theorem 2.2 and remark 2.3 (i), we have $B = A \cap D = \{d \in D : \chi d \subset G'\}$ where G' is a subgroup containing G_0. Since remark 2.3 (ii) shows trivially that BfB is contained in B if and only if χf is contained in G' the condition 3 of the introduction implies that A is precisely the set of those elements from F.

The condition is sufficient because, if $A = \{f : \chi f \subset G'\}$, remark 2.3 (i) and (ii) show that $A^2 \subset A$, and that $A \cap D = B$ is a stable subset which satisfies the conditions of theorem 2.2 and $BAB = B$. Thus the conditions 1 and 2 are satisfied and since, as above, BfB is contained in B only if $\chi f \subset G'$, the maximality condition 3 is also verified.

As a consequence we have

Corollary 2.5. If A belongs to $\mathfrak{R}(F)$ and if the index m in G of the subgroup G' defined above is finite, at least one positive power $f^{m'}$, $m' \leq m$ of each f belongs to A.

Proof. Let us observe that, for any f, $f' \in F$, one has $G'\chi(ff') = G'\chi f'$ if and only if $G'\chi f' = G'$, that is, by theorem 2.4, if and only if f belongs to A.

Since, by hypothesis, not all the $m+1$ cosets

$$G',\ G'\chi f,\ G'\chi f^2,\ \ldots,\ G'\chi f^m$$

are distinct, one must have $G'\chi f^{m''} = G'\chi f^{m'+m''}$, i. e. $f^{m'} \in A$ for some positive m' at the most equal to m.

§ 3. The conditions U_x.

In this § we use the following conditions U_x $(x = d, r, l, k)$ for characterizing $\mathfrak{R}(F)$. We recall that U_d is defined by

$$(U_d):\quad fA \cap Af \cap A \neq \varnothing \quad \text{only if } f \in A.$$

Thus, if A is a nonempty stable subset satisfying U_d, it is a submonoid (i. e. it contains the neutral element e of F) because $eA \cap Ae \cap A \neq \varnothing$. It is readily verified that, when A is stable, equivalent forms of U_d are

$$fA \cap A \neq \varnothing \text{ and } Af \cap A \neq \varnothing \text{ only if } f \in A\ ;$$

(because a, $af = a_1 \in A$, and a', $fa' = a_1' \in A$ imply $(a_1' a) f = f(a' a_1) = a_1 a_1' \in A$)

and, also

a, af, $fa \in A$ only if $f \in A$.

We define U_r by

$$(U_r):\qquad Af \cap A \neq \varnothing \text{ only if } f \in A.$$

Then, U_r (or the symmetric condition U_l, $fA \cap A \neq \varnothing$ only if $f \in A$) implies U_d. As it is easily checked (Cf. the beginning of 4 below), U_r is equivalent to the condition 2′ of the introduction.

When A is a submonoid, the conjunction U_k of the conditions U_r and U_l is more expeditiously written as

$$(U_k):\qquad A \cap AfA \neq \varnothing \quad \text{only if } f \in A.$$

A theory of the subsets, which satisfy U_x $(x = r, l, k)$ (''les complexes unitaires'') is due to P. DUBREIL [6].

We first verify the following

Remark 3.1. When the submonoid A of F satisfies M_k, N_d and U_d, the condition N_r (respectively N_l) is a necessary and sufficient condition that it satisfies U_r (respectively U_l).

Proof. Because of M_k we can assume without loss of generality that F itself admits minimal right and left ideals and we use freely the notations of § 2. The condition N_d can be taken as the hypothesis that $A \cap K_{1,1}$ is not empty.

Let us first verify that $B = A \cap D$ satisfies the condition (i) of theorem 2.2 (i. e. $G' = \sigma(A \cap K_{1,1})$ is a subgroup of G). Indeed, if $g = \sigma a \in G'$, for some

$a \in A \cap K_{1,1}$, the element $b = \sigma^{-1}g^{-1}$ satifies the relation $ba^2 = a^2b = a$, that is, $bA \cap Ab \cap A \neq \varnothing$. Thus, by U_d, $b \in A$ and, finally, $g^{-1} = \sigma b \in G'$.

(Reciprocally if F reduces to the union of D and a neutral element e, it is easily checked that for any B satisfying the conditions of theorem 2.2 the submonoid union of B and e, satisfies the condition U_d. Indeed, for any $d \in D$,

$$dB \cap B \neq \varnothing \quad \text{and} \quad Bd \cap B \neq \varnothing \quad \text{imply} \quad d \in K_{i,j} \text{ with } (i, j) \in I_B \times J_B$$

and, then b, $db \in B$ implies $\sigma' d \in G'$).

Now we have :

N_r implies U_r.

Because of our hypothesis, N_r is equivalent to the requirement that every $e_{i,1}$ $(i \in I)$ belongs to A, or, in the notations of theorem 2.2, that $I = I_B$. It follows that for any $d \in D$, if $bd \in A$ for some $b \in B$, then d belongs to A ; indeed, $bd \in A$ implies $d \in L_j$, where $j \in J_B$ and $\sigma' d \in G'$ since G' is a subgroup which contains all the elements $g_{j,i}$ with $(i, j) \in I \times J_B$.

This practically ends the proof because if a, $af \in A$, the element $d = fe_{1,1}$ from D satisfies the condition $bd \in A$ with $b = e_{1,1}a \in B$. Thus we have, a, af, $e_{1,1}$, $fe_{1,1} \in A$ and, by U_d, we conclude that a, $af \in A$ only if $f \in A$, that is, U_r.

The reciprocal statement (U_r implies N_r) is contained in the following slightly less special implication which will be needed later :

When M_r, N_d and U_r imply N_r.

We assume that F itself contains an element r which is such that the ideal rF is minimal ; thus, because of N_d, A contains at least one element $b \in FrF \cap A$ which is such that bF is a minimal right ideal.

Let us show that $A \cap fF \neq \varnothing$ for all $f \in F$, (i. e., N_r) ; indeed since bF is minimal, there exists at least one f' which is such that $b = bff'$. Because of U_r the product ff' belongs to A and this concluedes the proof.

Theorem 3.2. *If the submonoid A satisfies M_k, necessary and sufficient conditions that it belongs to $\mathfrak{R}(F)$ are U_d and N_k or U_r and N_l or U_k and N_d.*

Proof. Let us assume that A belongs to $\mathfrak{R}(F)$ and use the notations of theorem 2.4 ; by corollary 2.5 every idempotent of F belongs to A, and consequently A satisfies N_k ; the fact that $AfA \cap A \neq \varnothing$ only if $f \in A$ (i. e. U_k) has already been verified in the proof of theorem 2.4.

Reciprocally, we observe that, according to remark 3.1, the three conditions "U_x and N_x," are equivalent to "U_k and N_k" when A satisfies M_k. Using the notations of remark 3.1, the condition N_k imposes that $B = A \cap D$ intersects every $K_{i,j}$ and consequently $B = \{d \in D : \chi d \subset G'\}$ where G' is a subgroup containing G_0 ; once more, since $BfB \subset B$ only if χf is contained in G' we finally obtain $A = \{f : \chi f \not\subset G'\}$ and the result is entirely proved.

§ 4. The set of conditions (1′, 2′, 3′)

In order to make the proof clearer we recall first the following well known result (Cf. [19]) :

Theorem 4.1. *To any nonempty subset X of F there corresponds one quotient monoid $\gamma_X F$ which is characterized by the following properties*

(i) *The homomorphism γ_X is compatible with X ;*

(ii) *If γ' is any homomorphism of F compatible with X, $\gamma_X F$ is a homomorphic image of $\gamma' F$.*

Proof. Let us consider the mapping λ_X of F to the subsets of F that is defined by

$$\lambda_X f = \{f' \in F : ff' \in X\}$$

(Cf. [5]).

We have

1. if $x \in X$, $\lambda_X f = \lambda_X x$ only if $f \in X$ (because $\lambda_X f$ contains e if and only if $f \in X$) ;

2. if $\lambda_X f = \lambda_X f'$, then $\lambda_X (ff'') = \lambda_X(f'f'')$ for all $f'' \in F$.

Consequently, if S denotes the set of all $\lambda_X f (f \in F)$, we can define a representation $(S, F) \to S$ by

$$(\lambda_X f) f' = \lambda_X(ff') \text{ for any } \lambda_X f \in S \text{ and } f' \in F.$$

We denote the corresponding homomorphism of F by γ_X and we observe that the congruence relation $\gamma_X f = \gamma_X f'$ (i. e. $\lambda_X(f''f) = \lambda_X f(f''f')$ for all $f'' \in F$) can be expressed in the symmetrical form:

for all $f_1, f_2 \in F$, $f_1 f f_2 \in X$ if and only if $f_1 f' f_2 \in X$.

This shows instantly that γ_X is compatible with X since $efe \in X$ if and only if $f \in X$.

Let now $\gamma' : F \to \bar{F}$ be any homomorphism and define $\bar{X} = \gamma' X$; we can construct in the same manner as above a quotient monoid $\gamma_{\bar{X}} \bar{F}$ and for any $f, f' \in F$ we have $\gamma_{\bar{X}} \gamma' f = \gamma_{\bar{X}} \gamma' f'$ only when for all $f_1, f_2 \in F$

$\gamma' f_1 f f_2 \in \gamma' X$ if and only if $\gamma' f_1 f' f_2 \in \gamma' X$.

Consequently, when $\gamma'^{-1} \gamma' X = X$, we have $\gamma_{\bar{X}} \gamma' f = \gamma_{\bar{X}} \gamma' f'$ only if $\gamma_X f = \gamma_X f'$ and the result is proved.

Incidentally, the notations introduced provide the formal verification that U_r is equivalent to the condition 2′ of the introduction, because on the one hand, if A is stable and if it satisfies U_r, we have $e \in A$ and $\lambda_A a = A$ for any $a \in A$; thus $\lambda_A e = \lambda_A f$ if and only if $f \in A$ and A is precisely the submonoid which lets $\lambda_A e$ invariant in the representation $(S, F) \to S$ described above. On the other hand, if S' is any set and $(S', F) \to S'$ a representation, for any given $s \in S'$, the submonoid $A' = \{f \in F : sf = s\}$ satisfies U_r because of the associativity.

Theorem 4.2. *If the stable subset A of F satisfies M_r, N_l and U_r, it belongs to $\mathfrak{R}(F)$.*

Proof. Since N_l is stronger than N_d, we already know by the last part of the proof of remark 3.1 that A satisfies N_r and we shall repeatedly use this fact.

Without loss of generality we shall assume that $F = \gamma_A F$; consequently, because of theorem 2.1 and M_r, the monoid F itself admits minimal right ideals ; it will be enough to verify that it admits also minimal left ideals, because, then, by remark 3.1, U_l is a simple consequence of M_k, N_k and U_r.

The verification involves three steps.

i. Let $b \in A$ be a fixed element such that bF is a minimal right ideal (such an element exists because of N_d). We verify that for any $f \in F$ there exists at least one $f' \in F$ which is such that $\lambda_A fb = \lambda_A f' b^2$.

Indeed, by N_r, $fbf_1 \in A$ for some f_1 ; by N_l, $f'b^2f_1 \in A$ for some f' ; by the hypothesis that bF is minimal, $bf_1f_1' = b$ for some f_1'. Thus

$$\lambda_A fbf_1 = \lambda_A f' b^2 f_1 = A$$

because of the hypothesis that A satisfies U_r. Finally, multiplying by f_1' we get

$$\lambda_A fb = \lambda_A fbf_1 f_1' = \lambda_A f' b^2 f_1 f_1' = \lambda_A f' b^2$$

and our remark is proved.

ii. Let us keep the same notations and define $\bar{b}$ by the condition that $b^2\bar{b} = b$.

From the relation $\lambda_A fb = \lambda_A f'b^2$, we deduce by multiplication by $\bar{b}b$ that

$$\lambda_A fb\bar{b}b = \lambda_A f' b^2\bar{b}b = \lambda_A f' b^2 = \lambda_A fb .$$

Since this holds for each $f \in F$, it follows from the hypothesis $\gamma_A F = F$ that $b\bar{b}b = b$. Consequently $b\bar{b}$ is an idempotent.

The last step is classical (cf. [3], [15], [16]) but we include its proof here for the sake of completeness :

iii. If F contains an idempotent c which is such that cF is a minimal right ideal, then, Fc is a minimal left ideal.

Indeed, for any $f_1 \in F$, we have $cf_1cf_2 = c$ for some f_2 and $cf_2cf_3 = c$ for some f_3 because of the minimal character of cF. Multiplying the last equality by cf_1 we get

$$cf_1cf_2cf_3 = cf_1c, \quad \text{that is} \quad ccf_3c = cf_1c.$$

Consequently, $cf_2cf_1c = c^2 = c$ and the result is proved since we have shown that c belongs to any left ideal Ff_1c.

Remark. Counter examples (cf. [13]) show that it is not possible to dispense entirely with some requirement on the minimal ideals in the various implications between the conditions N_x and U_x, described here.

For example, let F be the monoid of permutations of the set of integers generated by the translation $n \to n + 1$, and $n \to n - 1$ and the permutation which lets invariant the negative integers and which consists of the cycles

$$(1,2)\,(3,4,5)\,(6,7,8,9)\,\ldots\,\left(\frac{n\cdot n-1}{2}, \frac{n\cdot n-1}{2}+1, \ldots, \frac{n\cdot n+1}{2}-1\right)\ldots$$

Let A be the submonoid of F that lets 0 invariant. It is easily checked that $\gamma_A F = F$, that F has no minimal ideals and that A satisfies N_k and U_k.

We conclude by giving a simple characterization of $\gamma_A F$ for any A from $\mathfrak{R}(F)$ (cf. [14]).

The notations are that of §§ 2 and 3.

Remark. 4.3. If A belongs to $\mathfrak{R}(F)$, a necessary and sufficient condition that $\gamma_A F = F$ is that $f = f'$ if and only if

$$\pi\sigma(e_{1,j}\, fe_{i,1}) = \pi\sigma(e_{1,j}\, f'\, e_{i,1})$$

for all $(i, j) \in I \times J$ where π is a homomorphism of G whose kernel, E, is the largest normal subgroup of G contained in G'.

Proof. Let us observe that because of U_k and N_k, the relation $f_1 f f_2 \in A$ is equivalent to $e_{1,1} f_1 f f_2 e_{1,1} \in A$ for any three elements f_1, f and f_2 of F. With the help of the mapping $\tau_{i,j}$ (cf. § 1) can we write $e_{1,1}f_1$ and $f_2 e_{1,1}$ as $(\sigma^{-1} g_1)e_{1,i}$ and $e_{j,1}(\sigma^{-1} g_2)$ respectively, for suitable $g_1, g_2 \in G$ and idempotents $e_{1,i}$ and $e_{j,1}$.

Thus, $f_1 f f_2 \in A$ is equivalent to $g_1 \sigma(e_{1,i} f e_{j,1}) g_2 \in G'$ where g_1, g_2 and (i, j) do not depend upon f. It follows from the definitions of γ_A that $\gamma_A f'$ if and only if for each $(i,j) \in I \times J$ and, then, for all g_1, $g_2 \in G$, one has $g_1 \sigma(e_{1,i} f e_{j,1}) g_2 \in G'$ when and only when $g_1 \sigma(e_{1,i} f' e_{j,1}) g_2 \in G'$. Since for each (i,j) this relation between $g = \sigma(e_{1,i} f e_{j,1})$ and $g' = \sigma(e_{1,i} f' e_{j,1})$ is precisely $\gamma_{G'} g = \gamma_{G'} g'$ and since E is, trivially, the kernel of $\gamma_{G'}$ the result is proved.

It follows that a set of necessary and sufficient conditions that $D = \gamma_{G'} D$ is :

i. the only normal subgroup of G contained in G' is $\{e_G\}$;

ii. the $J \times I$ matrix $(g_{j,i})$ has all its rows and columns distinct.

As an application we can display the following example which shows that, even if F is finitely generated, the condition that for some fixed finite m, f^m belongs to A for all $f \in F$ does not insure that $\gamma_A F$ has only finitely many minimal quasi ideals.

Example. Let F consist of e, all the powers a^m of a certain element a and of a minimal two-sided ideal D of the type described in § 1. The group G will be the symmetric group on three elements generated by α and β satisfying the relations $\alpha^2 = \beta^3 = (\alpha\beta)^2 = e_G$; I, and J will be the set of positive integers.

The element a is entirely defined by the rules :

$$e_{1,j}\, a = \begin{cases} \tau_{1,j+1}(e_G) & \text{if } j \text{ is not a power of 2,} \\ \tau_{1,j+1}(\alpha) & \text{if } j \text{ is a power of 2.} \end{cases}$$

We define the right ideals R_i by $R_1 = e_{1,1}F$, $R_{i+1} = aR_i$ and, accordingly, the matrix $(g_{j,i})$ has all its entries in the subgroup $G_0 = \{e_G, \alpha\}$.

Finally, $A = \{f \in F : \chi f = G_0\}$ contains the sixth power of every element of F and it belongs to $\mathfrak{R}(F)$.

By considering for each value of m $\geqq 0$ the sub-block of the matrix $(g_{j,i})$ determined by $1 \leqq i \leqq 2^m$, $1 + 2^m \leqq j \leqq 2^{m+1}$, one easily checks that no two rows of this matrix are the same and that consequently, it also contains infinitely many distinct columns.

Thus $\gamma_A F$ is not finite and, since F is generated by a and $b = \tau_{1,1}(\beta)$, the example has all the properties stated.

(Received May 28, 1961.)

REFERENCES

[1] Chevalley, C.: *Fundamental concepts of Algebra* (N. Y. 1956) chap. 1.
[2] Clifford, A. H.: *Am. J. Math.* (66), 1942. 327.
[3] Clifford, A. H.: *Am. J. Math.* (70), 1948. 521.
[4] Clifford, A. H.: *Am. J. Math.* (82), 1960. 430.
[5] Dubreil, P.: *Mem. Acad. Institut France.* (63), 1941. 1.
[6] Dubreil, P.: *Rendiconti Circ. Mat. Palermo* 1951. 183.

[7] Dubreil, P.: *Bull. Soc. Math. France.* (82), 1953. 289.
[8] Levy, F. W.: *Bull. Calcutta Math. Soc.* (36), 1944. 191.
[9] Miller, D. D. and Clifford, A. H.: *Trans. Am. Math. Soc.* (92), 1956. 270.
[10] Preston, G. B.: *Quart. J. Math.* (9), 1958. 169.
[11] Rees, D.: *Proc. Cambridge Phil. Soc.* (36), 1958. 169.
[12] Schützenberger, M. P.: *Trans. Institute of Radio Engineers.* vol. IT2. 1956. 47.
[13] Schützenberger, M. P.: *Publ. Scien. Univ. Alger.* (6), 1959. 85.
[14] Schützenberger, M. P.: *C. R. Acad. Sci. Paris* (244), 1957. 2219.
[15] Schwartz, St.: *Czeckoslovak Math. J.* (76), 1951. 229.
[16] Steinfeld, O.: *Publ. Math. Debrecen.* (4), 1956. 262.
[17] Stoll, R. R.: *Duke Math. J.* (11), 1944. 251.
[18] Suschkewitsch, A.: *Math. Annalen.* (99), 1928. 30.
[19] Teissier, M.: *C. R. Acad. Sci. Paris* (232), 1951. 1987.

ОБ ОДНОМ СЕМЕЙСТВЕ ПОДМОНОИДОВ

M. P. SCHÜTZENBERGER

Резюме

В этой заметке описывается некоторое семейство $K(F)$ подмоноидов моноида F, имеющих свойства, возможно близкие к свойствам подгрупп некоторой группы. Если F — свободный моноид, тогда подмоноиды семейства $K(F)$ имеют приложения к некоторым вопросам кодирования как особому случаю свободных подмоноидов моноида F. Характерно, что если A принадлежит $K(F)$, то для каждого $f \in F$ найдется хотя бы один f' такой, что $ff'f \in A$ (существование слабого обратного элемента) и, наоборот, если f и $ff'f$ принадлежат A, то f' тоже принадлежит A (каждый слабый обратный некоторого элемента подмоноида $A \in K(F)$ принадлежит A).

Большая часть статьи посвящена дисскуссии того заслуживающего внимания факта, что при обычных ограничениях относительно существования минимальных идеалов эти двухсторонние условия содержатся в еще более слабых аналогичных односторонных условиях.

SPECIAL ARTICLE

REPORT ON MATHEMATICS IN THE MEDICAL SCIENCES*

DAVID D. RUTSTEIN, M.D.,† MURRAY EDEN, PH.D.,‡ AND MARCEL P. SCHÜTZENBERGER, M.D.§

BOSTON

THIS conference on "Mathematics in the Medical Sciences" was called because medical scientists have begun to ask questions that entail the development of new mathematical theory and the application of more complex mathematical reasoning than has been the case in the past. Also, the effective use of new instruments now becoming available to medicine demands an understanding of their underlying physical theory and its mathematical application. The meeting was concerned with both the nature of mathematics and the technics for its application that have been useful to biology and medicine in the past or may be applicable in the near future. In this report — which is not a complete summary of the meeting — the lectures and discussions are focused on the principles governing the interrelation of mathematical theory and the biologic and medical sciences. After the summary there is a general statement by Professor William G. Cochran that epitomizes the present status of "The Role of Mathematics in the Medical Scienoes."

For the purposes of this report, biomathematics is defined as the development of new mathematical theories or technics under the stimulus of unsolved biologic problems and the application of existing mathematical theory and technics for describing and interpreting biologic and medical phenomena.

There is considerable overlap between the fields of biomathematics and statistics. To identify biomathematics more clearly, the field of statistics is briefly discussed, and relations to biomathematics are indicated. Biomathematics as considered at the Conference will then be presented, with particular reference to the construction of hypotheses or "model building."

STATISTICS

Much of the statistical work in the medical sciences in the recent past has been concerned with the systematic application of logical reasoning to the solution of particular quantitative problems. Well known are such applications as drug assays, testing of laboratory procedures and sample and mass surveys. As this area of statistics has developed in relation to medical research, there has been increasing interest in the principles governing the design and analysis of experiments. An appropriate design appears to be crucially important, especially in problems involving large numbers of investigators. Moreover, precise analysis and interpretation of the data are the last important steps of a successful experiment. This kind of statistics has become classic and was intentionally omitted from the program.

But before this use of statistics is dismissed, one relation to biomathematics is worth mentioning. The use of human subjects imposes limitations on the design of experiments or makes human experimentation impossible. Sometimes questions can be restated so that controlled experiments involving human beings can be done. For instance, it may be impossible to perform such an experiment to determine whether a new drug is better than no treatment in a disease like rheumatic fever, but it is possible to perform a controlled experiment on whether the new drug is better than the generally accepted treatment. But when human experiments dependent upon classic statistics cannot be performed, as in experiments involving human genetics, mathematical elaborations of Mendelian theory may make it possible to predict human genetic distributions.

Another relation between statistics and biomathematics is concerned with the application of statistics in the exploration of a new field of investigation or a new set of problems. In this sense, statistics has a use analogous to that of the microscope or dye permitting the visualization of details not perceptible to the naked eye. Three speakers at the Conference presented illustrations of this use of statistics. Pipberger described a method of principal component analysis of electrocardiographic tracings and a procedure for discriminating between tracings according to some particular property. Thus, with the help of time integrals and spatial rotations, it was possible to identify the elements of the electrocardiogram that correlated independently with such abnormalities as ventricular hypertrophies and myocardial

*Based on a conference held under the auspices of the Department of Preventive Medicine, Harvard Medical School, Boston, January 16 and 17, 1961. Program titles and participants were as follows: "Order of Amino Acids in a Protein," Sidney A. Bernhard, National Institute of Mental Health; "Genetic Mapping," Cyrus Levinthal, Massachusetts Institute of Technology; "Mathematical Models for Muscular Contraction," Richard Podolsky, United States Naval Medical Research Institute; "Mathematical Models for the Study of Physiological Systems," Norman Z. Shapiro, National Institutes of Health; "The Role of Mathematics in the Medical Sciences," William G. Cochran, Harvard University; "Mathematical Models and Computational Technics for the Analysis of Neuroelectric Activity," Walter A. Rosenblith, Massachusetts Institute of Technology; "Collection, Storage, Analysis and Use of Electrocardiographic Data," H. V. Pipberger, Georgetown University School of Medicine and United States Veterans Administration; "Potential Applications of Mathematics to the Diagnosis of Illness," Murray Eden, Massachusetts Institute of Technology; and "Use of Mathematics in the Study of Biologic Transport Systems," John L. Stephenson, National Heart Institute.

We are grateful to Professor Norman F. Ramsey, of Harvard University, Professor John R. Pappenheimer, of Harvard Medical School, and Dr. James A. Shannon, of the National Institutes of Health, for their contributions to the Conference as chairmen of the sessions.

†Professor of preventive medicine and head, Department of Preventive Medicine, Harvard Medical School.

‡Associate professor of electrical engineering, Center for Communications Sciences, Research Laboratory of Electronics, Massachusetts Institute of Technology; lecturer on preventive medicine, Harvard Medical School.

§Maître de conférences, Faculty of Sciences, University of Poitiers, Poitiers, France; visiting professor, Department of Statistics, University of North Carolina.

Reprinted from the *New England Journal of Medicine*
265:172-176 (July 27), 1961

infarction, and eliminate those that were dependent and introduced no new evidence of such involvement.

Rosenblith, in a problem of pattern recognition, described the use of the statistical theory of communication as a means of exhibiting regularities in electroencephalographic tracings and in recordings of electrical events observed in the nervous system.

Eden presented a third application of this use of statistics. This concerned the striking of a proper balance between information retrieval and the automation of various aspects of medical diagnosis and treatment. Thus, methods exist for the retrieval of such items of information as laboratory reports and other quantitative data, whereas the defining, storing and retrieving of descriptive data such as that found in the history and physical examination are as yet unsolved. Eden emphasized the fact that, in the final analysis, machine diagnosis of illness will depend on information concerning the probability of specific symptoms, signs or laboratory tests appearing in certain diseases, and that to be applicable, these data will have to be collected in many kinds of population groups.

These newer applications of statistics would be prohibitively time consuming were it not for recent technologic development in the computation art. First of all, very large amounts of data can be processed rapidly. Secondly, it may be possible to build a simple, special-purpose computer that will perform the preliminary processing of the data for the clinician. If further analysis is desirable, this first step will have been helpful in preparing the data for introduction into a large, general-purpose computer.

It was made clear throughout the Conference that computers can be helpful only when the investigator has carefully defined his terms, meticulously planned the design of his experiments, established a logical method of classification and collected his information in consistent fashion.

The Construction of Hypotheses or "Model Building"

Much of the time of the Conference was devoted to the use of mathematics in the construction of hypotheses useful in medical and biologic research. This involves the analysis of existing data with the help of mathematical theory so that more useful hypotheses may be created than would otherwise be the case. This use of mathematics in constructing biologic or medical hypotheses is often referred to as model building. Indeed, the creation of mathematical models has been implicit in all the statistical applications mentioned above because the particular probability theory upon which the statistics are based in itself represents a mathematical model.

Mathematical models are particularly useful when the investigator by intuitive means cannot see the relation between two or more facts or sets of data. In such a situation the mathematician may establish a deductive relation through a mathematical equation that can then be tested by experiment. An example in a physical system is Kundt's deduction that the velocity of sound in a gas is related to its specific heats. This was later verified by experiment. An example in biology proposed by Shapiro concerns the use of labeled compounds in the study of metabolism. If one uses a mathematical model involving simultaneous linear differential equations with constant coefficients for a description of the kinetics of a labeled system, one can mathematically deduce the form of relation between amounts of labeled compounds measured and time — a form that is subject to empirical verification.

Physical Models

Mathematicians may also be helpful by pointing out that certain well studied physical systems may be useful in the investigation of particular biologic systems. The extensive use of physical models is supported by the ideas that physicists have been successful in explaining nature, that alternatives to the vitalist approach must be physical ones and that the mathematical knowledge of the average biologist is limited to that used in simple physical systems. Where a physical system is applicable, this approach may be successful. Thus, the Hodgkin–Huxley model of nerve transmission is based on electromagnetic theory and that of the transmission of electric power through cables.

There are, however, limitations to the use of physical systems in biologic research. Biologic measurements often cannot be made with the great precision of physical ones. Moreover, the clearly defined conditions of equilibrium in physical systems often cannot be approximated in biology, even when static measurements are precise. Thus, the Russians could calculate the movements of the Earth and Venus, the trajectory of the carrier rocket and the point, direction and time of firing of the interplanetary rocket on its way to Venus. In biologic research, the chloride concentration of the blood may be measured with some precision at a particular time, but because of the many known and unknown factors that may influence it, it may not be possible to predict its value from moment to moment, or in different compartments of the body.

Furthermore, physical systems can often be simplified without disturbing the realistic relation of the simplified to the original system. In contrast, it may be very difficult in biology to simplify a system without making it totally artificial. Thus, the Donnan equilibrium may be measured precisely in a simplified physical system, but its direct application to the formation of cerebrospinal fluid might be unrealistic.

Where it is impossible to find a simple physical model that is directly applicable to a biologic system, mathematical theory may be helpful. Thus, Stephenson showed that all the kinds of data needed on the basis of a physical model to study the complex prob-

lem of fat absorption cannot be collected in a human subject. But it was possible with fatty acid tracer data and the use of integral equations to work out the absorption, splitting and recirculation of fatty acids and of triglycerides. This approach may be very useful in the study of metabolism and in other biologic and medical problems. In this connection, it is instructive that the type of equation used by Stephenson is known to mathematicians as the Volterra equation and has a broad range of application to biologic and physical systems. Actually, Volterra was a mathematician who developed this equation for the study of another biologic problem — that concerned with the competition of species.

Monte Carlo Methods

Until recently, mathematical models of complex biologic systems could not be verified because the mathematics required to get to the experimental implementations of the models is remarkably complicated and the computations inordinately long. Such a model was presented by Podolsky with reference to the kinetics of muscular contraction. He could study his model because the large, general-purpose digital computer can by reason of its great speed do what a human computer would find prohibitively time consuming. But even this approach is limited. Some biomathematical models, although made up of simple elements, become complex because they have so many interactions that even the largest computer cannot by analytic methods produce a specific answer for the entire model. Thus, when Podolsky's model simulated the transition from one steady state of muscular contraction to another, a new approach was necessary. Here, Podolsky and Shapiro used Monte Carlo methods, by which, in terms of a statement of probability, they ran a series of mathematical experiments on a digital computer and were able to converge on a numerical solution to the entire problem. Monte Carlo methods are presently under study by a number of mathematicians, and are applicable to many biologic and medical problems.

Stochastic Processes

In many biologic systems there is no present possibility of making a physical model, however complicated. Instead, certain factors must be regarded as having values controlled by chance. Processes of this kind are referred to as stochastic. Stochastic mathematical models are models in which the behavior of the system under study is modified by the addition of a variable with a known probability of occurrence and distribution. At the Conference, Rosenblith demonstrated the use of such a model by the effect of a repetitive standardized auditory stimulus on the neuroelectric activity of the human brain. Such models have also been used in many kinds of biologic studies, including epidemiology by Bailey, in cancer induction by Neyman and evolutionary genetics assessing a long-range possibility of eugenics by Dahlberg. This approach, supported by computer facilities, will enormously broaden the opportunities of the application of mathematics to the study of biologic and medical phenomena.

Finite Mathematics

In applying mathematics to biology, biologists must not limit their conception of mathematics to a kind of glorified arithmetic concerned only with the relation between quantities of things. Actually, as illustrated at the Conference, a good deal of modern mathematics is concerned with the relations between objects rather than with numerical values. Bernhard applied this kind of mathematics to the sequencing of amino acids in a protein. In collaboration with Duda, he developed a logical procedure for determining the sequence by digital computation. This approach has come to be called finite mathematics and has been used in other biologic research — for example, by McCulloch and Pitts in their hypothesis of the random nerve net to explain the mechanism of the nervous system. This procedure can also be used for other than biologic problems, as was demonstrated by Eden at the Conference in his synthesis of human handwriting by mathematical methods. The stress in finite mathematics is on combinatorial technics. Levinthal discussed the combinatorial procedures involved in genetic mapping and posed certain unanswered questions relating to the verification of the assumption of the linearity of the genetic map. This same mathematical approach has been used by the team of Levi-Strauss, a biologist, and Weil, a mathematician, in studying laws governing marriage relations in primitive societies, and illustrates the application of this method to the study of anthropology.

Specificity of Hypotheses

Shapiro made a significant point when he declared that a mathematical model (hypothesis) to be useful had to be stated in specific and restrictive terms. If the hypothesis when tested by experiment can be satisfied by many different sets of data, it is too general and has little differential or predictive value. As an example Shapiro showed that the multilinear receptor theory of color vision* was satisfied by many sets of data collected when observers matched lights in the entire range of the visible spectrum.

Information Theory and Cybernetics

Except for a brief reference on neuroelectric phenomena by Rosenblith, the application to biology of the mathematics concerned with information

*This theory states that color vision in the human eye depends upon a finite number (three?) of receptors, each of which follows Beer's law.

theory and cybernetics was not covered in the Conference but should be mentioned briefly here. At a time when biologists, attempting to bring rigor into their field of interest, felt obliged to adhere to quantitative physical models, it remained for the mathematician Norbert Wiener to point out that the way the organism processes energy may be much less important than the mechanisms by which these processes are controlled. The concept of information theory and the feedback of cybernetics have also been applied to the study of perception by Reichardt in the vision of the bee, and by Stark in studies of muscular and pupillary contractions.

Application of Standard Models

Because of limitations of time, the program and the meeting were devoted almost entirely to questions requiring both biologic and mathematical research. Only brief mention was made of the application of standard mathematical methods to problems in biology, in medicine and in medical care. If more time had been available, presently unexploited applications of standard mathematical methods would have been discussed. For example, scientists in the Department of Statistics of the University of North Carolina have recently applied mathematical technics used in the "inventory control problem" to the operation of a blood bank — that is, the minimum number of units of blood in a bank required to meet all essential needs.

Epilogue

Some of the ways in which the life sciences and mathematics can interact were outlined at the Conference. The notions that mathematics can only be used as a service tool in biology or that biology was either too vague or too trivial to engage the interest of the mathematician were dispelled. It also became obvious that most medical and biologic investigators know too little mathematics of either the classical or modern variety to work productively with the mathematician. On the other hand, it became evident that mathematical theory and technics need extensive development for the effective study of biologic problems. Ideally, prospective life scientists should be well grounded in both biology and in mathematics, but such persons will always be rare. In spite of this, fruitful collaboration can go on between the two disciplines if biologic and medical investigators will learn enough mathematics, electrical engineering and physics to be able to ask the proper questions of the mathematician, and if the mathematician will learn enough biology to be able to develop the mathematical theory and technics necessary to the solution of important biologic and medical problems.

Suggested Bibliography

General

Oncley, J. L. *Biophysical Science.* 609 pp. New York: Wiley, 1959.

Wiener, N. *Cybernetics.* 194 pp. New York: Wiley, 1948.

Contributions to Biology and Problems of Health: Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability, v.4: Held at the Statistical Laboratory, University of California, December, 1954, July and August, 1955. Edited by J. Neyman. 187 pp. Berkeley: Univ. of California Press, 1956.

Contributions to Biology and Problems of Health: Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, v.5: Held at the Statistical Laboratory, University of California. Edited by J. Neyman. Berkeley: Univ. of California Press (in press).

Rashevsky, N. *Mathematical Biophysics: Physico-mathematical foundations of biology.* Third edition. 2 vol. New York: Dover, 1959-1960.

Symposium on Information Theory, London, 1955. *Information Theory: Papers read at Symposium held at the Royal Institute, London, Sept. 12th-16th, 1955.* Edited by C. Cherry. 401 pp. London: Academic Press, 1956.

Symposium on Information Theory, London. *Information Theory: Papers read at Symposium held at the Royal Institute, London, Aug. 30-Sept. 4, 1960.* Edited by C. Cherry (in press).

Interdisciplinary Conference on Self-Organizing Systems, Chicago, 1959. *Self-Organizing Systems: Proceedings.* Edited by M. C. Yovits and S. Cameron. 322 pp. New York: Pergamon, 1960. (*International Tracts in Computer Science and Technology and Their Application.* Vol. 2.)

Symposium on Information Theory in Biology: Held at Gatlinburg, Tennessee, October 29, 1956. Edited by H. P. Yockey, H. Quastler and R. Platzman. 418 pp. New York: Pergamon, 1958.

Bailey, N. T. J. *Mathematical Theory of Epidemics.* 194 pp. New York: Hafner, 1957.

Genetics and Protein Structure

Benzer, S. On topography of genetic fine structure. *Proc. Nat. Acad. Sc.* **47**:403, 1961.

Levinthal, C. Coding aspects of protein synthesis. *Rev. Mod. Physics* **31** (1):249-255, 1959.

Levinthal, C. Genetic and chemical studies with alkaline phosphatase of *E. coli. Brookhaven Symposia in Biol.* **12**:76-85, 1959.

Levinthal, C., Garen, A., and Rothman, F. Relationship of gene structure to protein structure: studies on alkaline phosphatase of *E. coli.* In *Fifth International Congress of Biochemistry: Moscow, 1961* (in press).

Bernhard, S. A. Simple model of molecular specificity in enzyme-substrate systems. I. Theory and applications to system acetylcholinesterase-substrate. *J. Am. Chem. Soc.* **77**:1966-1972, 1955.

Bernhard, S. A. Simple model of molecular specificity in enzyme-substrate systems. II. Correlation of Michaelis constant with inhibition constant. *J. Am. Chem. Soc.* **77**:1973, 1955.

Bernhard, S. A. Some mathematical consequences of interaction of enzyme with two reactant species. *Disc. Faraday Soc.* **20**:267, 1956.

Bernhard, S. A. Symposium on protein microstructure. *J. Polymer Sc.* (in press).

Bernhard, S. A., Katchalski, E., Berger, A., and Sela, M. Polyfunctional catalysis in polypeptides and enzymes. *J. Am. Chem. Soc.* (in press).

Mathematical Models (Physiology)

Podolsky, R. J. Nature of contractile mechanism of muscle. In *Biophysics of Physiological and Pharmacological Techniques.* Edited by A. M. Shanes. Washington, D. C.: American Association for the Advancement of Science (in press).

Shapiro, N., and Vreenegoor, H. Generalized important event technique. *J. A. Comp. Machinery* (in press).

Stephenson, J. L. Theory of measurement of blood flow by dilution of indicator. *Bull. math. Biophys.* **10**:117-121, 1948.

Stephenson, J. L. Theory of measurement of blood flow by dye dilution technique. *IRE Tr.*, pp. 81-88, December, 1958. (PGME-12.)

Stephenson, J. L. Theory of transport in linear biological systems. I. Fundamental integration equation. *Bull. math. Biophys.* **22**:1-17, 1960.

Stephenson, J. L. Theory of transport in linear biological systems. II. Multiflux problems. *Bull. math. Biophys.* **22**:113-138, 1960.

Medical Diagnosis

Proceedings of conference on diagnostic data processing. *IRE Tr.*, pp. 232-368, October, 1960. (ME-7.)

Neurophysiology

Communications Biophysics Group of Research Laboratory of Electronics, and Siebert, W. M. *Processing Neuroelectric Data.* Cambridge, Massachusetts: Massachusetts Institute of Technology, July 7, 1959. (Tech. Rep. No. 351.)

Barlow, J. S. Rhythmic activity induced by photic stimulation in relation to intrinsic alpha activity of brain in man. *Electroencephalog. & Clin. Neurophysiol.* **12**:317-326, 1960.

Clark, W. A., et al. Average response computer (ARC): digital device for computing averages and amplitude and time histograms of electrophysiological response. *IRE Tr.*, pp. 46-51, January, 1961. (BME-8.)

Gerstein, G. L., and Kiang, N. Y.-S. Approach to quantitative analysis of electrophysiological data from single neurons. *Biophys. J.* **1**:15-28, 1960.

Goldstein, M. H. Statistical model for interpreting neuro-electric responses. *Information & Control* **3**:1-17, 1960.

Peake, W. T. *An Analytical Study of Electric Responses at the Periphery of the Auditory System.* Cambridge, Massachusetts, Research Laboratory of Electronics, March 17, 1960. (RLE Tech. Rep. No. 365.)

Farley, B. G., Frishkopf, L. S., Clark, W. A., Jr., and Gilmore, J. T., Jr. *Computer Techniques for the Study of Patterns in the Electroencephalogram.* Technology Press, Cambridge, Massachusetts. Research Laboratory of Electronics, November 6, 1957. (RLE Tech. Rep. No. 337, & Lincoln Lab. Tech. Rep. No. 165.)

Année 1962

Bibliographie

[1] Marcel-Paul Schützenberger. Finite counting automata. *Information and Control*, 5 :91–107, 1962.

[2] Marcel-Paul Schützenberger. Certain infinite formal products and their combinatorial applications. In *Colloquium Comb. Methods Probab. Theory*, pages 58–63. Aarhus, 1962.

[3] Marcel-Paul Schützenberger. On a theorem of R. Jungen. *Proc. Amer. Math. Soc.*, 13 :885–890, 1962.

[4] Murray Eden and Marcel-Paul Schützenberger. Remark on a theorem of Dénes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 7 :353–355, 1962.

[5] Roger C. Lyndon and Marcel-Paul Schützenberger. The equation $a^m = b^n c^p$ in a free group. *Michigan Math. J.*, 9 :289–298, 1962.

[6] Marcel-Paul Schützenberger. On probabilistic push-down storages. In *Self-Organizing Systems, Proceedings*, pages 205–213. Spartan Books, Washington, 1962.

[7] Marcel-Paul Schützenberger. On an abstract machine property preserved under the satisfaction relation. Technical Report NC-167, IBM Thomas Watson Research Center, 1962.

[8] Marcel-Paul Schützenberger. On the minimum number of elements in a cutting set of words. Technical Report NC-173, IBM Thomas Watson Research Center, 1962.

[9] Marcel-Paul Schützenberger. On a family of formal power series. 11 pages, manuscrit, mars 1962.

Reprinted from INFORMATION AND CONTROL, Volume 5, No. 2, June 1962

INFORMATION AND CONTROL **5**, 91–107 (1962)

Finite Counting Automata

M. P. SCHÜTZENBERGER*

Harvard Medical School, Boston, Massachusetts

I. INTRODUCTION

The purpose of this note is to define a family $\mathfrak{R}_*$ of sets of words that is, in some sense, the simplest natural generalization of the family $\mathfrak{R}_0'$ of Kleene's (1956) regular events (cf, also, Bar-Hillel and Shamir (1960) and Shepherdson (1959) and below for an abstract definition). However, even if this point of view constitutes the main motivation and if it suggests the terminology, our treatment of the question will be entirely algebraic. In fact this paper can be considered as an attempt towards a classification of the (infinite) monoids of finite dimensional rational matrices which are the semidirect sum of finite monoids. A discussion of these points is to be found in Schützenberger (1962).

The set of F of the so-called *input words* (that is, the *free monoid* F generated by the finite set $X = \{x_i\}$) is assumed to be fixed. We recall that according to Bar-Hillel and Shamir (1960) a *regular event* F' is a subset F' of F such that $\varphi^{-1}\varphi F' = F'$ for some homomorphism φ of F into a finite monoid and our construction hinges upon the algorithm described in the following definition.

DEFINITION. A *finite counting automation* β of order q is the integral valued function of F that is given by:

(i) A finite set of $(q_j + 1)$-tuples $(\alpha_j) = (F'_{j,1}, F'_{j,2}, \cdots, F'_{j,q_j+1})$ of regular events $F'_{j,i}$ $(1 \leqq j \leqq M; q_1, q_2, \cdots, q_M \leqq q)$.

(ii) A polynomial $\bar{\beta}$ (with integral coefficients) in the variates $\alpha_1, \alpha_2, \cdots, \alpha_M$.

For each word f of F, $\beta f = \bar{\beta}(\alpha_1 f, \alpha_2 f, \cdots, \alpha_M f)$ where for each j, $\alpha_j f$ denotes the number of factorizations $f = f_1 f_2 \cdots f_{q_j+1}$ of f into $q_j + 1$ words such that $f_1 \in F'_{j,1}, f_2 \in F'_{j,2}, \cdots, f_{q_j+1} \in F'_{j,q_j+1}$.

(*) This work was done in part at the Department of Statistics of the University of North Carolina, under Contract AF 49 (638)-213 of the United States Air Force, and supported in part by a grant from the Commonwealth Fund.

The functions α_j themselves will be called *counters* and we shall say that β is a *linear finite counting automaton* if $\bar{\beta}$ reduces to a linear combination of the α_j's.

For instance, a counter α of order zero is defined by a single regular event F' and by the rule $\alpha f = 1$ if $f \in F'$, $\alpha f = 0$, otherwise. Hence, here α is, in fact, the *characteristic function* of F'.

Reciprocally, we define the *support* $F'(\beta)$ of the finite counting automaton β as the set of words $F'(\beta) = \{f \in F : \beta f \neq 0\}$.

It is easily verified that any finite counting automaton is equal to a linear one of sufficiently higher order and, denoting by $\mathfrak{R}_q$ the family of the supports of the linear finite counting automata of order q, we shall finally define $\mathfrak{R}_*$ as the union $\bigcup_{q \geq 0} \mathfrak{R}_q$.

Clearly $\mathfrak{R}_0'$ (the family of regular events) is a subset of $\mathfrak{R}_*$ and it can be shown without difficulty that $\mathfrak{R}_0' = \mathfrak{R}_0$. In order to show that the finite counting automata allow operations exceeding the power of the conventional *one way one tape automata* of Rabin and Scott (1959) it suffices to consider the following example (cf. Elgot (1956)):

Let the regular event F_{x_i} be defined by the condition that the word f belongs to F_{x_i} if and only if its last letter is x_i. The counter of order one α_i, defined by the pair (F_{x_i}, F), enumerates the number of times x_i appears in the input word f. Taking, for instance, $\bar{\beta} = \alpha_1 - \alpha_2$, the corresponding linear counting automaton β is such that f belongs to $F'(\beta)$ if and only if it does not contain as many x_1's as x_2's. Obviously with the same type of counters, but with a polynomial $\bar{\beta}'$ of order three or more, the problem of deciding if $F'(\beta')$ is or is not equal to F (or if it is or is not the complement of a finite set of words) leads to the classical difficulties of diophantine analysis. Hence, there is some interest in obtaining an independent characterization of the parameter q. For this purpose let us say that $\overline{\deg}\ \beta = q'$ if q' is the least integer such that for all nonempty words f the absolute value of βf is bounded by a constant multiple of the q'th power $|f|^{q'}$ of the length $|f|$ of f. It is trivial that $\overline{\deg}\ \beta$ is finite for any finite counting automaton because for any $q'' > 0$ the total number of factorizations of a word f into $q'' + 1$ factors is itself bounded by a constant multiple of $|f|^{q''}$.

Our main result (to be proved in Section II) is that for *any* finite counting automaton β: (i) $\overline{\deg}\ \beta$ is equal to the greatest lower bound of the (not necessarily integral) numbers $r \geq 0$ such that $\overline{\lim}_{|f| \to \infty} |f|^{-r} |\beta f| = 0$. (ii) There exists a *linear* finite counting automaton identically equal to β whose order is precisely $\overline{\deg}\ \beta$.

In fact each linear finite counting automaton β with $\bar{\deg}\,\beta = q > 0$ is closely associated with an extension by a finite monoid of a free nilpotent group of class at most q (of a free abelian group if $q = 1$). We intend to examine the special case of $\mathfrak{R}_1$ and $\mathfrak{R}_2$ in another paper.

In the last section of this paper we verify that $\mathfrak{R}_*$ is closed with respect to the operations of union, intersection, and set product and, by way of counterexamples, we show that nothing more of Kleene's (1956) theorem remains valid for $\mathfrak{R}_*$.

It may be mentioned that the family $\mathfrak{R}_*$ is a special case of the more general family of sets of words defined in Schützenberger (1961) and that it could be partially characterized by adding the following restriction to (a), (b), $\cdots$ (e) of Schützenberger (1961, p. 245).

(f). The ratio of the amount of information stored in the internal memory to the amount brought to the machine tends to zero with the length of the input word.

In the remainder of this section we reduce our original definition to a simpler form and we prove a few elementary results needed in Section II.

I.1. Every finite counting automaton is equal to a linear one.

PROOF. It is sufficient to prove that if α and α' are two counters, the function β defined by the identity $\beta f = \alpha f \alpha' f$ is equal to a linear finite automaton, and to use induction on the degree of the polynomial $\bar{\beta}$.

Let us recall first that to any finite family $\{F_j'\}$ of regular events F_j' there corresponds an homomorphism φ of the monoid F onto a finite quotient monoid $H = \varphi F$, and a collection $\{H_j'\}$ of subsets of H such that $f \in F_j'$ if and only if $\varphi f \in H_j'$ (Bar-Hillel and Shamir (1960)).

Hence to any counter α defined by a $(q + 1)$-tuple $(F_1', F_2', \cdots, F_{q+1}')$ of regular events contained in the family $\{F_j\}$, we can associate the finite set of all the $(q + 1)$-tuples $(\alpha_i) = (h_{i_1}, h_{i_2}, \cdots, h_{i_{q+1}})$ of elements of H which are such that

$$h_{i_1} \in H_1', h_{i_2} \in H_2', \cdots, h_{i_{q+1}} \in H_{i_{q+1}}'$$

Then α is equal to the linear finite counting automaton $\sum_i \alpha_i$, where for each i and input word f the counter α_i enumerates the number of distinct factorizations $f = f_1 f_2 \cdots f_{q+1}$ such that

$$\varphi f_1 = h_{i_1}, \varphi f_2 = h_{i_2}, \cdots, \varphi f_{q+1} = h_{i_{q+1}}'$$

Let us say that this factorization is *proper* if none of the words f_i is the empty word. For any $(q + 1)$-tuple $(h_1, h_2, \cdots, h_{q+1})$ of elements of $H = F$, we say that the function $\bar{\alpha}$ of F is a φ-counter if it

enumerates the number of proper factorizations with

$$\varphi f_1 = h_1 , \varphi f_2 = h_2 , \cdots , \varphi f_{q+1} = h_{q+1}$$

or, as we shall say, if it enumerates the number of α-factorizations of the input word. It is clear that any of the counters α_i above is equal to the sum of the φ-counters defined by the same $(q + 1)$-tuple and of all the $(2^{q+1} - 2)$ φ-counters defined by the $(q' + 1)$-tuples $(q' < q)$ which result from the deletion of one or of several h_{i_j}'s in $(h_{i_1} , h_{i_2} , \cdots , h_{i_{q+1}})$.

Consequently, it suffices to prove the statement for the special case in which both α and $\bar{\alpha}$ are φ-counters. Then, in fact, the function β enumerates for each word f the number of pairs consisting of a α- and of a α'-factorization of f.

Let us consider an arbitrary monoid G, a $(q + 1)$-tuple $(d) = (g_1 , g_2 , \cdots , g_{q+1})$ and a $(q' + 1)$-tuple $(d') = (g_1' , g_2' , \cdots , g_{q'+1}')$ of elements of G.

We say that $(d'') = (g_1'', g_2'', \cdots , g_{q''+1}'')$ is a *refinement* of (d) and (d') if it has the following properties:

Every g_i of (d) is equal to a product $g_i = g_k''$ or $g_k'' g_{k+1}'' \cdots g_{k'}''$, of consecutive elements of (d''); the same is true for every g_i' of (d'); every $g_{i''}''$ of (d'') is the first factor of a product corresponding to an element of (d) or of (d').

Because of this last condition $q'' \leqq q + q'$ and, consequently, (d) and (d') have only finitely many distinct refinements when G is finite. The same is true when $G = F$, a free monoid, because, e.g., any relation of the form $f = f'f''$ uniquely determines f'' for given f, $f' \in F$.

For instance, if $(d) = (g_1 , g_2)$ and $(d') = (g_1', g_2')$, the set $\{(d'')\}$ of their refinements is empty unless $g_1 g_2 = g_1' g_2'$. If this condition is met, $\{(d'')\}$ consists of the triples (g_1 , g_2'', g_2') with $g_1 g_2'' = g_1'$, $g_2'' g_2' = g_2$ and of the triples (g_1', g_2'', g_2) with $g_1' g_2'' = g_1$, $g_2'' g_2 = g_2'$. If $g_1 = g_1'$ and $g_2 = g_2'$, the set $\{(d'')\}$ contains also the pair (g_1 , g_2).

This definition concludes the proof because β enumerates all the α''-factorizations of f where (α'') is a refinement of (α) and (α') and, consequently, $\beta = \sum \alpha''$ where the summation is over all the φ-counters corresponding to these refinements.

We shall have to consider now and in the following section matrix-valued functions of F. It will always be assumed that the matrices under consideration are finite dimensional matrices with rational entries. The matrix-valued function μ is a *representation* of F if the square matrices μf are such that $\mu ff' = \mu f \mu f'$ identically for all words f and f'; it is a

finite representation if the set $\{\mu f : f \in F\}$ is finite; it is a *representation with bounded denominator* if $p\mu f$ is an integral matrix for some integer p and all words f.

Any entry $\nu_{ii'}$ of a matrix-valued function determines a numerical function $\beta f = (\nu f)_{ii'}$ of F, and the quantity $\overline{\deg}\ \nu$ will be defined as the supremum of the same symbol over all functions determined by the entries of ν.

I.2. To every finite counting automaton β there corresponds at least one representation μ such that β is determined by one of its entries.

PROOF: According to I.1 it is sufficient to prove the statement for a linear finite automaton β of order q.

Let φ be a fixed homomorphism of F onto a finite monoid H, and consider the set α_j $(1 \leqq j \leqq N)$ of all the φ-counters of order at most q. Let the jth coordinate at the vector $v(f)$ be equal to $\alpha_j f$ for each j and $f \in F$. We verify that for each letter x of X there exists a $N \times N$ integral matrix μx such that $v(fx) = v(f)\mu x$ identically.

This is trivial if $q = 0$ and we consider a fixed counter α of order $q > 0$ with $(\alpha) = (h_1, h_2, \cdots, h_{q+1})$. We denote by α' the φ-counter $(h_1, h_2, \cdots, h_q)$ of order $q - 1$ and by α_k'' $(1 \leqq k \leqq K)$ the set of all the φ-counters α_k'' of order q with $(\alpha_k'') = (h_1, h_2, \cdots, h_q, h'')$ where h'' satisfies the condition $h''\varphi x = h_{q+1}$. Since $\alpha f x = \alpha' f + \sum_k \alpha_k'' f$ or $\alpha f x = \sum_k \alpha_k'' f$ according to $\varphi x = h_{q+1}$ or $\varphi x \neq h_{q+1}$, our preliminary result is proved.

Since β is a linear combination of the α_j's, this shows that, in fact, β is determined by an automaton of the family $\mathfrak{A}$ described in the definition 1 of Schützenberger (1961) and the complete result follows from an elementary construction explained in detail in the same paper.

Let μ be a given representation of F. A representation $\bar{\mu}$ of the same dimension will be called a *finite part of order q* of μ if it is a finite representation and if for any $(2q - 1)$-tuple of words $(f_1, f_2, \cdots, f_{2q-1})$ the product

$$\bar{\mu}f_1\mu f_2\bar{\mu}f_3 \cdots \bar{\mu}f_{2i-1}\mu f_{2i}\bar{\mu}f_{2i+1} \cdots \mu f_{2q}\bar{\mu}f_{2q-1}$$

where $\bar{\mu}f_i = \mu f_i - \bar{\mu}f_i$, is identically zero. Thus the hypothesis that $\bar{\mu}$ is a finite part of μ implies that every matrix $\bar{\mu}$ belongs to the radical of the algebra generated by the matrices μx $(x \in X)$. We define Ord μ as the lowest possible order of a finite part of μ with the convention that Ord μ is infinite if μ has no finite part of finite order.

I.3. If Ord μ is finite, every function of F determined by an entry of μ

is equal up to a constant factor to a linear finite counting automaton whose order is at most Ord μ.

PROOF: If Ord $\mu = 0$, the statement is trivial because the hypothesis amounts to the assumption that μ itself is finite and we can assume now that $0 < q = \text{Ord } \mu$.

We consider an homomorphism φ of F onto the quotient monoid H that is defined by the following conditions:

(i) $\varphi f \neq \varphi e$ if f is not the empty word e.

(ii) For all $f, f' \in F$ and $x, x' \in X$, $\varphi fx = \varphi f'x' = h$, say if and only if $\bar{\mu}f\mu x = \bar{\mu}f'\mu x'$ and $\bar{\mu}fx = \bar{\mu}f'x'$.

By hypothesis H is a finite monoid. We define a representation μ (with finite part $\bar{\mu}$) of $H = \{h\}$ by setting $\mu h = \bar{\mu}f\mu x$ and $\bar{\mu}h = \bar{\mu}fx$. To every $(q+1)$-tuple $(\alpha) = (h_1, h_2, \cdots, h_{q'+1})$ of elements of H we associate the matrix

$$\mu\alpha = \bar{\mu}h_1\bar{\mu}h_2 \cdots \bar{\mu}h_i \cdots \bar{\mu}h_{q'}\bar{\mu}h_{q'+1}$$

where, of course, $\bar{\mu} = 0$ and $\bar{\mu}e =$ the unit matrix. Let now $f = x_1x_2 \cdots x_n$ be an arbitrary word expressed as a product of the generators. Since $\mu = \bar{\mu} + \bar{\mu}$ we have

$$\mu f = (\bar{\mu}x_1 + \bar{\mu}x_1)(\bar{\mu}x_2 + \bar{\mu}x_2) \cdots (\bar{\mu}x_n + \bar{\mu}x_n).$$

Developing this expression and observing that on the one hand $\bar{\mu}$ is a representation and on the other hand any product containing $q+1$ matrices $\bar{\mu}$ is zero, we obtain μf as a sum of terms of the form

$$\bar{\mu}f_1\bar{\mu}x_1\bar{\mu}f_2\bar{\mu}x_2 \cdots \bar{\mu}f_{q'}\bar{\mu}x_{q'}\bar{\mu}f_{q'+1}$$

with $q' \leqq q$. Clearly each of these terms is equal to some matrix $\mu\alpha$ as defined above and, more accurately, we have the identity $\mu f = \sum \alpha f \mu\alpha$ where the summation is over all the counters α of order at most q defined above.

Since the set of all these matrices $\mu\alpha$ is finite, it is trivial that $K\mu f$ is an integral matrix for all f of F and some fixed integer K and the result is proved.

If φ is any homomorphism of F onto a finite monoid we say that the $(2p+1)$-tuple $(s) = (f_1, f_2, \cdots, f_{2p+1})$ of elements of F is *φ-special* if for each $i = 1, 2, \cdots, p$

$$\varphi f_{2i-1}f_{2i} = \varphi f_{2i-1}, \varphi f_{2i}^2 = \varphi f_{2i}, \varphi f_{2i}f_{2i+1} = f_{2i+1}.$$

Since φf is finite, there corresponds to any $(s') = (f_1', f_2', \cdots, f_{2p+1}')$

a finite positive inter a such that

$$(s) = (f_1 f_2'^b, f_2'^a, f_2'^b f_3' f_4^b, f_4'^a, \cdots, f_{2p}'^a, f_{2p}'^b f_{2p+1}')$$

is φ-special for all large enough b.

Also we shall use the abbreviation $s^{(k)}$ for denoting the word $f_1 f_2^k f_3 f_4^k \cdots f_{2p}^k f_{2p+1}$.

I.4. If Ord $\mu = q$ is finite and if (s) is φ-special, there exists $q+1$ matrices ${}_0\mu s, {}_1\mu s, \cdots, {}_q\mu s$ such that for all k

$$\mu s^{(k)} = \sum_{0 \leqq j \leqq q} k^j {}_j\mu s$$

Proof. By straightforward computation using the development of $\mu s^{(k)}$ as a product of matrices $\bar{\mu}$ and $\bar{\bar{\mu}}$.

I.5. If $0 < Ord\ \mu < \infty$, μ is equivalent to a representation of the form $\begin{pmatrix} \mu' & \nu \\ 0 & \mu'' \end{pmatrix}$ where μ' is a finite representation and μ'' a representation with $Ord\ \mu'' = Ord\ \mu - 1$.

Reciprocally, if the representation μ is in the semireduced form

$$\mu = \begin{pmatrix} \mu' & \nu \\ 0 & \mu'' \end{pmatrix}$$

then $Ord\ \mu \leqq Ord\ \mu' + Ord\ \mu'' + 1$.

Proof: Let V denote the set of all the vectors v such that for all words f', $f''\bar{\mu}f'\mu f''v = 0$.

Because of the hypothesis that μ admits a finite part $\bar{\mu}$ of order $q = \text{Ord}\ \mu < \infty$, V is not empty and, after performing a suitable linear transformation, we can assume that V consists of all the vectors having their M last coordinates zero. Then, since for all f', $f'' \in F$ and $v \in V$ one has $\mu f''v \in V$ and $\bar{\mu}f'v = 0$, μ and $\bar{\bar{\mu}}$ have, respectively, the forms

$$\mu = \begin{pmatrix} \mu' & \nu \\ 0 & \mu'' \end{pmatrix} \quad \text{and} \quad \bar{\bar{\mu}} = \begin{pmatrix} 0 & \nu' \\ 0 & \bar{\mu}'' \end{pmatrix}$$

where μ' and μ'' are representations and where $\dim \mu'' = M$.

It follows that $\bar{\mu} = \mu - \bar{\mu}$ has the form

$$\begin{pmatrix} \mu' & \nu - \nu' \\ 0 & \mu'' - \bar{\mu}'' \end{pmatrix}$$

and, since it is a finite representation, the same is true of μ'.

Observe now that the module V^* spanned by all the row vectors of

all the matrices $\bar{\mu}f'\mu f''(f', f'' \in F)$ consists of the vectors having their first $N-M$ coordinates zero where $N = \dim \mu$ is strictly larger than M because of the hypothesis that $0 < \mathrm{Ord}\ \mu$.

Direct computation shows that any product $\bar{\mu}f_1\mu f_2\bar{\mu}f_3 \cdots \mu f_{2q-2}\bar{\mu}f_{2q-1}$ has the form $\begin{pmatrix} 0 & \mathbf{n} \\ 0 & \mathbf{m} \end{pmatrix}$ where $\mathbf{n} = \nu' f_1\mu'' f_2\,\mathbf{m}'$, $\mathbf{m} = \bar{\mu}'' f_1\mu' f_2\mathbf{m}'$ and $\mathbf{m}' = \bar{\mu}'' f_3\mu''\bar{\mu}f_4''f_5 \cdots \mu'' f_{2q-2}''\bar{\mu}f_{2q-1}$. Because of our remark above on V^*, the condition that $\mathbf{n} = \mathbf{m} = 0$ identically, which is implied by $\mathrm{Ord}\ \mu = q$, implies itself that $\mathbf{m}' = 0$ identically, that is, finally, that $\mathrm{Ord}\ \mu'' = q - 1$ and the direct part of the statement if proved.

With respect to the second part of the statement, it suffices to prove it for $\mathrm{Ord}\ \mu' = 0$ and to apply induction on $\mathrm{Ord}\ \mu$. However, if $\mathrm{Ord}\ \mu' = 0$ we can use the notations introduced above and the hypothesis that $\mathrm{Ord}\ \mu'' = q - 1$ implies that the matrix $\mathbf{m}'$ is identically zero. Hence, the matrices $\mathbf{m}$ and $\mathbf{n}$ are also identically zero and consequently $\mathrm{Ord}\ \mu \leqq q$. This concludes the proof of I.5.

II. VERIFICATION OF THE MAIN PROPERTY

Let ν be any matrix valued function of F. If

$$\overline{\lim}_{k\to\infty} |s^{(k)}|^{-1} |\nu s^{(k)}| = 0$$

for any $(2p+1)$-tuple (s) (and $s^{(k)}$ defined as in I.4), we write $\underline{\deg}\ \nu = 0$. If it is not so, there exists a largest integer q (possibly $q = \infty$) such that there exists an integer p and a $(2p+1)$-tuple (s) for which $\overline{\lim}_{k\to\infty} |s^{(k)}|^{-q} |\nu s^{(k)}| \neq 0$. Then we write $\underline{\deg}\ \nu = q$ and we say that (s) is *effective* for ν. Necessarily $\underline{\deg}\ \nu \leqq \overline{\deg}\ \nu$ and, if these two parameters are equal, their common values is the greatest lower bound of the numbers $r \geqq 0$ such that $\overline{\lim}_{|f|\to\infty} |f|^{-r} |\nu f| = 0$.

As for the symbols $\overline{\deg}$ and Ord, it is trivial that $\underline{\deg}\ \mu = \underline{\deg}\ \mu'$ for any representation μ' equivalent to μ.

Finally, let it be observed that (with the notations of I.4) $\underline{\deg}\ \mu$ can be defined as the largest q' such that ${}_{q'}\underline{\mu s} \neq 0$ for some $\bar{\mu}$-special $(2p+1)$-tuple (s). Indeed, under these last conditions $\overline{\lim}_{k\to\infty} |s^{(k)}|^{-q'} |\mu s^{(k)}|$ is proportional to ${}_{q'}\underline{\mu s}$. Reciprocally, given any effective $(2p+1)$-tuple (s') we can choose the integers a and b in a such a way that $(s) = (f', f_2'^b, f_2'^a, \cdots f_{2p}'^a, f_{2p}'^a f_{2p+1}')$ is both effective and φ-special for any fixed φ and in particular for $\varphi = \bar{\mu}$.

In order to simplify the proof of the main property II.4, we verify it separately in the special case of $\overline{\deg}\ \mu = 0, 1$. Our first and fundamental preliminary result is a modified version of a classical theorem of Burnside (1911, note j).

II.1. The three following conditions on a rationally irreducible representation μ with bounded denominator are equivalent:

(i) *Ord* $\mu = 0$.

(ii) *For all* $f, f', f'' \in F$ *and* $\epsilon > 0$, $\overline{\lim}_{k\to\infty}(1 - \epsilon)^k \mid \mu f' f^k f'' \mid = 0$.

(iii) *The set* $\{\mathrm{Tr}\ \mu f : f \in F\}$ *is finite.*

Proof: (i)⇒(ii). The condition (i) is equivalent to the condition that μ is a finite representation. Hence, it implies that $\underline{\deg}\ \mu = \overline{\deg}\ \mu = 0$. Since, trivially, $\underline{\deg}\ \mu = 0$ implies (ii) the result is proved.

(ii)⇒(iii). For any $f \in F$ and k, $\mathrm{Tr}\ \mu f^k$ is the sum of the kth powers of the characteristic roots ρ_j of μf. Hence, (ii) implies that for all $\epsilon > 0$, $\overline{\lim}_{k\to\infty} \sum_j \rho_j^k (1 - \epsilon)^k = 0$ and, consequently, that $\mid \rho_j \mid \leqq 1$ for every root ρ_j. It follows that $\mid \mathrm{Tr}\ \mu f \mid \leqq \sum_j \mid \rho_j \mid \leqq \dim \mu$, a bounded quantity. Since by hypothesis $\mathrm{Tr}\ \mu f$ is a rational number with bounded denominator, the implication (ii)⇒(iii) is proved.

(iii)⇒(i). Let $\{f_j\}(1 \leqq j \leqq N' \leqq N^2)$ be a basis of the module $\mathfrak{M}$ over the rationals spanned by all the matrices $\mu f (f \in F)$ and write $f' \equiv f''$ if and only if for all $j = 1, 2, \cdots, N'$ one has $\mathrm{Tr}\ \mu f' f_j = \mathrm{Tr}\ \mu f'' f_j$. The condition (iii) implies that the equivalence $\equiv$ has only finitely many classes and it suffices to verify that in fact $f' \equiv f''$ only if $\mu f' - \mu f'' = 0$.

Indeed, let $\mathfrak{M}'$ be the module of all matrices m' of $\mathfrak{M}$, such that for all $m \in \mathfrak{M}$, $\mathrm{Tr}\ m'm = 0$. By definition, for any $m' \in \mathfrak{M}'$, $m \in \mathfrak{M}$ and k, $\mathrm{Tr}\ m'^k m = 0$ and, consequently, all the characteristic roots of $m'm$ are zero. Hence, for given $m' \in \mathfrak{M}'$, there exists no $m \in \mathfrak{M}$ such that the first row of $m'm$ is the vector $(1, 0, 0, \cdots, 0)$. Since the representation μ is assumed to be irreducible, this shows that the first row of m' is the zero vector. The same remark applies to any row of any matrix of $\mathfrak{M}'$ and it shows that this set reduces to the zero matrix. By definition, $f' \equiv f''$ only if $\mu f' - \mu f'' \in \mathfrak{M}'$ and the proof of (iii)⇒(i) (and, consequently, of I.1) is completed.

Let us consider a representation $\mu = \begin{pmatrix} \mu' & \nu \\ 0 & \mu'' \end{pmatrix}$ such that $\bar{\mu} = \begin{pmatrix} \mu' & 0 \\ 0 & \mu'' \end{pmatrix}$ is a finite representation. If v is any vector, it follows from I.5 that the vector valued function μv (defined as $\mu f v$ for each f of F) satisfies the

inequality $\underline{\deg}\ \mu v \leqq \overline{\deg}\ \mu v \leqq \operatorname{Ord} \mu = 0$ or $= 1$. There is no loss in generality in assuming that after a suitable linear transformation $V = \{v : \underline{\deg}\ \mu v = 0\}$ consists of the vectors having their last M coordinates zero where, possibly, $M = 0$. We say then that μ is in *standard form*.

II.2. Under the hypothesis stated

$$\mu'' = \begin{pmatrix} \mu_0'' & \nu_0'' \\ 0 & \mu_1 \end{pmatrix} \quad \text{and} \quad \mu = \begin{pmatrix} \mu' & \nu_0' & \nu_1' \\ 0 & \mu_0'' & \nu_0'' \\ 0 & 0 & \mu_1 \end{pmatrix}$$

where both $\mu_0 = \begin{pmatrix} \mu' & \nu_0' \\ 0 & \mu_0'' \end{pmatrix}$ *and* μ_1 *are finite representations with* $\dim \mu_1 = M$.

PROOF. By hypothesis the monoid $\bar{\mu}F = \{\bar{\mu}f : f \in F\}$ has a finite number H of elements. If the triple of words $(t) = (f', f, f'')$ is such that $\bar{\mu}f'f = \bar{\mu}f'$, $\bar{\mu}ff'' = \bar{\mu}f''$, $|f| > 0$ we write $(t) \in T$ (or $\in T_{\bar{\mu}}$).

Trivially, any word g of F of length $|g| > H^2$ admits at least one factorization $g = g_1g_2g_3$ with $(g_1, g_2, g_3) \in T$. Direct computation shows that if $(t) = (f', f, f'') \in T$, the matrix $\mu t^{(k)} = \mu f' f^k f''$ is equal to $\mu t^{(0)} = \mu f' f''$ plus k times the matrix $\begin{pmatrix} 0 & \nu t \\ 0 & 0 \end{pmatrix}$ where, by definition, $\nu t = \mu' f' \nu f \mu'' f''$.

It follows that either $\nu t' = 0$ for all (t') of T and, then, the monoid μF contains at most H^2 distinct elements or, otherwise, $\nu t' \neq 0$ for at least one $(t') \in T$ (which is an *effective* triple) and $\underline{\deg}\ \mu = \overline{\deg}\ \mu = \operatorname{Ord} \mu = 1$. More generally, for any fixed vector w, either $\begin{pmatrix} 0 & \nu t \\ 0 & 0 \end{pmatrix} w = 0$ for all (t) of T and then the set $\{\mu f w : f \in F\}$ contains at most H^2 distinct vectors, or, otherwise, $\underline{\deg}\ \mu w = 1$.

Thus, we can assume now that $M > 0$ and, trivially, $M \leqq \dim \mu''$ since μ' is a finite representation. According to the definition of V, it follows that μ has the form

$$\begin{pmatrix} \mu' & \nu_0' & \nu_1' \\ 0 & \mu_0'' & \nu_0'' \\ 0 & \mu'' & \mu_1 \end{pmatrix}$$

where the following conditions are satisfied:

(i) $\dim \mu_1 = M$;

(ii) The module spanned by all the row vectors of all the matrices

$\nu_1' t$ $((t) \in T)$ has rank equal to its dimension M;

(iii) $\nu_0' t = 0$ for all (t) of T.

Observe now that for any $f''' \in F$ and $(t) = (f', f, f'') \in T$ the triple $(f') = (f', f, f''f''')$ also belongs to T. Consequently, since $\nu_0' t' = \nu_1' t \mu''' f''' = 0$ by (iii), it follows from (ii) that μ''' is identically zero. Since we have seen that $\{\nu_0' f : f \in F\}$ is a finite set, the proof of II.2 is completed.

II.3. Let the representation

$$\mu = \begin{pmatrix} \mu_0 & \nu_1 & \nu_3 \\ 0 & \mu_1 & \nu_2 \\ 0 & 0 & \mu_2 \end{pmatrix}$$

be in the standard form of II.2 with:

$$\mu_0 \to \begin{pmatrix} \mu' & \nu_0' \\ 0 & \mu_0'' \end{pmatrix}, \quad \nu_1 \to \begin{pmatrix} \nu_1' \\ \nu_0'' \end{pmatrix}, \quad \mu_1 \to \mu_1$$

and satisfy the conditions:

$$\underline{\deg} \begin{pmatrix} \mu_0 & \nu_1 \\ 0 & \mu_1 \end{pmatrix} = \operatorname{Ord} \begin{pmatrix} \mu_0 & \nu_1 \\ 0 & \mu_1 \end{pmatrix} = 1; \quad \underline{\deg}\, \mu_2 < \underline{\deg} \begin{pmatrix} \mu_1 & \nu_2 \\ 0 & \mu_3 \end{pmatrix} (= q < \infty).$$

Then $\underline{\deg}\, \mu \geqq q + 1$.

Proof: According to the remarks made at the beginning of this section, (ii) implies that $\begin{pmatrix} \mu_1 & \nu_2 \\ 0 & \mu_3 \end{pmatrix}$ admits at least one effective $(2p+1)$-tuple $(s) = (f_1, f_2, \cdots, f_{2p+1})$ which is $\bar{\mu}$-special. Thus, $\underline{\deg}\, \mu \geqq q + 1$ unless ${}_j\mu s = 0$ for all $j \geqq 0$ as we shall assume now. Then, by hypothesis,

$${}_q\mu s = \begin{pmatrix} 0 & \mathbf{n}_1' & \mathbf{n}_3' \\ 0 & 0 & \mathbf{n}_2' \\ 0 & 0 & 0 \end{pmatrix}$$

with $n_2' \neq 0$.

Because of the hypothesis that $\begin{pmatrix} \mu_0 & \nu_1 \\ 0 & \mu_1 \end{pmatrix}$ is in standard form, there exists at least one triple $(t') = (g_1, g_2', g_3)$ satisfying the conditions of II.2 which is such that $\nu t' \mathbf{n}_2 \neq 0$. By taking c large enough we can deduce from (t') a triple $(t) = (g_1, g_2 = g_2'^c, g_3)$ which is $\bar{\mu}$-special

and for which we have

$$ {}_1\underline{\mu}t = \begin{pmatrix} 0 & \mathbf{n}_1 & \mathbf{n}_3 \\ 0 & 0 & \mathbf{n}_2 \\ 0 & 0 & \mathbf{m}_1 \end{pmatrix} $$

with $\mathbf{n}_1\mathbf{n}_2 = c\nu_1' t\mathbf{n}_2 \neq 0$.

We claim that by a suitable choice of $a, b > 0$ the $(2p+3)$-tuple $(u) = (g_1, g_2^a, g_3 f_1, f_2^b, f_3, \cdots, f_{2i}^b, \cdots, f_{2p+1})$ which is $\bar{\mu}$-special by construction, satisfies the inequality ${}_{q+1}\underline{\mu}u \neq 0$ from which the result instantly follows.

Indeed, we have $\mu u^{(k)} = \sum_{0 \leqq j''} k^{j''} \left(\sum_{j+j'=j''} a^j b^{j'} {}_j\underline{\mu}t_{j'}\underline{\mu}s \right)$ and, because of the linear independence of the monomials $a^j b^{j'}$, it suffices to show that ${}_j\underline{\mu}t_{j'}\underline{\mu}s \neq 0$ for at least one pair (j, j') such that $q+1 = j + j'$. Since for $j = 1$ and $j' = q$ we have

$$ {}_1\underline{\mu}t_q\underline{\mu}s = \begin{pmatrix} 0 & 0 & \mathbf{n}_1\mathbf{n}_2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq 0, $$

the statement II.3 is proved.

II.4. If μ is a representation with bounded denominator, then $\underline{\deg}\ \mu = \overline{\deg}\ \mu = \mathrm{Ord}\ \mu$. Furthermore, if $q = \mathrm{Ord}\ \mu$ is finite, μ has an effective $(2q+1)$-tuple which is $\bar{\mu}$-special for some finite part $\bar{\mu}$ of order q of μ.

Proof: Since $\underline{\deg} \leqq \overline{\deg} \leqq \mathrm{Ord}$, trivially, we have only to prove that $\underline{\deg} \geqq \overline{\deg}$ and $\overline{\deg} \geqq \mathrm{Ord}$. The proof is by induction on $\dim \mu$, the initial case being trivial.

If μ is irreducible, the result has already been proved in II.1 since this remark shows that $\underline{\deg}\ \mu = \overline{\deg}\ \mu = \mathrm{Ord}\ \mu = 0$ or $=$ infinity. In the latter case, the condition (ii) of II.1 shows that there exists an effective triple.

Consequently, we can assume now that $\mu = \begin{pmatrix} \mu_0 & \nu_0 \\ 0 & \mu_1 \end{pmatrix}$ with $\dim \mu_0 > 0$ where μ_0 is irreducible and where (by I.5 and the induction hypothesis) $\underline{\deg}\ \mu_1 = \overline{\deg}\ \mu_1 = \mathrm{Ord}\ \mu_1 = \overline{\deg}\ \mu$ or $= \overline{\deg}\ \mu - 1$. Again the result is trivial unless $q = \deg \mu$ is finite as we shall always assume it now.

If μ_1 is a finite representation (in particular if it is irreducible), the result is already proved by II.2 which shows that $\underline{\deg}\ \mu = \overline{\deg}\ \mu = \mathrm{Ord}\ \mu = 1$ or $= 0$ according as there exists or not an effective triple.

Consequently, we can assume that Ord $\mu_1 = q_1 > 0$ and, by I.5, that

$$\mu_1 = \begin{pmatrix} \mu_{10} & \nu_{11} \\ 0 & \mu_{11} \end{pmatrix}$$

where Ord $\mu_{10} = 0$ and Ord $\mu_{11} = q_1 - 1$.

Then, μ has the form

$$\begin{pmatrix} \mu_0 & \nu_{00} & \nu_{01} \\ 0 & \mu_{10} & \nu_{11} \\ 0 & 0 & \mu_{11} \end{pmatrix}$$

and, by applying II.2, we can bring $\begin{pmatrix} \mu_0 & \nu_{00} \\ 0 & \mu_{10} \end{pmatrix}$ into standard form. Thus, finally, μ has the form

$$\mu = \begin{pmatrix} \mu_0 & \nu_{000} & \nu_{001} & \nu_{011} \\ 0 & \mu_{100} & \nu_{100} & \nu_{110} \\ 0 & 0 & \mu_{101} & \nu_{111} \\ 0 & 0 & 0 & \mu_{11} \end{pmatrix} = \begin{pmatrix} \mu_0' & \nu_0' \\ 0 & \mu_1' \end{pmatrix}$$

where dim $\mu_{101} = 0$ if and only if $\underline{\deg} \begin{pmatrix} \mu_0 & \nu_{00} \\ 0 & \mu_{10} \end{pmatrix}$ is 0. In any case, we have $\mu_0' = \begin{pmatrix} \mu_0 & \nu_{000} \\ 0 & \mu_{100} \end{pmatrix}$ = a finite representation, $\mu_1' = \begin{pmatrix} \mu_{101} & \nu_{111} \\ 0 & \mu_{100} \end{pmatrix}$ = a representation with Ord $\mu_1' = q_1$ or $q_1 - 1$.

Let us distinguish the two possibilities:

(i) μ_1' admits a finite part $\bar{\mu}_1'$ of order $q_1 - 1$. Then $\begin{pmatrix} \bar{\mu}_0' & 0 \\ 0 & \bar{\mu}_1' \end{pmatrix}$ is a finite part of order q_1 of μ. Since Ord $\mu \geqq \overline{\deg}\ \mu = q$, this shows that $q_1 = q$. Hence, trivially, Ord $\mu = \overline{\deg}\ \mu$ and by the induction hypothesis $\underline{\deg}\ \mu_1 = q$ with an effective $(2q + 1)$-tuple of the required type. Since $\underline{\deg}\ \mu$ is at most equal to q and at least equal to $\underline{\deg}\ \mu_1$, the double equality is proved.

(ii) Ord $\mu_1' = q_1$. Since Ord $\mu_{11} = q_1 - 1$, by construction, we have surely dim $\mu_{101} \neq 0$ and we can apply II.3 with the correspondence $\mu_0' \to \mu_0$, $\mu_{101} \to \mu_2$ and $\mu_{11} \to \mu_3$. This shows that $\underline{\deg}\ \mu \geqq q_1 + 1$ and consequently, $q_1 = q - 1$. It follows that $\underline{\deg}\ \mu = \overline{\deg}\ \mu = q$ with the required type of effective $(2q + 1)$-tuple. Furthermore, μ_1 admits a finite part $\bar{\mu}_1$ of order $q - 1$ and, consequently, $\begin{pmatrix} \mu_0 & 0 \\ 0 & \bar{\mu}_1 \end{pmatrix}$ is a finite part

of μ of order at most q. This proves the double equality in the second case and it concludes the proof of the main property.

III. TWO COUNTEREXAMPLES

It has been seen in I.1 that any finite counting automaton is equal to a linear one of sufficiently higher order. Our first counter example is intended to show that, of course, the converse proposition is not true.

III.1. There exists at least one linear counting automaton of order two such that its support cannot be the support of any finite counting automaton of order one.

PROOF: Let the regular events Fx_i and x_iF be defined by the condition that f belongs to Fx_i (respectively to x_iF) if and only if its last letter (respectively its first letter) is x_i . Assume for simplicity that $X = \{y, z\}$ and define $\beta = \alpha - \alpha'$ where $(\alpha) = (Fy, F, zF)$ and $(\alpha') = (Fz, F, yF)$. Direct computation shows that if $f_i = y^iz$, $f_j = y^jz$, $i \neq j$ there exists for every $p > 0$ at least one word f' and integer k such that $\beta f_i f_i^{kp} f_j^{kp} f_j f' = 0$ and $\beta f_i f_j^{kp} f_i^{kp} f_j f' \neq 0$.

Let now β' be a finite counting automaton of order one defined by a polynomial function of the linear finite counting automata $\beta_i (i = 1, 2, \cdots, M)$ of order one. Using the notations of II.2, we assume that each β_i is determined by an entry of a representation μ_i admitting a finite part $\bar{\mu}_i$ of order one. Hence, there exists an homomorphism φ such that $\varphi f = \varphi f'$ if and only if $\bar{\mu}_i f = \bar{\mu}_i f'$ for $i = 1, 2, \cdots, M$ and φF is a finite monoid.

Trivially, if $\varphi f = \varphi f'$, there exists a finite p such that for each i, $\mu_i f^p$ is idempotent and direct computation shows that $\mu_i f f^{kp} f' f^{kp} f' = \mu_i f f'^{kp} f^{kp} f'$, hence $\beta f f^{kp} f'^{kp} f' = \beta f f'^{kp} f^{kp} f'$, for every k.

However, the words f_i considered above constitute an infinite family of words which, pairwise, do not satisfy this relation whence the conclusion follows instantly.

We now prove the closure properties of $\mathfrak{R}_*$.

III.2. The family of the supports of the finite counting automata of order at most q is closed under intersection and union.

PROOF: Let β and β' be two finite counting automata of order at most q. According to the very definition of this algorithm, the function β'' and β''' of F defined respectively by the identities $\beta'' f = \beta f \beta' f$ and $\beta''' f = (\beta f)^2 + (\beta' f)^2$ are also finite counting automata of order at

most q and we have

$$F'(\beta'') = \{f \in F : \beta f \beta' f \neq 0\} = F'(\beta) \cap F'(\beta')$$

$$F'(\beta''') = \{f \in F : (\beta f)^2 + (\beta' f)^2 \neq 0\} = F'(\beta) \cup F'(\beta'').$$

III.3. The family $\mathfrak{R}_*$ is closed under set product.

PROOF: Observe that if $(\alpha') = (F_1', F_2', \cdots, F_{q'}')$ and $(\alpha'') = (F_1'', F_2'', \cdots, F_{q''}'')$ define the two counters α' and α'', the $(q' + q'')$-tuple (α''') of regular events $(\alpha''') = (F_1', F_2', \cdots, F_{q'}', F_1'', F_2'', \cdots, F_{q''}'')$ defines a counter α''' which satisfies the convolution identity $\alpha'''f = \sum \{\alpha' f' \alpha'' f'' : f'f'' = f\}$.

Since the convolution product is distributive over the addition, we can associate to any pair of finite counting automata β' and β'' a third finite counting automaton β''' such that $\beta'''f = \sum \{(\beta' f')^2 (\beta'' f'')^2 : f'f'' = f\}$, identically, and the result is proved since, by construction, $F'(\beta''') = F'(\beta')F'(\beta''')$.

It has been shown elsewhere (Schützenberger (1961, counterexamples II.2 and II.3)) that the family of the sets of the form

$$\bar{F}(\beta) = F - F'(\beta) = \{f \in F : \beta f = 0\}$$

is distinct from $\mathfrak{R}_*$ and that it is *not* closed under the formation of set products.

III.4. For each $q > 0$, $\mathfrak{R}_{q-1} \neq \mathfrak{R}_q$ and, consequently, $\mathfrak{R}_*$ is not closed under Kleene's star operation $*$.

PROOF: Let again $X = \{y, z\}$ and define the following regular events:

$$Y^* = \{y, y^2, \cdots, y^n, \cdots\}, \qquad Z^* = \{z, z^2, \cdots, z^n, \cdots\}$$

$$G_q = (Y^*Z^*)^q \text{ (with } G_0 = \{e\}).$$

Hence, $f \in G_q$ if and only if $f = y^{k_1} z^{k_1'} \cdots y^k p z^{k_p'} \cdots y^{k_q} z^{k_q'} = g^{(K)}$, say, where all the coordinates $k_1, k_1', \cdots, k_q, k_q'$ of the vector K are positive integers.

If (α_p) denotes the pair $(G_p(Y^* \cup \{e\}), G_{q-p})$ of regular events, it is clear that the corresponding counter α_p of order one is such that $\alpha_p f = k_p$ if $p \in G_q$, $\alpha_p f = 0$, otherwise. A similar construction holds for k_p' and it follows that the following function β_q is a linear finite counting automata of order q:

$\beta_q f = 0$ if f is not in G_q ;

$$\beta_q f = (k_1 - k_1')(k_2 - k_2') \cdots (k_q - k_q') \quad \text{if} \quad f = g^{(K)}.$$

Hence, f belongs to $F'(\beta_q)$ in all cases except if $f \in G_q$ and if $k_p = k_p'$ for some pair of coordinates of K. We show that $F'(\beta)$ does not belong to $\mathfrak{R}_{q-1}$.

Indeed, by II.4 any linear finite counting automaton β' of order q' is determined by some entry of a matrix representation μ of F admitting a finite part $\bar{\mu}$ of order q'.

For suitable integers a and b the $(4q + 1)$-tuple $(s) = (y^b, y^a, y^b z^b, z^a, z^b y^b, \cdots, y^b z^b, z^a, z^b)$ is $\bar{\mu}$-special and for any vector $\bar{K}$ the word $s^{(K)} = y^{2b+a\bar{k}_1} z^{2b+a\bar{k}_1'} \cdots y^{2b+a\bar{k}_q} z^{2b+a\bar{k}_q'}$ is equal to $g^{(K)}$ where $K = 2bU + a\bar{K}$ with $U = (1, 1, \cdots 1)$.

Consequently, according to I.3, $\beta' s^{(K)}$ is a polynomial, say $b'(\bar{K})$, of degree at most Ord β' in the coordinate of $\bar{K}$. Now, if $F(\beta') = F(\beta_q)$ they have the same intersection with the set $\{s^{(\bar{K})}\}$ and, consequently, $b'(\bar{K})$ must be zero whenever $k_i = k_i'$ for some $i \leqq q$. Hence $b'(\bar{K})$ has degree at least q since it admits the product $(\bar{k}_1 - \bar{k}_1')(\bar{k}_2 - \bar{k}_2') \cdots (\bar{k}_q - \bar{k}_q')$ as a factor.

This concludes the proof that if $F(\beta') = F(\beta_q)$ then $F(\beta')$ is not contained in $\mathfrak{R}_{q-1}$.

We have seen that $G_1 = Y^* Z^*$ belongs to $\mathfrak{R}_1$ and by definition $G_1^* = \cup\{G_q : q > 0\}$.

By the same argument as above it follows that $F(\beta')$ cannot be equal to G_1^* if β' has a finite order since this would imply that $\bar{b}'(\bar{K})$ has infinite degree. Thus, $\mathfrak{R}_*$ is not closed under Kleene's star operation.

Of course for any set F' of $\mathfrak{R}_*$ it is possible to construct a *finite* dimensional integral representation μ of F such that a word f belongs to F'^* if and only if some fixed entry of μf is not zero [cf. Schützenberger (1961), p. 258 and 265]. Thus, as a byproduct, we have obtained the result that $\mathfrak{R}_*$ is a proper subfamily of the family of the sets words accepted by the automata of $\mathfrak{A}$.

RECEIVED January 23, 1962

REFERENCES

BAR-HILLEL, Y., AND SHAMIR, E. (1960). *Bull. Research Council Israel* **8F**, 155.

BURNSIDE, W. (1911). "Theory of Groups of Finite Order," 2nd ed. Cambridge Univ. Press.

ELGOT, C. C. (1960). *Trans. Am. Math. Soc.* **9º**, 61.

KLEENE, S. C. (1956). *In* "Automata Studies." Princeton Univ. Press, Princeton, New Jersey.

RABIN, M., AND SCOTT, D. (1959). *I.B.M. J. Research* **3,** 114.

SCHÜTZENBERGER, M. P. (1961). *Information and Control* **4,** 245.

SCHÜTZENBERGER, M. P. (1962). Certain elementary families of automata. *Proc. Polytech. Inst. Brooklyn. Symposium on Math. Theory of Automata.*

SHEPHERDSON, J. C. (1959). *I.B.M. J. Research* **3,** 198.

Matematisk Institut
Aarhus Universitet
Danmark

Colloquium on

COMBINATORIAL METHODS IN PROBABILITY THEORY

August 1 - 1o, 1962

CERTAIN INFINITE FORMAL PRODUCTS AND THEIR COMBINATORIAL APPLICATIONS

M. Schützenberger
Boston

58.

Colloquium
Aarhus, August 1-1o, 1962.

CERTAIN INFINITE FORMAL PRODUCTS AND THEIR COMBINATORIAL APPLICATIONS

M. Schützenberger
Boston

I. Introduction.

This note is concerned with certain relations between properties of factorization of a free monoid (in a manner similar to that considered in group theory) and properties of words modulo a cyclic permutation of their letters.

From these we deduce identities involving infinite formal product in non-commuting variables that are related to combinatorial questions.

II. Notations.

1) For any set X we denote by $\mathcal{F}_X$ the free monoide generated by X (with neutral element 1) and $\mathcal{F}_X^+ = \{ f \in \mathcal{F}_X : f \neq 1 \}$.

A factorization $\mathcal{F}_X = \prod \{ F_i : i \in I \}$ of $\mathcal{F}_X$ is a collection of submonoids $\mathcal{F}_i \subset \mathcal{F}_X$ indexed by the elements of a totally ordered set I such that the following is true.

To any $f \in \mathcal{F}_X$ there corresponds a unique finite subset $I_f \subset I$ and to each $i \in I_f$ a unique element, say, $\alpha_i f \in F_i^+$ such that $f = \alpha_{i_1} f \alpha_{i_2} f \ldots \alpha_{i_m} f$ where $I_1 = \{ i_1 < i_2 < \ldots < i_m \}$.

For instance, if $I = \{1,2\}$, $\{ F_1, F_2 \}$ form a factorization of $\mathcal{F}_X$ if every $f \in \mathcal{F}_X$ has one and only one factorization $f = f_1 f_2$ with $f_1 \in F_1$, $f_2 \in F_2$. The less straight forward definition above is needed for covering the case where I is not finite.

2) Let $f \sim f'$ (f and f' are conjugate) if and only if there exist $f'', f''' \in \mathcal{F}_X$ such that $f = f''f'''$, $f' = f'''f''$. In fact, since $\mathcal{F}_X$ is a submonoid of the free group generated by X, $f \sim f'$ if and only if they are conjugate in the usual sense because then, $f \equiv f''f'f''^{-1}$.

3) Let X and Y be two sets and $\varphi : \mathcal{F}_Y \to \mathcal{F}_X$ be a homomorphism. We shall say that φ is a c-homomorphism if and only if

(i) $\varphi^{-1}1 = 1$

(ii) for all $g, g' \in \mathcal{F}_Y$ and $f \in \mathcal{F}_X \setminus \varphi\mathcal{F}_Y$ $\varphi g\ f \neq f \varphi g'$

Then the subset φY of $\mathcal{F}_X$ will be called "c-free". It is clear that a c-free set $A \subset \mathcal{F}_X$ generates a free submonoid (noted $\{1\} \cup A^*$) or in equivalent fashion that a c-homomorphism is a monomorphism because the condition (ii) above is stronger than the condition

U_d) For all $f \in \mathcal{F}_Y \setminus \mathcal{F}_X$

$$(\varphi\mathcal{F}_Y)f \cap f(\varphi\mathcal{F}_Y) \cap \varphi\mathcal{F}_Y = \emptyset$$

which, as it is well known, insures that φ is 1-1 (into).

III. A preliminary result.

Let us denote by $\widetilde{A}^*$ for any $A \subset \mathcal{F}_X$ the set of all conjugates of the words belonging to the least stable subset A^* that contains A. We have

Property 1. If the three submonoids F_1, F_2, F_3 of $\mathcal{F}_X$ form a factorization of $\mathcal{F}_X$ and are generated by A_1, A_2, and A_3 respectively, then

(i) A_1, A_2 and A_3 are c-free

(ii) $\{\tilde{A}_1^*, \tilde{A}_2^*, \tilde{A}_3^*\}$ is a partition of $\mathcal{F}_X^+$

The proof which is not difficult is based upon the remark that (i) and (ii) are trivially satisfied when $\{A_1 \cap X, A_2 \cap X, A_3 \cap X\}$ is not a proper partition of X. On the contrary when, e.g.

$A_1 \cap X = X'$ with $\emptyset \neq X' \neq X$ there exists a factorization of F into the three monoids $\{1\} \cup X'^*$, $\{1\}$, $\{1\} \cup X'' \mathcal{F}_X$ $(X'' = X \setminus X')$ where the last one is generated by the c-free set $X'' \cup X''X'^*$. Observe that the uniqueness of the factorization implies that no word of $A_1 \setminus X' \cup A_2 \cup A_3$ begins with a letter from X'. This allows to "eliminate" X', by considering a set Y and a monomorphism $\psi : \mathcal{F}_Y \to \{1\} \cup X'' \mathcal{F}_X$. Then $\mathcal{F}_Y$ has a factorization $\{G_1, G_2, G_3\}$ such that $\psi G_2 = F_2$, $\psi G_3 = F_3$ and $\psi G_1 = F_1 \setminus X'^*$. Furthermore one can show that ψ is a c-homomorphism and that the truth of (i) and (ii) for $\{G_1, G_2, G_3\}$ implies the truth of (i) and (ii) for $\{F_1, F_2, F_3\}$. Iterating 2n times this construction gives a free monoid $\mathcal{F}_Z$, with a factorization $\{H_1, H_2, H_3\}$ and a c-homomorphism $\bar{\psi} : \mathcal{F}_Z \to \mathcal{F}_Y$ such that $\bar{\psi} H_2 = F_2$ and that all the words of degree less than n in $\bar{\psi}\mathcal{F}_Z$ belong to $\bar{\psi} H_2$. Since n is arbitrarily large, this gives the possibility of proving that A_2 is c-free and the rest of the proof is rather straight forward.

The method is essentially that of Lazard [2]. The same technique shows:

Property 2. To any partition $\{F'_1, F'_2, F'_3\}$ of $\mathcal{F}^+_X$, there corresponds one and only one triple of subsets $A_1 \subset F_1$, $A_2 \subset F'_2$, $A_3 \subset F'_3$ satisfying the hypothesis of Property 1. Finally let $\{F_i\}$ $(i \in I)$ be a factorization of F such that the sets A_i $(i \in I)$ generating the submonoids F_i have the two properties

(i)' The sets A_i $(i \in I)$ are c-free

(ii)' The sets A_i^* $(i \in I)$ form a partition of $\mathcal{F}^+_X$.

Taking any one of the F_i's, say F_j we can construct a factorization $F_{j_1}, F_{j_2}, F_{j_3}$ of this monoid satisfying the hypothesis of Property 1.

The same argument shows that the collection $\{F_{i'}\}$ $i' \in I'$ with I' obtained by replacing in I the element j by the triple $(j,1)$, $(j,2)$, $(j,3)$ still satisfies (i)' and (ii)'.

Now let us consider an injective mapping μ of $\mathcal{F}^+_X$ into the interval $[0,1]$, each number of this interval being represented by its ternary expansion

$$r = \sum \{r_n 3^{-n},\ n>0\} \qquad r_n = 0,1,2 \ .$$

The first digit of μf gives a partition $\mathcal{F}'_j = \{f \in \mathcal{F}_X^+ : (\mu f)_1 = j\}$ (j=0,1,2) from which we can derive a factorization F_1, F_2, F_3 of $\mathcal{F}_X$ by Property 2. Then, using the second digit in an obvious fashion we obtain a factorization of each of the three monoids F_1, F_2, F_3.

Passing to the limit we obtain a subset $H = \{h\}$ of elements of $\mathcal{F}_X$ and a total order $<$ on H having the following properties:

1) $h<h'$ if and only if $\mu h < \mu h'$

2) The collection $\mathcal{H}$ of all the monoids $\{1\} \cup h^*$ form a factorization of F.

3) Every $f \in \mathcal{F}_X^+$ is conjugate to some power of one and only one $h \in \mathcal{H}$.

Because of the property expressed by 3 we shall say that H is a "cyclic transversal" of $\mathcal{F}_X$.

IV Formal products.

Let us now consider $\mathcal{A}_X$ the large algebra of $\mathcal{F}_X$ over $\mathbb{Z}$. Any subset $A \subset \mathcal{F}_X$ has a (non-commutative) generating function $\bar{A} = \{f : f \in A\} \in \mathcal{A}_X$.

As is well known the group $\mathcal{J}_X$ of invertible elements of $\mathcal{A}_X$ consists of the elements of the form $1-a$ when a belongs to $\mathcal{A}_X^+$, the module spanned by $\mathcal{F}_X^+$. Further, if this is so $(1-a)^{-1} = 1+\sum\{a^n ; n>0\}$, so that if $\bar{A}$ is the generating function of a subset A of $\mathcal{F}_X^+$, $(1-\bar{A})^{-1}$ is the generating function of the submonoid generated by A if and only if A is free.

Thus with these new notations the hypothesis of property 1 take the form of the identity $(1-\bar{X})^{-1} = (1-\bar{A}_1)^{-1}(1-\bar{A}_2)^{-1}(1-\bar{A}_3)^{-1}$

In fact, if $X' = A_1 \cap X$ and $A'_1 = A_1 \setminus X'$, $X'' = X \setminus X'$. The "elimination" of X' is expressed by the formal computation

$$(1-\bar{X}'-\bar{X}'')^{-1} = (1-\bar{X}'-\bar{A}'_1)^{-1}(1-\bar{A}_2)^{-1}(1-\bar{A}_3)^{-1}$$

$$(1-\bar{X}''(1-\bar{X}')^{-1})^{-1} = (1-\bar{A}'_1(1-\bar{X}')^{-1})^{-1}(1-\bar{A}_2)^{-1}(1-\bar{A}_3)^{-1}$$

or $$(1-\bar{X})^{-1} = (1-\bar{X}')^{-1}(1-\bar{A}'_1(1-\bar{X}')^{-1})^{-1}(1-\bar{A}_2)^{-1}(1-\bar{A}_3)^{-1}$$

62.

Let us now define infinite formal products. Given a collection $\{a_i , i \in I\}$ of elements of $\mathcal{O}_X^+$ totally ordered by a relation $<$ we assume that for each $f \in \mathcal{F}_X$ there exists only a finite number of elements a_i such that f has a non zero coefficient $\langle a_i, f\rangle$ in them. Then for each f we define $\langle p, f\rangle$, the coefficient of f in p as the sum

$$\sum \langle a_{i_1}, f_{j_1}\rangle \langle a_{i_2}, f_{j_2}\rangle \ldots \langle a_{i_m}, f_{j_m}\rangle$$

extended to all factorizations $f = f_{j_1} f_{j_2} \ldots f_{j_m}$ into an arbitrary number $m>0$ of factors and for each such factorization to all m-tuples $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ such that $a_{i_1} < a_{i_2} < \ldots < a_{i_m}$ and $\langle a_{i_1}, f_{j_1}\rangle , \langle a_{i_2}, f_{j_2}\rangle, \ldots, \langle a_{i_m}, f_{j_m}\rangle \neq 0$

Clearly $1 + \sum\{\langle p, f\rangle f : f \in \mathcal{F}_X\} = p$ is a well defined element of $\mathcal{O}_X$ which we can consider as the infinite formal product of the elements $1+a_i$ with respect to $<$. Simple computation shows that p^{-1} is the infinite product of the elements $(1+a_i)^{-1}$ with respect to the opposite order $>$.

Applying our last remark of the previous section we obtain thus for each μ the identities

$$(*) \qquad (1-\bar{X})^{-1} = \prod_{<} \{(1-h)^{-1} : h \in H\} \qquad \text{or}$$

$$(**) \qquad 1-\bar{X} = \prod_{>} \{1-h : h \in H\}$$

where H is a cyclic tranversal. A special case of this construction has been given in [3] . A slight modification of the argument gives an identity of [5,1] . A commutative version of (**) has been used by Sherman [4] .

References.

1. K.T. Chen, R.H. Fox and R.C. Lyndon, Free differential calculus IV, Ann. Math., 68 (1958), 81-95.
2. M. Lazard, Groupes anneaux de Lie, et problème de Burnside, Roma. C.I.M.E. Inst. Math. (1960).
3. M.P. Schützenberger, Sur une proprietè combinatoire, Paris Seminaire P. Dubreil (1958).
4. S. Sheman, Combinatorial aspects of the Ising model, J. of Math. Phys., 1 (1960), 202-207.
5. A.I. Širšov, On free Lie rings, Math. Chornik, 45 (1958), 113-122.

ON A THEOREM OF R. JUNGEN

M. P. SCHÜTZENBERGER

Let us recall the following elementary result in the theory of analytic functions in one variable.

THEOREM (R. JUNGEN [7]). *If a is rational and b algebraic their Hadamard product c is algebraic; if, further, b is rational, c also is rational.*

For several variables, Jungen's proof shows that the theorem is still true for the Bochner-Martin [2] Hadamard product. It does not hold for the Cameron-Martin [3] and for the Haslam-Jones [6] Hadamard products. In this note we give a version of Jungen's theorem which is valid for a restricted interpretation of the notions involved when a and b are formal power series in a finite number of noncommuting variables.

1. **Notations.** Let R be a fixed not necessarily commutative ring with unit 1. For any finite set Z, $F(Z)$ is the free monoid generated by Z and $R_{\mathrm{pol}}(Z)$ is the free module on $F(Z)$ over R. An element a of $R_{\mathrm{pol}}(Z)$ will usually be written in the form $a=\sum\{(a,f)\cdot f: f\in F(Z)\}$ where the coefficients (a,f) are in R; $R_{\mathrm{pol}}(Z)$ is graded in the usual manner and $\pi_n a=\sum\{(a,f)\cdot f: f\in F(Z),\ \deg f\leqq n\}$. We identify R with $\pi_0 R_{\mathrm{pol}}(Z)$. $R_{\mathrm{pol}}(Z)$ is also a ring with product $aa'=\sum\{(a,f')(a',f'')\cdot f: f,f',f''\in F(Z), f=f'f''\}$.

It is well known (cf., e.g., [4; 3]) that these notions extend to the ring $R(Z)$ of the formal power series (with coefficients in R) in the noncommuting variables $z\in Z$; $R(Z)$ is topologized in the same manner as a ring of commutative formal power-series and $aa'=\lim_{n,n'\to\infty}(\pi_n a)(\pi_{n'}a')$. Any $b\in R^*(Z)=\{a\in R(Z):\pi_0 a=0\}$ has a quasi-inverse $(-b)^*=\lim_{n\to\infty}\sum_{n'<n}(-b)^{n'}$. If a is invertible, $a^{-1}=(1+b^*)(\pi_0 a^{-1})$ where $b=-(\pi_0 a^{-1})(a-\pi_0 a)\in R^*(Z)$. We shall say that $S^*\subset R^*(Z)$ is *rationally closed* if $r, r'\in R$, $b, b'\in S^*$ imply $rb+b'r', bb', b^*\in S^*$. If this is so, the set of those elements a of $R(Z)$ such that $a-\pi_0 a\in S^*$ is a ring containing the inverses of its invertible elements.

DEFINITION 1. $R^*_{\mathrm{rat}}(X)$ is the least rationally closed subset (of $R(X)$) containing X.

Now let $Y=\{y_s\}$ be a set of a finite number M of new variables and $R^M(X\cup Y)$ (resp. $R^M_{\mathrm{pol}}(X\cup Y)$) the cartesian product of M copies

Received by the editors December 6, 1961.

of the R-module $R(X \cup Y)$ (resp. $R^{M}_{\mathrm{pol}}(X \cup Y)$). For each $q = (q_1, \cdots, q_m) \in R^{M}(X \cup Y)$, $\pi_n q = (\pi_n q_1, \cdots, \pi_n q_m)$. If $q \in R^{*M}(X \cup Y)$ (i.e., if $\pi_0 q = 0$) let λ_q be the homomorphism of the monoid $F(X \cup Y)$ into the multiplicative monoid structure of $R(X \cup Y)$ that is induced by $\lambda_q x = x$ if $x \in X$ and $\lambda_q y_j = q_j$ if $y_j \in Y$. Since $\pi_0 q = 0$, λ_q can be extended to an endomorphism of the R-module $R(X \cup Y)$ by $\lambda_q a = \sum\{(a, f)\lambda_q f : f \in F(X \cup Y)\}$; also, $\lambda_q p = (\lambda_q p_1, \cdots, \lambda_q p_M)$ for any $p \in R^{M}(X \cup Y)$.

We shall say that $p \in R^{*M}(X \cup Y)$ is a *proper system* if $(p_j, y_{j'}) = 0$ for all $j, j' \leqq M$. Then, if $q \in R^{*M}(X)$, $\lambda_q p \in R^{*M}(X)$ and $\pi_{n+1}\lambda_q p = \pi_{n+1}\lambda_{\pi_n q} p$ for all n. Consider now the infinite sequence $p(0) = 0$, $p(1) = \lambda_{p(0)} p, \cdots, p(m+1) = \lambda_{p(m)} p, \cdots$. Trivially, $\pi_{m'} p(m') = \pi_{m'} p(m' + m'') \in R^{*M}(X)$ for $m' = 0$ and all m''. If these relations hold for $m' \leqq m$, they still hold for $m+1$ because

$$\pi_{m+1} p(m+1) = \pi_{m+1}\lambda_{p(m)} p = \pi_{m+1}\lambda_{\pi_m p(m)} p = \pi_{m+1}\lambda_{\pi_m p(m+m'')} p$$
$$= \pi_{m+1}\lambda_{p(m+m'')} p = \pi_{m+1} p(m+1+m'').$$

Hence, $p(\infty) = \lim_{m \to \infty} p(m)$ exists and it satisfies $p(\infty) \in R^{*M}(X)$, $\pi_0 p(\infty) = 0$, $p(\infty) = \lambda_{p(\infty)} p$. In fact, $p(\infty)$ is the only element to satisfy these equations because if $\pi_0 p' = 0$ and $p' = \lambda_{p'} p$, any relation $\pi_m p(\infty) = \pi_m p'$ implies $\pi_{m+1} p' = \pi_{m+1}\lambda_{\pi_m p'} p = \pi_{m+1}\lambda_{\pi_m p(\infty)} p = \pi_{m+1} p(\infty)$. For this reason we call $p(\infty)$ *the solution* of p.

DEFINITION 2. $R^{*}_{\mathrm{alg}}(X)$ is the least subset (of $R^{*}(X)$) that contains every coordinate of the solution of any proper system having its coordinates in $R^{*}_{\mathrm{pol}}(X \cup Y)$.

(REMARK. It can easily be shown that $R^{*}_{\mathrm{alg}}(X)$ is rationally closed and that it contains every coordinate of the solution of any proper system having its coordinates in $R^{*}_{\mathrm{alg}}(X \cup Y)$.)

DEFINITION 3. For any

$$a, b \in R(X), \quad a \odot b = \sum\{(a, f)(b, f) \cdot f : f \in F(X)\}.$$

2. **Main result.**

Property 2.1. The element a of $R^{*}(X)$ belongs to $R^{*}_{\mathrm{rat}}(X)$ if and only if there exists a finite integer $N \geqq 2$ and a homomorphism μ of $F(X)$ into the multiplicative monoid of $R^{N \times N}$ (the ring of the $N \times N$ matrices with entries in R) such that $a = \sum\{\mu f_{1,N} \cdot f : f \in F(X)\}$ (abbreviated as $\sum \mu f_{1,N} \cdot f$).

PROOF. (1) *The condition is necessary.* This is trivial if $a = \pi_1 a$. Hence it suffices to show that for any $r, r' \in R$, $a = \sum \mu f_{1,N} \cdot f$ and $a' = \sum \mu' f_{1,N'} \cdot f$ one can construct suitable homomorphisms giving $ra + a'r'$, aa' and a^{*}. This is done below, defining the homomorphisms by their restriction to X.

Addition. Let $N''=N+N'+2$ and $\mu''x\in R^{N''\times N''}$ defined for each $x\in X$ by

$$\mu''x_{i,1} = \mu''x_{N'',i} = 0 \qquad \text{for } 1 \leqq i \leqq N'';$$

$$\mu''x_{1,i+1} = r\mu x_{1,i} \quad \text{and} \quad \mu''x_{i+1,N''} = \mu x_{i,N} \qquad \text{for } 1 \leqq i \leqq N;$$

$$\mu''x_{1,i+N+1} = \mu' x_{1,i} \quad \text{and} \quad \mu''x_{i+N+1,N''} = \mu' x_{i,N'}\cdot r' \qquad \text{for } 1 \leqq i \leqq N';$$

$$\mu''x_{i,i'} = \text{the direct sum of } \mu x \text{ and } \mu' x \qquad \text{for } 2 \leqq i, i' \leqq N''-1;$$

$$\mu''x_{1,N''} = r\mu x_{1,N} + \mu' x_{1,N'}r'.$$

The verification is trivial.

Product. Let $N''=N+N'$ and define $\nu f\in R^{N''\times N''}$ for each $f\in F(X)$ by $\nu f_{i,i'}=\mu f_{i,N}$ if $f\neq 1$, $1\leqq i\leqq N$, $i'=N+1$; $\nu f_{i,i'}=0$, otherwise. Then, if $\mu''x=\bar{\mu}x+\nu x$ where $\bar{\mu}x$ is the direct sum of μx and $\mu' x$, one has for each $f = x^{(1)}x^{(2)}\cdots x^{(n)}$, $\mu''f = \bar{\mu}f + \sum\{\bar{\mu}f'\nu x^{(j)}\bar{\mu}f'' : f'x^{(j)}f'' = f\}$. Since $\nu f x^{(j)}=\bar{\mu}f\nu x^{(j)}$ and $(\nu f'''\bar{\mu}f'')_{1,N''}=0$ when $f''=1$, one has $\mu''f_{1,N''}=\sum\{(\mu f'_{1,N})(\mu' f''_{1,N'}) : f'f''=f\}$. Hence, $\sum\mu''f_{1,N''}\cdot f=aa'$.

Quasi-inverse. Let $N''=N$ and define $\nu f\in R^{N\times N}$ for each $f\in F(X)$ by $\nu f_{i,i'}=\mu f_{i,N}$ if $f\neq 1$, $1\leqq i\leqq N$, $i'=1$; $\nu f_{i,i'}=0$, otherwise. Then $\mu''x = \mu x + \nu x$ and since $\mu f\nu x = \nu f x$ identically one has $\mu''f = \sum\nu f^{(1)}\nu f^{(2)}\cdots\nu f^{(k)}\mu f^{(k+1)}$ where the summation is over all the factorisations $f=f^{(1)}f^{(2)}\cdots f^{(k+1)}$ of f in an arbitrary number of factors. The $(1, N)$ entry of any of these products is zero unless all its factors are different from 1 and under this condition, it is equal to $\mu f^{(1)}_{1,N}\,\mu f^{(2)}_{1,N}\cdots\mu f^{(k+1)}_{1,N}$. Hence, $\sum\mu''f_{1,N}\cdot f=\sum_{n>0}a^n=a^*$ and the first part of the proof is completed.

(2) *The condition is sufficient.* We say that the proper system p is *linear* if for each $j\leqq M$, $p_j=q_{j,0}+\sum_{j'}q_{j,j'}y_{j'}$ where all the q's belong to $R^*_{\mathrm{rat}}(X)$ and we verify that all coordinates of the solution of such a system belong to $R^*_{\mathrm{rat}}(X)$.

This is trivial if $M=1$ because $p(\infty)=(1-q_{1,1})^{-1}q_{1,0}(=(1+q^*_{1,1})q_{1,0})$. If it is true for $M'<M$ it is still true for M. Indeed, because $p(\infty)_M = (1-q_{M,M})^{-1}(q_{M,0}+\sum_{j<M}q_{M,j'}p(\infty)_{j'})$, the proper linear system p' defined by $p'_j = p_j - q_{j,M}y_M + q_{j,M}p_M$ for $j<M$ and $p'_M = (1-q_{M,M})^{-1}(p_M - q_{M,M}y_M)$ is such that $p(\infty)=p'(\infty)$. Since its first $M-1$ coordinates do not involve y_M the result follows from the induction hypothesis.

Now, given a homomorphism μ of $F(X)$ into $R^{M\times M}$, the M elements $a_j=\sum\{\mu f_{j,M}\cdot f : f\in F(X),\ f\neq 1\}$ are such that $(a_j,\ xf) = \sum_{j'}\mu x_{j,j'}(a_{j'}, f)$. Hence $(a_1, \cdots, a_M)$ is the solution of the linear proper system such that $q_{j,0} = \sum\{\mu x_{j,M}\cdot x : x\in X\}$, $q_{j,j'} = \sum\{\mu x_{j',j}\cdot x : x\in X\}$ for each j, j' and 2.1 is proved.

We now consider two subrings R' and R'' of R that commute element-wise.

Property 2.2. If $a=\sum\mu' f_{1,N}\cdot f\in R'^{*}_{\mathrm{rat}}(X)$ where μ' is a homomorphism into $R'^{N\times N}$ and if $b=p(\infty)_1\in R''^{*}_{\mathrm{alg}}(X)$ where the proper system p has its coordinates in $R''^{*}_{\mathrm{pol}}(X\cup Y)$, then $a\odot b\in R^{*}_{\mathrm{alg}}(X)$. If, further, $b\in R''^{*}_{\mathrm{rat}}(X)$ then $a\odot b\in R^{*}_{\mathrm{rat}}(X)$.

PROOF. We verify first the case of $b\in R''^{*}_{\mathrm{rat}}(X)$, i.e., of $b=\sum\mu'' f_{1,N''}\cdot f$ for some N'' and μ''. Then $a\odot b=\sum(\mu'\otimes\mu'')f_{1,NN'}\cdot f$ where the kroneckerian product $\mu'\otimes\mu''$ is a homomorphism of $F(X)$ into $R^{NN''\times NN''}$ because R' and R'' commute and the result is proved.

For the general case we denote by $K(Z)$ for any set Z the ring of the $N\times N$ matrices with entries in $R(Z)$. We shall have to consider several homomorphisms of module $\sigma\colon R^M(Z')\to K^M(Z'')$ where Z' and Z'' are two finite sets. In each case σ is defined by a mapping $Z'\to K(Z'')$ which is extended in a natural fashion to a homomorphism of the monoid $F(Z')$ into the multiplicative structure of $K(Z'')$. Then for each

$$a=(a_1,\cdots,a_M)\in R^M(Z'),\quad \sigma a_j=\sum\{(a_j,g)\cdot\sigma g\colon g\in F(Z')\}$$

and $\sigma a=(\sigma a_1,\cdots,\sigma a_M)$.

More specifically, $\mu\colon R^M(X)\to K^M(X)$ is induced by a mapping $\mu\colon X\to K(X)$ such that the entries of each μx belong to $R'^{*}(X)$.

For each $q\in R''^{*M}(X)$, $\lambda_{\mu q}\colon R(X\cup Y)\to K^M(X)$ is induced by $\lambda_{\mu q}f=\mu f$ if $f\in F(X)$ and $\lambda_{\mu q}y_j=\mu q_j$ if $y_j\in Y$. Hence, since R' and R'' commute element-wise, $\mu\lambda_q g=\lambda_{\mu q}g$ for each $g\in F(X\cup Y)$ (with λ_q as previously defined). Consequently, $\mu\lambda_q p=\lambda_{\mu q}p$ for any $p\in R''^{M}(X\cup Y)$.

Let now $Z=\{z_{j,i,i'}\}(1\leqq j\leqq M;\ 1\leqq i,\ i'\leqq N)$, a set of $M\times N\times N$ new variables and $\nu\colon R^M(X\cup Y)\to K^M(X\cup Z)$ induced by $\nu f=\mu f$ if $f\in F(X)$, $\nu y_j=$ the $N\times N$ matrix with entries $z_{j,i,i'}$ if $y_j\in Y$. Also $\lambda_{\nu q}\colon R(X\cup Z)\to R(X)$ is induced by $\lambda_{\nu q}f=f$ if $f\in F(X)$ and $\lambda_{\nu q}z_{j,i,i'}=(\nu q_j)_{i,i'}$ if $z_{j,i,i'}\in Z$. We extend $\lambda_{\nu q}$ to a homomorphism $K^M(X\cup Z)\to K^M(X)$ by defining $\lambda_{\nu q}\boldsymbol{m}$ for any $\boldsymbol{m}\in K(X\cup Z)$ as the $N\times N$ matrix with entries $\lambda_{\nu q}(\boldsymbol{m}_{i,i'})$.

Because R' and R'' commute, $\lambda_{\mu q}g=\lambda_{\nu q}\nu g$ for each $g\in F(X\cup Y)$ and, consequently, $\lambda_{\mu q}p=\lambda_{\nu q}\nu p$ for each $p\in R''^{*M}(X\cup Y)$. Hence, if p is a proper M-dimensional system with coordinates in $R''^{*}(X\cup Y)$ we have $\mu p(\infty)=\mu\lambda_{p(\infty)}p=\lambda_{\mu p(\infty)}p$. Since μ and ν coincide on $R''^{*M}(X)$, we have also $\mu p(\infty)=\nu p(\infty)=\lambda_{\mu p(\infty)}p=\lambda_{\nu p(\infty)}\nu p$.

However, the $M\times N\times N$ elements $p'_{j,i,i'}=(\nu p_j)_{i,i'}$ all belong to $R^{*}(X\cup Z)$ and they constitute a proper system p' of dimension MN^2. Thus, by construction, $(\mu p(\infty)_j)_{i,i'}=p'(\infty)_{j,i,i'}$ identically. If, fur-

ther, $p \in R''^{*M}_{\mathrm{pol}}(X \cup Y)$ all the entries appearing in vp belong to $R^{*}_{\mathrm{pol}}(X \cup Z)$ and then finally $(\mu p(\infty)_j)_{i,i'} \in R^{*}_{\mathrm{alg}}(X)$.

This completes the proof because

$$a \odot b = \sum \{(b, f)\mu' f_{1,N} \cdot f : f \in F(X)\}$$
$$= \sum \{(b, f)\mu f_{1,N} : f \in F(X)\} = \mu b_{1,N}$$

where for each $x \in X$, μ is defined by $\mu x_{i,i'} = \mu' x_{i,i'} \cdot x$.

REMARK 1. Definitions 1, 2, and 3 and the computations of this section used only the structure of monoid of the additive groups considered. Hence, the results are still valid when an arbitrary *semiring* S is taken in place of R. For S consisting of two Boolean elements, Jungen's theorem and its special case for b rational have been obtained in a different form by Y. Bar-Hillel, M. Perles and E. Shamir [1] (also by S. Ginsburg and G. F. Rose [5]) and by S. Kleene [8] respectively as by-products of more sophisticated theories.

REMARK 2. Let $R = C$, the field of complex numbers; and p a proper system of dimension M. Introducing $4M$ new symbols z_j and replacing each y_j by $z_{4j} + iz_{4j+1} - z_{4j+2} - iz_{4j+3}$ in the p_js we can deduce from p a new system of dimension $4M$ in which all the coefficients are non-negative real numbers and whose solution is simply related to $p(\infty)$.

Assume now that $p \in C^{*M}_{\mathrm{pol}}(X \cup Y)$ has only real non-negative coefficients and denote by α a homomorphism of $C_{\mathrm{pol}}(X \cup Y)$ into C. Because of the assumption that $(p_j, y_{j'}) = (p_j, 1) = 0$, identically, we can find an $\epsilon > 0$ such that $|\alpha p_j| < \epsilon$ for all j when $|\alpha x| \leqq \epsilon$ and $|\alpha y| \leqq 2\epsilon$ for all $x \in X$ and $y \in Y$. Since the sequence $\alpha p(0)$, $\alpha p(1)$, $\cdots$, $\alpha p(n)$, $\cdots$ is monotonically increasing it converges to a finite solution (cf., e.g., [10]).

Hence, the canonical epimorphism of $C_{\mathrm{pol}}(X \cup Y)$ onto the ring of the ordinary (commutative) polynomials can be extended to an epimorphism of $C_{\mathrm{alg}}(X)$ onto the ring of the Taylor series of the algebraic functions.

Acknowledgment. Acknowledgment is made to the Commonwealth Fund for the grant in support of the visiting professorship of biomathematics in the Department of Preventive Medicine at Harvard Medical School.

REFERENCES

1. Y. Bar-Hillel, M. Perles and E. Shamir, *On formal properties of simple phrase structure grammars*, Technical Report No. 4. Information System Branch, Office of Naval Research, 1960.

2. S. Bochner and W. T. Martin, *Singularities of composite functions in several variables*, Ann. of Math. **38** (1938), 293–302.

3. R. H. Cameron and W. T. Martin, *Analytic continuation of diagonals*, Trans. Amer. Math. Soc. **44** (1938), 1–7.

4. K. T. Chen, R. H. Fox and R. C. Lyndon, *Free differential calculus*. IV, Ann. of Math. (2) **68** (1958), 81–95.

5. S. Ginsburg and G. F. Rose, *Operations which preserve definability*, System Development Corporation, Santa Monica, Calif., SP-511, October, 1961.

6. U. S. Haslam-Jones, *An extension of Hadamard multiplication theorem*, Proc. London Math. Soc. II. Ser. **27** (1928), 223–232.

7. R. Jungen, *Sur les series de Taylor n'ayant que des singularités algébrico-logarithmiques sur leur cercle de convergence*, Comment. Math. Helv. **3** (1931), 226–306.

8. S. Kleene, *Representation of events in nerve nets and finite automata*, Automata Studies, Princeton Univ. Press, Princeton, N. J., 1956.

9. M. Lazard, *Lois de groupes et analyseurs*, Ann. Sci. Ecole Norm. Sup. (4) **72** (1955), 299–400.

10. A. M. Ostrowski, *Solutions of equations and systems of equations*, Academic Press, New York, 1960.

HARVARD MEDICAL SCHOOL

REMARK ON A THEOREM OF DÉNES

by

MURRAY EDEN[1] and M. P. SCHÜTZENBERGER[2]

The aim of this note is to give a slightly more explicit form to the correspondence between labelled trees and cycles which was found by Dénes (1).

Let $X = \{x_i\}$ $(1 \leqq i \leqq n)$ be a set of vertices and

$T = \{t_j = (x_{i_j}, x_{i'_j}) = (x_{i'_j}, x_{i_j})\}$, $(1 \leqq j \leqq n-1)$ be a set of edges so that (X, T) is a tree.

Considering T as an abstract alphabet, we denote by G the set of all words in the letters $t \in T$ that contain each t at most once and we define a mapping from G into the symmetric group of permutations S_n on elements $1, 2, \ldots, n$, by associating the transposition $\bar{t}_j = (i_j, i'_j)$ with each $t_j = (x_{i_j}, x_{i'_j})$ and the product $\bar{g} = \bar{t}_{j_1} \bar{t}_{j_2} \ldots \bar{t}_{j_k}$ with each $g = t_{j_1} t_{j_2} \ldots t_{j_k}$. It is trivial that any two $\bar{t}_j$, s corresponding to disjoint edges commute. Thus, unless (X, T) is a „bush" not all the associated cycles are distinct.

For any permutation $s \in S_n$ and triple (a, b, c) of distinct elements, we define the indicator $\sigma(s; a, b, c)$ with value zero if a, b, and c do not belong to the same cycle of s and with value ± 1 depending upon whether b is or is not between a and c. Formally,

$\sigma(s; a, b, c) = \sigma(s; b, c, a) = -\sigma(s; a, c, b) = 1$ if $s^n a = b$, $s^{n'} a = c$ with $n \leqq n'$ and $s^{n''} a \neq c$ for all $n'' < n'$.

Property 1. *For any $g = g't \in G$ and triple (a, b, c) if $\sigma(\bar{g}'; a, b, c) \neq 0$ then $\sigma(\bar{g}'; a, b, c) = \sigma(\bar{g}; a, b, c)$.*

Proof. The subgraph $(X, T') \subset (X, T)$ corresponding to the factors of $\bar{g}'$ is a disjoint union of trees $T' = (T''_1, T''_2, \ldots, T''_k)$ some of which may be reduced to a single vertex.

By DÉNES' Theorem there is a one to one correspondence between the trees T''_j and the cycles which constitute the factors of the permutation $\bar{g}'$.

Now $T' \cup \{t\} \subset T$ is also a disjoint union of trees. The new edge $t = (x_i, x_j)$ connects two disconnected components of the graph (T', X). Furthermore, i and j belong to two different cycles of $\bar{g}'$ say, $(i\, i_2\, i_3 \ldots i_m)$ and $(j\, j_2\, j_3 \ldots j_p)$. Now $\bar{g} = \bar{g}' i$ is obtained by replacing these two cycles by the single cycle $(i\, j_2\, j_3 \ldots j_p\, j\, i_2 \ldots i_m)$. The result follows immediately. As a consequence, we have:

Property 2. *If $t = (x_i, x_j)$ and $t' = (x_j, x_k)$ are two distinct factors of the word $g = g_1 t g_2 t' g_3$, then $\sigma(\bar{g}; i, j, k) = 1$.*

Proof. Using the construction which has been given for property 1, it follows that i and j are in one cycle and k in another cycle of $\bar{g}_1 \bar{t} \bar{g}_2$, the particular cycles being $(i l_2 \ldots l_m j l_{m+2} \ldots l_{m+m})$ and $(k k_2 \ldots k_{m''})$, say. Thus

[1] Massachusetts Institute of Technology.
[2] Harvard University.

$\bar{g}_1 \bar{t} \bar{g}_2 \bar{t}'$ contains the cycle $(jk_2 \ldots k_{m'} kl_{m+2} \ldots l_{m+m'} il_2 \ldots l_m)$. Hence, $\sigma(\bar{g}_1 \bar{t} \bar{g}_2 \bar{t}'; i, j, k) = 1$ and the results follow from property 1.

Let $h_i \in G$ denote a product of the set of all edges t_{j_i} incident to x_i. Let us now consider $W = \{w = (h_1, h_2, \ldots, h_n)\}$ the set of all n-tuples of elements $h_1, \ldots, h_n$. Clearly the number of elements w in W is

$$\Delta(X, T) \overset{\text{def}}{=} \prod_{1 \leq i \leq n} (\deg x_i)!$$

For any given $w \in W$ if $g^* = \prod_{j=1}^{n-1} t_j$ is a product of maximal degree $n - 1$, and if for each i, the symbol t_j corresponding to edges incident to x_i appear in g^* in the same order as in h_i, then we shall say after LYNDON [2] that the word g^* is a *minimal infiltration* of $w(g^* \in ((w)))$. Clearly if $\deg x_i = 1$, h_i reduces to a single t and may as well be omitted from w.

For instance if $X = (x_1, x_2, x_3, x_4)$ and $T = \{t_1 = (x_1, x_2), t_2 = (x_2, x_3), t_3 = (x_3, x_4)\}$ we have

$$W = \{(t_1, t_1 t_2, t_2 t_3, t_3), (t_1, t_2 t_1, t_2 t_3, t_3), (t_1, t_2 t_1, t_3 t_2, t_3), (t_1, t_1 t_2, t_3 t_2, t_3)\} =$$
$$= \{(t_1 t_2, t_2 t_3), (t_2 t_1, t_2 t_3), (t_1 t_2, t_3 t_2), (t_2 t_4 t_3 t_2)\}$$

and

$$((t_2 t_1, t_2 t_3)) = \{t_2 t_1 t_3, t_2 t_3 t_1\}$$
$$((t_1 t_2, t_3 t_2)) = \{t_1 t_3 t_2, t_3 t_1 t_2\}$$
$$((t_2 t_1, t_3 t_2)) = \{t_3 t_2 t_1\}$$
$$((t_1 t_2, t_2 t_3)) = \{t_1 t_2 t_3\}.$$

We now wish to prove:

Property 3. *If $g \in ((w))$ and $g' \in ((w'))$ then $\bar{g} = \bar{g}'$ i. f. f. $w = w'$.*

Proof. By 2) we know that $w \neq w'$ implies $\bar{g} \neq \bar{g}'$ for every $g \in ((w))$ $g' \in ((w'))$ since there must be at least one triple such that $\sigma(g; i, j, k) = -\sigma(g'; i, j, k)$.

To prove the backward implication let us now assume that $f, f' \in ((w))$ and $f \neq f'$.

According to 3) it may be that $f = f'$ i.e. $f = t_{i_1} t_{i_2} \ldots t_{i_{n-1}}$, $f' = t_{i'_1} t_{i'_2} \ldots t_{i'_{n-1}}$ but that one can be reduced to the other by a certain number of exchanges of adjacent t's corresponding to disjoint edges in which case $\bar{f} = \bar{f}'$.

Assume there is no such reduction f^* of f to f'. Thus there exists some minimal n such that the left factor of f^* of degree n is different from the left factor of f', i.e.

$$f^* = g_1(x_j, x_k)\, g_2; \quad f' = g'_1(x_1, x_m)\, g'_2(x_j, x_k)\, g'_3.$$

Now g'_1 contains an edge with either x_j or x_k as end point since otherwise n is not minimal. Suppose this edge is (x_j, x_p). Then $f^* = g_1(x_j, x_k)\, g'_2(x_j, x_p)\, g_{3'}$ and $f' = g_1(x_1, x_m)\, g'_2(x_j, x_p)\, g'_3(x_j, x_k)\, g'_4$ in which case by property 2, $\sigma(f^*; j, k, p) \neq \sigma(f'; j, k, p)$.

This completes the proof and it follows immediately that the number of distinct cycles associated to the tree (X, T) is simply $\Delta(X, T)$.

(Received April 1, 1962)

REFERENCES

[1] Dénes, J.: „The Representation of a Permutation as the Product of a Minimal Number of Transpositions and its Connection with the Theory of Graphs." *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **4** (1959) 63—70.

[2] Lyndon, R. C.: „On Burnside's Problem. I." *Trans. Amer. Math. Soc.* 77 (1954) 207—215.

ЗАМЕЧАНИЕ ОБ ОДНОЙ ТЕОРЕМЕ DÉNES-A

M. EDEN и M. P. SCHÜTZENBERGER

Резюме

Авторы устанавливают соответствие между числом деревьев с n нумерированными вершинами и числом разложений цикла степени $n-1$ на произведений транспозиций. Они доказывают что эти два числа равны.

THE EQUATION $a^M = b^N c^P$ IN A FREE GROUP

R. C. Lyndon and M. P. Schützenberger

1. INTRODUCTION

The question of finding all solutions for the equation $a^M = b^N c^P$ in a free group is of interest only if none of the exponents is 0 or 1; we assume, then, that $M, N, P \geq 2$. The equation possesses obvious solutions for which a, b, and c are all powers of a common element; it will be shown that these are all solutions.

R. Vaught conjectured that $a^2 = b^2 c^2$ had only these obvious solutions, and R. C. Lyndon [3] verified this conjecture by a combinatorial argument. His result carries with it the case that all three exponents are even. That there are only the obvious solutions in the case where the three exponents have a common prime divisor was established independently by G. Baumslag [1], E. Schenkman [4], and J. Stallings [6], all of whom employed more characteristically group theoretic methods. The proof here, for general $M, N, P \geq 2$, is of a combinatorial nature.

In Section 2 we record some properties of the free monoid F of words representing elements in a free group G. In Section 3 we reduce the problem of finding all the solutions of the equation $a^M = b^N c^P$ in G to that of finding all solutions of each of two equations in F. In Sections 4 and 5 we show in turn that each of these equations has only the obvious solutions.

The greater part of the argument deals with the case that one of the exponents is 2 or 3. This suggests that arbitrary equations in powers of elements from a free group have only more or less obvious solutions when the exponents are sufficiently large. More generally, one may expect that in some sense more complicated equations have fewer solutions, with only rather special equations possessing genuinely nondegenerate solutions. Thus the equation $a^M = b^N c^P d^Q$, which possesses a wealth of nontrivial solutions when all four exponents are 2, appears to have only obvious solutions when all exponents are large.

2. COMBINATORIAL LEMMAS

Let G be a group freely generated by a set X of generators x. Let F be the monoid freely generated by the set $X \cup \bar{X}$, where $\bar{X}$ is a set, disjoint from X, of elements $\bar{x}$ in one-to-one correspondence with the elements x of X. The elements of F are *words*. A word a *represents* the group element ϕa, where ϕ is the epimorphism from F onto G carrying x into x and $\bar{x}$ into x^{-1}. The *length* $|a|$ of a word a is the number of factors in its expression as a product of the *letters* x and $\bar{x}$. The *formal inverse* $\bar{a}$ of a word a is its image under the involutory antiautomorphism of F that interchanges x and $\bar{x}$. Clearly $\phi\bar{a} = (\phi a)^{-1}$.

Received May 18, 1962.

The work of the first author was done in part under NSF Grant G-24333. That of the second author was done in part at the University of Poitiers and in part at the University of North Carolina under Contract AF 49(638)-213 with the Air Force office of Scientific Research.

A word is *reduced* if it contains no factor $a\bar{a}$ for $a \neq 1$. Each element of G is represented by a unique reduced word.

A word is *cyclically reduced* if it is reduced and is not of the form $ab\bar{a}$ for $a \neq 1$. Each reduced word is uniquely representable in the form $ab\bar{a}$ for b cylically reduced.

A word is *primitive* if it is not of the form a^m for any $m > 1$.

Two words are *cyclically conjugate* if they are of the forms ab and ba, respectively. If one is cyclically reduced or primitive, then so is the other.

The first of the following more or less obvious and familiar properties of F is due to F. W. Levi [2].

LEMMA 1. *If* $ab = cd$ *and* $|a| \leq |c|$, *then* $c = ae$ *and* $b = ed$ *for some* e.

The proof is immediate.

LEMMA 2. *If* $ab = bc$ *and* $a \neq 1$, *then* $a = uv$, $b = (uv)^k u$, *and* $c = vu$ *for some* $u, v \in F$ *and* $k \geq 0$.

Proof. If $|b| \leq |a|$, then by Lemma 1, $a = bv$ and $c = vb$ for some v, and the conclusion holds with $u = b$ and $k = 0$. If $|a| < |b|$, then again by Lemma 1, $b = ab'$ for some b', whence $a^2 b' = ab'c$. Thus $ab' = b'c$. Since $a \neq 1$, $|b'| < |b|$, and the desired conclusion follows by induction on $|b|$, the initial case being trivial.

LEMMA 3. *If* $ab = ba$, *then* a *and* b *are powers of a common element.*

Proof. If $a = 1$, the conclusion is immediate. Otherwise, applying Lemma 2, from the relation $a = uv = vu$ we conclude by induction that u and v are powers of a common element.

LEMMA 4. *If* a *and* b *have powers* a^m *and* b^n *with a common initial segment of length* $|a| + |b|$, *then* a *and* b *are powers of a common element.*

Proof. The hypothesis implies that ba^m and b^{n+1} have a common initial segment of length $|a| + 2|b|$, whence ba^m and b^n have a common initial segment of length $|a| + |b|$. Similarly, ab^n and a^m have a common initial segment of length $|a| + |b|$. It follows that ab^n and ba^m have a common initial segment of length $|a| + |b|$, and hence that $ab = ba$. The conclusion now follows by Lemma 3.

COROLLARY 4.1. *If* $a^m = b^n$ *and* $m \geq 1$, *then* a *and* b *are powers of a common element.*

Proof. If $m = 1$ or $n \leq 1$ the conclusion is immediate; otherwise

$$|a^m| = |b^n| \geq |a| + |b|$$

and the conclusion follows by Lemma 4.

COROLLARY 4.2. *If* $a \neq 1$, *then there exists a unique primitive* b *and an integer* $k \geq 1$ *such that* $a = b^k$.

LEMMA 5. *If* a *and* b *are primitive and have powers* a^m *and* b^n *with a common initial segment of length* $|a| + |b|$, *then* $a = b$.

Proof. The conclusion immediately follows from Lemma 4 and Corollary 4.2.

LEMMA 6. *If* a *is a reduced word and* $\bar{a}$ *divides* ac *in the sense that* $ac = u\bar{a}v$ *for some* u *and* v, *then* $\bar{a}$ *divides* c.

Proof. Assume that $ac = u\bar{a}v$ and that $\bar{a}$ does not divide c. By Lemma 1 it follows, first, that $|c| < |\bar{a}v|$, whence $|u| < |a|$, and, second, that $a = uw$ and $wc = \bar{a}v$ for some $w \neq 1$. Now $wc = \bar{a}v = \bar{w}\bar{u}v$ implies that $w = \bar{w}$, which is impossible for $w \neq 1$ and w reduced.

COROLLARY 6.1. *If* a *is reduced and* $\bar{a}$ *divides a power of a word of the form* $a^m b$, *then* $\bar{a}$ *divides* b.

COROLLARY 6.2. *If* a *is a reduced word and* $\bar{a}$ *divides a power of* a, *then* $a = 1$.

LEMMA 7. *If* a *is a reduced word and both* a *and* $\bar{a}$ *divide some power of a cyclically reduced word* b, *then* b *is cyclically conjugate to a word of the form* $av\bar{a}u$.

Proof. Since a begins some cyclic conjugate of a power of b, and thus a begins some power of a cyclic conjugate b' of b, $a = b'^k b_1$, where $b' = b_1 b_2$ and $k \geq 0$. Since $\bar{a}$ is a factor of a power of b and hence of a power of $b'^{k+1} = ab_2$, it follows by Corollary 6.1 that $\bar{a}$ divides b_2. Thus b_2 has the form $b_2 = v\bar{a}u$. Since $a = b'^k b_1$ with $|a| \leq |b_2| < |b'|$, it follows that $k = 0$ and $a = b_1$. Consequently, $b' = b_1 b_2 = av\bar{a}u$.

3. REDUCTION OF THE PROBLEM

We consider now elements ϕa, ϕb, and ϕc of the free group G that satisfy an equation

$$(\phi a)^M = (\phi b)^N (\phi c)^P, \tag{3.1}$$

where M, N, $P \geq 2$, and we shall show that ϕa, ϕb, and ϕc are powers of a common element of G. Our aim in this section is to replace the hypothesis (3.1) by hypotheses on the elements a, b, and c of the monoid F. Clearly, we may suppose that a, b, and c are reduced and primitive. Under this assumption, it will suffice to show that one of a, b, and c is equal or inverse to another.

It is also clear that we may replace ϕa, ϕb, and ϕc by their conjugates under any element of G, and thus replace a, b, and c, by the corresponding reduced primitive words a', b', and c'. If $b = ub'\bar{u}$, where b' is cyclically reduced, then conjugation by ϕu replaces b by b', thereby justifying the assumption that

$$b \textit{ is cyclically reduced.} \tag{3.2}$$

We next justify the additional assumption that

$$bc \textit{ is reduced.} \tag{3.3}$$

If $b = \bar{c}$, the desired conclusion holds. Otherwise, by Lemma 5, there is a bound on the length of an initial segment common to powers of $\bar{b}$ and c. Therefore, for some m and n, neither $\bar{b}^m$ nor c^n is an initial segment of the other. It follows that there exist factorizations $b = b_1 b_2$ and $c = c_1 c_2$, with $b_1 \neq 1$, $c_2 \neq 1$, such that, $b_2 b^{m-1}$ and $c^{n-1} c_1$ are formal inverses and that $b_1 c_2$ is reduced. Conjugation by

$$\phi(c^{n-1} c_1) = \phi(\bar{b}_2 \bar{b}^{m-1})$$

now replaces b by $b' = b_2 b_1$ and c by $c' = c_2 c_1$. Now b' is cyclically reduced, since b is reduced. In addition, $b'c'$ is reduced, since $b_1 c_2$ is reduced and $b_1, c_2 \neq 1$.

We now show that we may assume further that one of the following two conditions holds: either

(Case I) a *is cyclically reduced,*

or

(Case II) c *is cyclically reduced.*

We argue by induction on $|a|$. The initial case that $|a| = 1$ falls under Case I. Suppose then that (3.2) and (3.3) hold, but that neither a nor c is cyclically reduced. Then a has the form $a = ua'\bar{u}$ for some letter $u \in X \cup \overline{X}$, and therefore the reduced word $ua'^M\bar{u}$ representing $(\phi a)^M$ begins with u and ends with $\bar{u}$. Since bc is reduced, it follows that the reduced word for $(\phi b)^N(\phi c)^P$ begins with the same letter as b and ends with the same letter as c. In view of (3.1), we see that b begins with u and that c ends with $\bar{u}$. Thus $b = ub_0$ for some b_0, and c, since it is not cyclically reduced, has the form $c = uc'\bar{u}$. Conjugation by ϕu now replaces a, b, and c by a', $b' = b_0u$, and c'. Here b' is cyclically reduced since b is cyclically reduced. Further, b'c' is reduced, since $b' = b_0u$ with $u \neq 1$, and, since $c = uc'\bar{u}$ is reduced, uc' is reduced. Thus (3.2) and (3.3) remain true, while a is replaced by a' with $|a'| = |a| - 2$. The desired conclusion therefore follows by induction.

In the remaining sections we treat Cases I and II in turn. We emphasize that the two cases are not disjoint; in fact, we treat Case II by reducing it to its intersection with Case I. For this common case, where both a and c are cyclically reduced, the argument of Section 4 could, of course, be substantially simplified.

4. CASE I

Write $c = \bar{g}c'g$, where c' is cyclically reduced and g is possibly 1. Since a and b are cyclically reduced and bc is reduced,

$$a^M = b^N\bar{g}c'^Pg,$$

where both members are reduced, and thus represent the same word. Dropping primes, we show that if

$$a^M = b^N\bar{g}c^Pg \tag{4.1}$$

in the monoid F, where the word represented by the two members is reduced and a, b, and c are primitive, then $g = 1$ and $a = b = c$. We proceed by induction on $|a|$, the initial case $|a| = 1$ being vacuous.

The equation (4.1) implies the analogous equation

$$a'^M = \bar{c}^P g\bar{b}^N\bar{g}, \tag{4.1'}$$

where a' is a cyclic conjugate of $\bar{a}$. If $|b^N| \geq |a| + |b|$, then a^M and b^N have a common initial segment of length $|a| + |b|$, and it follows by Lemma 5 that $a = b$. Thus we may assume that

$$|b^{N-1}| < |a| \tag{4.2}$$

and, symmetrically, in view of (4.1'), that

$$|c^{P-1}| < |a|. \tag{4.3}$$

By Lemma 7, from the occurrences of the factors $\bar{g}$ and g in a^M we may conclude that a has the form $a = u\bar{g}vg$ and that

$$b^N\bar{g} = a^{m_1}u\bar{g}, \quad c^P g = vga^{m_2}, \tag{4.4}$$

where $m_1 + m_2 + 1 = M$. Thus $|a^{m_1}| < |b^N|$. On the other hand (4.2) implies that $|b^N| \leq |a^2|$, and we conclude that $m_1 \leq 1$. Similarly, $m_2 \leq 1$, whence $M \leq 3$; that is, $M = 2$ or $M = 3$.

If $M = 2$, we may suppose by symmetry that $m_1 = 1$ and $m_2 = 0$. Now (4.4) implies that $b^N = au = u\bar{g}vgu$ and $c^P = v$, whence $b^N = u\bar{g}c^P gu$. Therefore there exists a cyclic conjugate b' of b for which $b'^N = u^2\bar{g}c^P g$. Since

$$|b'| = |b| \leq |b^{N-1}| < |a|,$$

we conclude by induction that $g = 1$ and $b' = u = c$, and therefore that $b = b' = c$.

If $M = 3$, then $m_1 = m_2 = 1$, and (4.4) implies that

$$b^N = au = u\bar{g}vgu \quad \text{and} \quad c^P = vgu\bar{g}v.$$

By symmetry we may suppose that $|c^P| \leq |b^N|$, and therefore $|v| \leq |u|$. By (4.2), $|b^{N-1}| < |a|$, and therefore $b^N = au$ implies that $|u| < |b|$. It follows from $b^N = u\bar{g}vgu$ that b both begins and ends with u. From Lemma 2 it follows that u and b have the forms $u = (pq)^k p$ and $b = (pq)^{k+1}p$ where, since b is primitive, $q \neq 1$. Now $b = uqp = pqu$, whence $b^N = u\bar{g}vgu$ implies that

$$\bar{g}vg = qpb^{N-2}pq.$$

If $|q| \leq |g|$, then g ends with both q and $\bar{q}$, which is impossible for $q \neq 1$. Therefore $|g| < |q|$, $q = \bar{g}q_1 g$ for some q_1, and

$$v = q_1 gpb^{N-2}p\bar{g}q_1.$$

Because $|v| \leq |u| < |b|$, this implies that $N = 2$ and $v = q_1 gp^2\bar{g}q_1$. It follows that

$$2|p| + |q| < |v| \leq |u| = (k+1)|p| + k|q|,$$

and therefore that $k \geq 2$.

The relation

$$c^P = vgu\bar{g}v = q_1 gp^2\bar{g}q_1 g(pq)^k pgq_1 gp^2\bar{g}q_1 = q_1 gp(pq)^{k+2}p^2\bar{g}q_1$$

implies that

$$c'^P = (pq)^{k+2}p^2\bar{g}q_1^2 gp$$

where c' is a cyclic conjugate of c. Now the inequality $k \geq 2$ implies that

$$|qp^2\bar{g}q_1^2gp| < 4|p| + 3|q| \leq |(pq)^{k+1}p|.$$

Therefore

$$|c'^P| = |(pq)^{k+1}p| + |qp^2\bar{g}q_1^2gp| < 2|(pq)^{k+1}p|$$

and $|c'| < |(pq)^{k+1}p|$. Thus c'^P has an initial segment $(pq)^{k+2}p$ of length greater than $|c'| + |pq|$ in common with $(pq)^{k+3}$, and it follows by Lemma 5 and Corollary 4.1, since c' is primitive, that pq is a power of c'. From the fact that $(pq)^{k+2}$ and $(pq)^{k+2}p^2\bar{g}q_1^2gp$ are both powers of c' it follows that $p^2\bar{g}q_1^2gp$ is a power of c', and therefore $p^3\bar{g}q_1^2g$ is a power of a cyclic conjugate c'' of c'. From the relation

$$|c''| = |c| \leq |pq| < |p^3\bar{g}q_1^2g|$$

it follows that $c''^Q = p^3\bar{g}q_1^2g$ for some integer $Q \geq 2$. Finally, since

$$|c''| = |c| \leq |c^{P-1}| < |a|,$$

we may conclude by induction that $g = 1$ and that p and $q_1 = q$ are powers of a common element, which contradicts the hypothesis that $b = (pq)^{k+1}p$ is primitive.

5. CASE II

Write $a = \bar{g}a'g$, where a' is cyclically reduced and possibly $g = 1$. Dropping primes, we obtain an equation

$$\bar{g}a^Mg = b^Nc^P \tag{5.1}$$

in the monoid F. Here the word represented by the two members is reduced, and a, b, and c are primitive. We shall show that $g = 1$ and $a = b = c$. In view of Case I, it will suffice to show that $g = 1$.

If $|c^P| \leq |g|$, then $g = hc^P$ for some h; and after cancelling a factor c^P from each member, we see that $\bar{c}^P\bar{h}a^Mh = b^N$. This equation falls under Case I and has a solution only if $\bar{c} = b$, which is contrary to the hypothesis that b^Nc^P is reduced. Thus we may assume that $|g| < |c^P|$, and, symmetrically, that $|g| < |b^N|$. It follows by Lemma 1 that

$$\begin{aligned} g = \overline{b^{n_1}b_1} &= c_2c^{p_2}, \\ a^{m_1}a_1 &= b_2b^{n_2}, \\ c^{p_1}c_1 &= a_2a^{m_2}, \end{aligned} \tag{5.2}$$

where $a = a_1a_2$, $b = b_1b_2$, and $c = c_1c_2$, and where $m_1 + m_2 + 1 = M$, $n_1 + n_2 + 1 = N$, and $p_1 + p_2 + 1 = P$. We remark that (5.2) implies (5.1).

The system (5.2) leads to systems

$$g' = \overline{a^{m_1}a_1} = \bar{b}_2(\bar{b}_1\bar{b}_2)^{n_2}, \quad c^{p_1}c_1 = a_2a^{m_2}, \quad (\bar{b}_1\bar{b}_2)^{n_1}\bar{b}_1 = c_2c^{p_2}, \tag{5.2'}$$

and

(5.3") $$g'' = \overline{\bar{c}^{p_2}\bar{c}_2} = \bar{b}_1\bar{b}^{n_1}, \quad \bar{a}_1^{m_2}\bar{a}_2 = \bar{c}_1\bar{c}^{p_1}, \quad \bar{b}^{n_2}\bar{b}_2 = \bar{a}_1\bar{a}^{m_1}.$$

Thus the hypothesis on the six exponents that there exist an element g and factorizations of three primitive words a, b, and c such that (5.2) holds is symmetric under cyclic permutation of the pairs (n_1, n_2), (m_1, m_2), (p_1, p_2), and also under interchange of (n_1, n_2) with (p_1, p_2) coupled with reversal of all three pairs.

We exploit this symmetry to reduce the discussion of (5.2) to three cases as follows. First, we choose Case A to be that where the two exponents in some one of the equations (5.2) both vanish. We choose Case B to be that where all of m_1, n_1, and p_1, or else all of m_2, n_2, and p_2, are positive. For Case C we may, in view of Case B, assume that some exponent vanishes, say $n_1 = 0$. Then $n_2 \neq 0$ in view of the equation $n_1 + n_2 + 1 = N$, while we may assume that $p_2 \neq 0$ in view of Case A. In view of Case B, we may now assume that $m_2 = 0$, while, in view of Case A, we may take $p_1 \neq 0$. We now treat these three cases in turn.

Case A. Here we suppose that the two exponents in the first of equations (5.2) vanish. This equation becomes $g = \bar{b}_1 = c_2$. Thus $b = \bar{g}b_2$, $c = c_1\bar{g}$, and substituting these right-hand members in (5.1) and cancelling the initial $\bar{g}$ and the final g, we obtain the equation

(5.3) $$a^M = (b_2\bar{g})^{N-1} b_2\, c_1 (gc_1)^{P-1}.$$

If $|(b_2\bar{g})^{N-1}b_2| \geq |b_2\bar{g}| + |a|$, it follows by Lemma 5 that $a = b_2\bar{g}$, and since $\bar{g}ag$ is reduced, that $g = 1$, as required. Thus we may assume that

$$|(b_2\bar{g})^{N-2} b_2| < |a|.$$

Also $|(b_2\bar{g})^{N-1}| = |a|$ would imply that $a = (b_2\bar{g})^{N-1}$, and, since $\bar{g}ag$ is reduced, g would equal 1; therefore we may assume that $a \neq (b_2\bar{g})^{N-1}$. We show that the inequality $|a| < |(b_2\bar{g})^{N-1}|$ is impossible. This inequality implies that the first factor a in the product (5.1) begins with b_2 and that the second begins with $h_2 b_2$, for some factorization $\bar{g} = h_1 h_2$ with $h_1, h_2 \neq 1$. By Lemma 2, there exist u, v, and k such that $b_2 = (uv)^k u$ divides a power of $h_2 = uv$. From (5.3), there exists a cyclic conjugate a' of a for which

$$a'^M = (\bar{g}b_2)^{N-1} b_2\, c_1 (gc_1)^{P-1} b_2.$$

By Lemma 7, a' has the form $a' = \bar{g}pgq$; and from the relations

$$|a'| = |a| < |(b_2\bar{g})^{N-1}| = |(\bar{g}b_2)^{N-1}|$$

we conclude that g divides b_2. Moreover, $\bar{h}_2$ divides $g = \bar{h}_2\bar{h}_1$. It follows that $\bar{h}_2$ divides a power of h_2, which, by Corollary 6.2, contradicts the fact that $h_2 \neq 1$. Thus we may assume, symmetrically, that

(5.4) $$|(b_2\bar{g})^{N-1}| < |a|, \quad |(gc_1)^{P-1}| < |a|.$$

In view of (5.4), cancelling the first and last factors a from (5.3) shows that a^{M-2} divides $b_2 c_1$. The inequalities (5.4) imply that $|b_2|, |c_1| < |a|$; hence $|b_2 c_1| < 2|a|$. It follows that $M - 2 < 2$, that is, $M = 2$ or $M = 3$.

If $M = 2$, we may suppose by symmetry that

$$|(b_2\bar{g})^{N-1}b_2| \geq |c_1(gc_1)^{P-1}|.$$

It follows that $(b_2\bar{g})^{N-1}b_2 = a_1a_2a_1$ and $a_2 = c_1(gc_1)^{P-1}$, whence

$$(b_2\bar{g})^N = a_1c_1(gc_1)^{P-1}a_1g \quad \text{and} \quad b'^N = (a_1\bar{g})^1a_1c_1(gc_1)^{P-1},$$

where b' is a cyclic conjugate of b. Since this equation is of the form (5.3) with $|b'| = |b_2g| < |a|$, we may conclude by induction on $|a|$ that $g = 1$.

If $M = 3$, then not both N and P can exceed 2, since then (5.4) would imply that $2|b_2|$, $2|c_1| < |a|$, and that

$$|b_2| + |c_1| < |a| \leq |b_2c_1|.$$

By symmetry, we assume that $N = 2$. Equation (5.3) now becomes

$$a^3 = b_2gb_2c_1(gc_1)^{P-1}.$$

It follows that there exist factorizations $b_2 = b_3b_4$ and $c_1 = c_3c_4$ with $b_3 \neq 1$ and $c_4 \neq 1$ such that

$$a = b_2\bar{g}b_3 = b_4c_3 = c_4(gc_1)^{P-1}.$$

Since a begins both with $b_2 = b_3b_4$ and with b_4, it follows by Lemma 2 that $b_3 = uv$ and $b_4 = (uv)^ku$ for some u, v and some k. Now $b_2\bar{g}b_3 = b_4c_3$ implies that $c_3 = vu\bar{g}uv$, while $b_4c_3 = c_4(gc_1)^{P-1}$ implies that

$$c_3b_4c_3 = c_1(gc_1)^{P-1} \quad \text{and} \quad c_1(gc_1)^{P-1} = vu\bar{g}(uv)^{k+2}u\bar{g}uv.$$

Since $|c_1| > |c_3| = vu\bar{g}uv$, there exist an $h \geq 1$ and a factorization $uv = w_1w_2$ with $w_1 \neq 1$ such that the initial occurrence of c_1 in this expression for $c_1(gc_1)^{P-1}$ has the form $c_1 = vu\bar{g}(uv)^hw_1$. Now c_1 ends both with $w_1w_2w_1$ and with $\bar{g}uv$. Since $|w_2w_1| = |uv|$, it follows that unless $g = 1$, $\bar{g}$ ends with the same letter as w_1, and hence as c_1, which is contrary to the hypothesis that c_1g is reduced. Therefore $g = 1$, as required.

Case B. Here we may assume that $m_1, n_1, p_1 \neq 0$ and, using cyclic symmetry, that $|c| \leq |a|$. From the equality

$$c^{p_1}c_1 = a_2a^{m_2} \qquad (p_1 \neq 0),$$

we conclude that c begins $(a_2a_1)^{m_2+1}$ and since $|c| \leq |a| = |a_2a_1|$, that c begins, and therefore divides, a_2a_1. From the equality

$$a^{m_1}a_1 = b_2b^{n_2} \qquad (m_1 \neq 0)$$

we conclude that a_2a_1 divides b^{n_2+1}, and hence that c divides b^{n_2+1}. It follows from the relation

$$b^{n_1}b_1 = \bar{c}^{p_2}\bar{c}_2 \qquad (n \neq 0)$$

that b begins $\bar{c}^{p_2+1}$. Thus b, and with it b^{n_2+1}, is a product of initial segments of $\bar{c}$. Now the factor c of b^{n_2+1} must end with a part of some initial segment $d \neq 1$ of

$\bar{c} = de$, and since $|d| \leq |\bar{c}| = |c|$, c must end with all of d, that is, $c = fd$. But then $c = fd = \bar{e}\bar{d}$ and $d = \bar{d}$, which is impossible for a part $d \neq 1$ of the reduced word c. This shows that Case B is impossible.

Case C. Here $n_1 = m_2 = 0$ with the remaining exponents positive. The equations (5.2) now take the form

$$b_1 = \bar{c}^{p_2}\bar{c}_2,$$

$$a^{m_1}a_1 = b_2 b^{n_2}, \tag{5.5}$$

$$a_2 = c^{p_1}c_1.$$

If a_2 divided b_1, then c, which divides a_2, would divide b_1; and hence c would divide a power of $\bar{c}$. By Corollary 6.2, this would imply that $c = 1$. We conclude, symmetrically, that neither of a_2 and b_1 divides the other. Now the inequality

$$|a^{m_1}a_1| \geq |a| + |b|$$

would imply by Lemma 5 that $a_1 a_2 = b_2 b_1$, whence one of a_2 and b_1 would divide the other. We conclude that $|a^{m_1}a_1| < |a| + |b|$. By symmetry, we assume that $|b| \leq |a|$, whence $|a^{m_1}a_1| < 2|a|$ and $m_1 = 1$. We now see that $|a_1 a_2 a_1| < |a| + |b|$, whence $|a_1| < |b|$. If $n_2 = 1$, it follows from the relation $a_1 a_2 a_1 = b_2 b_1 b_2$ that either a_2 divides b_1 or b_1 divides a_2. If $n_2 > 2$, from the equality $a_1 a_2 a_1 = b_2 b^{n_2}$, we see by cancelling the factors a_1 that $(b_2 b_1)^{n_2-2} b_2$ divides a_2, and therefore b_1 divides a_2. We conclude that $n_2 = 2$.

The second of the equations (5.2) now takes the form

$$a_1 a_2 a_1 = b_2 b_1 b_2 b_1 b_2. \tag{5.6}$$

Since it was established that $|a_1| < |b| = |b_2 b_1|$, it follows from (5.6) that $|a_2| > |b_2|$. Consequently b_2 is in the middle of a_2, that is, there exist u and v, with $|u| = |v|$, such that $a_2 = ub_2 v$. It follows from (5.6) that $a_1 u = b_2 b_1$ and $va_1 = b_1 b_2$. Since b_1 does not divide $a_2 = ub_2 v$, b_1 does not divide u, and, from $a_1 u = b_2 b_1$ we see that $|u| < |b_1|$. From the relations $|v| = |u| < |b_1|$ and $va_1 = b_1 b_2$ it follows that v divides $b_1 = \bar{c}^{p_2}\bar{c}_2$. If $|c|$ were no greater than $|v|$, then v would contain a cyclic conjugate of $\bar{c}$, which, since v divides $a_2 = c^{p_1}c_1$, would contradict Corollary 6.2. We conclude that $|u| = |v| < |c|$. But now $a_2 = ub_2 v$, and therefore also $ub_2 b_1$, begins with c, while

$$vb_2 b_1 = va_1 u = b_1 b_2 u$$

begins with $\bar{c}$. Since $|u| = |v| < |c|$ there exists a $d \neq 1$ for which $c = ud$ and $\bar{c} = vd$. This implies that $c = ud = \bar{d}\bar{v}$, which contradicts the hypothesis that c is cyclically reduced.

REFERENCES

1. G. Baumslag, *On a problem of Lyndon*, J. London Math. Soc. 35 (1960), 30-32.
2. F. W. Levi, *On semigroups*, Bull. Calcutta Math. Soc. 36 (1944), 141-146.
3. R. C. Lyndon, *The equation* $a^2b^2 = c^2$ *in free groups*, Michigan Math. J. 6 (1956), 89-95.
4. E. Schenkman, *The equation* $a^n b^n = c^n$ *in a free group*, Ann. of Math. 70 (1959), 562-564.
5. M. P. Schützenberger, *Sur l'équation* $a^{2+n} = b^{2+m} c^{2+p}$ *dans un groupe libre*, C. R. Acad. Sci. Paris 248 (1959), 2435-2436.
6. J. Stallings, *On certain relations in free groups* (Abstract 559-166) Notices Amer. Math. Soc. 6 (1959), 532.

The University of Michigan
and
Harvard Medical School

ON PROBABILISTIC PUSH-DOWN STORAGES

M. P. Schützenberger
Department of Preventive Medicine
Harvard Medical School
Boston, Massachusetts

1. INTRODUCTION

The aim of this note is to describe certain elementary problems of random walks arising in the study of a restricted family of finite automata with a push-down storage.

For any set $X = \{x\}$ we shall denote by $F(X)$ the set of all finite strings of elements of X and we shall refer to $F(X)$ as the *free monoid generated by X* (the operation being, as usual, the concatenation).

If X is considered as the input alphabet of an automaton $\mathcal{S}$ consisting of sets of states $S = \{s\}$ with transition function $(SxX) \to S$, we shall use the following notations:

$$\text{if } f = x_{i_1} x_{i_2} \dots x_{i_m} \in F(X) \text{ and } s \in S,$$

$$s.f = s_{i_m} \text{ is defined inductively}$$

$$\text{by } s_{i_o} = s,\ s_{i_1} = (s_{i_o}, x_{i_1}), \dots,$$

$$(s_{i_j}, x_{i_{j+1}}) = s_{i_{j+1}}, \text{ and } s.f = s \text{ if } f = e \text{ (the empty word).}$$

$$\phi_S f = \phi_S f' \text{ if and only if for all } s \in S,\ s.f = s.f'.$$

By definition a finite automaton with push-down storage $\mathcal{A}$ is given by:

(1) a finite input alphabet X

(2) two finite sets of states S and T; two distinguished elements $s_o \in S$, $t_o \in T$; a distinguished subset $\bar{S}$ of S.

(3) an internal alphabet X; a word $g_o \in F(Z)$; a finite subset $\bar{G}$ of $F(Z)$.

(4) the following mappings:
σ : (S,T,X) → S;
χ : (S,T,X) → (T,T) (the family of all sets of pairs of elements of T);
α : (S,T,X) → F(Z);
(T,Z) → T.

The unbounded part of the memory of $\mathfrak{A}$ consists of a tape on which words in the alphabet Z can be written or erased.

If $f = x_{i_1} x_{i_2} \ldots x_{i_n} \ldots$ is an input sequence, the automaton starts in the initial state (s_o, t_o, g_o) where g_o means that the word g_o is already stored in the memory of $\mathfrak{A}$.

The letters $x_{i_1}, x_{i_2} \ldots x_{i_n} \ldots$ are read sequentially and, for each of them, the following cycle of operations is performed:

If the state of $\mathfrak{A}$ is $(s,t,g) \in (S,T,F(Z))$ and the incoming input letter is $x_{i_n} = x$:

(1) the state s is changed to $\sigma(s,t,x) = s'$;
(2) the machine searches if there exists a factorisation $g = g'g''$ of the stored word g such that $(t_o g', t_o g'') \in \chi(s,t,x)$.
If at least one such factorisation exists and if the one for which the degree (= length) of g'' is minimal is g'.g'', the word g'' is erased from the memory. If no such factorisation exists, nothing is erased and g' = g.
(3) the word $\alpha(s,t,x) = g''$ is written in the memory to the right of g'.
(4) the state t is changed to $t_o.g'g''$ and the cycle is completed.

If after completing the cycle corresponding to the n-th letter of the input sequence $f = x_{i_1} x_{i_2} \ldots x_{i_n} \ldots$ the state is s and the stored word is g, we shall write $f_n = x_{i_1} x_{i_2} \ldots x_{i_n} \in F(X)$, $s = \sigma(f_n)$, $g = \alpha(f_n)$. By definition, $f_n \in \overline{K}$ if $(\sigma(f_n), \alpha(f_n)) \in (\overline{S}, \overline{G})$ and $f_n \in K$ if $f_n \in \overline{K}$, and for $n' < f_n' \notin \overline{K}$. In accordance with usual terminology, K may be called the set of the words accepted by $\mathfrak{A}$.

With respect to the probabilities we shall always assume that there exists a finite n, a fixed nxn matrix p and a representation μ of F(X) by nxn matrices such that (1) Tr p = 1; (2) for all $f \in F(X)$ $\Sigma \{\mathrm{Tr}\ p\mu fx : x \in X\} = \mathrm{Tr}\ p\mu f \geq 0$. Then, for each $f \in F(X)$, Tr $p\mu f$ can be interpreted as the probability measure of the set of all infinite input sequences which begin with f.

In particular, the case of n = 1 corresponds to the hypothesis that the letters of X are produced independently with constant probabilities $\bar{\mu}x = \mathrm{Tr}\ p\mu x$.

If there exists a one-to-one correspondence between the letters of X and the states of a Markov chain with initial probabilities $\Pr(x_i)$ and transitions $\Pr(x_i \mid x_j)$, one can take h equal to twice the number n' of elements of X and define for each $x_i \in X$ the matrix μx_i by

$$(\mu x_i)_{j,k} = \Pr(x_i \mid x_j) \text{ if } k = i,\ 1 \leq i,\ j \leq n';$$

$$= \Pr(x_i) \text{ if } k = i,\ j = n' + 1;$$

$$= 0 \text{ otherwise.}$$

2. EXAMPLES

Since these definitions are quite restrictive, it may be worthwhile to indicate how they relate to more familiar structures.

(1) Let $\alpha(s,t,x)$ be identically the empty word e. So, only the finite part δ plays a role and $\mathfrak{A}$ is a conventional one way one tape automaton (Rabin and Scott). The family of the sets $\overline{K}$ corresponding to these automata is usually called the family $\mathfrak{R}$ of *regular events* (Kleene).

(2) Let Z consist of a single letter. The word $\alpha(f_n) = z^m$ stored in the memory can be identified with the non-negative integer m and $\mathfrak{A}$ can be considered as a finite automaton with a (unbounded) counter. Clearly, given an integer valued function β of X, it is possible to choose S, T, $(S,X) \to S$, etc., so that for each $f_n = x_{i_1} x_{i_2} \ldots . x_{i_m}$:

(i) $\alpha(f_n) = z^{s_n}$ where s_n is the absolute value of the cumulative sum $\beta x_{i_1} + \beta x_{i_2} + \ldots . + \beta x_{i_m} = \beta f_n$.

(ii) the sign of βf_n can be determined from the knowledge of $\sigma(f_n)$.

The associated probabilistic problem is that of the elementary discrete one dimensional random walk.

(3) Let $X = \{x_{\pm i}\}$, $Z = \{z_{\pm i}\}$ $(1 \leq i \leq N)$ and assume that the automaton $\mathfrak{A}$ is constructed so that if the stored word $g = \alpha(f_{n-1})$ ends with the letter z_i and if the n-th input letter is x_j then the cycle consists of the following single operations:

if $i = -j$, z_i is erased,

if $i \neq -j$, x_i is written to the right of g.

If g is the empty word z_i is written on the tape.

Assuming that the initial word g_o is the empty word, the word $\alpha(f_n)$ stored after reading the input word f_n can be considered as being obtained by replacing the x_i's by the z_i's and by erasing every pair of consecutive letters with opposite indices.

For instance if $\bar{f} = x_1\, x_2\, x_3\, x_{-3}\, x_{-4}\, x_4\, x_{-2}\, \ldots$, the values of $\alpha(f_1)$, $\alpha(f_2)$, $\ldots$ $\alpha(f_7)$ are

$$x_1;\ x_1\, x_2;\ x_1\, x_2\, x_3;\ x_1\, x_2;\ x_1\, x_2\, x_{-4};\ x_1\, x_2;\ x_1;\ \ldots\ .$$

Clearly, if g_o is the empty word e, the set $D_N = \{f \in F(Z) : \alpha(f) = e\}$ depends only on N and it is in fact the kernel of the epimorphism γ of F(X) onto the free group generated by $\{x_i\}$ $(1 \leq i \leq N)$ that satisfies identically $1 = \gamma x_i\, \gamma x_{-1}$.

The theory of the associated random walk is due to Kesten.

3. CONTEXT FREE LANGUAGES

Let us consider after N. Chomsky two alphabets $\Xi = \{\xi\}$, $X = \{x\}$ and a finite collection G of pairs (ξ_i, f_i) where $\xi_i \in \Xi$ and where f_i belongs to the free monoid generated by the union of X and Ξ, but is not the empty word nor an element of Ξ. Taking an initial subset Ξ' of Ξ, we consider the least subset $L'_G \in F(X \cup \Xi)$ that satisfies the two conditions:

(1) $\Xi' \in L'_G$,

(2) if $f = f'\, \xi_i\, f'' \in L'_G$ and $(\xi_i, f_i) \in G$, then $f' f_i f'' \in L'_G$.

We denote by L_G the subset of L'_G which consists of the words containing only letter from X. L_G is the "context free language" generated by the grammar G. In a loose way, L'_G may be described as the set of all words which can be obtained starting from Ξ' by an arbitrary number of applications of the *rewriting rules* $\xi_i \rightarrow f_i$.

This formal construction is due to Post [12], but its special importance comes from its rediscovery by N. Chomsky who has founded upon it a general theory of natural languages.

As a matter of interest it must be mentioned that S. Ginsburg has recently observed that artificial programming languages like Algol are also context free languages. (Of course this remark should not be understood as implying that actual human languages are nothing more than glorified versions of Algol. Indeed, in Chomsky's theory the context-free level is only an initial germ out from which the true language emerges by the interplay of the higher structures ruled by the so called "transformations.")

As an example, let $\Xi = \{\xi\}$, $X = \{a,b\}$ and $G = \{(\xi, a), (\xi, b\,\xi\,\xi)\}$. The starting letter ξ can be replaced by a or by $b\,\xi\,\xi$. Since $a \in F(X)$, $a \in L$; in $b\,\xi\,\xi$, the first or the second ξ can be replaced by a or by $b\,\xi\,\xi$, giving the words $ba\,\xi$, $bb\,\xi\,\xi\,\xi$, $b\,\xi\,a$, baa, $bb\,\xi\,\xi a$, $b\,\xi\,b\,\xi\,\xi$, $bab\,\xi\,\xi$, $bb\,\xi\,\xi\,b\,\xi\,\xi$, ... on which the

process has to be repeated, etc. Thus, one finds that a, baa, bbaaa, babaa, form the set of the words of L of degree of most 5. In fact a word f belongs to L if and only if the number of a's contained in it is equal to 1 plus the number of b's and if any proper left factor of f contains at most as many a's as b's. Thus, L corresponds to the set of the well formed formula in the so-called parenthesis free notation.

An other less obvious property is enjoyed by L: in some well defined sense every word is produced in an essentially unique manner. In other terms there is a unique manner of inserting brackets in any word $f \in L$ so that these brackets indicate how the word has been produced (for instance, (b(a(bba))).

This possibility does not necessarily exist for an arbitrary context free language. Hence for a given grammar G one may find it advisable to attach an integer n_f to every word $f \in F(X)$ in such a way that:

$n_f = 0$ if $f \notin L_G$ and, if $f \in L_G$. n_f = the number of essentially different ways in which f is produced.

Clearly, it is a desirable quality of an artificial language that $n_f \leq 1$ since, otherwise, there would be strings which would admit several possible interpretations. Unfortunately as shown by Bar-Hillel, Shamir and Perles the question to decide if $n_f \leq 1$ identically for an arbitrary context free language is recursively unsolvable.

Returning to the definitions introduced previously we can state the following property:

For any context free language L one can find:

a regular event K,
a natural number N,
a mapping $\phi : X \to F(X)$,
such that $L = \{\phi f : f \in K \cap D_N\}$

where, as usual ϕf denotes the word obtained when replacing in f every letter x_i by the (eventually empty) word ϕx_i.

Moreover, the construction implies that every word is produced the same number of times by both processes. In fact the number n_f defined rather loosely above can be more accurately defined as the number of words $f \in K \cap D_N$ such that $f' = \phi f$.

Reciprocally given any finite automaton with push-down storage one can find a set Ξ and a grammar G such that the corresponding language L is precisely equal to $\overline{K}$ (or to K). Moreover the construction shows that every word is produced at most once so that the correspondence between the two processes is really one-to-one. (Cf. Chomsky 1962 a)

4. GENERATING FUNCTIONS

The construction carried out in the last section can be given another interpretation which will be best explained in the case of the language L described above.

Let us assume that the successive letters are produced randomly according to the hypothesis of section 1 and introduce a new auxiliary variable t.

We can associate with L the usual generating function: $r' = \Sigma \{t|^{f}|\Pr f : f \in L\}$ where the notation $|f|$ represents the degree (length) of f and, according to hypothesis, the probability Pr f is equal to Tr μf.

However, for the sake of convenience, let us consider another generating function: $r = \Sigma \{t|^{f}|\mu f : f \in L\}$ from which r' can be deduced by the operation Tr. As it is well known r satisfies the algebraic matrix equation $r = t\mu a + t\mu b r^2$ which simply expresses that any word $f \in L$ which is not the word a has a unique factorisation $f = bf'f''$ where both f' and f'' belong to L.

Thus, in particular if dim $\mu = 1$ we obtain the classical formula

$$2r = 1 - (1 - 4t^2 \bar{\mu}a\, \bar{\mu}b)^{1/2}$$

in which the right member can be expanded in a Taylor series converging for $|t|$ small enough.

If dim $\mu = N > 1$, this straight-forward method is not possible because the equation is one in non-commutative variables. However, for given μa and μb the matrix μr can be computed from the system of n^2 algebraic equation which results from the identification of the entries in both members.

Now clearly the same remark is valid for the context free language produced by any grammar G. Let us consider the letters $(\xi_1, \ldots . \xi_m)$ of Ξ as "unknown" and the letters $x \in X$ as (non commutative) coefficients. Then to the grammar G it corresponds in a one-to-one manner the system of equations

$$\xi_i = \Sigma \{f_i : (\xi_i, f_i) \in G\} \qquad (i = 1, 2, \ldots . .)$$

A simple discussion shows that because of our hypothesis on G such a system has always a formal solution $(r_1, r_2, \ldots . , r_m)$ in which each of the components r_i is a formal power series in the variables of X (and coefficients in the semi-ring of non-negative integer). In fact, for any word f the coefficient of f in r_i is exactly n_f, the number of times f is produced by G when the initial letter is ξ_i.

Hence, if G is such that $n_f \leq 1$ identically, and if we define the representation $\bar{\mu}$ by the condition that for each $x \in X$, $\bar{\mu}x = t\mu x$,

the image of r_i by $\bar{\mu}$ is precisely the desired generating function. As observed before this is the case for the sets K (or $\bar{K}$) defined by a finite automaton with a push-down storage.

For instance let us consider the set D_N introduced earlier. Clearly, D_N is the disjoint union of the sets $D_{N,i}$ $(-N \leq i \leq N)$ consisting of all words from D_N which begin with the letter x_i. Now is $f \in D_{N,i}$ the hypothesis that the first return to an empty memory occurs at the last letter of f implies that $f = x_i f' x_{-i}$ where f' is a word (possibly empty) of which the memory contains only x_i. Hence, introducing unknown ξ_i $(-N \leq i \leq N)$ in one-to-one correspondence with the letters x_i we have:

$$\xi_i = x_i (1 + \xi_{-i} - \Sigma \xi_i)^{-1} x_{-i}.$$

This can be brought to polynomial form by replacing each equation by the pair:

$$\xi_i = x_i x_{-i} + x_i \bar{\xi}'_i x_{-i}$$

$$\bar{\xi}'_i = \bar{\xi}_i \bar{\xi}'_i - \xi_i \quad \text{where} \quad \bar{\xi}_i = \sum_{j \neq -i} \xi_j .$$

For the case of an arbitrary finite automaton with push-down storage, the equations are slightly more complicated but their obtention is a quite straightforward matter [18].

5. PROBABILITIES OF ABSORPTION

Given a finite automaton with push-down storage A, let us call probability of absorption π_A the probability measures of the set of all infinite input sequences which have at least one left factor belonging to K. From the remarks of the last section, it follows instantly that π_A is an *algebraic function* of the entries of the matrices μx. This result is essentially due to Kesten. It is worthwhile to contrast it with the fact that for finite automata (without unbounded memory) the corresponding probability is always a rational function of the entries of μ.

In the opposite direction, probabilities attached to a more general type of unbounded memory usually fail to be algebraic. For instance, let A and A' be two finite automata with counter (i.e., let the internal alphabets of A and of A; consist of a single letter). Simple computation shows that the generating function of the set $K \cap K'$ may have a logarithmic singularity, hence it may be an (elementary) transcendant function [17]. From an analytic point of view, this corresponds to the well known fact that the Hadamard product of two algebraic functions is not necessarily algebraic.

Geometrically it is one way of interpreting the essentially deeper character of the two dimensional random walk over the one dimensional absorption problems.

A more directly probabilistic implication of these remarks can be obtained when it is assumed that $\overline{S}$ (the distinguished final set of states) is $\{s_o\}$ (the distinguished initial state) and that, similarily, $\overline{G} = \{g_o\}$ = the empty word. Then, K defines the support of a regular event in Feller's terminology and the process is indeed a recurrent one if the letters x_i are produced independently with constant positive probabilities. Under this hypothesis the character persistent or transient of the recurrent event depends only on the analytic nature of the singularity of the generating function that is nearest to 0.

Thus, if A is a strictly finite automaton, the recurrent event, if persistent, has necessarily a finite mean recurrence time. If A uses in nontrivial fashion its unbounded memory, the event has always an infinite mean recurrence time and it can be persistent only if A is in fact an automaton with a counter and if the probabilities themselves satisfy a certain equation. Again these results go back to Kesten. However, the techniques described here allow generalisation from the case of a free group to that of an extension of a free group by a finite monoid.

ACKNOWLEDGEMENT

Acknowledgement is made to the Commonwealth Fund for the grant in support of the visiting professorship of biomathematics in the Department of Preventive Medicine at Harvard Medical School.

REFERENCES

1. Bar-Hillel, Y., Perles, M., and Shamir, E.: On Formal Properties of Simple Phase Structure Grammars. Techn. Report No. 4, Applied Logic Branch, The Hebrew University of Jerusalem, July 1960.
2. Birkeland, R.: Sur la Convergence des Developments qui Expriment les Racines de l'Equation Algebrique Generale. C. R. Acad. Sciences, *171:* 1370-1372, 1920; *172:* 309-311, 1921.
3. Chomsky, N.: A Note on Phrase Structure Grammars. Information and Control, *2:* 393-395, 1959.
4. Chomsky, N.: On Certain Formal Properties of Grammar. Information and Control, *2:* 137-167, 1959.
5. Chomsky, N.: Context Free Grammars and Push-Down Storage. Quarterly Progress Reports, No. 65, Research Laboratory of Electronics, M.I.T., 1962.

6. Ginsburg, S.: Some Remarks on Abstract Machines. Trans. Am. Math. Soc., *96*: 400-444, 1960.
7. Ginsburg, S., and Rose, G. F.: Operations which Preserve Definability in Languages. Technical Memorandum. System Development Corporation, Santa Monica (Calif.), 1961.
8. Ginsburg, S., and Rice, H. G.: Two Families of Languages Related to Algol. Technical Memorandum. System Development Corporation, Santa Monica (Calif.), 1961.
9. Kesten, M.: Symmetric Random Walks on Groups. Trans. Am. Math. Soc., *92:* 336-354, 1959.
10. Kleene, S. C.: Representation of Events in Nerve Nets and Finite Automata. Automata Studies, Princeton University Press, pp. 3-41, 1956.
11. Parikh, R. J.: Language Generating Devices. Quarterly Progress Report, No. 60, Research Laboratory of Electronics, M.I.T., pp. 199-212, January 1961.
12. Post, E.: A Variant of a Recursively Unsolvable Problem. Bull. Am. Math. Soc., *52*: 264-268, 1946.
13. Rabin, M. O., and Scott, D.: Finite Automata and Their Decision Problems. I.B.M. Journal of Research, *3:* 115-125, 1959.
14. Raney, G. N.: Functional Composition Patterns and Power-Series Reversion. Trans. Am. Math. Soc., *94:* 441-451, 1960.
15. Redei, L.: Die Verallgemeinerung der Schreierschen Erweiterungs Theorie. Act. Sci. Math., Szeyed *14:* 252-273, 1952.
16. Scheinberg, S.: Note on the Boolean Properties of Context Free Languages. Information and Control, *3:* 372-375, 1960.
17. Schützenberger, M. P.: Un Probleme de la Theorie des Automates. Seminaire Dubriel Pisot (Paris), December 1959.
18. Schützenberger, M. P.: On a Family of Formal Power Series. Submitted to Proc. Am. Math Soc.

RESEARCH NOTE

NC-167 Thomas J. Watson Research Center, Yorktown Heights

ON AN ABSTRACT MACHINE PROPERTY PRESERVED UNDER THE SATISFACTION RELATION

by

M. P. Schützenberger *

11/12/62

ABSTRACT: Application of classical results on finite monoids to the Elgot-Rutledge theory gives a new property of machines that is preserved under the satisfaction relation.

---

* Presently at the University of Poitier in Poitier, France.

IBM

1.

I. INTRODUCTION

The present note is a straightforward application to the theory of C. C. Elgot and J. D. Rutledge of basic results of D. D. Miller and A. H. Clifford. We refer to these two papers for further motivation and bibliography. I gladly acknowledge my indebtedness to C. C. Elgot for most of my notions on this topic.

For simplicity it will be assumed once for all that all the machines considered here are finite, that all their transitions are defined and that all their states are accessible from the initial state. The input alphabet X and the output set $Y = \{y_j\}$ $(1 \leq j \leq n)$ are fixed finite sets. Thus a machine $\mathcal{M}$ can be identified with a triple $(S, \sigma, \bar{\beta})$ where:

i) S is a finite set of states;

ii) σ is a right regular mapping onto S of the free monoid F generated by X;

iii) $\bar{\beta}$ is a mapping of S into the set $\{0, 1, \ldots, n\}$ of integers.

We recall that σ is a right regular mapping iff for all $f, f', f'' \in F$, $\sigma ff'' = \sigma f'f''$ only if $\sigma f = \sigma f'$. With this notation the initial state is σe (e = the neutral element of F); the transition function $S \times X \to S$ is defined for all $s \in S$ and $x \in X$ by $s \cdot x = \sigma((\sigma^{-1} s)x)$ (and, more generally, for all $f \in F$, $s \cdot f = \sigma((\sigma^{-1} s)f)$); the behavior of $\mathcal{M}$ is the set of all pairs (f, y_i) where $f \in F$, $y_i \in Y$ are such that $i = \bar{\beta}\sigma f$.

2.

Thus another machine $\mathcal{M}' = (T, \tau, \bar{\delta})$ satisfies $\mathcal{M}$ iff, for all $f \in F$, $(\bar{\beta}\sigma f - \bar{\delta}\tau f)\cdot\bar{\beta}\sigma f = 0$.

Finally, a subset S' of states of $\mathcal{M}$ is compatible in the sense of C. C. Elgot and J. D. Rutledge iff, for all $s_1, s_2 \in S'$, $f \in F$, $s_3 = s_1 \cdot f$, $s_4 = s_2 \cdot f$, one has $(\bar{\beta}s_3 - \bar{\beta}s_4)\bar{\beta}s_3\bar{\beta}s_4 = 0$.

We shall associate to $\mathcal{M}$ a finite collection Gp $\mathcal{M}$ of finite groups and to each $G \in$ Gp $\mathcal{M}$ a finite collection Comp G of quotient groups G/N such that the following relation holds:

Main Property. If $\mathcal{M}'$ satisfies $\mathcal{M}$ there corresponds to each $G \in$ Gp $\mathcal{M}$ at least one $K \in$ Gp $\mathcal{M}'$ and at least one $G' \in$ Comp G such that G' is a homomorphic image of a subgroup of K.

It is not claimed that this property is useful for solving algorithmically the state minimization problem. On the other hand, it is vacuous only if there exists a natural number p such that for all f, the set $\sigma f^p F f^p$ is compatible.

It has been pointed out by C. C. Elgot that the property could be part of a proof showing that if the monoid of $\mathcal{M}$ is a group (cf. below), the same is true for any minimal machine satisfying $\mathcal{M}$.

Clearly the main property could be formulated without recourse to machine terminology in terms of two partitions

3.

$\{R_i\}_{0\leq i\leq n}$ and $\{R'_i\}_{0\leq i\leq n}$ of F into regular events such that, for all positive i, $R_i \subset R'_i$.

II. VERIFICATION OF THE MAIN PROPERTY

Let $\mathcal{M} = (S, \sigma, \bar{\beta})$ be a fixed machine and define as usual a homomorphism μ of F onto quotient monoid $M = \mu F$ by setting for all $f, f' \in F$

$\mu f = \mu f'$ iff, for all $s \in S$, $s \cdot f = s \cdot f'$.

Definition 1. Gp $\mathcal{M}$ is the family of all subsets G of M that have the following properties:

i) G is isomorphic to a group;

ii) G is not properly contained in another subset of $\mathcal{M}$ isomorphic to a group.

Definition 2. For each $G \in$ Gp $\mathcal{M}$, Comp G is the set of all quotient groups $G' = G/N$ for which the normal subgroup N of G is such that $\sigma\mu^{-1}N$ is a compatible subset of S in the sense of C. C. Elgot and J. D. Rutledge.

We reformulate in the following terms a fundamental result of A. A. Miller and A. H. Clifford.

Theorem 1. If the subset G of M is isomorphic to a group, it contains one and only one idempotent u; then the set $G_u = \{m \in M: m \in uMu;\ u \in Mmu \cap umM\}$ belongs to Gp $\mathcal{M}$ and it admits G as a subgroup.

4.

It follows that we can attach to each $f \in F$ a well defined group $G_f \in \underline{Gp}\ \mathcal{M}$ by the following construction based upon the remark that for all $f, f', f'' \in F$, one has $S.f'f \subset S.f$ and $\underline{Card}\ S \cdot f'ff'' \leq \underline{Card}\ S \cdot f$.

Let k be the least natural number such that $S.f^{k+1} = S.f^k$ ($k < \infty$ since $\underline{Card}\ S < \infty$); let $\bar{k}$ be the least natural number such that, for all $s \in S$, $s.f^{k+\bar{k}} = s.f^k$ ($\bar{k} < \infty$ since f determines a permutation of $S.f^k$); let $\bar{f} = f^{k+k'}$ where k' is the least natural number congruent to $-k$ modulo $\bar{k}$.

By construction, for all $s \in S$, $s.\bar{f}\bar{f} = s.\bar{f}$ and, thus, $u = \mu\bar{f}$ is an idempotent of M.

Consider any $m \in G_u$; since $umu = m$, the set $\mu^{-1}m \cap \bar{f}F\bar{f}$ is not empty. Hence $G_u \subset \mu(\bar{f}F\bar{f})$.

Now let $H_f = \{f' \in F: f' \in \bar{f}F\bar{f}, \underline{Card}\ S.f' = \underline{Card}\ S.\bar{f}\}$ and verify that $G_u = \mu H_f$ and that G_u is isomorphic to a group. Indeed, if $m_1 \in G_u$ and $f_1 \in \mu^{-1}m_1 \cap \bar{f}F\bar{f}$, the existence of $m_2 \in G_u$ such that $m_1m_2 = u$ implies the existence of at least one $f' \in F$ such that $S.\bar{f} = S.f_1f'$. However, since $f_1 \in \bar{f}F\bar{f}$ implies $S.f_1 \subset S.\bar{f}$, this gives $S.f_1 = S.\bar{f}$, proving $G_u \subset \mu H_f$.

Reciprocally, $f' \in H_f$ implies $S.f' = S.\bar{f}f' = S.\bar{f}$. Thus $\{\mu f'^p: p \geq 0\}$ is contained in H_f and it is isomorphic to a group having $u = \mu\bar{f}$ as its neutral element. Hence, setting $m' = \mu f'$

there exists a natural number p' such that $m'' = \mu f'^{p'}$ satisfies $m'm'' = m''m' = u$. Thus $m', m'' \in uMu$; $u \in Mmu \cap umM$ (since $m'' \in M$) proving $m = \mu f' \in G_u$ and concluding the verification.

Now let $\mathcal{M}' = (T, \tau, \bar{\delta})$ be another machine; the homomorphism π of F onto a quotient monoid P is defined for all $f, f' \in F$ by $\pi f = \pi f'$ iff, for all $t \in T$, $t.f = t.f'$.

We shall use repeatedly the fact that $\mu\pi^{-1}$ is a mapping of the family of all subsets of P into the family of all subsets of M such that for all $P', P'' \subset P$, one has $(\mu\pi^{-1}P')(\mu\pi^{-1}P'') \subset \mu\pi^{-1}P'P''$. Thus, particularly if P' is stable (i.e., if $P'^2 \subset P'$), its image $\mu\pi^{-1}P'$ is also a stable subset of M.

Similar properties hold for $\pi\mu^{-1}$.

Remark 1. For each $G \in \mathrm{Gp}\,\mathcal{M}$, there exists at least one $K \in \mathrm{Gp}\,\mathcal{M}'$ and a subgroup $\bar{K}$ of K such that $\bar{K} = K \cap \pi\mu^{-1}G$ and $G \subset \mu\pi^{-1}\bar{K}$.

Proof. By construction $\pi\mu^{-1}G$ is a finite stable subset of P. Hence we can find at least one element $\bar{f} \in \mu^{-1}G$ having the following properties:

i) $\mu\bar{f} = u$, the idempotent of G;

ii) $\pi\bar{f} = v$, an idempotent of P;

iii) For all $f' \in \mu^{-1}G$, $\mathrm{Card}\ T.\bar{f} \leq \mathrm{Card}\ T.f'$.

Thus by the construction recalled above, $G = G_u$; $K = K_v \in \mathrm{Gp}\,\mathcal{M}'$,

6.

and $K = \pi L_{\bar{f}}$ where $L_{\bar{f}} = \{f' \in F: f' \in \bar{f}F\bar{f}, \underline{\text{Card}}\ T.f' = \underline{\text{Card}}\ T.\bar{f}\}$.

Because of iii, $H_{\bar{f}} \subset L_{\bar{f}}$; $\mu H_{\bar{f}} = G$; $\pi\mu^{-1}G \cap K = \pi H_{\bar{f}} = \bar{K} \subset K$.

Since K is finite and $\bar{K}$ is stable, we have verified that $\bar{K}$ is a subgroup of K.

Let us now define a mapping ρ of $\bar{K}$ into the family of all subsets of G by setting for all $k \in \bar{K}$, $\rho k = G \cap \mu\pi^{-1}k$.

Remark 2. The mapping ρ is a homomorphism of $\bar{K}$ onto the quotient group G/N where $N = \rho v$ $(= G \cap \mu\pi^{-1}\bar{f})$.

Proof. This is a well-known computation: since G is a stable subset of M, ρ maps every stable subset of $\bar{K}$ (and in particular $\{v\}$) onto a stable subset of G, hence onto a subgroup since G is finite and $\rho k \neq \emptyset$ for all $k \in \bar{K}$.

Let k be any element of $\bar{K}$ and take $k' \in \bar{K}$; $q, q' \in G$ such that $kk' = v$, $g \in \rho k$; $g' \in \rho k'$. Since $k'k = v$, we have the relations:

$$(\rho k)g'g \subset (\rho kk')g = Ng \subset \rho kk'k = \rho k;$$

$$gg'\rho k \subset g\rho k'k = gN \subset \rho kk'k = \rho k.$$

Because G is a group, however, $(\rho k)g'g \subset \rho k$ implies $(\rho k)g'g = \rho k$ and, similarly, $gg'\rho k = \rho k$. Thus $\rho k = gN = Ng$, proving that N is a normal subgroup and ρ an epimorphism $\bar{K} \to G/N$.

This essentially concludes the verification of the main property. Indeed, if $\mathfrak{M}'$ satisfies $\mathfrak{M}$, the element $t = \tau\bar{f} \in T$ is

7.

a state of $\mathcal{M}'$ and, by a special case of Theorem 2 of C. C. Elgot and J. D. Rutledge, we know that the set $\sigma\tau^{-1}t$ (which contains $\sigma\mu^{-1}N$) is a compatible set of states.

Example. Let X = {x}, a single letter, i.e., let $\mathcal{M}$ be an input-free machine. (cf. C. C. Elgot and J. D. Rutledge.) Then taking f = x in the construction described after Th. 1, Gp $\mathcal{M}$ consists of a single cyclic group with $\bar{k}$ elements. Our main property asserts that any input-free machine which satisfies $\mathcal{M}$ has a loop with $\bar{k}'$ states where $\bar{k}'$ is some multiple of a divisor d' of $\bar{k}$ such that if $dd' = \bar{k}$ the set $\{\sigma f^{\bar{k}+i}: 0 \leq i < d\}$ of states of $\mathcal{M}$ is compatible.

Remark. Let us recall that $\mathcal{M}$ is minimal iff, for all $f, f' \in F$, the relation $\sigma f \neq \sigma f'$ implies that, for at least one $f'' \in F$, one has $\bar{\beta}\sigma ff'' \neq \bar{\beta}\sigma f'f''$. Under the hypothesis that both $\mathcal{M}$ and $\mathcal{M}'$ are minimal, the homomorphisms μ and π depend only upon the partitions $\{R_i\}$ and $\{R'_i\}$ of F where $R_i = (\bar{\beta}\sigma)^{-1}i$ and $R'_i = (\bar{\delta}\tau)^{-1}i$ for $0 \leq i \leq n$. Then, by an argument symmetric to the one used by C. C. Elgot and J. D. Rutledge, it is easily seen that the condition "$\sigma\mu^{-1}N$ is compatible" can be replaced by the two-sided condition: for all $f, f' \in \mu^{-1}N$ and $f_1, f_2 \in F$, the elements $f_3 = f_1 f f_2$ and $f_4 = f_1 f' f_2$ satisfy $(\bar{\beta}\sigma f_3 - \bar{\beta}\sigma f_4)\,\bar{\beta}\sigma f_3\,\bar{\beta}\sigma f_4 = 0$.

8.

REFERENCES

C. C. Elgot and J. D. Rutledge, "Machine properties preserved under state minimization," Switching Circuit Theory and Logical Design, Proceedings of the Third Annual AIEE Symposium, 61-70 (September 1962).

D. D. Miller and A. H. Clifford, "On regular D-classes in semigroups," Trans. Am. Math. Soc., 82, 270-280 (1956).

RESEARCH NOTE

NC-173 Thomas J. Watson Research Center, Yorktown Heights

ON THE MINIMUM NUMBER OF ELEMENTS IN A CUTTING SET OF WORDS

by

M. P. Schutzenberger
11/1/62

Let X be an alphabet of $k < \infty$ letters and F the set of all words in this alphabet. We shall say that a subset K of F is a cutting set iff there exists only a finite number of words f of F which have no factorization of the form $f = f_1 f_2 f_3$ with $f_2 \in K$, $f_1, f_3 \in F$. We shall limit our attention to the cutting sets consisting of words of a fixed length n and we intend to verify that the minimum number of words in such a set is $n^{-1}k^n (1 + \alpha(k,n))$ where $\alpha(k,n)$ tends to zero when $\mathrm{Max}(k,n) \to \infty$.

First inequality

Let us recall that a word $f \in F^+$ (the set of all nonempty words of F) is primitive (or aperiodic) iff $f = f'^p$ implies $p = 1$; every $f \in F^+$ can be written in one and only one manner as f'^p with $p > 0$ and f' primitive.

Two words f and f' are conjugate iff there exist $f_1, f_2 \in F$ such that $f = f_1 f_2$ and $f' = f_2 f_1$; then $fg = gf'$ for all $g = (f_1 f_2)^p f_1$, $p \geq 0$.

IBM

Reciprocally, if g is a right and a left factor of some word f'', i.e., if there exist f and f' such that $fg = gf' = f''$, induction on the length $|g|$ of g shows that for some $f_1, f_2 \in F$ and $p \geq 0$ one has $g = (f_1 f_2)^p f_1$, $f = f_1 f_2$, $f' = f_2 f_1$, $f'' = (f_1 f_2)^{p+1} f_1$.

Finally, the number of classes of conjugate primitive words of length n is $\psi_k(n) = n^{-1} \sum_{d|n} k^{n/d} \mu(d)$ where μ denotes the Möbius function [1].

Let f and f' be two primitive words whose lengths divide n and assume that for some positive p the words f^p and f'^p have a common factor f'' of length n. This implies $f'' = (f_2 f_1)^d$ and $f'' = (f'_2 f'_1)^{d'}$ where the words f_1, f_2, f'_1, f'_2 satisfy $f = f_1 f_2$ and $f' = f'_1 f'_2$. Hence, since f and f' are primitive, $d = d'$ and f and f' are conjugate. It follows that the minimum number of words in a cutting set is at least equal to $\sum_{d|n} \psi_k(d) = n^{-1} \sum_{d|n} k^{n/d} \phi(d)$ (where ϕ denotes Euler's function) and, consequently, that $\alpha(k, n) \geq 0$.

Second inequality

We exhibit a cutting set C having exactly $\sum_{m \leq n} \psi_k(m)$ words with the help of the following construction, which has been studied by K. T. Chen, R. H. Fox, and R. C. Lyndon [2].

Let X be totally ordered by $\leq$ and let $\leq$ also denote the induced lexicographic order on F. Define the subset H of F^+ by:

$f \in H$ iff, for all $f', f'' \in F^+$, $f = f'f''$ implies $f < f''f'$.

Clearly for each n, the set of all $h \in H$ of length n is a set of representatives of the classes of conjugate primitive words of this length. Further, it has been proved by the authors quoted above that $H = H'$ where $H' \subset F^+$ is defined by the seemingly more restrictive condition:

$f \in H'$ iff, for all $f', f'' \in F^+$, $f = f'f''$ implies $f < f''$.

We recall the proof for the sake of completeness.

Let $f \in H'$; $f = f'f''$; $f', f'' \in F^+$. Since $|f''| < |f|$ (where $|f|$ denotes the length of f), the condition $f < f''$ implies $f < f''f'''$ for all $f''' \in F$. Hence, $H' \subset H$.

In order to show $H \subset H'$ we verify first that H contains no word f such that there exist $f_1, f_2, f_3 \in F^+$ satisfying $f = f_1 f_2 = f_2 f_3$. Indeed, let $f = f_1 f_2 = f_2 f_3$; $f < f_2 f_1$ and $f < f_3 f_2$; either $|f_2| < |f_1|$, or $|f_1| \leq |f_2|$.

In the first case, $|f_2| < |f_1|$, there exists $f_4 \in F^+$ such that $f_1 = f_2 f_4, f_3 = f_4 f_2$ implying $f = f_2 f_4 f_2 < f_3 f_2 = f_4 f_2 f_2$ and, consequently, $f_2 f_4 < f_4 f_2$. Thus $f_2 f_2 f_4 < f_2 f_4 f_2 = f$, showing $f \notin H$.

In the second case, $|f_1| \leq |f_2|$, there exists $f_4 \in F$ such that $f_2 = f_1 f_4 = f_4 f_3$ implying $f = f_1 f_4 f_3 \leq f_2 f_1 = f_1 f_4 f_1$ and, consequently, $f_3 \leq f_1$. Thus (since $|f_1| = |f_3|$), $f_3 f_1 f_4 \leq f_1 f_4 f_3 = f$, showing again $f \notin H$.

Consider now $f \in H$ and any factorization $f = f'f''$ with $f', f'' \in F^+$. By our last remark, f'' can never be a right factor of f. Thus $f < f''f'$ implies $f < f''$. Hence $f \in H'$, and the proof is concluded.

Now let C be the set of all the left factors of length n of all the words of the form h^p with $p > 0$, $h \in H$, $|h| \leq n$.

We verify that C is a cutting set, i.e., that any infinite sequence $s = x_{i_1} x_{i_2} \ldots x_{i_j} \ldots$ of letters of X has at least one factor in C. Since $X \subset H$ we can assume $n > 1$, and since $\underline{\text{Card}}\ X < \infty$ we can also assume that the left factor $f = x_{i_1} x_{i_2} \ldots x_{i_n}$ of length n of s is $\leq$ any other factor $x_{i_j} x_{i_{j+1}} \ldots x_{i_{j+n}}$ of the same sequence s. Thus any factor f' of f satisfies $x_{i_1} x_{i_2} \ldots x_{i_{|f'|}} \leq f'$. This shows directly that $f \in H' \subset C$ when f admits no word of F^+ as a right and a left factor.

In the remaining cases, let $g = (f_1 f_2)^p f_1$ be the word of maximal length $< |f|$ which is a right and a left factor of f. We verify that $f_1 f_2 \in H$. Indeed, because of the maximality of $|g|$, the word $f_1 f_2$ is primitive, and we can define g_1 and g_2 by the conditions $g_1 g_2 = f_1 f_2$ and $g_2 g_1 \in H$.

If $|g_1| \leq |f_1|$ or if $p > 0$, the word $g_2 g_1$ itself is a factor of f since $f = (f_1 f_2)^{p+1} f_1$. Hence, $f_1 f_2 \leq g_2 g_1$ and $g_2 g_1 \leq f_1 f_2$ $(= g_1 g_2)$, showing that $f_1 f_2 = g_2 g_1 \in H$.

If $p = 0$ and $|f_1| \leq |g_1|$, there exists f_3 such that $g_1 = f_1 f_3$, $f_2 = f_3 g_2$, and the left factor $g_2 f_1$ of $g_2 g_1$ is a factor of $f = f_1 f_2 f_1 = f_1 f_3 f_2 f_1$. Hence, since $g_2 g_1 \leq g_1 g_2 = f_1 f_2$, the word $g_2 f_1$ is equal to a left factor of f, that is, $|g_2| = 0$, since by hypothesis $f_1 = g$ is the longest word to be a right and a left factor of f. It follows that $p = 0$ and $|f_1| \leq |g_1|$ imply $f_1 = g_1$ and the verification is concluded.

Now, $\underline{\text{Card}}\ C = n^{-1} k^n (1 + \alpha'(k,n)) = \sum_{0<m\leq n} \psi_k(m) \leq \sum_{0<m\leq n} m^{-1} k^m$ $= n^{-1} k^n \sum_{0\leq j<n} n(n-j)^{-1} k^{-j}$. Thus for each $\varepsilon > 0$ there exists a finite number k_ε such that, for all $k > k_\varepsilon$ and n, one has $\alpha(k,n) \leq \alpha'(k,n) < \varepsilon$.

Finally, let the set C' consist of all the words x^{n+1} $(x \in X)$ and of all the words $x'f$ where $x' \in X$, $f \in C$, and where the first letter $x'' \in X$ of f satisfies $x'' < x'$. C' is a cutting set because it contains $x_{i_1} x_{i_2} \ldots x_{i_{n+1}}$ if $x_{i_2} x_{i_3} \ldots x_{i_{n+1}}$ is $\leq$ any other factor of length n of the infinite sequence $x_{i_1} x_{i_2} \ldots$. Now, $\underline{\text{Card}}\ C' = (n+1)^{-1} k^{n+1} (1 + \alpha''(k, n+1)) \leq k + (k-1)\ \underline{\text{Card}}\ C$ $\leq k + (k-1) \sum_{0<m\leq n} m^{-1} k^m = (n+1)^{-1} k^{n+1} (1 + (n+1) k^{-n} + (n.n-1.k)^{-1}$ $+ \ldots + (2.1.k^n)^{-1} - k^{-n-1})$. Thus, for each $\varepsilon > 0$ and k, there exists a finite number $n_{k,\varepsilon}$ such that for all $n > n_{k,\varepsilon}$ one has $\alpha(k,n) \leq \alpha''(k,n) < \varepsilon$ and the property is entirely verified.

6

Remark. For $n = 1, 2$, or 3, one has Card $C' = \sum_{d|n} \psi_k(d)$. For $n = 4$, the same bound is attained by the set C'' consisting of all words $xx'x''x'''$ where x, x', x'', and x''' satisfy one of the following mutually exclusive conditions:

i) $x = x' = x'' = x'''$;
ii) $x' < x$ and $x' \leq x'' \leq x'''$;
iii) $x' < x$, $x''' < x''$ and $x'' < x$;
iv) $x' < x$, $x''' < x''$, $x'' = x$ and $x''' \leq x'$.

For $n = 5$ and $k = 2$ the minimum number of elements in a cutting set is $9 = 1 + \psi_2(1) + \psi_2(5)$.

REFERENCES

[1] C. Moreau, in E. Lucas, Théorie des nombres, Paris, 1891, pp. 501-503.

[2] K. T. Chen, R. H. Fox, and R. C. Lyndon, "Free differential calculus IV," Annals of Mathematics 68, 82-86 (1958).

March, 1962

On A Family of Formal Power-Series

M. P. Schützenberger

1. Introduction

In [6] we considered three modules $R_{pol}(X) \subset R_{rat}(X) \subset R_{alg}(X)$ of formal power-series (with coefficients in a unital ring R) in the non-commuting variables $x \in X$. These formal power-series are related to polynomials and to Taylor series expansions of rational and algebraic functions.

We recall that the family $\mathfrak{R}$ of the so-called regular events consists of all subsets of a finitely generated free monoid $F(Z)$ that are a finite union of sets of the form $\phi^{-1}h$ $(= \{g \in F(Z) : \phi g = h\})$ for some homomorphisms ϕ of $F(Z)$ onto a finite quotient monoid $\phi F(Z) = \{h\}$ ([2],[7]). It is trivial that:

(I.rat). The generating function $c_{F'} = \Sigma \{g : g \in F'\}$ of any $F' \in \mathfrak{R}$ belongs to $R_{rat}(X)$.

(I'.rat). Any $a \in R_{rat}(X)$ can be represented in the form $a = \theta c_{F'} = \lim_{n \to \infty} \Sigma \{\theta g : g \in F', \deg g < n\}$ for some suitable $F' \in \mathfrak{R}$ and homomorphism $\theta : F(Z) \to R_{pol}(X)$.

However, if one replaces the condition that $\phi F(Z)$ is finite by the condition that $\phi F(Z)$ is abelian in the definition of $\mathfrak{R}$, the generating function of $\phi^{-1}h$ does not necessarily belong to $R_{alg}(X)$ [5]. Then one may ask

what type of monoids give a family $\mathfrak{L}'$ of subsets of $F(Z)$ having the properties (I.alg) and (I'.alg) derived from (I.rat) and (I'.rat) by substituting R_{alg} to R_{rat}. We shall show that a partial answer is given by the extensions of a free group by a finite monoid [4]. This provides alternative proofs of some theorems of [1] and [3].

This note is part of common research with N. Chomsky.

2. Preliminary definitions.

Let there be: 1) Three finite sets Z, S and X; 2) a homomorphism ψ of $F(Z)$ onto a finite monoid K; 3) Three mappings α', σ' and χ of (K,S,X) into $F(Z)$, S and the family of all subsets of K, respectively. For our present purpose there is no loss of generality in assuming that $\psi g \neq \psi 1$ if $g \neq 1$ and that every $g \in F(Z)$ has at least one right factor in each $\chi(k,s,x)$.

For any $\bar{g} = (g,s) \in \bar{G} = (F(Z),S)$ and $x \in X$: $\alpha(\bar{g},x) = \alpha(g,s,x) = g'$, the word of highest degree such that $g\,\alpha'(\psi g,s,x) = g'g''$, with $g'' \in \chi(\psi g,s,x)$. $\sigma(\bar{g},x) = \sigma(g,s,x) = \sigma'(\psi g,s,x)$, $\bar{g}.x = (\alpha(\bar{g},x), \sigma(\bar{g},x)) \in \bar{G}$.

In the usual fashion we extend this mapping $(\bar{G},X) \longrightarrow \bar{G}$ to a representation of $F(X)$ (= the free monoid generated by X) by mappings of $\bar{G}$ into itself. If $f \in F(X)$, $\bar{g} \in \bar{G}$ and $\bar{g}' = \bar{g}.f$, we write $\bar{g}' = (\alpha(\bar{g},f), \sigma(\bar{g},f))$.

Let $f' < f$ denote that f' is a proper (i.e., $\neq f$) left factor of f.

For each 5-tuple $j = (j_i)$ with arbitrary j_1, $j_3 \in F(Z)$; j_2, $j_4 \in S$; $j_5 \in F(Z) \cup \{0\}$ we define:

$$C(j) = \{f \in F(X) : f \neq 1, (j_1,j_2).f = (j_3,j_4)\}, \text{ if } j_5 = 0;$$

$$= \{f \in F(X) : f \neq 1, (j_1,j_2).f = (j_3,j_4);\ j_5 < \alpha(j_1,j_2,f') \text{ for each } f',\ 1 < f' < f\}, \text{ if } j_5 \in F(Z).$$

2.1. The generating function $c(j)$ of any $C(j)$ with $j_5 = 0$ belongs to $R_{alg}(X)$.

Proof. Let J be the set of all j's which satisfy any one of the conditions $j_3 \leq j_1 = j_5$, $j_1 < j_3 = j_5$, $j_3 \leq j_5 < j_1$, or $j_5 = 0$. Let $Y = \{y(j) : j \in J\}$ be a set of new variables. For each $j \in J$, $p(j) \in R^*_{pol}(X \cup Y)$ is defined as follows:

If $j_3 \leq j_1 = j_5$ or $j_1 < j_3 = j_5$

$$p(j) = \Sigma\{x : x \in X \cap C(j)\} + \Sigma\{xy(\alpha(j_1,j_2,x),\sigma(j_1,j_2,x), j_3, j_4, j_5) : x \in X,\ j_5 < \alpha(j_1,j_2,x)\}.$$

If $j_3 \leq j_5 < 1_{j_1}$

$$p(j) = p(j_1,j_2,j_3,j_4,j_1) + \Sigma\{y(j_1,j_2,j_5g,s,j_5g)\, y(j_5g,s,j_3,j_4,j_5g') : s \in S,\ g \in g'Z,\ j_5g \leq j_1\}.$$

If $j_5 = 0$

$p(j) = p(j_1,j_2,j_3,j_4,j_3) + \Sigma \{y(j_1,j_2,g,s,0)\ y(g,s,j_3,j_4,j_3) : s \in S,\ g \leq j_3\}$.

Clearly, each equation expresses a unique factorisation property of the words of $C(j)$ as products of elements of X and words from other sets $C(j')$. Hence, each equation is an identity if $p(j) = y(j) = c(j)$ for all $j \in J$.

Let J_d denote the subset of all 5-tuples such that deg j_1, deg j_2, $\leq d$. If $j \in J_d$ the right member of the equation which defines $p(j)$ contains only variables $y(j')$ with $j' \in J_d$ or with j' of the form $j_3 \leq j_1 = j_5$. In this last case deg $j_3 \leq d$ and deg j_1 - deg $j_3 \leq \max \{\deg \alpha(\bar{g},x) : \bar{g} \in \bar{G},\ x \in X\}$.

Now let $\psi_2 g$ denote, for any $g \in F(Z)$, the subset $\{(\psi g', \psi g'') : g'g'' = g\}$ of (K, K). If $\psi_2 g_1 = \psi_2 g_4$, $\psi_2 g_2 = \psi_2 g_5$, $j = (g_1g_2g_3,s,g_1,s',g_1g_2)$, $f \in C(j)$, induction on deg f' shows that for each $f' < f$, $\alpha(g_1g_2g_3,s,f') = g_1g_2g'$, $\alpha(g_4g_5g_3,s,f') = g_4g_5g'$ with the same g' and $\sigma(g_1g_2g_3,s,f') = \sigma(g_4g_5g_3,s,f')$. Thus, $C(j) = C(g_4g_5g_3,s,g_4,s',g_4g_5)$. It follows that there exists a finite d^* such that for any fixed $d \geq d^*$ and $j \in J_d$, each $y(j')$ with $j' \notin J_d$ in the right member of $p(j)$ can be replaced by $y(j'')$ with $j'' \in J_d$ and $C(j') = C(j'')$. Making this substitution the set $(p(j))_{j \in J_d}$ becomes a proper system in the notation of [6] and 2.1. is verified.

3. Verification of (I.alg).

(I.alg). If $\bar{\gamma}$ is a homomorphism of $F(X)$ into an extension $\bar{G} = \{\bar{g}\}$ of a free group G by a finite monoid H, the generating function $c_{\bar{g}}$ of any $\bar{\gamma}^{-1}\,\bar{g}$ belongs to $R_{alg}(X)$.

Proof. Let G be generated by $\{z_i\}$ $1 \leq i \leq m$; $Z = \{z_{i'}\}$ $i' = \pm i$ and γ^* be the homomorphism of $F(Z)$ onto G such that $(\gamma^* z_i)^{-1} = \gamma^* z_{-i}$ for all $z_i \in Z$.

(i) Let us consider the special case of $\bar{G} = G$. Then $\bar{\gamma}$ is given by a homomorphism $\gamma : F(X) \longrightarrow F(Z)$ and $\bar{\gamma} f = (\gamma^* \circ \gamma) f$. Since γ itself is determined by its restriction to the finite set X we can assume $m < \infty$. If $\rho : F(Z) \longrightarrow F(Z)$ is such that $\rho g z_i z_{-i} g' = \rho g g'$ and $\rho g = g$ for all g having no factor of the form $z_i z_{-i}$, the word ρg is the so-called reduced form of g and $\gamma^* \rho g = \gamma^* g$ with $\rho g' \neq \rho g$ i.f.f. $\gamma^* g' \neq \gamma^* g$.

We consider the following special case of the representation defined in the preceding section:

1) K and S are identified with $F' = \{g \in F(Z) : \deg g \leq d\}$ where $d = \max\{\deg \gamma x : x \in X\}$. 2) $\psi g = g$ if $g \in F'$, ψg = the right factor of degree d of g if $g \notin F'$. 3) $\alpha'(k,s,x) = \rho(s\gamma x)$; $\chi(k,s,x) = k\,\rho(s\gamma x) = \sigma'(k,s,x)$.

Thus, if $\bar{g} = (g,s) \in (F(Z),F)$ one computes successively $s\gamma x$, $\rho(s\gamma x)$, $g\,\rho(s\gamma x)$; $\alpha(g,s,x)$ and $\sigma(g,s,x)$ are determined by $\alpha(g,s,x)\ \sigma(g,s,x) = g\,\rho\,(s\gamma x)$ and $\sigma(g,s,x) = \psi(g\,\rho(s\gamma x))$. Induction on deg f shows that for each f the word $\alpha(1,1,f)\ \sigma(1,1,f)$ is precisely equal to ρf and the result is a consequence of 2.1.

(ii) In the general case [4], $\bar{\gamma}$ is given by a homomorphism $\phi : F(Z) \longrightarrow H$, and a mapping $\gamma : (H,X,H) \longrightarrow F(Z)$. Then $\bar{\gamma}f = (\gamma^{*}o\gamma(\phi 1,f,\phi 1), \phi f) \in (G,H)$ where $\gamma : (H, F(X), H) \longrightarrow F(Z)$ is defined by the identities:

for all h, $h' \in H$, f, $f' \in F(X)$, $x \in X$

$\gamma(h,1,h') = 1$; $\gamma(h,fxf',h') = \gamma(h,f,(\phi xf')h')\ \gamma(h\phi f,x,(\phi f')h')\ \gamma(h\phi fx,f',h')$.

Let $X' = \{x'(h,x,h') : (h,xh') \in (H,X,H)\}$ be a set of new variables; ξ and γ' are the homomorphisms of $F(X')$ into $F(X)$ and $F(Z)$ induced by $\xi\, x'(h,x,h') = x$ and $\gamma' x'(h,x,h') = \gamma(h,x,h')$. If $\bar{H} = (H,H) \cup \{0\}$ we define a representation $(\bar{H},F(X')) \longrightarrow \bar{H}$ of $F(X')$ by the identities:

for all $x' \in X'$, $0.x' = 0$; for all $x'(h,x,h') \in X'$ and $(h_1,h_2) \in \bar{H}$,

$(h_1,h_2).x'(h,x,h') = (h_1\ \phi x,h')$ if $h = h_1$ and $h_2 = (\phi x)h'$; $(h_1,h_2).x'(h,x,h') = 0$, otherwise.

Thus, for any $h \in H$, the restriction of ξ to $F'_h = \{f' \in F(X') : (\phi 1,h).f' = (h, \phi 1)\}$ is a 1-1 mapping onto $\{f \in F(X) : \phi f = h\}$ and for each $f' \in F'_h$,

$\gamma'f' = \gamma(\phi 1, f', \phi 1)$. It follows that $c_{\bar{g}} = \xi c'_{g,\bar{h},\bar{h}'} = \Sigma\{\xi f' : f' \in F(X') : \rho\gamma'f' = g, \bar{h}.f' = \bar{h}'\}$ for suitable $g \in F(Z)$, $\bar{h}$, $\bar{h}'$, $\in H$.

Now let S, K, ψ, α', χ, σ' be the same as in (i), $\bar{S} = (S,\bar{H})$. For each $\bar{s} = (s,\bar{h}) \in \bar{S}$, $x' \in X'$, we define: $\bar{\alpha}'(k,\bar{s},x') = \alpha'(k,s,\xi x')$; $\bar{\chi}(k,\bar{s},x') = \{(\chi(k,s,\xi x'), \bar{h}') : \bar{h}' \in \bar{H}\}$; $\bar{\sigma}'(k,\bar{s},x') = (\sigma'(k,s,\xi x'), \bar{h}.x')$. It is trivial that $(g,(s,\bar{h})).x' = (\alpha(g,s,\xi x'), (\sigma(g,s,\xi x'), \bar{h}.x'))$, identically. Hence, $c'_{g,\bar{h},\bar{h}'}$ (or $c'_{g,\bar{h},\bar{h}'}$ -1) is a component of the solution of a proper system $p' \in R^{*M}_{pol}(X' \cup Y)$. Clearly, if one extends ξ to a homomorphism $R^{*M}_{pol}(X' \cup Y) \longrightarrow R^{*M}_{pol}(X \cup Y)$ by $\xi y = y$ for all $y \in Y$, $p = \xi p'$ is again a proper system and (in the notation of [6]) $p(n) = \xi p'(n)$ for all n. This concludes the verification of (I.alg).

4. Verification of (I'.alg).

Let Z, G, γ^* be the same as in section 3, $1 < m < \infty$; $\mathcal{R}' = \{F' \subset F(Z) : F' = (\gamma^{*-1}1) \cap F''$, $F'' \in \mathcal{R}\}$.

(I'.alg). Any $a \in R^*_{alg}(X)$ can be represented in the form $a = \lim_{n \to \infty} \Sigma\{\theta g : g \in F', \deg g < n\}$ for some suitable $F' \in \mathcal{R}'$ and homomorphism $\theta : F(Z) \longrightarrow R_{pol}(X)$.

Proof. (i) Let a be a component of the solution of the proper system $(p_j) = p \in R^{*M}_{pol}(X \cup Y)$. The support, Supp. b, of any formal power-series b is the set of all words having a non-zero coefficient in b. Since each p_j belongs to $R_{pol}(X \cup Y)$ there exists $d^* < \infty$ such that any $f \in \{f' \in$ Supp. p_j, $1 \leq j \leq M\}$

either belongs to $F(X)$ or has a factorisation

$f = f_1 y_{i_1} f_2 y_{i_2} f_3 \cdots f_d y_{i_d} f_{d+1}$ with $f_1, f_2, \ldots f_{d+1} \in F(X)$, $y_{i_1}, y_{i_2}, \ldots, y_{i_d} \in Y$, $1 \leq d = \deg_Y f < d^*$. We introduce a set $Z = \{z(j,f,d,\epsilon) : 1 \leq j \leq M,\ f \in \underline{\mathrm{Supp}}.\ p_j,\ 1 \leq d \leq d^*,\ \epsilon = + \text{ or } -\}$ of new variables and make the definitions:

If $f \in F(X) \cap \underline{\mathrm{Supp}}.\ p_j$, $\theta z(j,f,d,\epsilon) = < p_j, f >$ if $d = 1$ and $\epsilon = +$, $= 1$, otherwise; $\zeta_j f = z(j,f,1,+)\ z(j,f,1,-)\ z(j,f,2,+)\ z(j,f,2,-) \cdots z(j,f,d^*,+)\ z(j,f,d^*,-)$.

If $f = f_1 y_{i_1} f_2 \cdots y_{i_d} f_{d+1} \in \underline{\mathrm{Supp}}.\ p_j$ as above, $\theta z(j,f,d',\epsilon) = < p_j, f > f_1$ if $d' = 1$, and $\epsilon = +$; $= f_{d'}$ if $1 < d' \leq \deg_Y f + 1$ and $\epsilon = +$; $= 1$, otherwise; $\zeta_j f = z(j,f,1,+) y_{i_1} z(j,f,1,-)\ z(j,f,2,+) y_{i_2} z(j,f,2,-) \cdots$ $\cdots z(j,f,d,+) y_{i_d} z(j,f,d,-)\ z(j,f,d+1,+)\ z(j,f,d+1,-) \cdots z(j,f,d^*,+)\ z(j,f,d^*,-)$.

Thus, $\theta q_j = p_j$ where $q_j = \Sigma \{ \zeta_j f : f \in \underline{\mathrm{Supp}}.\ p_j \}$ and $q = (q_j) \in R_M^*(Z \cup Y)$ is a proper system such that (in the notation of [6]) $\lim_{n \to \infty} \theta q(n) = p(\infty)$, the solution of p. Moreover, if $Q_j = \underline{\mathrm{Supp}}.\ q_j$, $P_j(n) = F(Z) \cap \underline{\mathrm{Supp}}.\ q_j(n)$ and if η_n is the homomorphism of $F(Z \cup Y)$ into $R_{pol}(Z)$ induced by $\eta_n z = z$, $\eta_n y_j = P_j(n)$ for all $z \in Z$, $y_j \in Y$, it follows from the definitions that, for all n, $q_j(n+1)$ is the generating function of $\Sigma \{ \eta_n g : g \in Q_j \}$. Hence it suffices to show that the sets $P_j(\infty)$ have the desired form.

(ii) Let $V \subset F(Z)$ consist of:

all words $z(j,f,d,+)\, z(j',f',1,+)$ or $z(j',f',d^*,-)\, z(j,f,d+1,-)$

with $d \leq \deg_Y f$ and j' equal to the index i_d of the d-th factor y_{i_d} of f;

all words $z(j,f,d,+)\, z(j,f,d,-)$ with $d > \deg_Y f$;

all words $z(j,f,d,-)\, z(j,f,d+1,+)$.

We take H in 1-1 correspondence with $\{ 0, 1, Z, (Z,Z') \}$ and define the homomorphism $\phi : F(Z) \longrightarrow H$ by

$\phi g = h_g$ if $\deg g < 2$;

$\phi g = 0$ if g has at least one factor of degree two not belonging to V;

$\phi g = h_{z,z'}$ if $\phi g \neq 0$ and $g \in z\, F(Z) z'$.

$H_j = \{h_{z,z'} : z = z(j,f,1,+),\ z' = z(j,f,d^*,-),\ f \in \underline{\text{Supp}}.\ p_j\}$.

The homomorphism $\gamma^* : F(Z) \longrightarrow G$ is defined by $(\gamma^* z(j,f,d,+))^{-1} = \gamma^* z(j,f,d,-)$ for all elements of Z.

Induction on n shows that $P_j(\infty) \subset D = \{g \in F(Z) : \gamma^* g = 1,\ \phi g \neq 0\}$.

Let $\overline{P}(j,d,d') = \{g \in F(Z) : g \in \underline{\text{Supp}}.(\lim_{n \to \infty} \eta_n g') : g' \in Q'(j,d,d')\}$ where $Q'(j,d,d')$ denotes the set of all $g' \in F(Z \cup Y)$ of the form $z(j,f,d,+) g''\, z(j,f,d',-)$ that are a factor of some $g \in Q_j$; $\overline{P} = \bigcup \{P(j,d,d') : 1 \leq j \leq M,\ 1 \leq d < d' \leq d^*\}$.

Then $\overline{P}(j,d,d') \subset D$ and $P_j(\infty) = \overline{P}(j,1,d^*) = \overline{P} \cap \phi^{-1} H_j$.

Thus, it suffices to show that, conversely, every $g \in D$ belongs to $\overline{P}$. This is trivial if $\deg g \leq 2$. We assume the result proved for all words of degree $< n$ and we consider $g \in D$ of degree $n > 2$.

Let the factorisation $g = z\, g'\, z'\, g''$ of g be determined by the condition that $zg'z'$ is the left factor $\neq 1$ of lowest degree of g that satisfies $\gamma^* zg'z' = \gamma^* 1$. Since γ^* is a homomorphism into a free group, this implies $\gamma^* g'' = \gamma^* z\, z' = \gamma^* g' = \gamma^* 1$.

If $g'' \neq 1$, the induction hypothesis shows that $z = z(j,f,d,+)$, $z' = z(j,f,d',-)$, $g'' = z(j',f',d'',+)\, g'''\, z(j',f',d''',-)$ for some j,j', f,f', d,d', d'',d''' and $g''' \in F(Z)$. Because of $\phi g \neq 0$, we have $j = j'$, $f = f'$, $d'' = d'+1$ and the result is proved in this case.

If $g'' = 1$, the induction hypothesis shows that $1 \neq g' = z(j',f',d'',+)\, g'''\, z(j',f',d''',-)$. Because of $\phi g \neq 0$ and $\gamma^* z\, z' = \gamma^* 1$, we have $z = z(j,f,d,+)$, $d'' = 1$, $d''' = d^*$, (i.e. $g' \in \phi^{-1} H_{j'}$), $z' = z(j,f,d,-)$ and $z(j,f,d,+)\, y_{j'}$, $z(j,f,d,-)$ is a factor of a word of Q_j. Thus, $g \in \overline{P}(j,d,d)$ and the verification of (I'.alg) is completed.

Acknowledgement. Acknowledgement is made to the Commonwealth Fund for the grant in support of the visiting professorship of biomathematics in the Department of Preventive Medicine at Harvard Medical School.

References

1. Kesten, M. Symmetric random walks on groups. Trans. Am. Math. Soc. 92: 336-354, 1959.

2. Rabin, M. O., and Scott, D. Finite automata and their decision problems. I.B.M. Journal of Research. 3: 114-125, 1959.

3. Raney, G. N. Functional composition patterns and power-series reversion. Trans. Am. Math. Soc. 94: 441-451, 1960.

4. Redei, L. Die Verallgemeinerung der schreierschen Erweiterungs Theorie. Acta. Sc. Math. Szeged. 14: 252-273, 1952.

5. Schützenberger, M. P. Un probleme de la theorie des automates. Seminaire Dubreil Pisot, Paris, 13 eme annee, Nov. 1959, No. 3.

6. Schützenberger, M. P. On a theorem of R. Jungen. To appear in Proc. Am. Math. Soc.

7. Shepherdson, J. C. The reduction of two way automata. I.B.M. Journal of Research. 3: 198-200, 1959.

Année 1963

Bibliographie

[1] Marcel-Paul Schützenberger. Sur les contraintes définissant certains modèles formels de langage. *Mathématiques et sciences humaines*, 4 :3–8, 1963. aussi paru dans les Cahiers mathématiques en 1970.

[2] Marcel-Paul Schützenberger and Steven Sherman. On a formal product over the conjugate classes in a free group. *J. Math. Anal. Appl.*, 7 :482–488, 1963.

[3] Marcel-Paul Schützenberger. Quelques remarques sur une construction de Schensted. In *Séminaire Dubreil-Pisot, année 1962-63*, Exposé 15, 4 mars 1963, 12 pages. Inst. H. Poincaré, Paris, 1963.

[4] Marcel-Paul Schützenberger. Certain elementary families of automata. In *Proc. Sympos. Math. Theory of Automata (New York, 1962)*, pages 139–153. Polytechnic Press of Polytechnic Inst. of Brooklyn, Brooklyn, New York, 1963.

[5] Marcel-Paul Schützenberger. On context-free languages and push-down automata. *Information and Control*, 6 :246–264, 1963.

[6] Marcel-Paul Schützenberger. Quelques remarques sur une construction de Schensted. *Math. Scand.*, 12 :117–128, 1963.

[7] Noam Chomsky and Marcel-Paul Schützenberger. The algebraic theory of context-free languages. In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Systems*, pages 118–161. North-Holland, Amsterdam, 1963. Traduction : “Algebraische Theorie kontextfreier Sprachen”, Kibern. Sb., Nov. Ser. 3, 195-242 (1966).

Centre de mathématique sociale et de statistique
17, rue Richer - Paris 9e

MATHÉMATIQUES ET SCIENCES HUMAINES

M. P. SCHUTZENBERGER

SUR LES CONTRAINTES DEFINISSANT CERTAINS MODELES FORMELS DE LANGAGE

L'un des objectifs partiels de la linguistique est l'établissement de grammaires, c'est-à-dire, pour une langue naturelle donnée, l'établissement d'un système explicite de règles qui permettent, en théorie du moins, de décider pour toute suite arbitraire de phonèmes si cette suite appartient ou non à l'ensemble des phrases grammaticalement correctes de la langue considérée. L'expérience semble indiquer qu'à un certain niveau une partie de ces règles est d'une nature assez simple pour pouvoir être discutée par des procédés purement formels. C'est là le cas des règles d'emploi des parenthèses ou, si l'on préfère, le formalisme de l'emboîtement des syntagmes et le but de cet exposé est de présenter quelques résultats obtenus dans cette direction par N.CHOMSKY, ses amis et leurs élèves; je renvoie aux travaux de cet auteur et G.A. MILLER pour une bibliographie complète de la question. Comme les problèmes proprement linguistiques ne seront pas abordés ici, il sera commode d'utiliser les notations suivantes qui ne risquent pas d'entraîner de confusion bien que les termes de "lettres" et de "mots" y soient utilisés d'une façon qui apparaîtra sans doute incongrue aux linguistes.

Soit X un ensemble fini dont les éléments sont appelés conventionnellement des lettres, F l'ensemble de toutes les séquences finies que l'on peut former avec ces lettres, ces séquences étant elles-mêmes appelées des mots. Techniquement, F est le monoïde libre engendré par X. Ceci veut dire que le mot vide (noté e) appartient à F et qu'à toute paire de mots f, f' de F est associé leur produit ff', c'est-à-dire par définition, le mot formé de f suivi de f'. Nous supposerons désormais que les lettres de X sont indexées par les nombres $\pm i$ $(1 \leq i \leq n)$ c'est-à-dire que X est l'ensemble $\{x_1, x_{-1}, x_2, x_{-2}, \ldots, x_n, x_{-n}\}$. Un mot f sera dit réduit s'il ne contient aucune paire de lettres consécutives ayant des indices opposés, c'est-à-dire pour $f = x_{i_1} x_{i_2} \cdots x_{i_m}$ si $i_1 \neq -i_2$, $i_2 \neq -i_3$, $\ldots$, $i_{m-1} \neq -i_m$. Il est clair qu'à chaque mot f on peut associer un mot réduit, disons αf, obtenu en effaçant dans f toutes les paires de lettres consécutives dont les indices sont opposés et en répétant cette opération aussi longtemps qu'elle est possible. Par exemple

$$\alpha\ (x_1\ x_{-2}\ x_3\ x_{-3}\ x_2\ x_2\ x_{-1}\ x_1)$$

est égal à $x_1\ x_2$ à la suite des opérations suivantes : effacement de $x_3\ x_{-3}$ et de $x_{-1}\ x_1$ ce qui donne $x_1\ x_{-2}\ x_2\ x_2$; effacement de $x_{-2}\ x_2$ ce qui donne finalement $x_1\ x_2$. On peut vérifier que l'ordre dans lequel sont effectuées les opérations d'effacement est sans effet sur le résultat final αf qui est donc, pour chaque f, un mot bien défini. Plus généralement, f et f' étant deux mots quelconques, on a l'identité: $\alpha(ff') = \alpha(\alpha f \alpha f')$. Par exemple, pour $f = x_1\ x_{-2}\ x_3\ x_{-3}$, $f' = x_2\ x_2\ x_{-1}\ x_1$, on a :

$$\alpha f = x_1\ x_{-2},\ \alpha f' = x_2\ x_2$$

4.

et comme plus haut $\alpha\,(x_1\ x_{-2}\ x_2\ x_2) = x_1\ x_2$. Cette construction définit un ensemble D de mots privilégiés: ceux dont la forme réduite est le mot vide e ou, si l'on veut, ceux qui peuvent être réduits à rien par les opérations successives d'effacement. Par exemple le mot

$$f = x_1\ x_{-2}\ x_3\ x_{-3}\ x_2\ x_2\ x_{-1}\ x_1\ x_{-2}\ x_{-1}$$

appartient à D puisqu'avec des notations évidentes l'on a :

$$f = x_1\ (x_{-2}\ (x_3\ x_{-3})\ x_2)\ (x_2\ (x_2\ (x_{-1}\ x_1)\ x_{-2})\ x_{-2})\ x_{-1}$$

par contre le mot

$$f' = x_1\ x_{-2}\ x_3\ x_{-3}\ x_2\ x_2\ x_{-1}\ x_1\ x_{-1}\ x_{-2}$$

obtenu en permutant les deux dernières lettres de f n'appartient pas à D parce que $\alpha f'$ est le mot non vide $x_1\ x_2\ x_{-1}\ x_{-2}$. D est défini algèbriquement comme le noyau de l'homomorphisme β de F sur le groupe libre (engendré par l'ensemble X^+ des $x \in X$ d'indices positifs) qui satisfait pour chaque lettre x_i la condition que $\beta\ x_{-i}$ soit l'inverse de $\beta\ x_i$ (dans le groupe évidemment).

Soit d'autre part X_1 un sous-ensemble de X, V un ensemble de mots de deux lettres et R (X_1, V) l'ensemble de tous les mots de F qui commencent par une lettre de X_1 et dont aucune paire de lettres consécutives ne forme un mot de V. Par exemple pour $X_1 = X^+$ comme plus haut et $V = \{x_i\ x_j : i < 0 < j\}$, R (X_1, V) est l'ensemble des mots de la forme ff' où toutes les lettres de f ont des indices positifs (ce que l'on notera $f \in F_+$), et où toutes les lettres de f' ont des indices négatifs. Nous dirons que l'ensemble $L = D \cap R\ (X_1, V)$ des mots appartenant à la fois à D et à R (X_1, V) est un langage CF (Context Free) standard. Pour l'exemple qui vient d'être donné, L est formé de tous les mots de la forme $f\bar{f}$ où f appartient à F_+ et où $\bar{f}$ est obtenu en "retournant" le mot f, ce qui donne $\tilde{f}$, et en remplaçant dans $\tilde{f}$ chaque lettre x_i par la lettre d'indice opposé x_{-i} -(Par exemple pour $f = x_1\ x_1\ x_3\ x_2$, on pose $\tilde{f} = x_2\ x_3\ x_1\ x_1$ et $\bar{f} = x_{-2}\ x_{-3}\ x_{-1}\ x_{-1}$). Plus généralement, soit donné un homomorphisme φ de F dans lui-même, c'est-à-dire pour chaque lettre x_i de X un certain mot $\varphi\ x_i$ (éventuellement vide). Considérant un langage CF standard L on forme l'ensemble de tous les mots $\varphi\ f$ où f est dans L et ceci constitue ce que nous appellerons un langage CF général. Je rappelle que si

$$f = x_{i_1}\ x_{i_2}\ x_{i_3} \cdots x_{i_n}$$

le mot $\varphi\ f$ est défini comme le produit

$$\varphi\ x_{i_1}\ \varphi\ x_{i_2}\ \varphi\ x_{i_3} \cdots \varphi\ x_{i_n}$$

et φ n'est donc pas autre chose qu'une règle systématique de réécriture. Reprenant l'exemple donné plus haut de $L = \{f\bar{f} : f \in F_+\}$ et posant

$$\varphi\ x_i = \varphi\ x_{-i} = x_i$$

pour chaque $x_i \in X^+$, on obtient le langage $M = \{f\tilde{f} : f \in F_+\}$ formé des mots en miroirs (ou "palindromes) $f\tilde{f}$ dont la seconde moitié $\tilde{f}$ est précisément la première moitié "retournée".

La définition qui vient d'être donnée d'un langage CF général n'est pas la définition originale de CHOMSKY mais il est facile de voir que ces deux définitions sont à très peu près équivalentes. Cette définition et les raisons pour lesquelles ces langages sont dits "context free" sont données dans l'ouvrage de CHOMSKY et MULLER indiqué en références.

Ceci dit nous pouvons examiner les contraintes qu'imposent aux lettres d'un mot l'hypothèse que celui-ci appartient à un langage CF donné L et, afin de simplifier, nous supposerons que L est standard. Ces contraintes sont d'une double nature: les premières résultent de l'appartenance à R (X_1, V) et peuvent être considérées comme purement locales en ce sens qu'un être dont la mémoire se limiterait à la dernière lettre lue pourrait vérifier pour une séquence arbitraire de lettres si elles sont ou non satisfaites. Il n'est pas utile de rappeler les illustrations linguistiques ou pseudo-linguistiques bien connues de ce type extrêmement banal de contraintes (Cf. plus bas).

Par contre, les contraintes imposées par l'appartenance à D sont un peu moins simples: elles peuvent s'étendre aussi loin que l'on veut et elles nécessitent en général une mémoire non bornée. En effet, d'après l'identité

$$\alpha\,(\alpha f\ \alpha f') = \alpha f f',$$

deux mots f_1 et f_2 sont tels qu'il existe un mot f" pour lequel $f_1 f"$ et $f_2 f"$ appartiennent simultanément à D, si et seulement si on a $\alpha f_1 = \alpha f_2$. Réciproquement, quand cette relation est vérifiée pour tout f" tel que $f_1 f"$ soit dans D on a aussi $f_2 f"$ dans D. Par conséquent ayant déjà lu un segment initial f_1 d'un mot, il est nécessaire de conserver trace en mémoire au moins de sa forme réduite αf_1 afin de pouvoir déterminer si le mot complet $f_1 f"$ appartient ou non à D. Notons qu'il existe $((2n-1)^k - 1)\ n\ (n-1)^{-1}$ mots réduits distincts de longueur au plus k; donc pour vérifier par lecture séquentielle de gauche à droite l'appartenance à D d'un mot quelconque de longueur $\leq 2k$, il faut disposer d'une mémoire pouvant accumuler approximativement $k\ (1 + \mathrm{Log}_2 n)$ bits d'information. En outre, l'exemple donné plus haut de

$$L = x_{i_1}\ x_{i_2} \cdots x_{i_m}\ x_{-i_m} \cdots x_{-i_2}\ x_{-i_1}$$

montre qu'une fraction arbitrairement grande de ces informations doit être gardée en mémoire un temps arbitrairement long, à savoir depuis le début jusqu'à la fin du mot. Par contre, et c'est ce qui fait peut-être l'intérêt des langages CF, les contraintes existant entre les lettres d'un mot de D n'interviennent chacune essentiellement qu'une fois et ne se croisent pas les unes les autres. Soit par exemple un mot $f \in D$ de la forme $f_1\ x_i\ f_2$ où f_1 ne contient ni la lettre x_i ni la lettre x_{-i}. Le mot f_2 est soumis à la condition d'avoir la forme $f_3\ x_{-i}\ f_4$ où f_3 ne contient ni x_i ni x_{-i} et où en outre $\alpha f_3 = e$: c'est en ce sens que nous dirons que la contrainte imposée par la présence x_i ne joue essentiellement qu'une fois bien qu'évidemment il puisse y avoir un nombre arbitrairement grand de lettres x_i dont chacune impose l'existence d'une lettre x_{-i} qui en permet l'effacement ultérieur. De plus, dans ce même exemple, si

$$f_3 = f_5\ x_j\ f_6$$

où f_5 ne contient aucune des lettres $x_{\pm j}$, le fait que l'on doive avoir $\alpha f_3 = e$

6.

entraîne que f_6 soit égal à $f_7\ x_{-j}\ f_8$ où cette fois $\alpha f_7 = e$. On a donc nécessairement :

$$f = f_1\ x_i\ f_5\ x_j\ f_7\ x_{-j}\ f_8\ x_{-i}\ f_4$$

avec

$$\alpha f_7 = \alpha f_5\ f_8 = \alpha f_1\ f_4 = e$$

et, dans un sens assez évident, la contrainte associée à la lettre x_i "passe par dessus" la contrainte associée à la lettre x_j sans se croiser avec elle.

Par contre, l'ensemble des mots de la forme ff, avec f un mot quelconque de F^+, ne peut pas être un langage CF (standard ou non). En effet ici les contraintes consistant en ce que chaque lettre de la seconde moitié du mot correspond à une lettre de la première moitié s'entrecroisent et il peut y avoir un nombre arbitrairement grand de telles intersections (ceci n'est évidemment pas une preuve formelle que $\{ff: f \in F^+\}$ n'est pas CF). De même on montrerait que les ensembles de mots:

$$L_1 = \left\{ x_1^n\ x_2^n\ x_3^{n'} : n,n' > 0 \right\}$$

et

$$L_2 = \left\{ x_1^n\ x_2^{n'}\ x_3^{n'} : n,n' > 0 \right\}$$

sont tous deux des langages CF mais qu'il n'en est pas de même de leur intersection

$$L_1 \cap L_2 = \left\{ x_1^n\ x_2^n\ x_3^n : n > 0 \right\}$$

On pourrait illustrer à l'aide de règles tirées de la grammaire de langues naturelles ces deux contre-exemples et ceci montrerait, s'il le fallait, que les contraintes de type CF ne concernent qu'un horizon limité de la réalité linguistique. Une fois encore, je me bornerai à renvoyer aux travaux de CHOMSKY pour une discussion approfondie de ces problèmes, de la valeur explicative en linguistique des règles CF et enfin de leur rôle dans la construction de grammaires moins formelles. Cependant il convient de rappeler que dès le niveau des langages CF, les questions les plus évidentes sont, en général, indécidables au sens technique du terme. Ainsi, Bar Hillel, Perles et Shamir ont montré à l'aide du contre-exemple classique de Post qu'il n'existe aucun algorithme permettant de décider pour deux homomorphismes φ_1 et φ_2 quelconques si les langages CF

$$L_1 = \left\{ \varphi_1 f \bar{f} : f \in F_+ \right\}$$

et

$$L_2 = \left\{ \varphi_2 f \bar{f} : f \in F_+ \right\}$$

ont ou non au moins un mot en commun.

D'autre part, il résulte des définitions que la mémoire d'un dispositif destiné à tester séquentiellement l'appartenance des mots à un langage CF standard donné n'est utilisée que d'une manière extrêmement restreinte qui rappelle certains aspects élémentaires de la technique de programmation connue sous le nom de "push down storage". De façon schématique, (en laissant de côté la mémoire bornée qui vérifie l'appartenance à $R(X_1, V)$) si αf est le mot enregistré en mémoire après la lecture du segment initial f, et si la lettre suivante est x_i, il

suffit d'ajouter celle-ci à la fin de αf quand αf se termine par x_j où $j \neq -i$ et d'effacer x_j dans le cas contraire où $j = -i$. Cette procédure ne nécessite donc en fait que la considération de l'extrémité finale du mot mis en mémoire, d'où le terme "push-down".

Il existe bien entendu d'autres types de contraintes définissant des familles de langage et les plus étudiées d'entre elles constituent ce que l'on appelle après KLEENE les évènements réguliers, qui contiennent en particulier tous les langages n'ayant qu'un nombre fini de mots et les langages de la forme $R(X_1, V)$ utilisés plus haut. Avec nos notations, tout évènement régulier P est obtenu en prenant un sous-ensemble X_1 de X^+, un ensemble V contenant tous les mots de deux lettres $x_i\, x_j$ avec i négatif et j positif, ce qui donne un certain langage CF standard dont, finalement, P sera l'image par un homomorphisme φ tel que $\varphi\, x_i$ soit le mot vide pour chacun des i négatifs. Plus simplement, on peut définir P comme l'image homomorphique de $R(X'_1, V', X'_2)$ où l'homomorphisme φ n'est soumis à aucune restriction et où l'évènement régulier $R(X'_1, V', X'_2)$ est l'ensemble des mots commençant par une lettre du sous-ensemble X'_1 de X, n'ayant aucun facteur de longueur deux dans V' et se terminant par une lettre du sous-ensemble X'_2 de X. Donc, pour les évènements réguliers, il est possible de remplacer toutes les contraintes résultant de l'appartenance à D par la seule contrainte supplémentaire que le mot se termine par une lettre appartenant à un ensemble distingué. De façon plus directe, KLEENE avait défini un évènement régulier comme un ensemble P de mots tel qu'il soit possible de reconnaître si un mot arbitraire appartient à P par lecture séquentielle de ce mot, en ne gardant en mémoire à chaque lettre qu'une quantité d'information bornée par une valeur finie (ne dépendant que de P).

Par exemple, l'ensemble de tous les mots de longueur paire qui ne contiennent pas plus de trois fois la lettre x_2 forme un évènement régulier puisque l'appartenance à P peut être reconnue en ne gardant en mémoire pour chaque segment initial que la parité de la longueur de ce segment (soit 1 bit d'information) et le nombre (au plus égal à 3) de fois où x_2 est déjà apparu (soit 2 bits d'information).

Ceci équivaut à dire que P est un évènement régulier si et seulement si il est l'image inverse d'un sous-ensemble de l'image de F par un homomorphisme de ce monoïde dans un monoïde fini. Donc, ces questions qui étaient en règle générale indécidables pour les langages CF généraux deviennent ici susceptibles d'une solution algorithmique par exhaustion de tous les cas possibles dans un certain ensemble fini. Cette propriété essentielle qui semble avoir fasciné les esprits épris de finitude, permet de rattacher très simplement aux évènements réguliers les "grammaires probabilistes" n'utilisant qu'un nombre fini d'états (ce sont les "finite state sources" de SHANNON) c'est-à-dire en gros- les schémas markoviens finis) et propose le problème de discuter la façon dont une telle grammaire peut approcher un langage CF propre.

Intuitivement, il semble bien évident qu'un langage tel que l'ensemble $M = \{f\tilde{f} : f \in F_+\}$ des mots en miroir qui joue un rôle central dans l'étude des langages CF ne puisse être approché que de façon tout-à-fait triviale par un évènement régulier R. En effet, en dénotant par m_k, r_k et q_k le nombre des mots de longueur k de M, R et $Q = R \cap M$, on voit facilement que :

1) soit Q ne constitue qu'une fraction asymptotiquement nulle de M en ce sens que $\lim_{k \to \infty} q_k\, m_k^{-1} = 0$ (par exemple, si R est contenu dans M);

2) soit, au contraire, Q est une fraction asymptotiquement nulle de R, c'est-à-dire $\lim_{k\to\infty} q_k\, r_k^{-1} = 0$ (par exemple, si M est contenu dans R).

Plus généralement soit $A = \varphi L$ un langage CF (non nécessairement standard) satisfaisant la condition restrictive que chacun de ses mots soit l'image par φ d'un seul mot du langage standard L et cherchons à l'approximer par une séquence infinie strictement croissante d'évènements réguliers $\{R_i\}$ ($i = 1, 2, \ldots$) tous contenus dans A. Pour chaque i la valeur de cette approximation est fournie par la suite des différences $a_k - r_{i,k}$ (où a_k $r_{i,k}$ sont, comme plus haut, le nombre des mots de longueur k de A et de R_i), soit encore, de façon cendensée par le nombre $\overline{a} - \overline{r}_i$ où

$$\overline{a} = \sum_{k>0} a_k\, x^k,\quad \overline{r}_i = \sum_{k>0} r_{i,k}\, x^k$$

avec

$$x = (2n+1)^{-1}$$

ce qui revient à faire une moyenne (pondérée par les x^k) de toutes ces différences. Or, pour chaque i, $\overline{r}_i$ est une fraction rationnelle dont le dénominateur $\overline{p}_i$ peut être considéré comme une évaluation grossière de la complexité de R_i c'est-à-dire de la quantité d'information qu'il faut pouvoir garder en mémoire afin de tester l'appartenance d'un mot à R_i. Moyennant ces interprétations et le fait aisément vérifié que $\overline{a}$ est un nombre algébrique, des théorèmes classiques d'arithmétique montrent que pour chaque valeur ϵ de $\overline{a} - \overline{r}_i$, il n'existe qu'un nombre fini de R_i dont la "complexité" $\overline{p}_i$ soit inférieure à une certaine fonction (donnée par le théorème) de ϵ. C'est là le résultat cherché qui suggère comme une sorte de complémentarité entre les contraintes CF et les contraintes correspondant aux mémoires bornées.

Enfin, sur le plan des faits linguistiques observables, la nécessité d'une approximation des langues naturelles plutôt par des langues CF que par ces modèles à mémoire finie que sont les évènements réguliers semble avoir été admise implicitement par les techniciens de la traduction automatique eux-mêmes, puisque tous les programmes présentés jusqu'ici sont basés sur des méthodes qui sont essentiellement du type "push down" évoqué plus haut. Il me semble y avoir là une certaine ironie des choses sur laquelle je concluerai ces remarques.

REFERENCES :

N. CHOMSKY and G.A. MILLER. Introduction to the formal analysis of natural languages in "Handbook of Mathematical Psychology", Bush, Gallanter & Luce, Ed., 1962 (Wiley - N.Y.).

Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS Vol. 7, No 3, December 1963

On a Formal Product over the Conjugate Classes in a Free Group

M. P. SCHÜTZENBERGER* AND S. SHERMAN†

Harvard University Medical School
Wayne State University

Submitted by Richard Bellman

I. INTRODUCTION

In [1, Eq. (2)] a special case of [2, Theorem 1] (another special case of which was conjectured by Feynman in connection with the Ising model for ferromagnetism) is shown to be an analogue of an identity used by Witt, the special case playing a role relative to free groups analogous to that played by the Witt identity relative to free semigroups. Furthermore, in [1] the question is raised of establishing the special case by methods short of proving [2, Theorem 1]. In the current note a selfcontained proof of a non-commutative generalization of the special case is presented. More explicitly, given a set X_1 and a mapping ρ of X_1 into a certain algebra A we verify the identity of the formal products $\Pi\{(1 - \rho x)^2 : x \in X_1\}$ and $\Pi\{(1 - \rho g) : g \in C\}$ where the subset C of the free group G generated by X_1 is such that every $g \in G$, $g \neq e_G$ is conjugate of some positive power of one and only one element of C and where $\rho : G \to A$ is defined below. Since these definitions are not independent of the choice of X_1 it is convenient to introduce the set X consisting of the elements of X_1 together with their inverses so that A is the algebra over the ring R of the free monoid F generated by X. The anti-automorphism (of period 2) $\alpha : F \to F$, the idempotent endomorphism $\beta : F \to F$, the epimorphism $\nu : F \to G$ are defined by their restriction to X_1, that is

$$\text{for all } x \in X_1,\ x = \beta x = \beta\alpha x;\ (\nu x)^{-1} = \nu\alpha x.$$

The set $\bar{F}$ of the so called *reduced words* is the complement in F of the ideal generated by all words $x\alpha x (x \in X)$ and, as it is well known, the restriction of ν to $\bar{F}$ is a bijection. We abbreviate $\{f \in F : f \neq 1\}$ by F^+.

* The research of this author has been supported by the Commonwealth Fund and IBM.

† The research of this author has been supported by the National Science Foundation and the Michigan Institute of Science and Technology.

II. Weight Function

Let there be given a partial order $\leqslant$ (where $\not\leqslant$ means not $\leqslant$) on X_1 and a mapping $r: X_1 \times X_1 \to R$. Writing $s(x, x') = -1$ if $x \leqslant x'$, $= +1$ otherwise, we extend r to a mapping $X \times X \to R$ by the following rules:

For any $x, x' \in X$, $x \neq x'$

1. $r(\alpha x, \alpha x) = r(x, x) = -1$,
2. $r(x, x') = s(x, x')r(x, \alpha x) = -s(x', x)r(\alpha x, x') = -s(x, x')s(x', x)r(\alpha x, \alpha x')$
3. $r(x, \alpha x) = r(\alpha x, x) = 0$.

For any $f \in F^+$, $f = x_{i_1} x_{i_2} \cdots x_{i_m}$ $(x_{i_1}, x_{i_2}, \cdots, x_{i_m} \in X)$ we define

$$\rho' f = r(x_{i_1}, x_{i_2})\, r(x_{i_2}, x_{i_3}) \cdots r(x_{i_{m-1}}, x_{i_m})\, r(x_{i_m}, x_{i_1})$$

and $\rho f = \rho' f \cdot \beta f \in A$. For $g \in G$, $g \neq e_G$, we define $\rho g = \rho f$ where $f (= \nu^{-1} g \cap \bar{F})$ is the reduced word representing g.

In equivalent manner, let μ be any representation of F by $X \times X$ matrices conjugate to the representation μ' defined for all $x \in X$ by $(\mu' x)_{x', x''} = r(x, x'')$ if $x = x'$, $= 0$ otherwise.

Then, it is easily verified that for all $f \in F^+$, $\operatorname{Tr} \mu f = \rho' f$. In the case treated in [1, p. 226], the relation $x' \leqslant x$ corresponds to the relation "$x \neq x'$ and the loop x' is contained in the loop x." See Fig. 2. For each $x, x' \in X$, $r(x, x') = \pm 1$. This can be arranged as follows: By a homeomorphism of the plane onto itself the loops from the origin as center can all be displaced into the first quadrant. Each $x \in X$ corresponds to a counterclockwise loop and $x_i < x_j$ means that the initial tangent vector to x_i at the origin makes a smaller acute angle with the positive x-axis than the acute angle the initial tangent vector to x_j makes with the positive x-axis. Let $r(x_i, x_j) = s(x_j, x_i)$ for all $i < j$, $r(x_i, x_j) = -1$ for all $i > j$, and $r(x_i, x_i) = -1$ for all i. Thus in Fig. 1, $x \not\leqslant x'$ and $x' \not\leqslant x$ while $x < x'$.

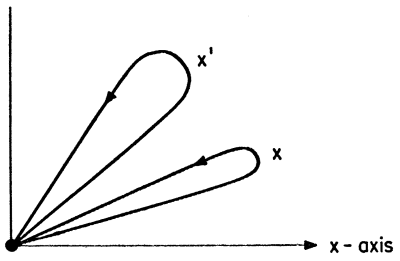


Fig. 1. $x \not\leqslant x'$, $x' \not\leqslant x$, $x < x'$

and $r(x, x')$ is given by Table I. For the case of Fig. 2 we have $x' \ominus x$, $x \oplus x'$, and $x < x'$ and r is given by Table II. A geometric interpretation

TABLE I

r	x	αx	x'	$\alpha x'$
x	−1	0	1	1
αx	0	−1	−1	−1
x'	−1	−1	−1	0
$\alpha x'$	1	1	0	−1

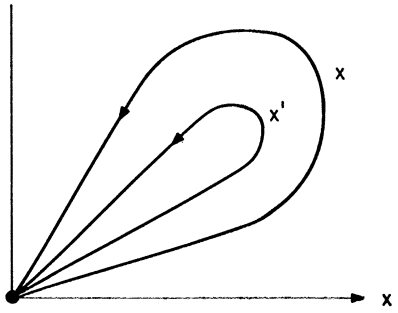


FIG. 2. $x' \ominus x$, $x \oplus x'$, $x < x'$

TABLE II

r	x	αx	x'	$\alpha x'$
x	−1	0	−1	−1
αx	0	−1	−1	−1
x'	−1	1	−1	0
$\alpha x'$	1	−1	0	−1

of $r(x, x')$ is that −1 corresponds to an odd number of traversals of the fourth quadrant by the tangent vector as it goes from the initial vector of x to the final vector of x to the initial vector of x' and +1 corresponds to an even number of traversals. Thus in the case of Fig. 1, for $r(x, x')$ we have Fig. 3 yielding $r(x, x') = 1$, while in the case of Fig. 1 for $r(x', \alpha x)$ we have Fig. 4 yielding $r(x', \alpha x) = -1$.

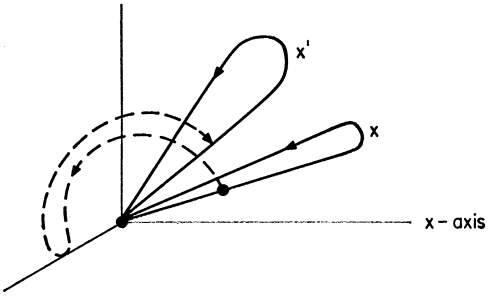


FIG. 3. $r(x, x') = 1$

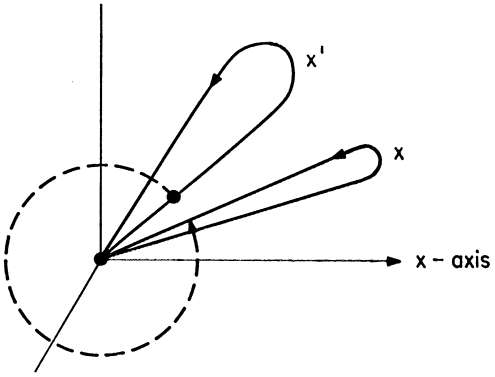


FIG. 4. $r(x', \alpha x) = -1$

III. LEXICOGRAPHIC STANDARD WORDS

Let $<$ be a lexicographic order on the free monoid F and define the subset H of F^+ ($= \{f \in F; f \neq 1\}$) by the condition that $f \in H$ if and only if $f = f'f''$; $f', f'' \in F^+$ implies $f < f''$.

LEMMA 1. [3] *To every $f \in F^+$ there corresponds a unique triple $f' \in F^+$, $f'' \in F$, $m > 0$ such that $f = (f''f')^m$ and $f'f'' \in H$.*

PROOF: Let the subset H' of F^+ be defined by the condition that $f \in H'$ if and only if $f = f'f''$; $f', f'' \in F^+$ implies $f < f''f'$. The verification that H' satisfies the conditions of Lemma 1 and that $H \subset H'$ is immediate. We show $H' \subset H$.

Assume that $f, f_1, f_2, f_3 \in F^+$ are such that $f = f_1f_2 = f_2f_3$ and $f < f_2f_1$. The first condition implies either $f_1 = f_2f_4$ or $f_2 = f_1f_4$ for some $f_4 \in F$. In the first case $f = f_2f_4f_2 < f_2f_1 = f_2f_2f_4$ implies $f_4f_2 < f_2f_4$. Hence $f_4f_2f_2 < f$

and $f \notin H'$. In the second case $f_2 = f_1 f_4 = f_4 f_3$ so that $f = f_1 f_2 < f_2 f_1$, $f_1 f_4 f_3 < f_1 f_4 f_1$ and $f_3 < f_1$. Since the degree of f_3 is equal to the degree of f_1, it follows that $f_3 f_1 f_4 < f$ and $f \notin H'$.

Thus if $f = f'f'' \in H'$; $f', f'' \in F^+$, then f'' is not a left factor of f and $f < f''f'$ implies $f < f''$, that is, $f \in H$ concluding the proof.

LEMMA 2 [4].[1] *Every $f \in F^+$ has one and only one factorization $f = h_{i_1} h_{i_2} \cdots h_{i_m}$ where the elements h_{i_j} belong to H and satisfy $h_{i_m} \leqslant h_{i_{m-1}} \leqslant \cdots \leqslant h_{i_2} \leqslant h_{i_1}$.*

PROOF: Let f and the h_{i_j}'s be as in the Lemma and further $f = f_1 h'$ where $h' \in H$ admits h_{i_m} as a right factor. By definition there exists $m'(1 \leqslant m' \leqslant m)$ and a right factor f' of $h_{i_{m'}}$ such that $h' = f'f''$, and consequently $f' \leqslant h'$. By Lemma 1, $h' \leqslant h_{i_m}$ and $h_{i_{m'}} \leqslant f'$. However, $h_{i_m} \leqslant h_{i_{m'}}$ by hypothesis so that $h_{i_m} = h'$ showing by induction that any f has at most one factorization of the type described.

Reciprocally, let ηf denote for each $f \in F^+$ the right factor $\in F^+$ of f that is minimal with respect to $<$. By Lemma 1, $\eta f \in H$ and, assuming $f \notin H$, $f = f''h'h$ where $h = \eta f$, $h' = \eta(f''h') \in H$, and, by construction, $h < h'h$. Hence the existence of at least one factorization of the prescribed type will follow by induction once it is verified that $h \leqslant h'$. For this, assume $h' \leqslant h$. Cancelling the common left factor of highest degree of h and h' in the two relations $h < h'h$ and $h' \leqslant h$ shows that $h' = h$, and this ends the proof.

IV. MAIN RESULT

Let $a(f)$ denote the coefficient of $f \in F$ in $a \in A$ and $A_u = \{a \in A: a(1) = 1\}$. Let $I = \{i\}$ be totally ordered by $<$ and the mapping $i \to p_i$ of I into A_u be such that for each $f \in F^+$ the subset $P_f = \{i \in I: p_i(f) \neq 0\}$ is finite. For each $f \in F^+$ we define

$$g(f) = \sum' p_{i_1}(f_{j_1})\, p_{i_2}(f_{j_2}) \cdots p_{i_m}(f_{j_m})$$

where the summation is over all factorizations $f = f_{j_1} f_{j_2} \cdots f_{j_m}$ of f into an arbitrary number of factors $f_{j_k} \in F^+$ and for each such factorization over all $p_{i_1} \in P_{f_{j_1}}$, $p_{i_2} \in P_{f_{j_2}}$, $\cdots$, $p_{i_m} \in P_{f_{j_m}}$ such that $i_1 < i_2 \cdots < i_m$. Thus, $1 + \Sigma\{g(f)f: f \in F^+\}$ is a well defined element of A_u which can be denoted by $\Pi\{p: p \in P; <\}$ where $P = \{p_i: i \in I\}$.

Since A_u is a group (with $a^{-1} = 1 + \Sigma_{n>0}(1 - a)^n$ for each $a \in A_u$) it is easily verified that, setting $p^{-1} > p'^{-1}$ iff $p' < p$, the inverse of

[1] We are indebted to P. M. Cohn for calling our attention to the Širšov paper.

$\Pi\{p: p \in P; <\}$ is precisely $\Pi\{p^{-1}: p \in P; >\}$. In particular (with the notations of Section III) if $(1 - h)^{-1} > (1 - h')^{-1}$ iff $h < h'$, Lemma 2 can be interpreted as the identity

$$\Pi\{(1 - h)^{-1}: h \in H; >\} = \sum\{f: f \in F\}.$$

Since the right member is equal to $(1 - \sum\{x: x \in X\})^{-1}$ it follows:

LEMMA 3 [2, Eq. 6].

$$\Pi\{1 - h: h \in H; <\} = 1 - \sum\{x: x \in X\}.$$

Let us assume that the lexicographic order $<$ is now extended from X to $X \cup \alpha X$ such that for all $x, x' \in X$, $x < x'$ implies $x < \alpha x < x' < \alpha x'$ and $x \circledast x'$. On this larger domain the interpretation of $<$ in terms of angular order of initial vectors is no longer valid.

Setting $\bar{H} = \{h \in H: h^2 \in \bar{F}\}$ it is easily verified from Lemma 1 that every $g \in G$, $g \neq e_G$ is conjugate to one and only one element of the form $(vh)^m$ with $m > 0$, $h \in \bar{H}$.

THEOREM. [1, Eq. (2).]

$$\Pi\{1 - \rho x)^2: x \in X_1; <\} = \Pi\{1 - \rho h: h \in \bar{H}; <\}.$$

PROOF: Since by definition $1 - \rho x = 1 - \rho\alpha x$ it is sufficient to verify for each $x \in X$

$$1 - \rho x = \Pi\{1 - \rho h: h \in \bar{H} \cap xF; <\}.$$

For this, let K_x be the set of all words of the form $xf \in \bar{F}$ where f has no factor $x' \leqslant x$ and does not belong to $F\alpha x$. By construction $K_x \subset \bar{H}$ and each $h \in \bar{H} \cap xF$ (more generally, each $f \in \bar{F} \cap xF \setminus F\alpha x$) has one and only one factorization as a product of elements of K_x. Also if $h = kf$, $h' = k'f'$, $k, k' \in K_x$, and both f and f' having x as a left factor, then the relation $h < h'$ implies $k \leqslant k'$. Thus one can find a lexicographically ordered free monoid F' and a monomorphism $\zeta: F' \to F$ such that F' is generated by $\zeta^{-1}K_x$ and that $\bar{H} \cap xF = \zeta H'$ where H' is defined for F' as H was defined for F in Section 3.

Using Lemma 3 and the remark that for any $h, h' \in \bar{H} \cap xF$, $\rho hh' = \rho h \rho h'$, it follows:

$\Pi\{1 - \rho h: h \in \bar{H} \cap xF; <\} = 1 - \sum\{\rho k: k \in K_x\}$. We verify that the right member reduces to $1 - \rho x$ by constructing an involutory mapping τ of $K_x \setminus \{x\}$ such that, identically, $\rho\tau k + \rho k = 0$.

For this, consider any element $k \neq x$ of K_x. It admits a factorization $k = x\, x_{i_1}^{m_1} x_{i_2}^{m_2} \cdots x_{i_p}^{m_p}$ where $p \geqslant 1$; $m_1, m_2, \cdots, m_p > 0$; $x_{i_1}, x_{i_2}, \cdots, x_{i_p} \in X$; $\beta x \neq \beta x_{i_1}$, $\beta x_{i_1} \neq \beta x_{i_2}$, $\cdots$, $\beta x_{i_{p-1}} \neq \beta x_{i_p}$, $\beta x_{i_p} \neq \beta x$.

Let $j^* = 1$ if $p = 1$ or if $p > 2$ and $\beta x_{i_2} \circledast \beta x_{i_1}$ and $j^* =$ the largest index such that $\beta x_{i_{j^*}} \circledcirc \beta x_{i_{j^*-1}} \cdots \circledcirc \beta x_{i_2} \circledcirc \beta x_{i_1}$ otherwise. Then we define τk as the element obtained when replacing in k, $x_{i_{j^*}}^{m_{j^*}}$ by $(\alpha x_{i_{j^*}})^{m_{j^*}}$.

Since $k \in H$, $\beta x < \beta x_{i_1}$ and because of our choice of $<$ this implies $\beta x \circledast \beta x_{i_1}$. Thus $\beta x_{i_{j^*}} \neq \beta x$ and consequently, $\tau k \in K_x$, $\tau\tau k = k$. For the same reason, $\beta x \circledast \beta x_{i_{j^*}}$ when $j^* = p$. Hence setting $x' = x_{i_{j^*-1}}$ ($= x$ if $j^* = 1$), $\bar{x} = x_{i_{j^*}}$, $x'' = x_{i_{j^*+1}}$ ($= x$ if $j^* = p$) one has always $\beta x' \circledast \beta\bar{x}$ and $\beta x'' \circledast \beta\bar{x}$. From the definition of r it follows that

$$r(x', \bar{x})\,(r(\bar{x}, \bar{x}))^{m_{j^*}}\, r(\bar{x}, x'') \;=\; -r(x', \alpha\bar{x})\;\; (r(\alpha\bar{x}, \alpha\bar{x}))^{m_{j^*}}\, r(\alpha\bar{x}, x'').$$

This shows that $\rho\tau k + \rho k = 0$ and concludes the proof.

References

1. Sherman, S., Combinatorial aspects of the Ising model for ferromagnetism, II. *Bull. Am. Math. Soc.* **68**, 225-229 (1962).
2. Sherman, S., Combinatorial aspects of the Ising model for ferromagnetism, I. *J. Math. Phys.* **1**, 202-207 (1960).
3. Chen, K. T., Fox, R. H., and Lyndon, R. H., Free differential calculus IV. *Ann. Math.* **68**, 81-95 (1958).
4. Širšov, A. I. On free Lie rings. *Mat. Sbornik* **45**, 113-122 (1958)

Séminaire Dubreil.
Algèbre et théorie
des nombres

Marcel P. Schützenberger

Quelques remarques sur une construction de Schensted

Séminaire Dubreil. Algèbre et théorie des nombres, tome 16, n° 2 (1962-1963), exp. n° 15, p. 1-12.

<http://www.numdam.org/item?id=SD_1962-1963__16_2_A4_0>

Séminaire DUBREIL-PISOT
(Algèbre et Théorie des nombres)
16e année, 1962/63, n° 15

15-01

4 mars 1963

QUELQUES REMARQUES SUR UNE CONSTRUCTION DE SCHENSTED

par Marcel P. SCHÜTZENBERGER

1. Introduction.

Soient A et B deux sous-ensembles de $\underset{\sim}{N}$ et $\alpha : A \to B$ une bijection. Dans un article récent ([1]), C. SCHENSTED a donné une construction remarquable associant de façon injective à $\alpha : A \to B$ une paire $(P(\alpha , A) , Q(\alpha , A))$ de tableaux standards de Young de même forme. Nous nous proposons ici d'apporter quelques précisions supplémentaires aux propriétés déjà établies par SCHENSTED en rattachant cette question au problème classique de S. NEWCOMB et en observant que la tableau $Q(\alpha , A)$ est en fait identique à $P(\alpha^{-1} , B)$. Nous supposerons que le lecteur est familier avec les définitions et les résultats fondamentaux de SCHENSTED.

Pour simplifier les notations, nous considérerons les tableaux standards de Young comme des éléments particuliers du module $\mathcal{C}$ des applications de $\underset{\sim}{N} \times \underset{\sim}{N}$ dans $\underset{\sim}{Z}$ bien que, de fait, seules interviennent les structures d'ordre de ces ensembles.

Pour tout $T \in \mathcal{C}$, la forme (ou support) de T sera l'ensemble

$$|T| = \{(i , j) \in \underset{\sim}{N} \times \underset{\sim}{N} : T_{i,j} \neq 0\}$$

et $\{T\}$ désignera l'image par T de $|T|$ dans $\underset{\sim}{Z}$.

Quand $T : |T| \to \{T\}$ est bijective (donc, en particulier quand T est standard) on dénotera par T^{-1} l'application inverse. Enfin pour $a \in \underset{\sim}{Z} \setminus \{0\}$ et $(i , j) \in \underset{\sim}{N} \times \underset{\sim}{N}$, $a_{ij} \in \mathcal{C}$ sera définie de façon évidente par $|a_{i,j}| = (i , j)$, $\{a_{i,j}\} = a$.

Pour tout sous-ensemble fini A' de A et tout entier non négatif m, on posera :

$$P(\alpha , A'_m) = 0 \quad \text{si} \quad m = 0$$

et, inductivement,

([1]) SCHENSTED (C.). — Longest increasing and decreasing subsequences, Canad. J. of math., t. 13, 1961, p. 179-192.

15-02

$$P(\alpha\ ,\ A'_m) = P(\alpha\ ,\ A'_{m-1}) \leftarrow \alpha\ a'_m \quad \text{si} \quad 0 < m \leqslant \underline{\text{Card}}\ A' \ ,$$

où a'_m dénote le m-ième élément de A' par ordre croissant ;

$$P(\alpha\ ,\ A'_m) = P(\alpha\ ,\ A'_n) = P(\alpha\ ,\ A') \quad \text{si} \quad m \geqslant \underline{\text{Card}}\ A' = n \ .$$

Il est logique de considérer ici $0 \leftarrow \alpha\ a'_1$ comme le tableau standard $(\alpha\ a'_1)_{1,1}$ et, par conséquent, $P(\alpha\ ,\ A')$ est exactement le P-symbole de SCHENSTED de la séquence $(\alpha\ a'_1\ ,\ \alpha\ a'_2\ ,\ \dots\ ,\ \alpha\ a'_n)$. De même :

$Q(\alpha\ ,\ A'_0) = 0$;

$Q(\alpha\ ,\ A'_m) = Q(\alpha\ ,\ A'_{m-1}) + (a'_m)_{i,j}$ pour $0 < m \leqslant \underline{\text{Card}}\ A'$ avec

$$(i\ ,\ j) = |P(\alpha\ ,\ A'_m)| \setminus |P(\alpha\ ,\ A'_{m-1})| \ ,$$

$Q(\alpha\ ,\ A'_m) = Q(\alpha\ ,\ A'_n) = Q(\alpha\ ,\ A')$ pour $m \geqslant \underline{\text{Card}}\ A' = n$.

Quand $A' = [1\ ,\ n]$ ceci est la définition même du Q-symbole de SCHENSTED de $(\alpha\ a'_1\ ,\ \alpha\ a'_2\ ,\ \dots\ ,\ \alpha\ a'_n)$, et pour A' quelconque, $Q(\alpha\ ,\ A')$ se déduit simplement de ce Q-symbole en remplaçant dans ce dernier tableau chaque $m \in [1\ ,\ n]$ par a'_m .

Posons $||0|| = 0$ et pour chaque $T \in \mathfrak{T}\setminus\{0\}$,

$$||T||^{-1} = \underline{\text{Min}}(i + j - 1 : \ (i\ ,\ j) \in |T|) \ .$$

Il résulte de la définition même de l'opération $\leftarrow$ que pour $A' = A$ on a

$$\lim_{m,m' \to \infty} ||P(\alpha\ ,\ A_m) - P(\alpha\ ,\ A_{m'})|| = \lim_{m,m' \to \infty} ||Q(\alpha\ ,\ A_m) - Q(\alpha\ ,\ A_{m'})|| = 0 \ .$$

On pourra donc toujours définir

$$P(\alpha\ ,\ A) = \lim_{m \to \infty} P(\alpha\ ,\ A_m) \quad \text{et} \quad Q(\alpha\ ,\ A) = \lim_{m \to \infty} Q(\alpha\ ,\ A_m) \ .$$

Toutes les notations qui viennent d'être introduites seront systématiquement utilisées dans tout ce qui suit.

2. L'opération Δ .

Soit $Q = Q(\alpha\ ,\ A_m) \neq 0$ $(m < \infty)$. On définit un autre tableau standard $Q' = \Delta Q(\alpha\ ,\ A_m)$ et une séquence $|U_m| = ((i_1\ ,\ j_1)\ ,\ (i_2\ ,\ j_2)\ ,\ \dots\ ,\ (i_p\ ,\ j_p))$ d'éléments de $|Q|$ par les conditions suivantes :

(1) $$(i_1\ ,\ j_1) = (1\ ,\ 1)$$

et, pour chaque $k \in [1\ ,\ p-1]$,

$$(i_{k+1}\ ,\ j_{k+1}) = (i_k + 1\ ,\ j_k) \quad \text{ou} \quad = (i_k\ ,\ j_k + 1) \ ;$$

15-03

(2) $Q'_{i',j'} = Q_{i',j'}$ si $(i' , j') \notin |U_m|$

$= Q_{i,j}$ avec $(i , j) = (i_{k+1} , j_{k+1}) \in |U_m|$

si $(i' , j') = (i_k , j_k) \in |U_m|$ et $k \in [1 , p - 1]$

$= 0$ si $(i' , j') = (i_p , j_p)$

le dernier élément de $|U_m|$;

(3) Q' est un tableau standard .

De façon plus explicite, connaissant déjà $(i' , j') = (i_k , j_k) \in |U_m|$, on pose $k = p$ (c'est-à-dire que l'on considère (i' , j') comme le dernier élément de $|U_m|$) si

$$Q_{i'+1,j'} = Q_{i',j'+1} = 0 .$$

Sinon on détermine $(i_{k+1} , j_{k+1}) \in |U_m|$ par les conditions

$$(i_{k+1} , j_{k+1}) = (i' , + 1 , j')$$

si $0 < Q_{i'+1,j'} < Q_{i',j'+1}$ ou si $0 = Q_{i',j'+1} < Q_{i'+1,j'}$

$= (i' , j' + 1)$

si $0 < Q_{i',j'+1} < Q_{i',+1,j'}$ ou si $0 = Q_{i'+1,j'} < Q_{i',j'+1}$ qui assurent automatiquement que (3) est satisfaite.

Donc, $\{\Delta Q\}$ est l'ensemble $\{Q\}$ privé de son plus petit élément $Q_{1,1}$. D'après les définitions mêmes, $|U_m| \subset |U_{m+1}|$, et par conséquent,

$$|U| = \lim_{m\to\infty} |U_m| , \text{ ainsi que } \Delta Q(\alpha , A) = \lim_{m\to\infty} \Delta Q(\alpha , A_m)$$

sont bien définis. Plus généralement, si $\overline{Q}$ est un autre tableau standard on a

$$||\Delta Q - \Delta\overline{Q}||^{-1} \geqslant ||Q - \overline{Q}||^{-1} + 1 .$$

Exemple. - Si $Q = \begin{smallmatrix} 248 \\ 679 \end{smallmatrix}$ avec les notations de SCHENSTED,

$\Delta Q = \begin{smallmatrix} 478 \\ 69 \end{smallmatrix}$, $\Delta^2 Q (= \Delta(\Delta Q)) = \begin{smallmatrix} 678 \\ 9 \end{smallmatrix}$, $\Delta^3 Q = \begin{smallmatrix} 78 \\ 9 \end{smallmatrix}$, $\Delta^4 Q = \begin{smallmatrix} 8 \\ 9 \end{smallmatrix}$, $\Delta^5 Q = 9$, $\Delta^6 Q = 0$.

Remarque 1. - Si $A \neq \emptyset$, on a identiquement

$$\Delta Q(\alpha , A) = Q(\alpha , A\backslash\{a_1\}) .$$

Démonstration. - Le résultat peut être vérifié directement pour Card $A < 2$ et, dénotant pour abréger par A' l'ensemble $A\backslash\{a_1\}$, il suffit de vérifier que, pour $m > 1$, l'égalité de $\Delta Q(\alpha , A_{m-1})$ et $Q(\alpha , A'_{m-2})$ entraîne celle de $\Delta Q(\alpha , A_m)$ et $Q(\alpha , A'_{m-1})$.

15-04

Par définition il existe (i , j) et $(i' , j') \in \underset{\sim}{N} \times \underset{\sim}{N}$ tels que

$Q(\alpha , A'_{m-1}) = Q(\alpha , A'_{m-2}) + (a_m)_{i',j'}$ et $Q(\alpha , A_m) = \Delta Q(\alpha , A_{m-1}) + (a_m)_{i,j}$.

Donc, d'après l'hypothèse d'induction,

$$\Delta Q(\alpha , A_m) = Q(\alpha , A'_{m-1}) + (a_m)_{i,j} - (a_m)_{i',j'}$$

et il ne reste qu'à vérifier $(i , j) = (i' , j')$.

Rappelant le résultat fondamental de SCHENSTED (lemme 6, loco citato p. 183) :

$$P(\alpha , A_m) = (\alpha\, a_1 \rightarrow P(\alpha , A'_{m-2})) \leftarrow \alpha\, a_m$$

$$= \alpha\, a_1 \rightarrow (P(\alpha , A'_{m-2}) \leftarrow \alpha\, a_m)$$

et utilisant l'identité de forme des P-symboles et des Q-symboles on a :

$$|Q(\alpha , A_{m-1})| \setminus |Q(\alpha , A'_{m-2})| = (i'' , j'')$$

et

$$|Q(\alpha , A_m)| \supset |Q(\alpha , A'_{m-2})| \cup \{(i , j) , (i' , j') , (i'' , j'')\} .$$

Distinguons maintenant deux cas :

1° $(i'' , j'') \neq (i , j)$ et, par conséquent,

$$|Q(\alpha , A_m)| = |Q(\alpha , A'_{m-2})| \cup \{(i , j) , (i'' , j'')\} .$$

La commutativité des opérations $\rightarrow$ et $\leftarrow$ implique la relation

$$Q(\alpha , A_m) - Q(\alpha , A_{m-1}) = Q(\alpha , A'_{m-1}) - Q(\alpha , A'_{m-2}) = (a_m)_{i,j} .$$

Donc, $(i , j) = (i' , j')$ puisque le tableau obtenu en remplaçant a_m par zéro dans $\Delta Q(\alpha , A_m)$ est égal à $\Delta Q(\alpha , A_{m-1})$, c'est-à-dire à $Q(\alpha , A'_{m-2})$ par l'hypothèse d'induction.

2° $(i'' , j'') = (i , j)$ et, par conséquent,

$$|Q(\alpha , A_m)| = |Q(\alpha , A_{m-2})| \cup \{(i , j) , (\bar{i} , \bar{j})\}$$

où $(\bar{i} , \bar{j}) = |P(\alpha , A_m)| \setminus |P(\alpha , A'_{m-1})|$.

Cette dernière relation implique $(\bar{i} , \bar{j}) = (i + 1 , j)$ ou $(i , j + 1)$ et par conséquent $|Q(\alpha , A'_{m-2})| \cup \{(\bar{i} , \bar{j})\}$ n'est pas la forme d'un tableau standard. Comme $\Delta Q(\alpha , A_m)$ est un tableau standard tel que

$$|\Delta Q(\alpha , A_m)| = |Q(\alpha , A'_{m-2})| \cup \{(i' , j')\} \subset |Q(\alpha , A_m)|$$

on a donc encore $(i' , j') = (i , j)$ et la vérification est achevée.

PROPRIÉTÉ 1. — La correspondance de Schensted associant la paire $(P(\alpha , A) , Q(\alpha , A))$ à $\alpha : A \rightarrow B$ est injective.

15-05

Démonstration. - Le résultat plus fort prouvant, pour A fini, le caractère bijectif de la correspondance est dû à SCHENSTED (lemme 3, loco citato p. 182). Nous considérons le cas de A infini, et nous vérifions que la donnée de $P = P(\alpha , A)$ et de $Q = Q(\alpha , A)$ détermine de façon univoque a_1 , $\alpha\, a_1$, $P(\alpha , A')$ et $Q(\alpha , A')$ (avec $A' = A\backslash\{a_1\}$ comme plus haut).

Pour tout m positif fini, SCHENSTED a montré (loco citato) qu'il existe une et une seule paire (b , P'_m) telle que P'_m soit un tableau standard satisfaisant les relations

$$|P'_m| = |\Delta Q(\alpha , A_m)| \quad \text{et} \quad b \to P'_m = P(\alpha , A_m) \ ;$$

la remarque 1 montre, qu'en fait, $b = \alpha\, a_1$, $P'_m = P(\alpha , A'_{m-1})$ et en outre

$$\{a_1\} = \{Q(\alpha , A_m)\}\backslash\{Q(\alpha , A'_m)\} \quad (= \{Q\}\backslash\{\Delta Q\}) \ .$$

De par la définition même de l'opération $\to$, on a

$$(P(\alpha , A_m))^{-1}\, b = (i'_m , 1) \ ,$$

où i'_m est au moins égal au nombre i_m défini par

$$(i_m , j_m) = |Q(\alpha , A_m)|\backslash|\Delta Q(\alpha , A_m)| \ .$$

Comme $\lim_{m\to\infty} i'_m = i'^{*} < \infty$, il en résulte

$$\lim_{m\to\infty} i_m = i^* < \infty \ .$$

Ainsi, puisque, identiquement, $(i_m , j_m) \in |U|$ et $i_m \leq i_{m+1}$, la valeur de i^* est déterminée de façon unique par Q . Plus précisément

$$i^* = \max_{d>0} \{i : \ (i , j) \in |U| \ , \ i + j = d\}$$

et les inégalités qui viennent d'être écrites montrent que ce nombre i^* est fini pour tout tableau standard qui est un Q-symbole.

Définissons maintenant pour chaque $d > i^*$ le tableau standard $P^{(d)}$ par les relations

$$P^{(d)}_{i,j} = P_{i,j} \quad \text{si} \quad i + j \leq d \ ; \ = 0 \quad \text{si} \quad i + j > d \ .$$

D'après le résultat de SCHENSTED rappelé plus haut, il existe pour chaque $d > i^*$ une et une seule paire $(b^{(d)} , P'^{(d)})$ satisfaisant

$$P^{(d)} = b^{(d)} \to P'^{(d)} \quad \text{et} \quad |P^{(d)}|\backslash|P'^{(d)}| = \{(i^* , d - i^*)\} \ .$$

C'est une propriété élémentaire de $\to$ que $b^{(d)} \leq b$, identiquement. Par conséquent $b = \lim_{d\to\infty} b^{(d)}$ et trivialement

15-06

$$P(\alpha\ ,\ A') = \lim_{d\to\infty} P'^{(d)}$$

ce qui achève la vérification.

Il est utile de noter que si $P = Q$, on a $b = P_{1,1}$ (et, par conséquent, $b = \{Q\}\backslash\{\Delta Q\}$) si et seulement si $i^* = 1$. Ceci résulte immédiatement de la remarque plus générale que S étant un tableau standard quelconque et $0 < s < S_{1,1}$, on a l'identité

$$\Delta(s \to S) = \Delta(S \leftarrow s) = S\ .$$

3. Relation avec le problème de Newcomb.

Soit $Q = Q(\alpha\ ,\ A)$ et pour m positif

$\eta_m = \operatorname{sgn}(i' + j' - i - j)$ où $(i\ ,\ j) = Q^{-1}\ a_m$ et $(i'\ ,\ j') = Q^{-1}\ a_{m+1}$.

Par exemple, pour $Q = \begin{matrix}248\\679\end{matrix}$ on trouve que la suite $\eta_1\ ,\ \eta_2\ ,\ \cdots\ ,\ \eta_5$ est égale à $+1\ ,\ -1\ ,\ +1\ ,\ +1\ ,\ -1$, les paires $(2\ ,\ 4)\ ,\ (6\ ,\ 8)\ ,\ (8\ ,\ 9)$ et $(4\ ,\ 6)$ illustrant respectivement les cas (1), (2), (3) et (4) énumérés plus bas. Le lien entre les Q-symboles de Schensted et le problème de Newcomb est fourni par la

Remarque 2. - Pour chaque m positif, $\eta_m = +1$ ou -1 selon que $\alpha\ a_m < \alpha\ a_{m+1}$ ou $> \alpha\ a_{m+1}$.

Démonstration. - Pour $a_m > a_1$ fixe, définissons

$\bar{\eta}_m = \operatorname{sgn}(\bar{i}' + \bar{j}' - \bar{i} - \bar{j})$ où $(\bar{i}\ ,\ \bar{j}) = (\Delta Q)^{-1}\ a_m$, $(\bar{i}'\ ,\ \bar{j}') = (\Delta Q)^{-1}\ a_{m+1}$

et vérifions d'abord que $\eta_m = \bar{\eta}_m$ identiquement. Pour cela, il est commode de distinguer quatre cas :

1° $i = i'\ ,\ j = j' - 1$;

2° $i > i'\ ,\ j < j'$;

3° $i = i' - 1\ ,\ j = j'$;

4° $i < i'\ ,\ j > j'$.

Ce sont les seuls possibles, car le fait que Q est standard et que $a_m < a_{m+1}$ exclut $i' \leqslant i$ et $j' \leqslant j$, et, d'autre part, le fait que a_m et a_{m+1} sont deux éléments consécutifs de $\{Q\}$ exclut les cas $i = i'$ et $j < j' - 1$, $i < i'$ et $j < j'$ ou $i < i' - 1$ et $j = j'$ qui entraîneraient l'existence d'un élément $a = Q_{i,j'-1}$, $= Q_{i,j'}$, ou $= Q_{i'-1,j}$, respectivement, tel que $a_m < a < a_{m+1}$.

15-07

Dans les cas (2) et (4) puisque $(\Delta Q)^{-1} a_m = (i , j)$, $= (i - 1 , j)$ ou $= (i , j - 1)$ et $(\Delta Q)^{-1} a_{m+1} = (i' , j')$, $= (i' - 1 , j')$ ou $= (i' , j' - 1)$ et puisque ΔQ est standard, on obtient directement l'égalité $\eta_m = \overline{\eta}_m$ cherchée.

Traitons en détails les sous-cas suivants du cas (1) :

(11) $i > 1$ et $(i - 1 , j + 1)(=(i - 1 , j')) \in |U|$. Dans ce cas

$$(\Delta Q)^{-1} a_m = Q^{-1} a_m \quad \text{et} \quad (\Delta Q)^{-1} a_{m+1} = (i - 1 , j') \text{ ou } = (i, j')$$

selon que (i , j') appartient ou non à $|U|$. Donc $\eta_m = \overline{\eta}_m$.

(12) $i > 1$ et $(i - 1 , j) \notin |U|$. Puisqu'il n'existe aucun élément de $\{Q\}$ entre a_m et a_{m+1} on a $Q_{i-1,j+1} < a_m$. Donc $(i - 1 , j + 1) \in |U|$ et on est ramené au cas précédent.

(13) $j > 1$ et $(i , j - 1) \in |U|$. Ou bien $(\Delta Q)^{-1} a_m = (i , j)$ et $(\Delta Q)^{-1} a_{m+1} = (i' , j')$ ou bien $(i , j) \in |U|$. Dans ce dernier cas le fait que a_m et a_{m+1} sont consécutifs entraîne $a_{m+1} < Q_{i+1,j+1}$ ou $0 = Q_{i+1,j+1}$; donc $(i' , j') \in |U|$, $(\Delta Q)^{-1} a_{m+1} = (i , j)$, et enfin $\eta_m = \overline{\eta}_m$.

(14) Dans tous les cas restants,

$$(\Delta Q)^{-1} a_m = (i , j) \quad \text{et} \quad (\Delta Q)^{-1} a_{m+1} = (i' , j') .$$

Donc $\eta_m = \overline{\eta}_m$.

Ceci achève l'examen du cas (1) et le cas (3) pouvant être traité de façon absolument analogue, nous ne répéterons pas la discussion.

Considérons maintenant le Q-symbole Q_m relatif à α et à l'ensemble $\{a_m , a_{m+1} , \ldots\}$. De par la définition même des Q-symboles on a

$$Q_m^{-1} a_m = (1 , 1) \quad \text{et} \quad Q_m^{-1} a_{m+1} = (1 , 2) \text{ ou } = (2 , 1)$$

selon que $\alpha\, a_m < \alpha\, a_{m+1}$ ou $\alpha\, a_m > \alpha\, a_{m+1}$. Puisque, d'après la remarque 1, Q_m est égal à $\Delta^{m-1} Q$ pour chaque m positif la remarque 2 résulte directement de $\eta_m = \overline{\eta}_m$ par induction sur m .

4. La formule $Q(\alpha , A) = P(\alpha^{-1} B)$.

Soient a_p et $a_{p'}$ deux éléments d'un sous-ensemble quelconque A' de A . Nous définissons le déplacement $Dp(a_p , a_{p'} , \alpha , A')$ de $\alpha\, a_{p'}$ par $\alpha\, a_p$ dans (la construction de) $P(\alpha , A')$ par les règles suivantes :

$$Dp(a_p , a_{p'} , \alpha , A') = 0$$

si $p' > p$ ou si $p' < p$ et si $P(\alpha , A'_{p-1})^{-1} a_{p'} = P(\alpha , A'_p)^{-1} a_{p'}$;

15-08

$$= ((0 , 0) , (i , j))$$

où $(i , j) = P(\alpha , A'_p)^{-1} a_p$ si $p = p'$;

$$= ((i' , j') , (i , j))$$

(où $(i' , j') = P(\alpha , A'_{p-1})^{-1} a_{p'}$ et $(i , j) = P(\alpha , A'_p)^{-1} a_{p'}$) si $p' < p$ et $i \neq i'$.

Dans les deux derniers cas on dira encore que $\alpha\, a_p$ déplace $\alpha\, a_{p'}$ de (i' , j') à (i , j) , la valeur $(0 , 0)$ de (i' , j') pour $a_p = a_{p'}$ étant évidemment purement conventionnelle. De par la définition même de l'opération $\leftarrow \alpha\, a_p$ si $(i' , j') \neq (0 , 0)$ on a nécessairement $i' = i + 1$ et $j < j'$; en outre on observera que ce déplacement se produit si et seulement s'il existe $a'' \in A'$ tel que $\alpha\, a_p \leq \alpha\, a'' < \alpha\, a_{p'}$ et que $\underset{\sim}{Dp}(a_p , a'' , \alpha , A') = ((i'' , j'') , (i' , j'))$. Donc, dans tous les cas,

$$\underset{\sim}{Dp}(a_p , a'_{p'} , \alpha , A')$$

$$= \underset{\sim}{Dp}(a_p , a_{p'} , \alpha , \{a_{p''} \in A' : \ p'' \leq p \ ; \ \alpha\, a_{p''} < \alpha\, a_{p'} \ ; \ \alpha\, a_p \leq \alpha\, a_{p''}\}) \ .$$

Ces notations sont étendues de façon évidente à la bijection $\alpha^{-1} : B \to A$ et aux sous-ensembles B' de B .

Remarque 3. - Pour tout $a , a' \in A$, on a

$$\underset{\sim}{Dp}(a , a' , \alpha , A) = \underset{\sim}{Dp}(\alpha\, a' , \alpha\, a , \alpha^{-1} , B) \ .$$

Démonstration. - Il suffit évidemment de vérifier l'énoncé pour tous les ensembles finis et, procédant par induction, nous supposerons $A = A_n$, $B = B_n$ $(n < \infty)$ et que le résultat est déjà établi pour chacun des sous-ensembles propres de A .

Soit $a^* = \alpha^{-1} b_n$ (où, comme toujours, $b_n = \max\{b : \ b \in B\}$). Si $a , a' \in A \backslash \{a^*\}$, on a rappelé plus haut que

$$\underset{\sim}{Dp}(a , a' , \alpha , A) = \underset{\sim}{Dp}(a , a' , \alpha , A \backslash \{a^*\})$$

et

$$\underset{\sim}{Dp}(\alpha\, a' , \alpha\, a , \alpha^{-1} , B) = \underset{\sim}{Dp}(\alpha\, a' , \alpha\, a , \alpha^{-1} , B_{n-1}) \ .$$

Donc, dans ce cas, la relation cherchée se déduit immédiatement de l'hypothèse d'induction. En raison de la symétrie de l'énoncé (entre $\alpha : A \to B$ et $\alpha^{-1} : B \to A$), il ne reste à discuter que les deux cas où

1° soit $a = a^*$, $a' = a_n$;

2° soit $a = a_n$, $a' = a^*$, avec $a_n \neq a^*$.

15-09

Cas (1). - D'après la définition même de $\leftarrow \alpha\, a^*$ et le fait que $\alpha\, a^*$ est plus grand que tous les éléments de $\{P(\alpha\ ,\ A)\}$, l'ensemble des éléments déplacés par $\alpha\, a^*$ se réduit à $\alpha\, a^*$ lui-même. Donc, si $a^* \neq a_n$, on a

$$0 = \underset{\sim}{Dp}(a\ ,\ a'\ ,\ \alpha\ ,\ A) = \underset{\sim}{Dp}(\alpha\, a'\ ,\ \alpha\, a\ ,\ \alpha^{-1}\ ,\ B)\ .$$

Au contraire, si $a^* = a_n$, $\underset{\sim}{Dp}(a\ ,\ a'\ ,\ \alpha\ ,\ A) = ((0\ ,\ 0)\ ,\ (i\ ,\ j))$ où j est le plus petit entier tel que $(1\ ,\ j) \notin |\underset{\sim}{P}(\alpha\ ,\ A_{n-1})|$. La même observation vaut pour $\underset{\sim}{Dp}(\alpha\, a'\ ,\ \alpha\, a\ ,\ \alpha^{-1}\ ,\ B)$ avec cette fois $(1\ ,\ j') \notin |P(\alpha^{-1}\ ,\ B_{n-1})|$ et l'égalité des deux déplacements résulte de l'hypothèse d'induction qui implique

$$|P(\alpha\ ,\ A_{n-1})| = |Q(\alpha^{-1}\ ,\ \alpha^{-1}\, A_{n-1})| = |P(\alpha^{-1}\ ,\ B_{n-1})|$$

puisque dans le cas examiné ici $\alpha\, A_{n-1} = B_{n-1}$.

Cas (2). - Considérons d'abord le cas où

$$\underset{\sim}{Dp}(a\ ,\ a'\ ,\ \alpha\ ,\ A) = ((i\ ,\ j)\ ,\ (i+1\ ,\ \bar{j}))\ .$$

Ceci implique $P(\alpha\ ,\ A_{n-1})^{-1}\, b_n = (i\ ,\ j)$ et, par conséquent, l'existence de $x \in A_{n-1}$ tel que $\alpha\, x$ ait déplacé b_n de $(i'\ ,\ j')$ (qui est éventuellement $(0\ ,\ 0)$) à $(i\ ,\ j)$.

En outre il doit exister $y \in A_{n-1}$ tel que $\alpha\, a_n \leqslant \alpha\, y < b_n$ et que $\alpha\, a_n$ déplace $\alpha\, y$ de $(i''\ ,\ j'')$ à $(i\ ,\ j)$. De fait $\alpha\, y$ est le plus grand des éléments de B_{n-1} qui soit déplacé par $\alpha\, a_n$. Appliquant l'hypothèse d'induction à $A\backslash\{a^*\} = a^{-1}\, B_{n-1}$, on en conclut que y est le dernier élément de B_{n-1} déplaçant $a = a_n$ dans $P(\alpha^{-1}\ ,\ B_{n-1})$ et que par conséquent,

$$P(\alpha^{-1}\ ,\ B_{n-1})^{-1}\, a = (i\ ,\ j)\ .$$

De façon analogue, l'hypothèse d'induction appliquée à A_{n-1} montre que $a' = \alpha^{-1}\, b_n$ déplace x de $(i'\ ,\ j')$ à $(i\ ,\ j)$ dans $P(\alpha^{-1}\ ,\ A_{n-1})$.

Comme $x < a_n$ il en résulte que l'opération $\leftarrow a'$ déplace $a = a_n$ de $(i\ ,\ j)$ en $(i+1\ ,\ \bar{j})$ ce qui achève la vérification dans ce cas puisque, trivialement, $\bar{\bar{j}} = \bar{\bar{j}}$, ces deux nombres ne dépendant que des formes

$$|P(\alpha\ ,\ A_{n-1}\backslash\{a^*\})| \quad \text{et} \quad |P(\alpha^{-1}\ ,\ B_{n-1}\backslash\{\alpha\, a_n\})|$$

qui sont identiques d'après l'hypothèse d'induction.

En raison de la symétrie, on a établi du même coup que $\underset{\sim}{Dp}(a_n\ ,\ a^*\ ,\ \alpha\ ,\ A) = 0$ si et seulement si $\underset{\sim}{Dp}(b_n\ ,\ \alpha\, a_n\ ,\ \alpha^{-1}\ ,\ B) = 0$ ce qui termine la vérification de la remarque.

Observons maintenant que la construction qui vient d'être discutée donne

$$P(\alpha^{-1} , \alpha A_n) = P(\alpha^{-1} , \alpha A_{n-1}) + (a_n)_{i,j}$$

où $(i , j) = |P(\alpha , A_n)| \setminus |P(\alpha , A_{n-1})|$. Donc, supposant déjà établi que $P(\alpha^{-1} , \alpha A_{n-1}) = Q(\alpha , A_{n-1})$, on a encore

$$P(\alpha^{-1} , \alpha A_n) = Q(\alpha , A_n)$$

et, par induction, dans tous les cas

$$P(\alpha^{-1} , B) = Q(\alpha , A)$$

ce qui est la formule cherchée.

Donnons une application de cette remarque au cas particulier de $A = B$.

PROPRIÉTÉ 2. - Une condition nécessaire et suffisante pour que $\alpha : A \to A$ soit une involution est que $P(\alpha , A) = Q(\alpha , A)$.

Démonstration. - Il est trivial que $A = B$ et $\alpha = \alpha^{-1}$ entraînent $P(\alpha^{-1} , B) = P(\alpha , A)$, c'est-à-dire $P(\alpha , A) = Q(\alpha , A)$ d'après la formule vérifiée dans cette section.

Réciproquement, supposons $P(\alpha , A) = Q(\alpha , A)$ et montrons qu'il en résulte $a_1 = b_1$, $\alpha a_1 = \alpha b_1$,

$$P(\alpha , A \setminus \{a_1 , \alpha a_1\}) = Q(\alpha , A \setminus \{a_1 , \alpha a_1\})$$

ce qui, par induction établit la propriété.

Revenant aux notations de la fin de la section 2, nous distinguons deux cas selon que $i^* > 1$ ou $i^* = 1$.

(1) $i^* > 1$. On a $\Delta Q(\alpha , A) = Q(\alpha , A \setminus \{a_1\})$ et l'on sait en déduire αa_1 et $P(\alpha , A \setminus \{a_1\})$. D'après la formule de la présente section

$$P(\alpha , A \setminus \{a_1\}) = Q(\alpha^{-1} , B \setminus \{\alpha a_1\}) .$$

Donc par la remarque 1 :

$$\Delta P(\alpha , A \setminus \{a\}) = \Delta Q(\alpha^{-1} , B \setminus \{\alpha a_1\}) = Q(\alpha^{-1} , B \setminus \{\alpha a_1 , b_1\}) .$$

Partons maintenant de $Q(\alpha^{-1} , B)$ qui, toujours d'après la même formule, est égal à $P(\alpha , A)$. Répétant le même calcul que plus haut, on en déduit $Q(\alpha , A \setminus \{a^{-1} b_1 , a_1\})$ qui est donc égal à $Q(\alpha^{-1} , B \setminus \{\alpha a_1 , b_1\})$ d'après l'hypothèse $P(\alpha , A) = Q(\alpha , A)$. Une troisième application de la formule donne

$$Q(\alpha^{-1} , B \setminus \{\alpha a_1 , b_1\}) = P(\alpha , A \setminus \{a_1 , \alpha^{-1} b_1\})$$

et le résultat est vrai dans ce cas.

(2) $i^* = 1$. Dans ce cas les observations faites à la fin de la section 2 et la formule de la présente section donnent directement

$$\alpha\, a_1 = a_1 \quad \text{et} \quad P(\alpha\,,\, A\backslash\{a_1\}) = Q(\alpha\,,\, A\backslash\{a_1\})\,.$$

La propriété est donc vérifiée dans tous les cas.

Examinons plus en détail le cas où $A = B \neq \emptyset$ est fini et $P(\alpha\,,\, A) = Q(\alpha\,,\, A)$ et, pour tout tableau standard S, dénotons par $\underline{\mathrm{Imp}}|S|$ le nombre des $j \in N$ tels qu'il existe un nombre impair de $i \in N$ pour lesquels $(i\,,\, j) \in |S|$. Il résulte des définitions que, dans le cas (2) discuté plus haut,

$$|Q(\alpha\,,\, A)|\backslash|\Delta Q(\alpha\,,\, A)| = (i\,,\, j^*)$$

et qu'il n'existe pas d'autre $i \in N$ tels que $(i\,,\, j^*) \in |Q(\alpha\,,\, A)|$. Donc

$$\underline{\mathrm{Imp}}|Q(\alpha\,,\, A\backslash\{a_1\})| = \underline{\mathrm{Imp}}|Q(\alpha\,,\, A)| - 1\,.$$

Dans le cas (1) soit $(i^*\,,\, j^*) = |Q(\alpha\,,\, A)|\backslash|\Delta Q(\alpha\,,\, A)|$ où par hypothèse $i^* > 1$. Soit $|U'|$ la séquence relative à l'opération Δ dans $P(\alpha\,,\, A\backslash\{a_1\})$. Il est facile de voir qu'il existe un entier k tel que $(i_{k'}\,,\, j_{k'}) \in |U|$ entraîne $(i_{k'}\,,\, j_{k'}) \in |U'|$ si $k' < k$ et $(i_{k'-1}\,,\, j_{k'}) \in |U'|$ si $k' \geq k$. Il s'en déduit que

$$|P(\alpha\,,\, A\backslash\{a_1\})|\backslash|\Delta P(\alpha\,,\, A\backslash\{a_1\})| = (i^* - i\,,\, j^*)$$

et par conséquent, d'après nos remarques antérieures

$$\underline{\mathrm{Imp}}|Q(\alpha\,,\, A\backslash\{a_1\,,\, \alpha^{-1}\, b_1\})| = \underline{\mathrm{Imp}}|Q(\alpha\,,\, A)|\,.$$

Par induction, l'on conclut de ces deux relations que si α est une involution sur l'ensemble fini A, le nombre des éléments laissés invariants par α est précisément égal à $\underline{\mathrm{Imp}}|Q(\alpha\,,\, A)|$.

5. L'opération I .

Soit Q un tableau standard tel que $0 < \underline{\mathrm{Card}}\{Q\} = n < \infty$. On définit $Q^I \in \mathcal{C}$ par l'équation

$$Q^I = \Sigma\,\{(a_k)_{i_k, j_k} : k \in [1\,,\, n]\}$$

où a_k désigne le k-ième élément de $\{Q\}$ par ordre croissant et où, ici $(i_k\,,\, j_k) = |\Delta^{n-k+1} Q|\backslash|\Delta^{n-k} Q|$ (avec $\Delta^0 Q = Q$). Trivialement, $\{Q\} = \{Q^I\}$, $|Q| = |Q^I|$ et $Q^{IT} = Q^{TI}$ où T indique la transposition. De plus si la bijection $\sigma : \{Q\}\backslash\{a_1\} \to \{Q\}\backslash\{a_n\}$ définie par $\sigma\, a_{k+1} = a_k$ pour $k \in [1\,,\, n-1]$ est étendue de façon naturelle à $\mathcal{C}$, on vérifie sans peine que

15-12

$$Q^I = \sigma(\Delta Q)^I + (a_n)_{i_n,j_n} .$$

On verra plus bas que Q^I est standard et $Q^{II} = Q$. Par exemple, pour $Q = \begin{smallmatrix}248\\679\end{smallmatrix}$ comme plus haut,

$$Q^I = \begin{smallmatrix}267\\489\end{smallmatrix} , \quad (\Delta Q)^I = \begin{smallmatrix}478\\69\end{smallmatrix} , \quad \sigma(\Delta Q)^I = \begin{smallmatrix}267\\48\end{smallmatrix} .$$

Soit maintenant $Q = Q(\alpha , A)$ où $0 < \underline{\text{Card}}\ A = n < \infty$. La bijection $\overline{\alpha} : A \to B$ étant définie par $\overline{\alpha}\, a_k = \alpha\, a_{n-k+1}$ pour $k \in [1 , n]$, il a été prouvé par SCHENSTED que

$$P(\overline{\alpha} , A) = P(\alpha , A)^T$$

(lemme 7 loco citato p. 186). Nous vérifions par induction sur n que $Q(\overline{\alpha} , A) = Q(\alpha , A)^{IT}$ en observant que le résultat est vrai pour $n = 1$ et en supposant qu'il est déjà établi pour $A' = A\backslash\{a_1\}$.

Compte tenu de la relation $\overline{\alpha}\sigma A' = A'$, et écrivant comme d'habitude A_{n-1} pour $A\backslash\{a_n\}$, ceci revient à supposer

$$Q(\overline{\alpha} , A_{n-1}) = \sigma\, Q(\alpha , A')^{IT} .$$

Maintenant, $Q(\overline{\alpha} , A) = Q(\overline{\alpha} , A_{n-1}) + (a_n)_{i,j}$ et, comme on l'a noté plus haut,

$$Q(\alpha , A)^{IT} = \sigma(\Delta Q(\alpha , A))^{IT} + (a_n)_{j_n,i_n} .$$

D'après la remarque 1, $\Delta Q(\alpha , A) = Q(\alpha , A')$ et par conséquent :

$$Q(\overline{\alpha} , A) = Q(\alpha , A)^{IT} - (a_n)_{j_n,i_n} + (a_n)_{i,j} .$$

Il suffit donc de vérifier $|Q(\overline{\alpha} , A)| = |Q(\alpha , A)^{IT}|$. Or ceci résulte immédiatement de $|Q(\alpha , A)^I| = |Q(\alpha , A)|$, de l'égalité de forme des P-symboles et des Q-symboles et de l'égalité $|P(\overline{\alpha} , A)| = |P(\alpha , A)^T|$ impliquée par l'identité de Schensted. La formule est donc établie.

CERTAIN ELEMENTARY FAMILIES OF AUTOMATA

M. P. Schutzenberger
Harvard Medical School, Harvard University

Reprinted from the
PROCEEDINGS OF THE SYMPOSIUM ON
MATHEMATICAL THEORY OF AUTOMATA

POLYTECHNIC INSTITUTE OF BROOKLYN
April 24, 25, 26, 1962

CERTAIN ELEMENTARY FAMILIES OF AUTOMATA

M. P. Schutzenberger*
Harvard Medical School, Harvard University

We attempt to relate the difficulty of the decision problem of certain algorithms (automata) with the underlying algebraic structure. In particular we discuss the connection between "push-down storage" and "extension of a free group by a finite monoid."

I. INTRODUCTION

This note is concerned with the definition of families of sets of words in a finite input alphabet X.

In contrast with the more usual approach, the motivation for this study is purely formal, the purpose being to obtain sets of words as near as possible to the family of the so-called "regular events" [25] by their definition and by their closure properties under the elementary set-theoretic operations.

As an illustration let us consider the case of X reduced to a single letter x. Then a set of words $F' = \{x^{n_1}, x^{n_2}, \ldots, x^{n_i}, \ldots\}$ can be identified with the function from the natural numbers into $\{0, 1\}$ which, for each n, takes values 1 or 0 according to $x^n \in F'$ or not.

However, if one considers a process which *produces* the words of F' (as opposed to a process which *recognizes* or *accepts* them), more detailed information than " x^n is produced at least once" may be of some significance. Accordingly one may want to consider a numerical function (which we denote by (a, x^n)) expressing how many different ways (eventually zero) each word x^n is produced. A strictly equivalent procedure is to consider the *generating function*

$$a = \sum_{n \geq 0} (a, x^n) x^n$$

of the sequence $(a, x^0), (a, x^1), (a, x^2), \ldots, (a, x^n), \ldots$ With this notation, F' is the set of those x^n such that $(a, x^n) = 0$, that is, the *support* of the function a.

* Presently with the Faculte des Sciences, Poitiers, France.

Presented at the *Symposium on Mathematical Theory of Automata*, Polytechnic Institute of Brooklyn, April 24, 25, 26, 1962

In the present case, where $(a, x^n) \geq 0$ for all n, simple relations exist between the algebraic operations on generating functions and the elementary set-theoretic operations on their supports. In the more general case where (a, x^n) can be a negative integer, some analytic properties of the function a of x are reflected in the combinatorial nature of its support.

For instance, Skolem has proved[43] that if a is a rational function of x, it has an *ultimately periodic* support,[30] i. e., its support is a regular event. No corresponding result is known for a algebraic, but when the coefficient ring has non-zero characteristics, examples[27] show that the support is not necessarily a regular event. Moreover, the classical *gap theorems* [15] show that sets like $\{x^{n^2} : n > 0\}$ cannot be the support of an algebraic function over the field of complex numbers.

For the general case of X consisting of a finite number of variables, we need to define a non-commutative counterpart of the algebraic functions. This is done in Section II under certain restrictive hypotheses.

Then the algorithms by which the successive coefficients are computed can be reformulated in terms of automata, i. e., of representation of the free monoid F generated by X. As expected, these representations are among the most elementary from the point of view of the theory of monoids.

Indeed, the "algebraic" generating functions are associated with homomorphisms of F into a free group, i. e., with a special case of a so-called "push down storage." [37, 31]

In Section IV we list several problems concerning the supports which have been proved to be unsolvable.

Another presentation of this material but with a definitely different emphasis is given elsewhere by N. Chomsky and myself. [13] In fact, most of the remarks developed here (and especially the ones dealing with push down storage) are results of this collaboration over a period of many years.

II. FORMAL POWER SERIES

Let X be the finite input alphabet, F be the set of all words in the letters of X, and denote by (a, f) a mapping from F into the rational integers. To this mapping one associates the formal power series $a = \Sigma\{(a, f).f : f \in F\}$, which is an element of the completion of the free module generated by F. The set $R(X)$ of all such power series is a ring with addition $a + a' = \Sigma\{((a, f)(a', f)).f : f \in F\}$; non-commutative multiplication $aa' = \Sigma\{(a, f)(a', f').f'' : f, f', f'' \in F; f'' = ff'\}$ (where ff' is the concatenation of f and f'); and multiplication by a scalar $n.a = \Sigma\{n(a, f).f : f \in F\}$. These are, of course, the usual operations when X consists of a single variable.

An element a of $R(X)$ is *quasi-regular* if the coefficient of the

empty word in it is zero. Then a has a *quasi-inverse*

$$a^* = \sum_{n>0} a^n$$

which satisfies $a^*a + a = a + aa^* = a^*$. $R(X)$ also admits a *Hadamard product* (cf. reference 39) $a \circ a' = \Sigma\{(a,f)(a',f).\, f : f \in F\}$. All these operations are continuous in the usual topology of $R(X)$. The subset $R^{pos}(X)$ of the power series having non-negative coefficients is a *semi-ring* (i. e., it is closed under addition, multiplication and multiplication by a non-negative scalar) which contains the quasi-inverse of its quasi-regular elements.

The support, supp. a, of any $a \in R(X)$ is $\{f \in F : (a,f) \neq 0\}$. For any $a, a' \in R^{pos}(X)$:

$$\text{supp. } (a + a') = \text{supp. } a \cup \text{supp. } a';$$

$$\text{supp. } (a \circ a') = \text{supp. } a \cap \text{supp. } a';$$

$$\text{supp. } (a\, a') = (\text{supp. } a)(\text{supp. } a')$$

(= the "set product" of supp. a and supp. a').

If, further, a is quasi-regular,

$$\text{supp. } (a^*) = (\text{supp. } a)^* \left(= \cup \{(\text{supp. } a)^n : n > 0\}\right)$$

where in the right member $(\text{supp. } a)^*$ denotes Kleene's *star operation*.[25]

In fact these relations express the existence of a natural homomorphism (of semi-ring) sending $R^{pos}(X)$ onto the semi-ring $B(X)$ of formal power series with boolean coefficients. For obvious reasons the direct study of $R^{pos}(X)$ is far more elementary than that of $B(X)$.

DEFINITION 1: $R_{pol}(X)$ is the ring of the integral power series having a finite support. In other words, $R_{pol}(X)$ is the free (associative) algebra generated by X.

DEFINITION 2: $R^{pos}_{nil}(X)$ denotes the least semi-ring which contains every power series of the form $\Sigma\{f : f \in F'\}$ where F' is an arbitrary regular event.

DEFINITION 3: $R^{pos}_{rat}(X)$ denotes the least semi-ring that contains X and the quasi-inverse of each of its quasi-regular elements.

Now let $Y = \{y_j\}$ $(1 \leq j \leq N)$ be a set of $N < \infty$ new variables and consider an N-tuple $p = (p_j)$ of elements of $R_{pol}(X \cup Y)$. It is a *proper positive system* if it satisfies the conditions that for all $j, j' \leq N$:

1) $p_j \in R^{pos}(X \cup Y)$;

2) p_j is quasi regular;

3) $(p_j, y_{j'}) = 0$.

If, further, each $g \in \text{supp. } p_j$ has the form f or $fy_{j''}f'$ with f and f' belonging to the monoid generated by X, then p is a (two-sided) *linear system*. A linear system in which every f' is the empty word is a *right linear system*.

Let $u = (u_j)$ be an N-tuple of quasi-regular elements of $R(X)$ and define a homomorphism $\lambda_u: R(X \cup Y) \to R(X)$ by $\lambda_u y_j = u_j$, $\lambda_u x = x$ $(x \in X, y_j \in Y)$. It is trivial that any proper positive system p determines a *unique* quasi-regular N-tuple u such that for all $j \leq N$, $u_j = \lambda_u p_j$. Hence, u can be called "the solution" of p and we note that its coordinates belong to $R^{pos}(X)$.

DEFINITION 4: $R^{pos}_{alg}(X)$ is the least semi-ring that contains X and the coordinates of the solution of every proper positive system.

Clearly this definition is equivalent to the definition of the context free languages of Chomsky,[8, 9] and the coefficients in the solution precisely express the number of ways in which a word can be produced by the grammar corresponding to the system p.[18, 19]

It is trivial that $R^{pos}_{alg}(X)$ contains the quasi-inverses of its quasi-regular elements. Hence, $R^{pos}_{rat}(X)$ is a sub-semi-ring of $R^{pos}_{alg}(X)$. More accurately, an element of $R^{pos}_{alg}(X)$ belongs to $R^{pos}_{rat}(X)$ if and only if it is a coordinate of a proper positive linear right system. Thus, for any $a \in R^{pos}_{rat}(X)$, supp. a is a regular event.[8] Furthermore, $R^{pos}_{nil}(X)$ is a sub-semi-ring of $R^{pos}_{rat}(X)$ and $a \in R^{pos}_{rat}(X)$ belongs to $R^{pos}_{nil}(X)$ if and only if for all $\epsilon > 0$ and $f, f', f'' \in F$ one has $\lim_{n \to \infty} (1 + \epsilon)^{-n} (a, f' f^n f'') = 0$. The subset of $R^{pos}_{alg}(X)$, corresponding to the two-sided linear system, is not a semi-ring and $R^{pos}_{alg}(X)$ contains as a proper subset the least semi-ring which includes all these elements and the quasi-inverse of each of its quasi-regular members.

It is clear that if λ is the endomorphism of $R(X)$ induced by a mapping $\lambda x_i = a_i$ with $a_i \in R^{pos}_{rat}(X)$, the restriction of λ to $R^{pos}_{alg}(X)$ $\left[\text{resp. to } R^{pos}_{rat}(X)\right]$ is an endomorphism. Hence, as a variant of Jungen's theorem,[13] one verifies that $R^{pos}_{nil}(X)$ and $R^{pos}_{rat}(X)$

are closed for the Hadamard product and that $a \in R^{pos}_{rat}(X)$, $a' \in R^{pos}_{alg}(X)$ implies $a \circ a' \in R^{pos}_{alg}(X)$. It is well known[23] that $R^{pos}_{alg}(X)$ is not closed for the Hadamard product (even with X reduced to a single letter or with a more classical definition of the Hadamard product.[21, 4, 7]) In fact one has even the stronger result that for some pairs $a, a' \in R^{pos}_{alg}(X)$ the intersection of the supports of a and a' cannot be the support of a power series of the form $a'' - a'''$ where $a'', a''' \in R^{pos}_{alg}(X)$. $\left(\text{Take for instance } a = \Sigma\left\{x_1^n x_2^n x_3^{n'} : n, n' > 0\right\}\right.$ $\left.a' = \Sigma\left\{x_1^n x_2^{n'} x_3^{n'} : n, n' > 0\right\}\right)$. However, as in the commutative case, if p is an N-tuple of elements of $R^{pos}_{alg}(X \cup Y)$ satisfying conditions (1), (2) and (3) above, the system $\left\{y_j = p_j\right\}$ has a unique quasi-regular "solution" whose coordinates still belong to $R^{pos}_{alg}(X)$.

DEFINITION 5: $R_{nil}(X)$ (resp. $R_{rat}(X)$, $R_{alg}(X)$) is the least ring containing $R^{pos}_{nil}(X)$ (resp. $R^{pos}_{rat}(X)$, $R^{pos}_{alg}(X)$).

It follows from the definition of the semi-rings considered that each element of these rings can be expressed (in infinitely many ways) as the difference of *two* elements of the corresponding semi-rings. Hence, one could obtain directly $R_{nil}(X)$, $R_{rat}(X)$ or $R_{alg}(X)$ by replacing in definitions 2 and 3 the word *semi-ring* by the word *ring* or by omitting condition (1) in the definition of a proper system.

I stress once more that the only motivation I can offer for introducing the rings $R_{nil}(X)$, $R_{rat}(X)$, and $R_{alg}(X)$ is the strictly personal opinion that their definition is, in a sense, as simple as possible and, accordingly, that some reasonable families of sets of words are likely to include (or be included in) the corresponding families of supports.

Let α be the canonical homomorphism sending $R_{pol}(X)$ onto the ring of the ordinary (i. e., commutative) polynomials with integer coefficients. Clearly, one can extend α to epimorphisms of $R_{rat}(X)$ and $R_{alg}(X)$ onto the ring of the Taylor series expansions (with integer coefficients) of the ordinary rational and algebraic functions because,[3] for any $a \in R^{pos}_{alg}(X)$, there exists a finite constant $K > 0$ such that the quantity $(a, f)K^{-|f|}$ (where $|f|$ denotes the length of f) remains bounded over all $f \in F$.

If X consists of a single letter, $R_{nil}(X) = \alpha R_{nil}(X)$ is a rather classical object of study. I have no direct characteri-

zation of $\alpha\, R_{nil}(X)$ in the general case.

More specifically, let $\bar{a}$ be an ordinary rational function of the (commuting) variables $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M$ $(1 < M < \infty)$ and assume that the coefficients of its Taylor series expansion $\bar{a} = \Sigma\, \bar{a}_{n_1, n_2, \ldots, n_M} \times \bar{x}_1^{n_1} \bar{x}_2^{n_2} \ldots \bar{x}_M^{n_M}$ are integers satisfying identically $|\bar{a}_{n_1, n_2, \ldots, n_M}| \leq (k + n_1 + n_2 + \ldots + n_M)!\, (n_1!)^{-1} (n_2!)^{-1} \ldots (n_M!)^{-1}$ for some fixed finite k. What supplementary conditions are needed to insure that $\bar{a} = \alpha\, a$ for at least one $a \in R_{nil}(X)$?

III. THE QUOTIENT MONOID OF AN AUTOMATON

Now we describe algorithms by which the coefficients in these formal power series can be computed.

For this purpose F (the set of all input words) is considered as the free monoid generated by the input alphabet X (with the concatenation as product). Then if an automaton $\bar{\sigma}$ is given by a set of states S and the so-called "next state" function[17] $(S, X) \to S$, one may identify $\bar{\sigma}$ with the representation of F by mappings $S \to S$ that extends $(S, X) \to S$ in a natural fashion. Thus $\bar{\sigma}$ determines a homomorphism σ of F onto a quotient monoid σF (the "quotient monoid of the automaton") by:

> for all $f, f' \in F$, $\sigma f = \sigma f'$ if and only if for all $s \in S$, $s.f = s.f'$ (i. e., if for any choice of an initial state $s \in S$, the states $s.f$ and $s.f'$ reached after reading f or f' are the same).

In other words, if $\sigma f = \sigma f'$ the automaton $\bar{\sigma}$ offers no possibility of distinguishing between them. Hence, the set F_σ of the words accepted by $\bar{\sigma}$ has the closure property $F_\sigma = \sigma^{-1}\, \sigma\, F_\sigma$.[2, 42]

For instance, Bar Hillel and Shamir have pointed out that the regular events are characterized by this closure property with respect to the homomorphisms σ of F into a *finite* monoid. This allows one to translate into algebraic language certain of the operations performed on the sets of words. Thus, trivially, if $F_1 = \sigma_1^{-1}\, \sigma_1\, F_1$ and $F_2 = \sigma_2^{-1}\, \sigma_2\, F_2$ are two subsets of F having the closure property with respect to the homomorphisms σ_1 and σ_2, their union and intersection are closed with respect to the homomorphism $\sigma_3 : F \to \sigma_1 F \times \sigma_2 F$.

As another example, $\bar{\sigma}$ being a given automaton, let $\{F'_i : 1 \leq i \leq N'\}$ and $\{F''_i : 1 \leq i \leq N''\}$ be two partitions of F into a finite number of regular events and let τ be an arbitrary mapping into F of the set of all triples (i', x, i'') $(1 \leq i' \leq N',\ x \in X,\ 1 \leq i'' \leq N'')$. Consider now a device $\bar{\tau}$ (see reference 22) which associates with each input word $f = x_1 x_2 \ldots x_m$ the word

$\tau f = \tau(i'_1, x_1, i''_1)\,\tau(i'_2, x_2, i''_2) \ldots \tau(i'_m, x_m, i''_m)$ where, for each j, i'_j and i''_j are determined by $x_1 x_2 \ldots x_{j-1} \in F'_{i'_j}$ and $x_{j+1} \ldots x_{m-1} x_m \in F''_{i''_j}$. Finally, we cascade $\bar{t}$ and $\bar{\sigma}$ in the sense that we take τf instead of f as the input of $\bar{\sigma}$. Denoting this composite automaton by $\bar{\sigma}'$, it is easily verified that the corresponding homomorphism σ' is a homomorphism into the *extension of* σF *by a finite monoid* in the sense of Redei.[36] (Let A and B be two monoids and denote by b^a ($a \in A$, $b \in B$) a representation of A by endomorphisms of B. If the mapping $\beta: (A, A) \to B$ is such that the product $(a, b)(a', b') = (aa', b^{a'}, \beta(a, a')b'^{a})$ on (A, B) is associative, the corresponding monoid is called an extension of B by A).

Let us now consider the simplest type of infinite monoid, i. e., the infinite cyclic group. An automaton $\bar{\sigma}$ such that σF is a submonoid of this group consists of a single "counter." It is described by associating with each $x \in X$ a positive or negative integer σx so that for each $f = x_1 x_2 \ldots x_m$, the counter records the total $\sigma x + \sigma x_2 + \ldots + \sigma x_m$. We say that an input word f is *accepted* if and only if σf does not belong to some specified *finite* subset of integers. Then, trivially, the set F_σ of the words accepted by $\bar{\sigma}$ is the support of a formal power series $a \in R_{nil}(X)$. It is easily shown that conversely:

If $a \in R_{rat}(X)$ is such that $|(a, f)| \,/\, 1 + |f|\,|^{-1}$ is bounded over all $f \in F$, then supp. a is the set of the words accepted by a finite automaton $\bar{\sigma}$ such that

1) σF is a submonoid of the extension of an infinite cyclic group by a finite monoid.

2) f is accepted only if it does not belong to a certain prescribed finite collection of cosets (i. e., if $s_o f \notin S'$ where the initial state $s_o \in S$ and the final *finite* subset S' of S are given).

With the same type of quotient monoid σF but with an opposite definition of the rule for accepting words (that is, f being accepted if and only if $s_o f$ does belong to some prescribed *finite* collection of integers), one easily shows that $\Sigma\{f: f \in F_\sigma\}$ belongs to $R^{pos}_{alg}(X)$. A well-known example of a set of this type is the set of all well-formed formula in parenthesis free notation. Several theorems on a similar but more general problem have been proved by Raney.[35]

I mention that this last example can be converted into a classical probabilistic problem, viz., finding the probability generating function corresponding to the usual random walk problem for an arbitrary (finite state) Markov process.

Following the customary hierarchy which ranks nilpotent groups next above abelian groups in order of simplicity, we have:

A necessary and sufficient condition that $F_\sigma \in \{\text{supp. } a\colon a \in R_{nil}(X)\}$ is that

1) σF be a submonoid of the extension by a finite monoid of a free nilpotent group;

2) f is accepted if and only if σf does not belong to some prescribed finite subset of σF.

As mentioned above this is a weak form of a well-known result in the theory of rational functions. A similar property holds for $\{\text{supp. } a\colon a \in R_{rat}(X)\}$ with (2) as before and (1) replaced by

1') σF is a submonoid of the ring Z_N of the NxN-integral matrices $(N < \infty)$.

In this more general case, the automaton $\overline{\sigma}$ consists of a finite part $\overline{\sigma}_o$ and of a "memory" in which an N-dimensional integral vector v can be stored. When reading the input word

$$f = x_{i_1} x_{i_2} \dots x_{i_j} \dots x_{i_n},$$

each successive letter x_{i_j} determines a bounded sequence of computation amounting to the multiplication of v by a certain NxN integral matrix μx_{i_j}. Thus, at the end of f, the memory contains the vector $v(f) = v_o \mu x_{i_1} \mu x_{i_2} \dots \mu x_{i_n}$, and f is accepted if and only if $v(f)$ does not satisfy a finite number of linear equations. These are the automata of the family A.

Let $|v(F)|$ denote the length of the vector $v(f)$. By construction, $|v(f)|$ does not grow faster than an exponential function of the length, $|f|$, of the word f, but it may grow exactly at this rate. Hence, since the number of distinct words of length k or less is an exponential function of k, the memory V may be so well employed that $v(f) \neq v(f')$ for any two distinct words f and f', that is σ may be an isomorphism. In other words, the mapping $f \to v(f)$ may involve no compression of the data. Hence, it may be interesting to define a subfamily by requirements implying $\sigma F \neq F$.

The simplest condition is that $|v(f)|$ does not grow as fast as an exponential function of $|f|$, that is:

(*) For any $\epsilon > 0$ there exists a finite K such that $|v(f)| < (1+\epsilon)^{|f|}$ for all $f \in F$ of length $> K$.

Of course, this condition is equivalent to the one defining $R_{nil}(X)$ and it can be verified that for each $a \in R_{nil}(X)$ there exists a natural number d, the "degree" of a, having the following properties:

1) $(a, f)|f|^{-d}$ is bounded over all $f \in F$;

2) There are infinitely many words f such that $|(a, f)|\,|f|^{-d} \geq 1$.

Thus, the smaller the "degree" of a, the "smaller" is the homomorphic image σF. In particular, $d = 0$ characterizes the (infinite) regular events and the finite automata; $d = 1$ characterizes the example described at the beginning of this section.

An unsolved problem is to replace the rather obvious condition (*) by one involving only the rate of growth of the number of distinct vectors $v(f)$ with the length of f.

Pushdown Storage

I am unable to construct a family of automata which would serve for $R_{alg}(X)$ in exactly the same fashion as the family A does for $R_{rat}(X)$. However, under a certain relaxation of the conditions, such a family has been devised by N. Chomsky[10] for $R^{pos}_{alg}(X)$. Related results have been obtained independently by R. G. Evey. The nearest approximation to the desired results involves the following definition.

DEFINITION: A subset F' or F is a D-event if and only if it is the intersection of a regular event with the kernel of a homomorphism of F into a free group. Then:

A necessary and sufficient condition that $a \in R_{alg}(X)$ is that $a = \Sigma \{\theta f: f \in F'\}$ for some homomorphism $\theta : F \to R_{pol}(X)$ and D-event F'.

(This, of course, provided that the infinite sum exists.)

The corresponding statement with $R_{rat}(X)$ instead of $R_{alg}(X)$ and F' a regular (instead of D-) event is trival. In both statements the regular event F'' used can belong to the special subfamily of the sets F'' defined by two subsets V' and V'' of (X, X) and the condition:

$f = x_{i_1} x_{i_2} \ldots x_{i_m} \in F''$ if and only if $(x_{i_1}, x_{i_m}) \in V'$

and, for all j, $(x_{i_j}, x_{i_{j+1}}) \in V''$.

In other words F'' is the intersection of the complement of a *two-sided ideal* with a *quasi-ideal*. It is to be noted that the results are valid for any *unital* coefficient ring R (commutative or not) and that $\theta : F \to R_{pol}(X)$ can be restricted by the condition $\theta x = r.f$ with $r \in R$, $f \in F$ for all $x \in X$. Also the condition "F' is a D-event" can be replaced by "F' is the inverse image of a finite set for some homomorphism of F into the extension of a free group by a finite monoid."

In a still more restricted manner the D-event F' can be given the form $F' = F'' \cap F'''$ with F'' as above and F''' defined in the following fashion:

Let $X = \{x_i\}$, $(i = \pm 1, \pm 2, \ldots, \pm N)$ where $N \geq 2$. Clearly, there exists a unique epimorphism γ of $F(X)$ onto the free group G generated by all x_i's with $i > 0$ that identically satisfies $\gamma x_i \gamma x_{-i} = 1$. Then F''' is precisely the kernel of γ.

The fact that $\Sigma\{f: f \in F'\}$ belongs to $R_{alg}^{pol}(X)$ for any D-event F' follows from the construction which we now describe.

Let $\bar{\sigma}$ be an automaton consisting of a finite set of states S and of a tape (the "memory" of $\bar{\sigma}$) on which both writing and erasing are possible. Let Z be the alphabet used on this tape and G be the free monoid generated by Z. The automaton $\bar{\sigma}$ is given by:

1) A homomorphism ψ of G onto a finite monoid K;

2) A mapping σ': $(S, K, X) \to S$;

3) A mapping χ of (S, K, X) into the set of all subsets of K,

4) A mapping α' : $(S, K, X) \to G$.

Thus, the state of $\bar{\sigma}$ is a pair $(s, g) \in (S, G)$, and if the incoming input letter is x, the following operations are performed:

i) The finite part goes to $s' = \sigma'(s, \psi g, x)$;

ii) The word g is factored into a product $g'g''$ where g is the right factor of minimal degree such that $\psi g'' \in \chi(s, \psi g, x)$. ($g''$ is the empty word if no such factor exists);

iii) g'' is erased and replaced by the word $\alpha'(s, \psi g, x)(= g''')$.

Thus, the "next state function" is $(s, g)\cdot x = (s', g'g''')$. No essential gain in generality would accrue if the factorization $g = g'g''$ was determined by *two* finite state automata (with set of states S' and S'') and a condition of the form:

g' and g'' are such that

1) $s'_0 \cdot g' \in \bar{S}'$, $s''_0 \cdot g'' \in \bar{S}''$ where $\bar{S}' \subset S'$ and $\bar{S}'' \subset S''$ are functions of $s \in S$, ψg, and x;

2) g'' has maximal or minimal length depending upon the triple $(s \in S, \psi g \in K, x \in X)$.

In some sense, $\bar{\sigma}$ can be considered as a very special case of a "pushdown storage." [31, 37] Here the essential restrictions are:

1) The memory consists of a single tape;

2) Any feedback from the memory to the finite part is via the

image of g by a fixed homomorphism into a *finite* monoid;

3) For each input letter only a *bounded number* of erasing and writing operations are permitted.

Now let $(s_o \in S,\ g_o \in G)$ be an initial state and $\overline{S}_1 = \{(s_{1,i}, g_{1,i})\}$ and $\overline{S}_2 = \{(s_{2,i}, g_{2,i})\}$ be two finite sets of states. We define

$$F_\sigma = \{ f \in F: (s_o, g_o) \cdot f \in \overline{S}_1,\ (s_o, g_o) \cdot f' \notin \overline{S}_2$$

for all left factor f' of $f\}$. Direct computation shows that

$$\Sigma\{ f: f \in F_\sigma\} \in R^{pos}_{alg}(X) .$$

Clearly the D-events (or the inverse image of a finite set for a homomorphism of F into the extension of a free group by a finite monoid) are special cases of such sets F_σ. This provides an independent verification of certain results of Kesten[24] concerning random walks over free groups.

It is conjectured that, conversely, if the inverse image F' of a finite set for some homomorphism ϕ of F into a group is such that $\Sigma\{ f: f \in F'\} \in R_{alg}(X)$, then ϕF is a submonoid of the extension by a *finite* group of a *free* group.

IV. SOME PROBLEMS CONCERNING THE SUPPORTS

We have defined six families of sets $\mathfrak{R}^i_j = \{ \text{supp. } a : a \in R^i_j(X)\}$ (i = nothing or pos, j = rat, nil, or alg), and we know that $\mathfrak{R}^{pos}_{nil} = \mathfrak{R}^{pos}_{rat}$ = the family $\mathfrak{R}_o$ of regular events. It can easily be proved that $\mathfrak{R}_{nil,\,d} = \{ \text{supp. } a: a \in R_{nil}(X),\ \text{Deg } a \leq d \}$ is a strictly increasing function of d and that $\mathfrak{R}_{nil,\,2}$ contains sets which do not belong to $\mathfrak{R}^{pos}_{alg}$ (e. g., $\{ x^n y^m z^p:\ n^2 \neq mp\}$). Conversely, $\{ x^n y^n:\ n > 0 \} \in R^{pos}_{alg}$ but it does not belong to $\mathfrak{R}_{rat}$ (cf. reference 16).

Simple examples show that none of these families (except $\mathfrak{R}^{pos}_{rat}(X)$, of course) is closed under complementation.[38]

The first question is to determine whether or not $F' = F$ for a given set F' described as a member of one of these families; or, in other words, whether or not the corresponding automaton accepts all the possible input words.

It is trivial that to any diophantine equation E of degree d there corresponds at least one element $a \in R_{nil}(X)$ of the same "degree" such that E has a solution if and only if $0 = (a, f)$ for at

least one $f \in F$, that is, if and only if supp. $a \subset F$.

Conversely, for $d = 0, 1$ the problem of determining if supp. $a = F$ admits elementary solutions. The same problem for $d = 2$ (corresponding to the quadratic case for diophantine equations) relates to a question on (infinite) nilpotent monoids of class 1. For $d \geq 3$ the problem is at least as unsolvable as the ordinary diophantine problem.

For $a \in R_{rat}(X)$, a theorem of Markov shows that the same problem is unsolvable.[28] A fortiori, a similar negative result holds for $R_{alg}(X)$ and even for $R^{pos}_{alg}(X)$. However, in this case many more undecidability properties can be established because of the following construction due to Bar Hillel, Perles and Shamir.[1]

For each $f = x_{i_1} x_{i_2} \dots x_{i_m} \in F$ we denote by $\tilde{F}$ the word $x_{i_m} \dots x_{i_2} x_{i_1}$ and, given a homomorphism $\alpha: F \to F$ we consider the two-sided linear equation $y = x_o + \Sigma\{ xy\, \alpha x : x \in X \}$ where x_o is a new letter not contained in X.

Its solution is the power series $a \in R^{pos}_{alg}(X \cup \{x_o\})$:

$$a = \Sigma\{ \tilde{f}\, x_o\, \alpha f : f \in F \}.$$

Repeating the same construction with a second homomorphism $\alpha' : F \to F$, we obtain the power series $a' = \Sigma\{ f\, x_o\, \alpha'\, f : f \in F \}$.

The power series $a + a'$ also belongs to $R^{pos}_{alg}(X \cup \{x_o\})$.

Clearly, $a + a'$ has at least one coefficient ≥ 2 if and only if the supports of a and a' have a non-empty intersection; this last question is equivalent to Post's correspondence problem.[33]

Hence, trivially, the problem of determining whether an arbitary $b \in R^{pos}_{alg}(X)$ is or is not the generating function of its support (i. e., if an arbitrary context free grammar is or is not ambiguous[32]) is unsolvable.

In fact, Post's problem can be translated in many ways into the terminology used in this paper. For example, to any (one-way) two tapes finite automaton, one can associate a linear system whose solution is $\Sigma\{ \tilde{f}\, x_o\, f' \}$ where the summation is over all accepted pairs (f, f') of words. An especially simple case is the following.

Let us consider the equation

$$y = x_o + \Sigma\{ \widetilde{\alpha' x}\, y\, \alpha x : x \in X \}$$

(with α, α' as above) whose solution is

$$a'' = x_o + \Sigma\{ \alpha'\, \tilde{f}\, x_o\, \alpha f : f \in F(X) \}$$

Let a''_o be the special case corresponding to $\alpha = \alpha' =$ the iden-

tity mapping. The support of a''_0 is the so-called *mirror-image language* of Chomsky.[8] Again $a'' + a''_0$ has at least one coefficient larger than one if and only if the Post's problem for α and α' has a solution

Assume now that α and α' are homomorphisms into the submonoid F' of F generated by $X' \subset X$ and that

1) All the words αx have the same fixed degree d.

2) For all $x, x' \in X$ if αx and $\alpha x'$ belong to the same proper left ideal of F then $\alpha' x = \alpha' x'$.

With this hypothesis, for any fixed word $\bar{f}' \in F'$, the problem of determining if the supports of a'' and $a''_0\bar{f}$ have a non-empty intersection is precisely the so-called "Tag problem."[29]

Many other unsolvability properties have been established by Bar Hillel, Perles and Shamir[1] using these constructions, and the remark that when α (or α') is a monomorphism of $F(X)$ into $F(X')$ (that is, when, e.g., the set $\{\alpha' x : x \in X\}$ is a set of *code words* having the *unique decipherability* property), then the generating function of the complement of supp. a'' in $F(X')$ is also the solution of a linear system.

REFERENCES

1. Y. Bar-Hillel, M. Perles, and E. Shamir, "On Formal Properties of Simple Phrase Structure Grammars," Tech. Report 4, Applied Logic Branch, Hebrew University of Jerusalem (July 1960).

2. Y. Bar-Hillel and E. Shamir, "Finite-State Languages," *Bull. Res. Council Israel*, Vol. 8F, pp. 155-166 (1960).

3. R. Birkeland, "Sur la Convergence des Devellopements qui Expriment des Racines de l' Equation Algebrique Generale...," *C.R. Acad. Sciences*, Vol. 171, pp. 1370-1372 (1920); Vol. 172, pp. 309-311 (1921).

4. S. Bochner and W.T. Martin, "Singularities of Composite Functions in Several Variables," *Ann. Math.*, Vol. 38, pp. 293-302 (1938).

5. A.W. Burks and H. Wang, "The Logic of Automata," *J. Assoc. Comp. Mach.*, Vol. 4, pp. 193-218, 279-297 (1957).

6. A.W. Burks, D.W. Warren, and J.B. Wright, "An Analysis of a Logical Machine Using Parenthesis-Free Notation," *Math. Tables and Other Aids to Computation*, Vol. 8, pp. 53-57 (1954).

7. R.H. Cameron and W.T. Martin, "Analytic Continuation of Diagonals," *Trans. Am. Math. Soc.*, Vol. 44, pp. 1-7 (1938).

8. N. Chomsky, "On Certain Formal Properties of Grammar," *Information and Control*, Vol. 2, pp. 137-167 (1959).

9. N. Chomsky, "A Note on Phrase Structure Grammars, *Information and Control*, Vol. 2, pp. 393-395 (1959).

10. N. Chomsky, "Context Free Languages and Push-Down Storage," Quart. Prog. Rep. No. 65, Research Laboratory of Electronics, M.I.T. (1962).

11. N. Chomsky and G.A. Miller, "Finite State Languages," *Information and Control,* Vol. 1, pp. 91-112 (1958).

12. N. Chomsky and G.A. Miller, "Mathematical Models of Language," in Luce, Bush and Galanter (Eds.) *Handbook of Mathematical Psychology*

13. N. Chomsky and M.P. Schutzenberger, "The Algebraic Theory of Context-Free Languages," in *Computer Programming and Formal Systems,* P. Braffort and D. Hirschberg, Eds. ["Studies in Logic Series"] (Amsterdam: North Holland Pub. Co.).

14. C.C. Elgot, "Decision Problems of Finite Automata Design and Related Arithmetics," *Trans. Am. Math. Soc.,* Vol. 48, pp. 21-51 (1961).

15. E. Fabry, "Sur les Points Singuliers d'une Fonction Donnee par Son Developpement de Taylor," Ann. Ecole Normale Sup. Paris (3), Vol. 13, pp. 367-399 (1896).

16. P. Fatou, "Sur les Series Entieres a Coefficients Entiers," *C.R. Acad. Sciences,* Vol. 138, pp. 342-344 (1904).

17. S. Ginsburg, "Some Remarks on Abstract Machines," *Trans. Am. Math. Soc.,* Vol. 96, pp. 400-444 (1960).

18. S. Ginsburg and H.G. Rice, "Two Families of Languages Related to Algol," Technical Memo., System Development Corporation, Santa Monica, Calif. (1961).

19. S. Ginsburg and G.F. Rose, "Operations which Preserve Definability in Languages, Technical Memo., System Development Corporation, Santa Monica, Calif. (1961).

20. J. Giveon, "Boolean Matrices and Their Application to Finite Automata," Technical Report No. 5, Applied Logic Branch, The Hebrew University of Jerusalem (September 1960).

21. U.S. Haslam-Jones, "An Extension of Hadamard Multiplication Theorem," *Proc. London Math Soc.,* Vol. II, Ser. 27, pp. 223-232 (1928).

22. D.A. Huffman, "Information Lossless Finite State Automata," *Nuevo Cimento Supp.,* Vol. 13, pp. 397-405 (1959).

23. R. Jungen, "Sur les Series de Taylor n'Ayant que des Singularites Algebrico-Logarithmiques sur Leur Cercle de Convergence," *Comm. Math. Helvetici,* Vol. 3, pp. 236-306 (1931).

24. M. Kesten, "Symmetric Random Walks on Groups," *Trans Am. Math. Soc.,* Vol. 92, pp. 336-354 (1954).

25. S.C. Kleene, "Representation of Events in Nerve Nets and Finite Automata," in *Automata Studies* (Princeton, N.J.: Princeton Univ. Press, 1956), pp. 3-41.

26. C.Y. Lee, "Automata and Finite Automata," *Bell Syst Tech J.,* Vol. 39, pp. 1267-1296 (1960).

27. K. Mahler, "On a Theorem of Liouville in Fields of Positive Characteristic," *Canadian J Math.,* Vol. 1, pp. 397-400 (1949).

28. M.A. Markov, "Ob odnoi nervazresimoi probleme," *Doklady Akad. Nauk: n s.,* Vol. 78, pp. 1089-1092 (1951).

29. M.L. Minsky, "Recursive Unsolvability of Post.s Problem of Tag," *Ann. Math.,* Vol. 74, pp. 437-455 (1961).

30. A. Nerode, "Linear Automaton Transformation," *Proc. Am. Math Soc.,* Vol. 9, pp. 541-544 (1958).

31. A.G. Oettinger, "Automatic Syntactic Analysis and the Pushdown Store," in *Proc. Symposia in Applied Math,* Vol. 12: *Structure of Language and Its Mathematical Aspects* (Am. Math. Soc., 1961), pp. 104-124.

32. R.J. Parikh, "Language Generating Devices," Quart. Prog. Report No. 60., Research Laboratory of Electronics, M.I.T., pp. 199-212 (January 1961).

33. E. Post, "A Variant of a Recursively Unsolvable Problem," *Bull. Am. Math. Soc.,* Vol. 52, pp. 264-268 (1946).

34. M.O. Rabin and D. Scott, "Finite Automata and Their Decision Problems," *I.B.M. J. Res.,* Vol. 3, pp. 115-125 (1959).

35. G.N. Raney, "Functional Composition Patterns and Power-Series Reversion," *Trans Am. Math. Soc.,* Vol. 94, pp. 441-451 (1960).

36. L. Redei, "Die Verallgemeinerung der Schreierschen Erweiterungs Theorie," *Act. Sci. Math., Szeyed,* Vol. 14, pp. 252-273 (1952).

37. K. Samelson and F.L. Bauer, "Sequential Formula Translation," *Comm. Assoc. Comp. Mach.* Vol. 3, pp. 76-83 (1960).

38. S. Scheinberg, "Note on the Boolean Properties of Context Free Languages," *Information and Control,* Vol. 31, pp. 372-375 (1960).

39. S. Schottlander, "Der Hadamardsche Multiplicationsatz und weitere Kompositionsatze der Functionentheorie," *Math. Nachr.,* Vol. 11, pp. 239-294 (1954).

40. M.P. Schutzenberger, "Some Remarks on Chomsky's Context Free Languages," Quart. Prog. Report No. 63, Research Laboratory of Electronics, M.I.T. (October 1961).

41. M.P. Schutzenberger, "On Context-Free Languages and Push-Down Automata," to appear in *Information and Control,* 1963.

42. J.C. Shepherdson, "The Reduction of Two-Way Automata to One-Way Automata," *IBM J.,* Vol. 3, pp. 198-299 (1959).

43. T. Skolem, *Comptes Rendus* (8-eme Congres des Math. Scandinaves, Stockholm, 1934), pp. 163-170.

44. G. Szego, "Uber Potenzreihen mit Endlich Vielen Verschiedenen Koeffigenten," *Sitzber. Preuss. Akad. Wiss., Math. Phys. Kl.,* pp. 88-91 (1922).

Reprinted from INFORMATION AND CONTROL, Volume 6, No. 3, September 1963

INFORMATION AND CONTROL **6,** 246–264 (1963)

On Context-Free Languages and Push-Down Automata

M. P. SCHÜTZENBERGER

International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York

This note describes a special type of one-way, one-tape automata in the sense of Rabin and Scott that idealizes some of the elementary formal features used in the so-called "push-down store" programming techniques. It is verified that the sets of words accepted by these automata form a proper subset of the family of the *unambiguous context-free languages* of Chomsky's and that this property admits a weak converse.

INTRODUCTION

This note is concerned with some relations observed by N. Chomsky and myself between *context-free languages* and what will be called *push-down automata*.[1]

Informally, a push-down automaton is a special type of *one-way*, *one-tape* automaton in the sense of Rabin and Scott in which the memory is a (potentially infinite) tape, used in a certain restricted manner. For each successive letter of the input word, the word stored on the tape is modified by deletion or adjunction at its right end. This is done under the control of a finite state device which can scan the word stored on the tape and carry some additional information. The input word is accepted iff after reading its last letter both the word written on the tape and the state of the finite part belong to a *finite* prescribed set.

This operation appears as an abstraction of some elementary features of the programming technique known as "*push-down store*" (Newell and Shaw, 1957).

I am indebted to C. C. Elgot for enlightening discussions which have lead to the clarification of many points and to a definition of push-down automata which may be less unrealistic than the ones I had previously considered.

[1] *Note added in proof:* Our definition does not coincide with the one introduced by Shepherdson and Sturgis (1963).

We recall that a context-free language on an alphabet X is a subset L of the set F of all words in this alphabet that can be obtained by the following procedure, which is a special type of a Post production:

Let $\Xi = \{\xi_j\}(1 \leqq j \leqq n)$ be another set of letters and H be the set of all words in the letters of $X \cup \Xi$. In Chomsky's terminology Ξ is the *nonterminal* alphabet. A *grammar* is an assignment to each $\xi_j \in \Xi$ of a finite set p_j of words of H that does not contain the empty word e or any word consisting of a single letter of Ξ. Both X and Ξ are finite.

Let L_1 be the least subset of H that contains ξ_1 and every word $h'hh''$ $(h, h', h'' \in H)$ if it contains $h'\xi_j h''$ $(1 \leqq j \leqq n)$ and if $h \in p_j$. Then by definition $L = L_1 \cap F$ is the context-free language produced by the grammar $\{p_j\}$.

In the first part of the paper we verify the equivalence of this definition with another one which relates context-free languages with algebraic formal power series in noncommuting variables. The treatment given here is more elementary than that of Ginsburg and Rice (1962) and the notation introduced in this part is needed later on. Furthermore, except for trivial changes, the main bulk of the notation carries over to the still simpler case of the power series.

In the second part of the paper we verify that the set of words accepted by any push-down automaton (as defined here!) is a context-free language and, as an example, we consider the simplest nondegenerate type of such automata. The corresponding languages are called "*standard context-free languages.*"

In the fourth part we verify a weak converse of the first property. More explicitly, let F and F' be the set of all words in the letters of the alphabets X and X' (i.e., the free monoids generated by these sets). A homomorphism (of monoid) $\theta: F \to F'$ is any mapping from F to F' given by a mapping θ_1 from X to F' and the rule that $\theta e = e$ ($e =$ the empty word) and that for all $f = x_{i_1}x_{i_2} \cdots x_{i_m} \in F$, $(m > 0)$, $\theta f = \theta_1 x_{i_1} \theta_1 x_{i_2} \cdots \theta_1 x_{i_m}$.

We verify that for any context-free language $L' \subset F'$ one can find a set X, a homomorphism $\theta: F \to F'$ and a standard context-free language $L \subset F$ such that $L' = \theta L$ $(= \{\theta f: f \in L'\})$.

A closer connection between push-down devices and context-free languages is to be found in (Chomsky, 1962).

I. DEFINITIONS

Let X, Ξ, F, H be as in the introduction. $\mathfrak{P}(H)$ is the collection of all subsets of H. We shall consider $\mathfrak{P}(H)$ as an algebraic system (in

fact, a semiring) with respect to the usual operations of set union and set product (the intersection is not used).

We shall reserve the notation π_m (resp. π_F) for the projection of $\mathfrak{P}(H)$ onto the union of the elements of H of length $\leqq m$ (resp. onto $\mathfrak{P}(F)$). Thus, for $H' \in \mathfrak{P}(H)$, $\pi_0 H' = \phi$ iff $e \notin H'$ and we say then, as usual, that H' is *quasi regular*.

For any n-tuple $q = (q_1, q_2, \cdots, q_n)$ of elements of $\mathfrak{P}(H)$, we denote by λ_q the homomorphism (of semiring) of $\mathfrak{P}(H)$ induced by the substitution $\{\xi_j\} \to q_j$ for $1 \leqq j \leqq n$.

Now let L be a (nonempty) context-free language with grammar $p = (p_1, p_2, \cdots, p_n)$ and nonterminal alphabet Ξ. We can assume that no p_j is empty and we recall that by hypothesis (1) p is quasi regular (i.e., $\pi_0 p_j = \phi$ for each j); (2) no p_j contains a word consisting of a single nonterminal letter.

Let $\mathcal{P}$ (resp. $\mathcal{P}_F$) be the set of all quasi-regular n-tuples of subsets of H (resp. of F). For any mapping ϕ of $\mathfrak{P}(H)$ and $r \in \mathcal{P}$, ϕr denotes the n-tuple $(\phi r_1, \phi r_2, \cdots, \phi r_n)$.

Now let q be any element of $\mathcal{P}$. The conditions (1) and (2) on p imply the following identity or all $m \geqq 0$:

$$(*) \qquad \pi_{m+1}\lambda_q p = \pi_{m+1}\lambda_{\pi_m q} p.$$

Indeed, by the very definition of λ, for each j and m, $m' \geqq 0$, $\lambda_{\pi_m q} p_j \subset \lambda_{\pi_{m+m'} q} p_j$ and so, with obvious notations $\pi_{m+1}\lambda_q p \supset \pi_{m+1}\lambda_{\pi_m q} p$.

On the other hand, any word f of length $|f| \geqq m + 1$ belonging to some component of q can only intervene in $\lambda_q p$ by being substituted for some $\xi \in \Xi$ which (by condition (2)) is a factor of a word of length at least two. Thus the word resulting from this substitution has length at least $|f| + 1 \geqq m + 2$, and it disappears after the application of of π_{m+1}.

Similar reasoning shows that $\pi_F \lambda_q p = \lambda_{\pi_F q} p$.

Let us consider the sequence $p(m)$ $(m \geqq 0)$ of elements of $\mathcal{P}$ defined inductively by $p(0) = (\phi)(= (\phi, \phi, \cdots, \phi))$, and, for each $m \geqq 0$, $p(m + 1) = \lambda_{p(m)} p$.

We prove that, for each m', $m'' \geqq 0$, $\pi_{m'} p(m') = \pi_{m'} p(m' + m'')$ and $p(m') \in \mathcal{P}_F$.

The relations are trivial for $m' \leqq 1$. Assume that they hold for $m' \leqq m$. Then: $\pi_{m+1} p(m + 1) = \pi_{m+1}\lambda_{p(m)} p$ (by definition) $= \pi_{m+1}\lambda_{\pi_m p(m)} p$ (by $(*)$) $= \pi_{m+1}\lambda_{\pi_m p(m+m')} p$ (by the induction hypothesis) $= \pi_{m+1}\lambda_{p(m+m')} p$

(by (*)) $= \pi_{m+1}p(m + m'' + 1)$ (by definition). This gives the first relation, and $p(m+1) \in \mathcal{P}_F$ follows by induction from $p(0) \in \mathcal{P}_F$.

Hence the limit for $m \to \infty$ of $p(m)$ is a well defined quasi-regular n-tuple of subsets of F which we denote by $p(\infty)$ and which satisfies the identity $p(\infty) = \lambda_{p(\infty)}p$. Let us verify that any $p' \in \mathcal{P}$ which satisfies $p' = \lambda_{p'}p$ is equal to $p(\infty)$. Indeed, if $\pi_m p' = \pi_m p(\infty)$ for some $m \geqq 0$, the same relation holds for $m + 1$ since $\pi_{m+1}p' = \pi_{m+1}\lambda_{\pi_m p'}p = \pi_{m+1}\lambda_{\pi_m p(\infty)}p = \pi_{m+1}p(\infty)$. Since the hypothesis of quasi regularity on p' implies $\pi_0 p' = \pi_0 p(\infty)$, the verification is completed.

It remains to show that $L = \pi_F L_1$ is equal to the first component of $p(\infty)$.

For this, let $\lambda_p{}^0 p = p$ and, inductively, $\lambda_p^{m+1}p = \lambda_p{}^m(\lambda_p p)$. The classical identity $\lambda_q\lambda_{q'} = \lambda_{\lambda_q q'}$ (valid for any q, $q' \in \mathcal{P}$) shows that, for all m, $\pi_{m+1}p(m + 1) = \pi_{m+1}\lambda_p{}^m p$. Thus $p(\infty)$ can also be defined as $\lim_{m\to\infty} \lambda_p{}^m p$.

Consider now $p' = p \cup (\xi)$ where (ξ) is the n-tuple $(\{\xi_1\}, \{\xi_2\}, \cdots, \{\xi_n\})$ and where $\cup$ is to be performed component-wise. By definition the relation

$\pi_F\lambda_{p'}^{m'+1}p' = \pi_F\lambda_p^{m'+1}p$ is true for $m' = 0$; if it is true for $m' = m$, it is still true for $m' = m + 1$ because, setting $p''' = \lambda_{p'}^m p'$ and $p'' = \lambda_p{}^m p$,

$$\pi_F\lambda_{p'}^{m+1}p' = \pi_F(\lambda_{p'}^m(\lambda_{p'}p')) = \pi_F\lambda_{p'''}p' = \lambda_{\pi_F p'''}p$$

$$= \lambda_{\pi_F p''}p = \pi_F\lambda_{p''}p = \pi_F\lambda_p^{m+1}p.$$

Hence $\lim_{m\to\infty} \pi_F\lambda_{p'}^m p' = p(\infty)$. Since our original definition amounts to the definition of L_1 as the first component of $\lim_{m\to\infty} \lambda_{p'}^m p'$ the result is entirely verified.

For the sake of completeness we recall the proof of the following theorem which is needed later on. To simplify notation, $+$ and Σ are used instead of $\cup$; $\mathfrak{R}(F)$ is the set of all *regular events* on F; for each $q \in \mathfrak{P}(F)$, $q^* = \Sigma\{q^n : n > 0\}$.

THEOREM (Chomsky and Miller, 1956). *If for all j $(1 \leqq j \leqq n)$*

$$p_j = q_{j,0} + \Sigma\{\xi_{j'}q_{j,j'} : 1 \leqq j' \leqq n\}$$

with $q_{j,j'} \in \mathfrak{R}(F)$ for all j, j' $(1 \leqq j \leqq n)$, $(0 \leqq j' \leqq n)$, then every component of $p(\infty)$ belongs to $\mathfrak{R}(F)$.

PROOF: The result follows by induction on n from the unicity property of $p(\infty)$ mentioned above. Indeed for $n = 1$, $p' = q_{1,0}q_{1,1}^*$ satisfies $\lambda_{p'}p = p'$ and thus $p(\infty) = p' \in \mathfrak{R}(F)$. Assume now the result proved

for $n < n'$ and let (p_j') $(1 \leqq j \leqq n)$ be defined by

$$p_n' = q_{n,0}q_{n,n}^* + \Sigma\{\xi_j\ ,\ q_{n,j'}q_{n,n}^* : 1 \leqq j' < n\}$$

and for $1 \leqq j < n$, p_j' obtained by substituting in p_j the right member of this last equation for ξ_n .

The hypothesis $p(\infty) = \lambda_{p(\infty)}p$ shows that $p'(\infty) = p(\infty)$. However, the grammar (p_j') $(1 \leqq j < n)$ has only $n - 1$ nonterminal letters and the result follows from the induction hypothesis.

II. PUSH-DOWN AUTOMATA

In all this section $X = \{x\}$ and $Y = \{y\}$ denote respectively the input alphabet and the internal alphabet (used for writing on the tape). The corresponding sets of words are F and G. It can be proved that there would be no loss in generality in *card* X = *card* $Y = 2$; it is not so for card $Y = 1$.

If $\mathfrak{R}$ is a finite automaton with input alphabet Y, we denote by χg the state reached after reading the word $g \in G$, the initial state being fixed. A standard argument shows that there is no loss of generality (for our present purpose) in taking χ as a *finite homomorphism*—that is, as a mapping χ of G onto a finite monoid K such that for all g, g', g'', g''' the relation $\chi g = \chi g'$ implies $\chi g''gg''' = \chi g''g'g'''$.

For any $K' \in \mathfrak{P}(K)$ (the set of all subsets of K) and $g \in G$, $\rho_{K'}g$ denotes the longest left factor g' of g such that $g = g'g''$ with $\chi g'' \in K' \cup \{\chi g\}$.

With this notation, an elementary χ-push-down mapping $\mu: G \to G$ is given by:

(1) a finite homomorphism $\chi: G \to K$,

(2) a mapping $\alpha: K \to G$;

(3) a mapping $\bar{p}: K \to \mathfrak{P}(K)$.

For each $g \in G$, $\mu g = \rho_{\bar{p}\chi g}(g \cdot \alpha\chi g)$. In more concrete manner, one may think of a device $\mathfrak{R}$ which performs the following cycle of operations:

(1′) $\mathfrak{R}$ reads g and determines $\chi g \in K$;

(2′) $\mathfrak{R}$ writes the word $\alpha\chi g \in G$ to the right of the word g;

(3′) $\mathfrak{R}$ reads from right to left the word $g \cdot \alpha\chi g$ erasing successively each letter (eventually none) till it reaches a state belonging to the subset $\bar{p}\chi g$ of K or until it has erased the whole word $g \cdot \alpha\chi g$.

The word left is μg. Thus, either g is a left factor of μg (in symbols $g \mid \mu g$) or $\mu g \mid g$; we shall refer to this situation by saying that g and μg are *comparable*.

Now let S be a finite set in which a subset $\bar{S}$ and an element $s_\infty \in \overline{\bar{U}}$ have been distinguished. We denote by U (resp. $\bar{U}$, resp. $\bar{U}_\infty$) the set of all pairs (s, g) with $s \in S$ (resp. $s \in \bar{S}$, resp. $s = s_\infty$) and $g \in G$. We call *states* the elements of U. It will be understood that for $u, u' \in U$, the notation $u \mid u'$ (resp. $u \nmid u'$) means that $u = (s, g)$ $u' = (s', g')$ with $g \mid g'$ (resp. with *not* $g \mid g'$).

DEFINITION 1. A χ-push-down mapping $\mu: U \to U$ is given by:

(1) A mapping $\sigma: (S, K) \to S$ such that for all $k \in K$, $\sigma(s_\infty, k) = s_\infty$;

(2) For each $s \in S$ an elementary χ-push-down mapping μ_s on G with the restriction that for all $u \in \bar{U}_\infty$, $\mu_{s_\infty} u = u$.

For each $u = (s, g) \in U$, $\mu u = (\sigma(s, \chi g), \mu_s g)$. For simplicity, we shall rather deal with the mapping μ^* defined as follows.

Let $\mu^1 u = \mu u$ and for each $i > 0$, $\mu^{i+1} u = \mu(\mu^i u)$. The largest positive i (possibly infinite) such that, for all positive $i' < i$, $\mu^{i'} u \notin \bar{U}$ will be denoted by $j(u)$. Then

$$\mu^* u = \mu^{j(u)} u \quad \text{if} \quad j(u) < \infty;$$

$$= (s_\infty, g) \quad \text{if} \quad j(u) = \infty \quad \text{and} \quad u = (s, g).$$

Thus $\mu^* = \mu$ iff for all $u \in U$, $\mu u \in \bar{U}$.

DEFINITION 2. A *push-down automaton* $\mathfrak{A}$ is given by:

1) The finite alphabets X and Y;

2) A finite homomorphism $\chi: G \to K$;

3) A finite set S (with $\bar{S}$, s_∞ and U as above) and a χ-push-down mapping $\mu: U \to U$;

4) A mapping $\beta: (\bar{S}, X) \to S$ with $\beta(s_\infty, x) = s_\infty$ identically;

5) An *initial state* $u_0 \in \bar{U} \backslash \bar{U}_\infty$ (i.e., $u_0 \in \bar{U}$ and $u_0 \notin \bar{U}_\infty$); a finite set $\bar{U}_{fin} \subset \bar{U} \backslash \bar{U}_\infty$ of *final states*.

For each $u = (s, g) \in U$ and $x \in X$, the "next state" $u \cdot x$ is $\mu^*(\beta(s, x), g)$. For each $f \in F$, $u \cdot f = u$ if f is the empty word and $u \cdot f = (u \cdot f') \cdot x$ if $f = f'x$ ($f' \in F$, $x \in X$). The set *Acc* $\mathfrak{A}$ of the words accepted by $\mathfrak{A}$ is *Acc* $\mathfrak{A} = \{f \in F: u_0 \cdot f \in \bar{U}_{fin}\}$.

Finally $\mathfrak{A}$ is *simple* iff $\bar{S} = S$, and then, clearly, $\mu = \mu^*$.

According to *Definition* 2 the cycle of the automaton that is initiated by each input letter consists of two successive operations: the mapping $\beta: (\bar{S}, X) \to S$ and the mapping $\mu^*: U \to \bar{U}$; further, any state $u \in \bar{U}_\infty$ is a *sink*, i.e. for all $u \in \bar{U}_\infty$ and $f \in F$, one has $u \cdot f = u$. Intuitively, one might think of a device $\mathfrak{A}_0$ acting in the following manner for each state $(s, g) \in \bar{U}$ and incoming input letter $x \in X$:

$\mathfrak{A}_0$ goes first to the state $u = (\beta(s, x), g)$ and it performs the push-down mapping $\mu_{\beta(s,x)}$ which brings it to $u' = (s', g')$, say. If s' belongs to $\bar{S}$ the cycle initiated by x is already completed and $\mathfrak{A}_0$ reads the next input letter (this is always the case when $\mathfrak{A}$ is simple). If s' does not belong to $\bar{S}$, $\mathfrak{A}_0$ performs $\mu_{s'}$ and goes to $u'' = \mu_{s'}u'$. Again, if $u'' \in \bar{U}$ the cycle is completed; if it is not so, $\mathfrak{A}_0$ goes on performing a succession of push down mappings till, eventually, it reaches a state of $\bar{U}$. Clearly, in the general case, $\mathfrak{A}_0$ may never reach this subset and consequently it may happen that $\mathfrak{A}_0$ does not read the input word further than x. Our more formal definition of $\mathfrak{A}$ by the mapping μ^* is intended to obviate this minor notational difficulty.

PROPERTY 1. *For any push-down automaton $\mathfrak{A}$ there exists a simple push-down automaton $\mathfrak{A}'$ such that Acc $\mathfrak{A}$ = Acc $\mathfrak{A}'$.*

PROOF: The property amounts to the statement that for any given push-down mapping $\mu: U \to U$ there exists a finite set S', a surjection (i.e., mapping onto) $\zeta: S' \to S$ and a push-down mapping $\mu: (S', G') \to (S', G')$ that have the following properties:

(i) $\mu'^* = \mu'$;

(ii) for all $g \in G$, $s' \in S'$, the relation $\mu'(s', g) = (s'', g')$ implies $\mu^*(\zeta s', g) = (\zeta s'', g')$.

To simplify notation we first verify that there is no loss in generality in assuming that μ satisfies the conditions (1), (2), and (3) below. For $u = (s, g) \in U$, we write $\chi u = (s, \chi g)$ and $\chi U = \{(s, k): s \in S, k \in K\}$; $\chi \bar{U} = \{(s, k): s \in \bar{S}, k \in K\}$.

CONDITION (1). There corresponds to each $(s, k) \in \chi U$ a subset $K(s, k)$ of K (eventually empty) such that $g \in \chi^{-1}K(s, k)$ (i.e., $\chi g \in K(s, k)$) iff for all $g' \in G$ one has $\rho_{\bar{p}(s,k)}g'g\alpha(s, k) = g'$ (where $\bar{p}(s, k)$ denotes the subset of K defined by the function $\bar{p}$ associated with the elementary push-down mapping μ_s).

Indeed let the quotient monoid K_1 of G and the epimorphism (= homomorphism onto) $\chi_1 : G \to K_1$ be defined by the condition that, for any $g, g' \in G$, $\chi_1 g = \chi_1 g'$ iff for all $K', K'' \subset K$ one has:

$$\rho_{K'}g \mid \rho_{K''}g \text{ is equivalent to } \rho_{K'}g' \mid \rho_{K''}g'.$$

Clearly K_1 is finite and there exists a homomorphism $\bar{\chi}: K_1 \to K$ such that $\bar{\chi}\chi_1 = \chi$. Hence, any of the χ-push-down mappings entering in the definition of $\mathfrak{A}$ could have been defined as well as a χ_1-push-down mapping. Thus we can assume that χ has been replaced by χ_1 in the definition of $\mathfrak{A}$ and condition (1) is trivially satisfied.

CONDITION (2). For each $u \in U$, if $u \mid \mu u$ then $\mu u \in \bar{U}$.

Indeed let $j^+(u)$ denote the largest number i (possibly infinite) such that, for all positive $i' < i$, one has $\mu^{i'}u \notin \bar{U}$ and $u \mid \mu^{i'}u$. By construction $j^+(u) \leqq j(u)$.

Let $\mu' : U \to U$ be defined by:

$$\mu' u = \mu^{j^+(u)} u \text{ if } j^+(u) \text{ if finite;}$$

$$= \mu^* u \text{ if } j^+(u) \text{ is infinite.}$$

Clearly $\mu'^* = \mu^*$ and it suffices to verify that μ' is a χ-push-down mapping since, by construction, μ' satisfies Condition (2).

Consider $u = (s, g)$ such that $j^+(u) > 1$. Thus if $j' < j^+(u)$, the state $\mu^{j'}u$ has the form $(s', g\bar{g})$ for a certain $\bar{g} \in G$. Induction on j' shows easily that if $\chi g = \chi g'$ then $\mu^{j'}(s, g') = (s', g'\bar{g})$ with the same s' and the same $\bar{g}$. Hence $j^+(u)$ is a function of χu only.

It follows that the mapping $\mu'' : U \to U$, defined for all $u \in U$ by:

$$\mu'' u = \mu u \text{ if } j^+(u) \text{ is finite;}$$

$$\mu'' u = \mu^* u \text{ if } j^+(u) \text{ is infinite,}$$

is a χ-push-down mapping and, since $\mu''^* = \mu^*$, we can assume from now on that $\mu = \mu''$, i.e., that, for all $u \in U$, $j^+(u)$ is finite.

Then, as we have seen, for each $u = (s, g) \in U$, the state $\mu' u$ $(= \mu^{j^+(u)}u)$ has the form $(s', \rho_{K'} g\bar{g})$ where $s' \in S$, $K' \subset K$, and $\bar{g} \in G$ are functions of $\chi u \in \chi U$ only, and the verification is completed.

We point out that Condition (2) implies that, for each $(s, g) \in U$, the number $j(s, g) - 1$ is at most equal to the length $|g|$ of g. Indeed, if $\mu(s, g) \notin \bar{U}$, the state $(s', g') = \mu(s, g)$ is such that g is not a left factor of g' and, since g and g' are comparable, this implies $g' \mid g$ and $g' \neq g$, hence $|g'| < |g|$.

CONDITION (3). There exists a χ-push-down mapping $\bar{\mu}$ such that $\mu^* = \mu\bar{\mu}$.

We assume (1) and (2). Let $U_1 = \{u : j(u) = 1\} = \{u : \mu u \in \bar{U}\} = \{u : \mu u = \mu^* u\}$. By the definition of a χ-push-down mapping there exists a subset χU_1 of χU such that $U_1 = \{u : \chi u \in \chi U_1\}$. When $u = (s, g) \notin U_1$, we have seen above that the state $(s', g') = \mu(s, g)$ is such that $g' \mid g$. Hence, introducing a new mapping $\bar{\mu} : U \to U_1$ by the rule that for all $u \in U$:

$$\bar{\mu} u = u \quad \text{if} \quad u \in U_1,$$

$$\bar{\mu} u = \mu^{j(u)-1} u \quad \text{if} \quad u \notin U_1,$$

we have identically $\mu^* = \mu\bar{\mu}$ and the verification of (3) amounts to the verification that $\bar{\mu}$ can be defined as a χ'-push-down mapping for a suitable finite homomorphism χ'.

For showing this, let V denote the set of all quadruples $v = (s, k, s', k')$ with $s, s' \in S, k, k' \in K$. $V_1 \subset V$ is defined by the restriction $(s, k) \in \chi U_1$.

For $v = (s, k, s', k') \in V_1$ (resp. $\in V$) and $n > 0$, we define $\tilde{G}_n(v)$ (resp. $G(v)$) as the set of all $g' \in G$ satisfying $k\chi g' = k'$ which are such that for some $g \in \chi^{-1}k$ one has $j(s', gg') = n + 1$ and $\mu^n(s', gg') = (s, g)$ (resp. $j(s, gg')$ arbitrary and $\mu(s', gg') = (s, g)$). Thus for $v \in V_1$, $\tilde{G}_1(v) = G(v)$.

Because of the definition of $\bar{\mu}$ it is easily seen that if $g' \in G_n(v)$ (resp. $\in G(v)$) then for all $g \in \chi^{-1}k$ one has $j(s', gg') = n + 1$ and $\mu^n(s, gg') = (s, g)$ (resp. $\mu(s', gg') = (s, g)$).

For $v \in V_1$ ($v = (s, k, s', k')$) let $\tilde{G}(v)$ be the union of the sets $\tilde{G}_n(v)$ over all positive n; also let $(v', v'') \in \omega(v)$ mean that there exist $(s'', k'') \in \chi\bar{U}$ ($= \{\chi u : u \in \bar{U}\}$) such that $v' = (s, k, s'', k'')$ (hence $v' \in V_1$) and $v'' = (s'', k'', s', k')$. Thus, by construction, $G(v) \subset \tilde{G}(v)$ and, for $n > 0$, $g' \in \tilde{G}_{n+1}(v)$ iff there exist $(v', v'') \in \omega(v)$ and a factorization $g' = g''g'''$ such that $g'' \in \tilde{G}_n(v')$ and $g''' \in G(v'')$.

Introduce now a set $\{\xi_v\}$ of new letters indexed by the elements of V_1 and, for each $v \in V_1$, define

$$p_v = G(v) + \Sigma\{\xi_{v'}G(v'') : (v', v'') \in \omega(v)\}.$$

The system $(p) = (p_v)_{v \in V_1}$ defines a grammar and, by construction, $p_v(\infty) = \tilde{G}(v)$ for each $v \in V_1$.

However, for arbitrary $v = (s, k, s', k')$, the condition that a word g' belongs to $G(v)$ can be explicitly stated as:

$$s = \sigma(s', k'); \qquad k' = k\chi g'; \qquad \text{for all} \quad g \in \chi^{-1}k, \quad g = \rho_{\tilde{p}(s,k)}gg'\alpha(s, k).$$

Thus, because of Condition (1), each of the sets $G(v)$ ($v \in V_1$) is a regular event. By the theorem recalled at the end of the first section, it follows that each of the sets $\tilde{G}(v)$ ($v \in V_1$) is also a regular event. In other terms, there exists a homomorphism χ' of G onto a finite monoid K' and for each $v \in V_1$ a subset $\bar{K}'(v)$ of K' such that $G(v) = \chi'^{-1}\bar{K}'(v)$.

Since χ' may be chosen so that K' admits K as a homomorphic image, we may argue as in the verification of Condition (1) that in fact $\chi = \chi'$. Under this hypothesis it is immediate that $\bar{\mu}$ can be defined as a χ-push-down mapping and Condition (3) is verified.

The rest of the proof (i.e., that under the conditions (1), (2), and (3),

the mapping $\mu^* = \mu\bar{u}$ can be defined as a push-down mapping) is trivial and it is omitted.

PROPERTY 2. *For any push-down automaton* $\mathfrak{A}$ *the set Acc* $\mathfrak{A}$ *is a context-free language.*

PROOF: By Property 1 we can assume that $\mathfrak{A}$ is simple. Hence for each $s \in S$, $x \in X$, $g \in G$, the operation performed by $\mathfrak{A}$ consists of a transition $s \to s{\cdot}x \in S$ and of an elementary push-down mapping $\mu_{s,x}$ on G.

For simplicity we shall speak of the states as if they consisted only of a word. Thus the *length* $|u|$ of $u = (s, g)$ is the length $|g|$ of g; u is a *left factor* of $u' = (s'\ g')$ if g is a left factor of g' etc.

Let $f = x_{i_1}x_{i_2} \cdots x_{i_m}$ be an input word of length $|f| = m \geqq 2$. For any state u we consider the $m - 1$ *intermediate states* $u_1 = u{\cdot}x_{i_1}$, $u_2 = u{\cdot}x_{i_1}x_{i_2}$, $\cdots$, $u_{m-1} = u{\cdot}x_{i_1}x_{i_2} \cdots x_{i_{m-1}}$ and we define $\min(u; f)$ to be the minimum of their length. If u_j $(1 \leqq j \leqq m - 1)$ is such that $|u_j| = \min(u{\cdot}f)$ and if further either $j = m - 1$ or $|u_{j'}| > |u_j|$ for $j < j' \leqq m - 1$, we call u_j the *critical state* (of f at u); $f' = x_{i_1}x_{i_2} \cdots x_{i_j}$ and $f'' = x_{i_{j+1}}x_{i_{j+2}} \cdots x_{i_m}$ are the *critical factors* (of f at u). Clearly the critical state always exists (when $f \notin X$) and it is uniquely determined.

REMARK 1. *The critical state* u_j *of* f *at* u *is a left factor of all the intermediate states and it is comparable with both* u *and* $u' = u{\cdot}f$.

PROOF: The statement is trivial if f has length two, i.e., if $f = x_{i_1}x_{i_2}$, because there is only one intermediate state, viz. $u_1 = u{\cdot}x_{i_1}$, which, by force, is the critical one. The fact that u and u_1 and u_1 and $u' = u_1{\cdot}x_{i_2}$ are comparable is a direct consequence of the definitions.

Assume now the property verified for all words of length $< m$ $(m > 2)$ and consider $f = \bar{f}x_{i_m}$ of length $m = |\bar{f}| + 1$. The intermediate states of f at u are those of $\bar{f}$ at u plus the state $\bar{u} = u_{m-1} = u{\cdot}\bar{f}$. Hence we distinguish two cases:

(i) $|\bar{u}| > \min(u, \bar{f})$. Then, u_j $(1 \leqq j \leqq m - 2)$ is at the same time the critical state of f and $\bar{f}$.

(ii) $|\bar{u}| \leqq \min(u, \bar{f})$. Then $\bar{u}$ is the critical state of f. In case (i), because of the induction hypothesis we have only to prove that u_j is a left factor of $\bar{u}$ comparable with $u' = u{\cdot}f$. The first statement follows directly from the hypothesis that $\bar{u}$ is comparable with u_j and that $|u_j| < |\bar{u}|$. The second statement follows from the first, the fact that $u' = \bar{u}{\cdot}x_{i_m}$ and the remark that if a is a left factor of b and if b is comparable with c then in turn a is comparable with c. In case (ii), the in-

duction hypothesis and $|\bar{u}| \leqq |u_j|$ show that $\bar{u}$ is a left factor of u_j, hence a left factor of every intermediate state of f. By our last remark above it follows that $\bar{u}$ is comparable with u and since $\bar{u}$ is comparable with $u' = u \cdot x_{i_m}$, the verification is completed.

Let us introduce the notation $C(u, u', n)$ for denoting the set (eventually empty) of all input words $f \neq e$ such that $u \cdot f = u'$ and either $|f| = 1$ or $\min(u, f) \geqq n$. The symbols $+$ and Σ denote disjoint union of sets. Finally $|a|$ is the maximum of the lengths of the words $\alpha(s, k)$ used in the definition of the push-down mappings μ. Thus for any u and $f \in F$ of length at least two, we have $\min(u, f) \leqq |a| + |u|$ since this is an upper bound to the length of the first intermediate word. Our definition of the critical factors is summarized by the following relations:

REMARK 2. *For any triple* (u, u', n)

$$C(u, u', n) = X \cap C(u, u'n) + \Sigma\{C(u, u'', |u''|)C(u'', u', |u''| + 1) : u'' \in U, \; n \leqq |u''| \leqq |u| + |a|\}$$

where $\Sigma\{\ \}$ *is understood to be* ϕ *unless* $n \leqq |u| + |a|$.

PROOF: The left member is contained in the right member because any $f \in C(u, u', n)$ of length two or more has critical factors f^* and f'' and a critical state u'' satisfying the condition indicated.

Conversely if $f' \in C(u, u'', |u''|)$ and $f'' \in C(u'', u, |u''| + 1)$ for some u'' such that $n \leqq |u''| \leqq |u| + a$, the product $f = f'f''$ belongs to $C(u, u', n)$ and has u'' as its critical state. Finally, the right member is a disjoint union of sets because of the unicity of the critical factorization.

Let J_l (resp. J_r) be the set of all triples (u, u', n) such that $C(u, u', n)$ is a nonempty set of words which can be left (resp. right) critical factors of other words. According to the definition and to Remark 1, $(u, u', n) \in J_l$ (resp. $\in J_r$) iff $C(u, u', n) \neq \phi$, u and u' are comparable, $n = |u'|$ (resp. $n = |u| + 1$).

For fixed $d > 0$ we consider the following subsets of $J_l \cup J_r$:

$$J_l^{+} = \{(u, u', |u'|) : |u| \leqq |u'|\}$$
$$J_l^{-d} = \{(u, u', |u'|) : |u| = |u'| + d\}$$
$$J_r^{-} = \{(u, u', |u| + 1) : |u| \geqq |u'|\}$$
$$J_r^{+d} = \{(u, u', |u| + 1) : |u'| = |u| + d\}.$$

REMARK 3. *For each finite d there exists only a finite number of distinct sets $C(u, u', n)$ with $(u, u', n) \in J(d) = J_l^+ \cup J_l^{-d} \cup J_r^- \cup J_r^{+d}$.*

PROOF: For any quadruple of states of the form $u = (s, gg')$, $\bar{u} = (s, \bar{g}g')$, $u' = (s', gg'')$, $\bar{u}' = (s', \bar{g}g'')$ with $\chi g = \chi\bar{g}$, the definition of pushdown mappings and the hypothesis that χ is a homomorphism shows that

$$C(u, u', |\, g\,| + n') = C(\bar{u}, \bar{u}', |\, \bar{g}\,| + n')(n' \geqq 0).$$

This proves directly the statement for $J_l^{-d} \cup J_r^{+d}$. For J_l^+ we need to observe first that, if $|\, u\,| < |\, u'\,|, C(u, u', |\, u'\,|) \neq \phi$ only if $|\, u'\,| \leqq |\, u\,| + |\, a\,|$. For J_r^- it suffices to check that (with χ_1 as defined in the proof of Property (1)) the relations $\chi_1 q = \chi_1\bar{q}$, $\chi_1 q' = \chi_1\bar{q}'$ imply identically

$$C((s, gg'), (s', g), |\, gg'\,| + 1) = C((s, \bar{g}\bar{g}'), (s', \bar{g}), |\, \bar{g}\bar{g}'\,| + 1).$$

REMARK 4. *Each set $C(u, u', n)$ is a context-free language.*

PROOF: Let the triple $(u, u', n) = j_0$ be fixed. Let $d = \max(|\, u\,| - n, |\, u'\,| - n, |\, a\,|)$ and consider a minimal set J^* of triples (containing j_0) such that any set $C(j)$ with $j \in \bar{J}(d) = J(0) \cup J(1) \cdots \cup J(d)$ is equal to one and only one set $C(j')$ with $j' \in J^*$. By Remark 3 we know that J^* is finite. Furthermore, by construction, any critical (left or right) factor of a word from a set $C(j)$ with $j \in \bar{J}(d)$ or $j = j_0$ belongs itself to some set $C(j')$ with $j' \in \bar{J}(d)$—hence to some set $C(j'') = C(j')$ with $j'' \in J^*$. Hence, by Remark 2, there corresponds to each $j \in J^*$ an equation

$$C(j) = X_j + \Sigma\{C(j')C(j'') : (j', j'') \in \omega(j)\}$$

where $X_j = X \cap C(j)$ and where $\omega(j)$ denote a finite set of pairs of triples $j', j'' \in J^*$.

We introduce a set $\Xi = \{\xi_j\}(j \in J^*)$ of new letters and, for each $j \in J^*$ we define the set p_j as the union of X_j and of the products $\xi_{j'}\xi_{j''}$ where $(j', j'') \in \omega(j)$. This reduces the problem to the proof of equivalence of the two definitions of a context-free language described in the first part of the paper.

Taking into account that the set of all context-free languages is closed under (finite) union, Property 2 is verified.

III. EXAMPLE

Let us recall briefly the definition of the free group Γ generated by a set $Y' = \{y_{i'} : 1 \leqq i' \leqq n\}$. Let Y consist of $2n$ letters y_i $(i = \pm i'$,

$1 \leqq i' \leqq n$) and say that a word of G (the set of all words in the letters of Y) is *reduced* if it does not contain a pair of adjacent letters having opposite indices. Clearly to each $g \in G$ is associated a unique reduced word τg obtained by successive cancellation of such pairs of adjacent letters. For instance,

$$\tau(y_1 y_{-2} y_2 y_{-1}) = \tau(y_1 y_{-1}) = e$$

and

$$\tau(y_1 y_2 y_{-1} y_{-2}) = y_1 y_2 y_{-1} y_{-2}\,.$$

The homomorphism $\gamma_0 : G \to \Gamma$ is defined by $\gamma_0 g = \gamma_0 g'$ iff $\tau g = \tau g'$. We shall identify Γ with the set τG of all reduced words of G endowed with the (associative) multiplication $\tau(\tau g \cdot \tau g')(= \tau(gg'))$.

Thus, in particular, $\gamma_0 y_{-i} = (\gamma_0 y_i)^{-1}$ for all $y_i \in Y$. With this notation, any homomorphism $\gamma\colon F \to \Gamma$ is given by a homomorphism $\gamma' : F \to G$ and the rule $\gamma f = \gamma f'$ iff $\gamma_0 \gamma' f = \gamma_0 \gamma' f'$ (i.e., iff $\tau\gamma' f = \tau\gamma' f'$).

We verify that the inverse image $\gamma^{-1}\Gamma'$ by γ of any *finite* subset Γ' of Γ is a context-free language, by constructing a pushdown automaton $\mathfrak{A}$.

For this, consider a state s_0 and, for each word $\gamma' x_j \in \gamma X$, introduce n_j new states $s_{j,1}$, $s_{j,2}$, $\cdots$ s_{j,n_j} where n_j is the length of the word $\gamma' x_j$. Let S be the union of s_0 and all the states $s_{j,k}$ and $\bar{S}$ consist of s_0 only. The state s_∞ is not needed.

Now, in the notation of Definition 2 we define for each $x_j \in X$, $\beta(s_0, x_j) = s_{j,1}$ and to each state $s_{j,j'}$ we associate (1) a mapping $\sigma\colon S \to S$ defined by

$$\begin{aligned} \sigma s_{j,j'} &= s_{j,j'+1} && \text{if } j < n_j\,; \\ &= s_0 && \text{if } j = n_j\,. \end{aligned}$$

(We do not need to introduce K explicitly.) (2) A push-down mapping μ_i on G where $i(-n \leqq i \leqq n)$ is determined by $y_i = y_{j,j'}$ (i.e., where y_i is the j'th letter of $\gamma' x_j$) and where for each $y \in G$:

$$\begin{aligned} \mu_i g &= g' \quad \text{if} \quad g = g' y_{-i} && (g' \in G); \\ &= g y_i && \text{otherwise.} \end{aligned}$$

For instance, if $\gamma' x_1 = y_{-1} y_2$ the resulting transformation on G is $\mu_2 \mu_{-1}$ and it has the following effect: $g \to g'$ if $g = g' y_1 y_{-2}$; $g \to g' y_2$ if $g = g' y_1$ and $g' \notin G y_{-2}$; $g \to g y_{-1} y_2$ in any other case.

Induction of the length of the input word f shows that if the initial

word stored in the memory is a reduced word $g \in \tau G$, the word stored after reading f is $\tau(g\gamma' f)$. Thus, in particular, taking $g = e$ (the empty word), Property 2 shows that for any finite subset Γ' of Γ the set $\gamma^{-1}\Gamma' = \{f \in F : \gamma f \in \Gamma'\}$ is a context language.

Clearly, for any regular event R on F, it is possible to add enough new states to the finite part of the automaton that it accepts a set of the form $R \cap \gamma^{-1}\Gamma'$.

Let us consider the special case where $X = \{x_i : i = \pm i', 1 \leqq i' \leqq n\}$ and $\gamma' x_i = y_i$ identically. Taking the empty word e as initial word and the set $\{e\}$ as set of final word, it is clear that the set of words accepted by the automaton is $\gamma^{-1}e$, the kernel of the homomorphism γ of F onto Γ that satisfies identically $(\gamma x_i)^{-1} = \gamma x_{-i}$. Thus we have $(\mathfrak{u}_n)$.

For all $f, f', f'' \in F$, any two of the following relations imply the third one: $f \in \gamma^{-1}e; f'f'' \in \gamma^{-1}e; f'ff'' \in \gamma^{-1}e$.

In fact, it can be proved that $\gamma^{-1}e$ can be defined abstractly as the least subset of F that satisfies $(\mathfrak{u}_n)$ and that contains e and every product $x_i x_{-i}$ $(1 \leqq i \leqq n)$.

We construct explicitly the grammar defining $\gamma^{-1}e$. For each i $(i = \pm i', 1 \leqq i' \leqq n)$ let D_i denote the set of all $f \in x_i F \cap \gamma^{-1}e$ which are such that $f = f'f''; f', f'' \in \gamma^{-1}e$ implies f' or $f'' = e$. Intuitively, $f \in D_i$ iff the first letter of f is x_i and if the first return of the internal memory to the empty word occurs at the last letter of f. Thus $f \in D_i$ implies $f = x_i f' x_{-i}$ and either $f' = e$ or $f' = f_1 f_2 \cdots f_m$ where $f_1 \in D_{j_1}$, $f_2 \in D_{j_2} \cdots f_m \in D_{j_m}$ or in more concrete manner, where the words f_k are characterized by the fact that at that end of their last letter (when reading f) the word stored in the memory is reduced to the letter x_i. Clearly this factorization is unique and our hypothesis that the memory is never empty before the end of f implies that $j_1, j_2, \cdots, j_m \neq -i$. Conversely, if $D_i' = \Sigma\{D_j : j \neq -i\}$, any word of the form $x_i f' x_{-i}$ belongs to D_i if f' is a product of words from D_i'. Hence:

$$D_i = x_i x_{-i} + x_i D_i'^* x_{-i}$$

where $D_i'^*$ is defined by $D_i'^* = D_i' + D_i'^* D_i'$.

Introducing $2n$ new variables ξ_i and $\xi_i'^*$ we deduce the $2n$ equations:

$$\xi_i = x_i x_{-i} + x_i \xi_i'^* x_{-i}\,; \qquad \xi_i'^* = (e + \xi_i'^*)\Sigma\{x_j x_{-j} + x_j \xi_j'^* x_{-j} : j \neq -i\}.$$

Finally, denoting by D^* the set of the nonempty words of $\gamma^{-1}e$, we have $D^* = (e + D^*)\Sigma D_i$ from which we deduce the equation

$$\xi^* = (e + \xi^*)\Sigma\{x_i x_{-i} + x_i \xi_i'^* x_{-i} : i = \pm i',\ 1 \leqq i' \leqq n\}$$

in the new variable ξ^* corresponding to D^*. The fact that D^* is related to an algebraic system of equations goes back to Kesten (Kesten, 1959). When the internal alphabet consists of a single letter, γ becomes a homomorphism into a cyclic group and the memory can be identified with an unbounded counter. The corresponding theory is due to Raney (Raney, 1960) and it relates to the enumeration of well formed formulas (free notation); an approach similar to the present one has been given in (Schützenberger, 1959). As a point of marginal interest it may be mentioned that the set D^* (which we shall call a *Dyck set*) and, more generally, the standard context free languages defined below, can also be defined as the complement of the support of certain rational (non commutative) formal power series. This results instantly from the existence of isomorphic representations of the free group Γ by integral finite dimensional matrices (cf. e.g. Sanov, 1957).

For further reference we introduce the following definition in which it is assumed that $X = \{x_{\pm i}\}$ as above.

DEFINITION. A *standard context-free language* is a set $L = D^* \cap R(X_1, V)$ where D^* is a Dyck set and where the regular event $R(X_1, V)$ is given by a subset X_1 of X, a subset V of X^2 and the relation $R(X_1, V) = X_1F \setminus FVF$ (= the set of all words that begin with a letter from X_1 and that have no factor belonging to V).

IV. A WEAK CONVERSE PROPERTY

PROPERTY 3. *Each context-free language L can be represented as the homomorphic image of some standard context-free language.*

PROOF: Let L be produced by the grammar $(p_j)(1 \leqq j \leqq n)$, the notation being as in the first section of the paper.

Each word $h \in p_j$ has a unique factorization $h = f_1'\xi_{i_1}f_2'\xi_{i_2}f_3' \cdots f'_{\delta h}\xi_{i_{\delta h}}f'_{\delta h+1}$ where $f_1', f_2', \cdots, f'_{\delta h}, f'_{\delta h+1} \in F$, $\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_{\delta h}} \in \Xi$, $0 \leqq \delta h < d^*$ and where $d^* = 1 + \max\{\delta h : h \in p_j, 1 \leqq j \leqq n\}$. Eventually $h \in F$ in which case $\delta h = 0$ and $h = f_1'$.

Let us introduce a set X' of new letters $x(j, h, d, \epsilon)$ indexed by quadruples with $1 \leqq j \leqq n$; $h \in p_j$; $0 \leqq d \leqq d^*$; $\epsilon = \pm 1$. For given j and $h \in p_j$ the writing $q(h, d, d)$ denotes the word $x(j, h, d, +1)\xi_{i_d}x(j, h, d, -1)$ if $1 \leqq d \leqq \delta h$ and the word $x(j, h, d, +1)x(j, h, d, -1)$ if $\delta h < d \leqq d^*$; finally,

$$q(h, 0, 0) = x(j, h, 0, -1)q(h, 1, 1)q(h, 2, 2) \cdots q(h, d^*, d^*)x(j, h, 0, +1).$$

Let F' be the free monoid generated by $\Xi \cup X'$ and define the homomorphism $\varphi: H' \to H$ in the obvious way: for all $\xi \in \Xi$, $\varphi\xi = \xi$; for

$1 \leqq j \leqq n$, $h \in p_j$, $1 \leqq d \leqq 1 + \delta h$, $\epsilon = +1$, $\varphi x(j, h, d, \epsilon) = f_j'$; $\varphi x(j, h, d, \epsilon) = e$ for all the other elements of X'.

Thus for all $h \in p_j$, we have $h = \varphi q(h, 0, 0)$. It follows that if the grammar $(p_j')(1 \leqq j \leqq n)$ is defined by $p_j' = \{\zeta q(h,0,0) : h \in p_j\}$, we have for all $m > 0$ the identity $\varphi p'(m) = p(m)$. Hence $L = \varphi p_1'(\infty)$ and, without loss of generality, we shall assume henceforth that L *is* $p_1'(\infty)$ itself, that is, $X = X'$; $H = H'$, $(p_j) = (p_j')(1 \leqq j \leqq n)$. Under this assumption we shall write $x(h, d, \epsilon)$ instead of $x(j, h, d, \epsilon)$ since every word $h \in P = \bigcup\{p_j : 1 \leqq j \leqq n\}$ appears in one and only one set p_j $(1 \leqq j \leqq n)$.

For $1 \leqq d \leqq d' \leqq n$ we define:

$$q(h, d, d') = q(h, d, d)q(h, d + 1, d + 1) \cdots q(h, d', d').$$

(Thus $q(h, d, d')$ is defined only when none or both of d and d' are 0.) Finally:

$$Q = \{q(h, d, d') : h \in P, \quad d = d' = 0 \text{ or } 1 \leqq d \leqq d' \leqq d^*\}$$

$$T = \{\lambda_{p(\infty)} q(h, d, d') : q(h, d, d') \in Q\}.$$

We now define the standard (right) context-free language $D \cap R \cap H_j$ by:

(i) D is the Dyck set $D = \{f \in F : f \neq e; \gamma f = e\}$ where the homomorphism $\gamma : H \to \Gamma$ is defined by:

$$\text{for all } \xi \in \Xi, \qquad \gamma\xi = e$$

$$\text{for all } x(h, d, \epsilon) \in X, \qquad (\gamma x(h, d, \epsilon))^{-1} = \gamma x(h, d, -\epsilon).$$

(ii) R is the set of the words $f \in F$ such that each of their factors of length two belongs to $\bar{V} = \bar{V}' \cup \bar{V}'' \cup \bar{V}'''$ where:

$\bar{V}'$ is the set of all products $x(h, d, +1)x(h', 0, -1)$ and $x(h', 0, +1)$ $x(h, d, -1)$ for which d, h, and h' are such that $q(h, d, d) = x(h, d, +1)\xi_{j_d} x(h, d, -1)$ with $h' \in p_{j_d}$.

$\bar{V}''$ is the set of all products $x(h, d, +1)\ x(h, d, -1)$ with $h \in P$ and $\delta h < d \leqq d^*$.

$\bar{V}'''$ is the set of all words $x(h, d, -1)x(h, d', -1)$ with $h \in P$ and either $0 \leqq d < d' = d + 1 \leqq d^*$ or $d = d^*$ and $d' = 0$.

(iii) $H_j (1 \leqq j \leqq n)$ is the set of all words of H whose first (left) letter has the form $x(h, 0, -1)$ with $h \in p_j$.

We shall use repeatedly the fact that if the n-tuple $a = (a_1, a_2, \cdots, a_n)$ of elements of H is such that $a_1, a_2, \cdots, a_n \in \gamma^{-1}e$

(for short, if $a \subset (\gamma^{-1}e)$) then, for any $h \in H$, the two relations $\gamma h = e$ and $\lambda_a h \subset \gamma^{-1}e$ are equivalent.

Now we have:

(1) For each j $(1 \leqq j \leqq n)$, $p_j(\infty) = T \cap H_j$.

This is a direct consequence of $Q \cap H_j = P$ and the definition of T.

(2) $T \subset D \cap R$.

It is clear that $p(0) \subset (D \cap R \cap H_j)$ (i.e., for each j, $p_j(0) \subset D \cap R \cap H_j$). Assume $p(n') \subset (D \cap R \cap H_j)$ proved for $n' \leqq n$. Since, trivially, $Q \subset \gamma^{-1}e$, it follows that $\lambda_{p(n)}q \subset D$ for each $q \in Q$; hence, in particular, $p_j(n+1) \subset D$ for each j $(1 \leqq j \leqq n)$. Further, every factor of length two of a word of the set $\lambda_{p(n)}q (q \in Q)$ is a factor of a word contained in one of the sets $p_j(n)$ or belongs to $\bar{V}$. Hence $\lambda_{p(n)}q \subset R$ and, in particular, $p_j(n+1) \subset R$ for all j $(1 \leqq j \leqq n)$. The fact that $p_j(n) \subset H_j$, identically is trivial and the result follows by induction on n.

Since (1) and (2) show $L \subset p_1(\infty) \subset D \cap R \cap H_1$ and $T \cap H_1 \subset L$, the verification of Property 3, i.e., of $L = D \cap R \cap H_1$, will follow from:

(3) $D \cap R \subset T$.

Let $f \in D \cap F$. It is trivial that $f \in T$ for $|f| \leqq 2$. Hence we can assume the result proved for all $f' \in F$ of length $<n$ and $|f| = n > 2$.

We consider the factorization $f = g'g''(g', g'' \in F)$ where g' is defined as the shortest left factor of f that belongs to D. Since $f \in D$, this implies $\gamma g'' = (\gamma g')^{-1}\gamma f = e$ and we distinguish two cases:

(i) $g'' \neq e$. Then $g'' \in D$; $|g'|, |g''| < n$, and, by the induction hypothesis, there exist two elements $q', q'' \in Q$ such that $g' \in \lambda_{p(\infty)}q'$ and $g'' \in \lambda_{p(\infty)}q''$. Let $x(h', d', \epsilon')$ and $x(h'', d'', \epsilon'')$ denote respectively the last (right) letter of q' and the first (left) letter of q''. By the definition of Q, we have

$$1 \leqq d' \leqq d^* \quad \text{and} \quad \epsilon' = 1 \quad \text{or} \quad d' = 0 \quad \text{and} \quad \epsilon' = +1; \quad \text{similarly:}$$

$$1 \leqq d'' \leqq d^* \quad \text{and} \quad \epsilon'' = +1 \quad \text{or} \quad d'' = 0 \quad \text{and} \quad \epsilon'' = -1.$$

However, $g'g'' \in R$ implies $v = x(h', d', \epsilon')x(h'', d'', \epsilon'') \in \bar{V}$. Obviously $v \notin V' \cup V''$. Hence $v \in \bar{V}'''$ and, thus, $h = h' = h''$, $d'' = d' + 1$, $\epsilon' = -1$, $\epsilon'' = +1$. This means that $q'q'' = q(h, d_1, d_2) \in Q$ for some d_1, d_2. Since, now, $f \in \lambda_{p(\infty)}q(h, d_1, d_2)$, the result is verified in this case.

(ii) $g'' = e$, that is, $f = g'$.

Since $|f| = n > 0$, this implies $f = af'b$ with $a, b \in X, f' \in F, |f'| > 0$.

Because of the definition of γ, $\gamma f = e$ and the hypothesis that f has no left factor in D imply $\gamma f' = \gamma ab = e$. Thus $f' \in D \cap R$ and, by the induction hypothesis, $f' \in \lambda_{p(\infty)} q(h', d, d')$ for some $h' \in P$, $0 \leqq d \leqq d' \leqq d^*$.

It follows that $f' = x(h', d, \epsilon) f'' x(h', d', \epsilon)$ either with $1 \leqq d < d' \leqq d^*$, $\epsilon = +1$, $\epsilon' = -1$ or with $d = d' = 0$, $\epsilon = -1$, $\epsilon' = +1$. Let $v = ax(h', d, \epsilon)$, $v' = x(h', d', \epsilon')b$. Since $f \in R$, both belong to $\bar{V}$. Clearly $v, v' \notin \bar{V}''$. Further, $v, v' \notin \bar{V}'''$ because, for instance, $v \in \bar{V}'''$ would imply $1 \leqq d < d' \leqq d^*$, $\epsilon = +1$, $\epsilon = -1$. $a = x(h', d - 1, -1)$, hence $b = x(h', d - 1, +1)$ (since $\gamma b = (\gamma a)^{-1}$) and, finally, $v' \in \bar{V}'''$ giving $d' = d - 2$, in contradiction of $d < d'$. Thus, $v, v' \in \bar{V}'$, that is, $f = x(h, d'', +1) f' x(h, d'', -1)$ and $f' = x(h', 0, -1) f'' x(h', 0, +1)$, where h, h' and d'' are such that $q(h, d'', d'') = x(h, d'', +1)\xi_j x \cdot (h, d'', -1)$ with $h' \in p_j$. Thus $f \in \lambda_{p(\infty)} q(h, d'', d'')$, concluding the verification of Property 3.

REMARK. In the case of formal power series over a ring A, the sets p_j $(1 \leqq j \leqq n)$ defining the grammar are replaced by the elements $\bar{p}_j = \Sigma\{a_{j,h} h : h \in p_j\}$ (with $a_{j,h} \in A$) of the free algebra over A generated by $X \cup \Xi$. It is trivial that Property 3 and its proof remain valid provided that the homomorphism φ is replaced by a homomorphism $\bar{\varphi}$ sending the large algebra (over the integers) of the free monoid generated by $X' \cup \Xi$ into the large algebra (over A) of the free monoid generated by $X \cup \Xi$. For this, it suffices to replace φ by $\bar{\varphi}$ in all the definitions except for the conditions $\varphi x(j, h, 0, -1) = e$ $(1 \leqq j \leqq n, h \in p_j)$ which have to become $\bar{\varphi} x(j, h, 0, -1) = a_{j,h}$.

REFERENCES

CHOMSKY, N., AND MILLER, G. A., (1958), Finite state languages. *Information and Control* **1,** 91–112.

CHOMSKY, N., (1962), Context-free grammars and push-down storage. Quarterly Progress Report No. 65, Research Laboratory of Electronics, M.I.T.

CHOMSKY, N., AND SCHÜTZENBERGER, M. P., (1962), The algebraic theory of context-free language, to appear in "Computer Programming and Formal Systems," P. BRAFFORT AND D. HIRSCHBERG, eds. North Holland, Amsterdam.

GINSBURG, S., AND RICE, H. G., (1961), Two families of languages related to Algol. *J. Assoc. Computing Mach.* **9,** 350–370.

KESTEN, H., (1959), Symmetric random walks on groups. *Trans. Am. Math. Soc.* **92,** 336–354.

NEWELL, A., AND SHAW, J. C., (1957), Programming the logic theory machine. *Proc. Western Joint Computer Conf.*, p. 230.

RABIN, M. O., AND SCOTT, D., (1959), Finite automata and their decision problems. *IBM J. Research Develop.* **3,** 115–125.

RANEY, G. N., (1960), Functional composition patterns and power-series reversion, *Trans. Am. Math. Soc.* **94,** 441–451.

REDEI, L., (1950), Die Anwendung des Schiefen Produktes in der Gruppentheorie. *J. reine u. angew. Math.* **188,** 201–227.

SANOV, I. N., (1947), A property of a representation of a free group. *Dokl. Akad. Nauk SSSR* **57,** 657–659.

SCHÜTZENBERGER, M. P., (1959), "Un problème de la théorie des automates." Séminaire Dubreil-Pisot, Institut H. Poincaré, Paris.

SHEPHERDSON, J. C., (1959), The reduction of two-way automata to one-way automata. *IBM J. Research Develop.* **3,** 198–200.

SHEPERDSON, J. C., AND STURGIS, H. E., (1963), Computability of recursive functions. *J. Assoc. Computing Mach.* **10,** 217–255.

MATH. SCAND. 12 (1963), 117–128

QUELQUES REMARQUES SUR UNE CONSTRUCTION DE SCHENSTED

M. P. SCHÜTZENBERGER

1. Introduction.

Soit N l'ensemble des entiers positifs et $\alpha\colon A \to B$, une bijection d'un sous-ensemble $A \subset \mathsf{N}$ sur un autre sous-ensemble $B \subset \mathsf{N}$. C. Schensted [1] a découvert une construction remarquable qui associe de façon injective à $\alpha\colon A \to B$ une paire $\big(P(\alpha,A), Q(\alpha,B)\big)$ de tableaux standards de même forme.

Nous nous proposons de montrer ici que $Q(\alpha,B)$ est en fait égal à $P(\alpha^{-1},B)$ et qu'il existe une relation simple entre $Q(\alpha,B)$ et la factorisation de la suite $\alpha a_1, \alpha a_2, \ldots, \alpha a_i, \ldots$ (où $A = \{a_1 < a_2 < \ldots < a_i < \ldots\}$) en séquences croissantes maximales.

Pour simplifier les notations nous considérerons les tableaux standards comme des éléments particuliers du module $\mathscr{T}$ des applications de $\mathsf{N} \times \mathsf{N}$ dans Z bien que seules interviennent réellement les structures d'ordre de ces ensembles. Pour tout $P \in \mathscr{T}$ on définira la *forme* $|P|$ de P comme l'ensemble des $(i,j) \in \mathsf{N} \times \mathsf{N}$ tels que $P_{i,j} \neq 0$; le *contenu* $\{P\}$ de P sera l'image par P de $|P|$ dans Z. L'element P est *standard* si les conditions suivantes sont satisfaites:

1. Si (i,j) n'appartient pas à $|P|$, alors ni $(i+1,j)$ ni $(i,j+1)$ n'appartiennent à $|P|$.
2. La restriction de P à $|P|$ est non décroissante en chacun de ses arguments et $\{P\}$ est un ensemble d'entiers positif.
3. La restriction de P à $|P|$ est une bijection sur $\{P\}$.

Quand $P\colon |P| \to \{P\}$ est une bijection on dénotera par P^{-1} l'application inverse; pour tout $a \in \mathsf{Z} \setminus \{0\}$ et tout $(i,j) \in \mathsf{N} \times \mathsf{N}$, $a_{i,j}$ sera l'élément de $\mathscr{T}$ défini par les conditions $|a_{i,j}| = (i,j)$ et $\{a_{i,j}\} = a$.

Rappelons la construction de Schensted. Si P est un tableau standard et si $a \in \mathsf{N} \setminus \{P\}$, on montre qu'il existe un et un seul tableau standard P' (désigné par $P \leftarrow a$) qui satisfasse les conditions suivantes:

4. $|P| \subset |P'|$ et $\{P'\} = \{P\} \cup \{a\}$.

Reçu le 12. février, 1963.

5. $P'_{1,j_1} = a$ pour un certain $j_1 \in \mathbf{N}$; pour chaque $i \in \mathbf{N}$ il existe au plus un $j \in \mathbf{N}$ tel que $P_{i,j} \neq P'_{i,j}$; en outre, dans ce cas, il existe un $j' \leqq j$ tel que $P_{i,j} = P'_{i+1,j'}$.

De façon analogue, Schensted note $a \to P$ le tableau standard $(P^T \leftarrow a)^T$, où ^{T} indique la transposition.

Soit maintenant $\alpha : A \to B$ comme plus haut. Pour tout sous-ensemble fini A' de A et tout entier non négatif m, on posera

$$P(\alpha, A_m') = 0 \quad \text{si} \quad m = 0 ,$$

et, inductivement,

$$P(\alpha, A_m') = P(\alpha, A'_{m-1}) \leftarrow \alpha a_m' \quad \text{si} \quad 0 < m \leqq \operatorname{Card} A'$$

où a_m' dénote le m-ième élément de A' par ordre croissant;

$$P(\alpha, A_m') = P(\alpha, A_n') = P(\alpha, A') \quad \text{si} \quad m \geqq \operatorname{Card} A' = n .$$

Il est logique de considérer ici $0 \leftarrow \alpha a_1'$ comme le tableau standard $(\alpha a_1')_{1,1}$ et, par conséquent, $P(\alpha, A')$ est exactement le P-symbole de Schensted de la séquence $(\alpha a_1', \alpha a_2', \ldots, \alpha a_n')$. De même:

$$\begin{aligned} Q(\alpha, A_0') &= 0; \\ Q(\alpha, A_m') &= Q(\alpha, A'_{m-1}) + (a_m')_{i,j} \quad \text{pour} \quad 0 < m \leqq \operatorname{Card} A' \end{aligned}$$

avec $(i,j) = |P(\alpha, A_m')| \setminus |P(\alpha, A'_{m-1})|$,

$$Q(\alpha, A_m') = Q(\alpha, A_n') = Q(\alpha, A') \quad \text{pour} \quad m \geqq \operatorname{Card} A' = n .$$

Quand $A' = [1, n]$ ceci est la définition même du Q-symbole de Schensted de $(\alpha a_1', \alpha a_2', \ldots, \alpha a_n')$, et pour A' quelconque, $Q(\alpha, A')$ se déduit simplement de ce Q-symbole en remplaçant dans ce dernier tableau chaque $m \in [1, n]$ par a_m'.

Posons $\|0\| = 0$ et pour chaque $T \in \mathscr{T} \setminus \{0\}$,

$$\|T\|^{-1} = \operatorname{Min}\big(i + j - 1 : (i,j) \in |T|\big) .$$

Il résulte de la définition même de l'opération $\leftarrow$ que pour $A' = A$ on a

$$\lim_{m, m' \to \infty} \|P(\alpha, A_m) - P(\alpha, A_{m'})\| = \lim_{m, m' \to \infty} \|Q(\alpha, A_m) - Q(\alpha, A_{m'})\| = 0 .$$

On pourra donc toujours définir

$$P(\alpha, A) = \lim_{m \to \infty} P(\alpha, A_m) \quad \text{et} \quad Q(\alpha, A) = \lim_{m \to \infty} Q(\alpha, A_m) .$$

Les notations qui viennent d'être introduites seront systématiquement utilisées dans tout ce qui suit.

2. L'opération Δ.

Soit $Q = Q(\alpha, A_m) \neq 0$, $m < \infty$. On définit un autre tableau standard $Q' = \Delta Q(\alpha, A_m)$ et une séquence $|U_m| = \big((i_1,j_1), (i_2,j_2), \ldots, (i_p,j_p)\big)$ d'éléments de $|Q|$ par les conditions suivantes:

$$(1) \qquad (i_1,j_1) = (1,1)$$

et, pour chaque $k \in [1, p-1]$,

$$(i_{k+1},j_{k+1}) = (i_k+1, j_k) \quad \text{ou} = (i_k, j_k+1);$$

$$(2) \qquad \begin{aligned} Q'_{i',j'} &= Q_{i',j'} \text{ si } (i',j') \notin |U_m|; \\ &= Q_{i,j} \text{ avec } (i,j) = (i_{k+1},j_{k+1}) \in |U_m| \end{aligned}$$

si $(i',j') = (i_k,j_k) \in |U_m|$ et $k \in [1, p-1]$;

$$Q'_{i',j'} = 0 \text{ si } (i',j') = (i_p,j_p),$$

le dernier élément de $|U_m|$;

$$(3) \qquad Q' \text{ est un tableau standard .}$$

De façon plus explicite, connaissant déjà $(i',j') = (i_k,j_k) \in |U_m|$, on pose $k = p$ (c'est-à-dire que l'on considerè (i',j') comme le dernier élément de $|U_m|$) si

$$Q_{i'+1,j'} = Q_{i',j'+1} = 0 .$$

Sinon on détermine $(i_{k+1},j_{k+1}) \in |U_m|$ par les conditions

$$(i_{k+1},j_{k+1}) = (i'+1, j')$$

si $0 < Q_{i'+1,j'} < Q_{i',j'+1}$ ou si $0 = Q_{i',j'+1} < Q_{i'+1,j'}$;

$$(i_{k+1}, j_{k+1}) = (i', j'+1)$$

si $0 < Q_{i',j'+1} < Q_{i'+1,j'}$ ou si $0 = Q_{i'+1,j'} < Q_{i',j'+1}$ qui assurent automatiquement que (3) est satisfaite.

Donc, $\{\Delta Q\}$ est l'ensemble $\{Q\}$ privé de son plus petit élément $Q_{1,1}$. D'après les définitions mêmes, $|U_m| \subset |U_{m+1}|$, et par conséquent,

$$|U| = \lim |U_m|, \quad \text{ainsi que} \quad \Delta Q(\alpha, A) = \lim \Delta Q(\alpha, A_m)$$

sont bien définis. Plus généralement, si $\bar{Q}$ est un autre tableau standard on a

$$\|\Delta Q - \Delta \bar{Q}\|^{-1} \geqq \|Q - \bar{Q}\|^{-1} + 1 .$$

EXEMPLE. Si $Q = \frac{248}{679}$ avec les notations de Schensted,

$$\Delta Q = \tfrac{478}{69},\ \Delta^2 Q = \Delta(\Delta Q) = \tfrac{678}{9},\ \Delta^3 Q = \tfrac{78}{9},\ \Delta^4 Q = \tfrac{8}{9},\ \Delta^5 Q = 9,\ \Delta^6 Q = 0.$$

REMARQUE 1. *Si* $A \neq \emptyset$, *on a identiquement*

$$\Delta Q(\alpha, A) = Q(\alpha, A \setminus \{a_1\}) .$$

DÉMONSTRATION. Le résultat peut être vérifié directement pour $\mathrm{Card}\, A < 2$ et, dénotant pour abréger par A' l'ensemble $A \setminus \{a_1\}$, il suffit de vérifier que, pour $m > 1$, l'égalité de $\Delta Q(\alpha, A_{m-1})$ et $Q(\alpha, A'_{m-2})$ entraîne celle de $\Delta Q(\alpha, A_m)$ et $Q(\alpha, A'_{m-1})$.

Par définition il existe (i,j) et $(i',j') \in \mathbf{N} \times \mathbf{N}$ tels que

$$Q(\alpha, A'_{m-1}) = Q(\alpha, A'_{m-2}) + (a_m)_{i',j'}$$

et

$$Q(\alpha, A_m) = \Delta Q(\alpha, A_{m-1}) + (a_m)_{i,j} .$$

Donc, d'après l'hypothèse d'induction,

$$\Delta Q(\alpha, A_m) = Q(\alpha, A'_{m-1}) + (a_m)_{i,j} - (a_m)_{i',j'}$$

et il ne reste qu'à vérifier $(i,j) = (i',j')$.

Rappelant le résultat fondamental de Schensted [1, lemme 6, p. 183]

$$\begin{aligned} P(\alpha, A_m) &= \big(\alpha a_1 \to P(\alpha, A'_{m-2})\big) \leftarrow \alpha a_m \\ &= \alpha a_1 \to \big(P(\alpha, A'_{m-2}) \leftarrow \alpha a_m\big) \end{aligned}$$

et utilisant l'identité de forme des P-symboles et des Q-symboles on a

$$|Q(\alpha, A_{m-1})| \setminus |Q(\alpha, A'_{m-2})| = (i'', j'')$$

et

$$|Q(\alpha, A_m)| \supset |Q(\alpha, A'_{m-2})| \cup \{(i,j), (i',j'), (i'',j'')\} .$$

Distinguons maintenant deux cas:

1° $(i'',j'') \neq (i,j)$. Par conséquent,

$$|Q(\alpha, A_m)| = |Q(\alpha, A'_{m-2})| \cup \{(i,j), (i'',j'')\} .$$

La commutativité des opérations $\to$ et $\leftarrow$ implique la relation

$$Q(\alpha, A_m) - Q(\alpha, A_{m-1}) = Q(\alpha, A'_{m-1}) - Q(\alpha, A'_{m-2}) = (a_m)_{i,j} .$$

Donc, $(i,j) = (i',j')$ puisque le tableau obtenu en remplaçant a_m par zéro dans $\Delta Q(\alpha, A_m)$ est égal à $\Delta Q(\alpha, A_{m-1})$, c'est-à-dire à $Q(\alpha, A'_{m-2})$ par l'hypothèse d'induction.

2° $(i'',j'') = (i,j)$. Par conséquent,

$$|Q(\alpha, A_m)| = |Q(\alpha, A_{m-2})| \cup \{(i,j), (\bar{i},\bar{j})\}$$

où $(\bar{i},\bar{j}) = |P(\alpha, A_m)| \setminus |P(\alpha, A'_{m-1})|$. Cette dernière relation implique $(\bar{i},\bar{j}) = (i+1, j)$ ou $(i, j+1)$ et par conséquent $|Q(\alpha, A'_{m-2})| \cup \{\bar{i},\bar{j})\}$ n'est pas

la forme d'un tableau standard. Comme $\Delta Q(\alpha, A_m)$ est un tableau standard tel que

$$|\Delta Q(\alpha, A_m)| = |Q(\alpha, A'_{m-2})| \cup \{(i', j')\} \subset |Q(\alpha, A_m)|$$

on a donc encore $(i', j') = (i, j)$ et la vérification est achevée.

PROPRIÉTÉ 1. *La correspondance de Schensted associant la paire* $\big(P(\alpha, A), Q(\alpha, A)\big)$ *à* $\alpha : A \to B$ *est injective.*

DÉMONSTRATION. Le résultat plus fort prouvant, pour A fini, le caractère bijectif de la correspondance est dû à Schensted [1, lemme 3, p. 182]. Nous considérons le cas de A infini, et nous vérifions que la donnée de $P = P(\alpha, A)$ et de $Q = Q(\alpha, A)$ détermine de façon univoque a_1, αa_1, $P(\alpha, A')$ et $Q(\alpha, A')$, avec $A' = A \setminus \{a_1\}$ comme plus haut.

Pour tout m positif fini, Schensted a montré (loco citato) qu'il existe une et une seule paire (b, P_m') telle que P_m' soit un tableau standard satisfaisant les relations

$$|P_m'| = |\Delta Q(\alpha, A_m)| \quad \text{et} \quad b \to P_m' = P(\alpha, A_m);$$

la remarque 1 montre, qu'en fait, $b = \alpha a_1$, $P_m' = P(\alpha, A'_{m-1})$ et en outre

$$\{a_1\} = \{Q(\alpha, A_m)\} \setminus \{Q(\alpha, A_m')\} \quad (= \{Q\} \setminus \{\Delta Q\}) .$$

De par la définition même de l'opération $\to$, on a

$$\big(P(\alpha, A_m)\big)^{-1} b = (i_m', 1)$$

où i_m' est au moins égal au nombre i_m défini par

$$(i_m, j_m) = |Q(\alpha, A_m)| \setminus |\Delta Q(\alpha, A_m)| .$$

Comme $\lim i_m' = i'^* < \infty$, il en résulte

$$\lim i_m = i^* < \infty .$$

Ainsi, puisque, identiquement, $(i_m, j_m) \in |U|$ et $i_m \leqq i_{m+1}$, la valeur de i^* est déterminée de façon unique par Q. Plus précisément

$$i^* = \max_{d>0} \{i : (i, j) \in |U|, i + j = d\}$$

et les inégalités qui viennent d'être écrites montrent que ce nombre i^* est fini pour tout tableau standard qui est un Q-symbole.

Définissons maintenant pour chaque $d > i^*$ le tableau standard $P^{(d)}$ par les relations

$$\begin{aligned} P^{(d)}_{i,j} &= P_{i,j} \quad \text{si } i + j \leqq d ; \\ &= 0 \quad \text{si } i + j > d . \end{aligned}$$

D'après le résultat de Schensted rappelé plus haut, il existe pour chaque $d > i^*$ une et une seule paire $(b^{(d)}, P'^{(d)})$ satisfaisant

$$P^{(d)} = b^{(d)} \to P'^{(d)} \quad \text{et} \quad |P^{(d)}| \setminus |P'^{(d)}| = \{(i^*, d - i^*)\} .$$

C'est une propriété élémentaire de $\to$ que $b^{(d)} \leqq b$, identiquement. Par conséquent $b = \lim b^{(d)}$ et, trivialment,

$$P(\alpha, A') = \lim P'^{(d)}$$

ce qui achève la vérification.

Il est utile de noter que si $P = Q$, on a $b = P_{1,1}$ (et, par conséquent, $b = \{Q\} \setminus \{\Delta Q\}$) si et seulement si $i^* = 1$. Ceci résulte immédiatement de la remarque plus générale que S étant un tableau standard quelconque et $0 < s < S_{1,1}$, on a l'identité

$$\Delta(s \to S) = \Delta(S \leftarrow s) = S .$$

3. Factorisation en séquences croissantes.

Soit $Q = Q(\alpha, A)$ et pour m positif

$$\eta_m = \operatorname{sgn}(i' - j' - i + j)$$

où

$$(i, j) = Q^{-1} a_m \quad \text{et} \quad (i', j') = Q^{-1} a_{m+1} .$$

Par example, pour $Q = \frac{248}{679}$ on trouve que la suite $\eta_1, \eta_2, \ldots, \eta_5$ est égale à $+1$, -1, $+1$, $+1$, -1, les paires $(2,4)$, $(6,8)$, $(8,9)$ et $(4,6)$ illustrant respectivement les cas (1), (2), (3) et (4) énumérés plus bas. Le lien entre les Q-symboles de Schensted et le problème de Newcomb est fourni par la

REMARQUE 2. *Pour chaque m positif, $\eta_m = +1$ ou -1 selon que $\alpha a_m < \alpha a_{m+1}$ ou $> \alpha a_{m+1}$.*

DÉMONSTRATION. Pour $a_m > a_1$ fixe, définissons

$$\bar{\eta}_m = \operatorname{sgn}(\bar{i}' - \bar{j}' - \bar{i} + \bar{j})$$

où

$$(\bar{i}, \bar{j}) = (\Delta Q)^{-1} a_m, \qquad (\bar{i}', \bar{j}') = (\Delta Q)^{-1} a_{m+1}$$

et vérifions d'abord que $\eta_m = \bar{\eta}_m$ identiquement. Pour cela, il est commode de distinguer quatre cas:

1° $i = i'$, $j = j' - 1$;
2° $i > i'$, $j < j'$;
3° $i = i' - 1$, $j = j'$;
4° $i < i'$, $j > j'$.

Ce sont les seuls possibles, car le fait que Q est standard et que $a_m < a_{m+1}$ exclut $i' \leqq i$ et $j' \leqq j$, et, d'autre part, le fait que a_m et a_{m+1} sont deux éléments consécutifs de $\{Q\}$ exclut les cas $i = i'$ et $j < j' - 1$, $i < i'$ et $j < j'$ ou $i < i' - 1$ et $j = j'$ qui entraîneraient l'existence d'un élément $a = Q_{i,j'-1}$, $= Q_{i,j'}$, ou $= Q_{i'-1,j}$, respectivement, tel que $a_m < a < a_{m+1}$.

Dans les cas 2° et 4°, puisque $(\Delta Q)^{-1}a_m = (i,j)$, $(i-1,j)$ ou $(i,j-1)$ et $(\Delta Q)^{-1}a_{m+1} = (i',j')$, $(i'-1,j')$ ou $(i',j'-1)$ et puisque ΔQ est standard, on obtient directement l'égalité $\eta_m = \bar{\eta}_m$ cherchée. Traitons en détails les sous-cas suivant du cas 1°:

1.1° $i > 1$ et $(i-1,j+1)\big(=(i-1,j')\big) \in |U|$. Dans ce cas $(\Delta Q)^{-1}a_m = Q^{-1}a_m$ et $(\Delta Q)^{-1}a_{m+1} = (i-1,j')$ ou $=(i,j')$ selon que (i,j') appartient ou non à $|U|$. Donc $\eta_m = \bar{\eta}_m$.

1.2° $i > 1$ et $(i-1,j) \notin |U|$. Puisqu'il n'existe aucun élément de $\{Q\}$ entre a_m et a_{m+1} on a $Q_{i-1,j+1} < a_m$. Donc $(i-1,j+1) \in |U|$ et on est ramené au cas précédent.

1.3° $j > 1$ et $(i,j-1) \in |U|$. Ou bien $(\Delta Q)^{-1}a_m = (i,j)$ et $(\Delta Q)^{-1}a_{m+1} = (i',j')$ ou bien $(i,j) \in |U|$. Dans ce dernier cas le fait que a_m et a_{m+1} sont consécutifs entraîne $a_{m+1} < Q_{i+1,j+1}$ ou $0 = Q_{i+1,j+1}$; donc $(i',j') \in |U|$, $(\Delta Q)^{-1}a_{m+1} = (i,j)$, et enfin $\eta_m = \bar{\eta}_m$.

1.4° Dans tous les cas restants,

$$(\Delta Q)^{-1}a_m = (i,j) \quad \text{et} \quad (\Delta Q)^{-1}a_{m+1} = (i',j')\,.$$

Donc $\eta_m = \bar{\eta}_m$.

Ceci achève l'examen du cas 1° et le cas 3° pouvant être traité de façon absolument analogue, nous ne répéterons pas la discussion.

Considérons maintenant le Q-symbole Q_m relatif à α et à l'ensemble $\{a_m, a_{m+1}, \ldots\}$. De par la définition même des Q-symboles on a

$$Q_m{}^{-1}a_m = (1,1) \quad \text{et} \quad Q_m{}^{-1}a_{m+1} = (1,2) \text{ ou } = (2,1)$$

selon que $\alpha a_m < \alpha a_{m+1}$ ou $\alpha a_m > \alpha a_{m+1}$. Puisque, d'après la remarque 1, Q_m est égal à $\Delta^{m-1}Q$ pour chaque m positif la remarque 2 résulte directement de $\eta_m = \bar{\eta}_m$ par induction sur m.

Il resulte de cette remarque que si $a_i < a_{i+1} < \ldots < a_{i+m}$ est une séquence d'éléments consécutifs de A tels que $\alpha a_i < \alpha a_{i+1} < \ldots < \alpha a_{i+m}$, ces éléments figurent dans des colonnes *distinctes* de $Q(\alpha, A)$. En conjonction avec la formule $Q(\alpha, A) = P(\alpha^{-1}, B)$ vérifiée ci-dessous, ceci montre que les « modified standard tables » de Schensted [1, part II] n'ont aucune colonne possédant deux entrées positives égales.

4. La formule $Q(\alpha, A) = P(\alpha^{-1}, B)$.

Soient a_p et $a_{p'}$ deux éléments d'un sous-ensemble quelconque A' de A.

Nous définissons le déplacement $\mathrm{Dp}(a_p, a_{p'}, \alpha, A')$ de $\alpha a_{p'}$ par αa_p dans la construction de $P(\alpha, A')$ par les règles suivantes:

$$\mathrm{Dp}(a_p, a_{p'}, \alpha, A') = 0$$

si $p' > p$ ou si $p' < p$ et si $P(\alpha, A'_{p-1})^{-1} a_{p'} = P(\alpha, A_{p'})^{-1} a_{p'}$;

$$= \big((0,0), (i,j)\big)$$

$\big($où $(i,j) = P(\alpha, A_{p'})^{-1} a_p\big)$ si $p = p'$;

$$= \big((i',j'), (i,j)\big)$$

$\big($où $(i',j') = P(\alpha, A'_{p-1})^{-1} a_{p'}$ et $(i,j) = P(\alpha, A_{p'})^{-1} a_{p'}\big)$ si $p' < p$ et $i \neq i'$.

Dans les deux derniers cas on dira encore que αa_p *déplace* $\alpha a_{p'}$ *de* (i',j') *à* (i,j), la valeur $(0,0)$ de (i',j') pour $a_p = a_{p'}$ étant évidemment purement conventionnelle. De par la définition même de l'opération $\leftarrow \alpha a_p$ si $(i',j') \neq (0,0)$ on a nécessairement $i' = i+1$ et $j < j'$; en outre on observera que ce déplacement se produit si et seulement s'il existe $a'' \in A'$ tel que $\alpha a_p \leqq \alpha a'' < \alpha a_{p'}$ et que $\mathrm{Dp}(a_p, a'', \alpha, A') = \big((i'',j''), (i',j')\big)$. Donc, dans tous les cas,

$$\mathrm{Dp}(a_p, a_{p'}, \alpha, A') = \mathrm{Dp}\big(a_p, a_{p'}, \alpha, \{a_{p''} \in A' : p'' \leqq p;\ \alpha a_{p''} < \alpha a_p;\ \alpha a_p \leqq \alpha a_{p''}\}\big)\,.$$

Ces notations sont étendues de façon évidente à la bijection $\alpha^{-1}: B \to A$ et aux sous-ensembles B' de B.

Remarque 3. *Pour tout* $a, a' \in A$, *on a*

$$\mathrm{Dp}(a, a', \alpha, A) = \mathrm{Dp}(\alpha a', \alpha a, \alpha^{-1}, B)\,.$$

Démonstration. Il suffit évidemment de vérifier l'énoncé pour tous les ensembles finis et, procédant par induction, nous supposerons $A = A_n$, $B = B_n$ $(n < \infty)$ et que le résultat est déjà établi pour chacun des sous-ensembles propres de A.

Soit $a^* = \alpha^{-1} b_n$ où, comme toujours, $b_n = \max\{b : b \in B\}$. Si a, $a' \in A \setminus \{a^*\}$, on a rappelé plus haut que

$$\mathrm{Dp}(a, a', \alpha, A) = \mathrm{Dp}(a, a', \alpha, A \setminus \{a^*\})$$

et

$$\mathrm{Dp}(\alpha a', \alpha a, \alpha^{-1}, B) = \mathrm{Dp}(\alpha a', \alpha a, \alpha^{-1}, B_{n-1})\,.$$

Donc, dans ce cas, la relation cherchée se déduit immédiatement de l'hypothèse d'induction. En raison de la symétrie de l'énoncé (entre $\alpha: A \to B$ et $\alpha^{-1}: B \to A$), il ne reste à discuter que les deux cas où

1° soit $a = a^*$, $a' = a_n$;

2° soit $a = a_n$, $a' = a^*$, avec $a_n \neq a^*$.

Cas 1°. D'après la définition même de $\leftarrow \alpha a^*$ et le fait que αa^* est plus grand que tous les éléments de $\{P(\alpha, A)\}$, l'ensemble des éléments déplacés par αa^* se réduit à αa^* lui-même. Donc, si $a^* \neq a_n$, on a

$$0 = \mathrm{Dp}(a, a', \alpha, A) = \mathrm{Dp}(\alpha a', \alpha a, \alpha^{-1}, B) .$$

Au contraire, si $a^* = a_n$, $\mathrm{Dp}(a, a', \alpha, A) = \big((0,0), (1,j)\big)$ où j est le plus petit entier tel que $(1,j) \notin |\mathrm{P}(\alpha, A_{n-1})|$. La même observation vaut pour $\mathrm{Dp}(\alpha a', \alpha a, \alpha^{-1}, B)$ avec cette fois $(1,j') \notin |P(\alpha^{-1}, B_{n-1})|$ et l'égalité des deux déplacements résulte de l'hypothèse d'induction qui implique

$$|P(\alpha, A_{n-1})| = |Q(\alpha^{-1}, \alpha^{-1} A_{n-1})| = |P(\alpha^{-1}, B_{n-1})|$$

puisque dans le cas examiné ici $\alpha A_{n-1} = B_{n-1}$.

Cas 2°. Considérons d'abord le cas où

$$\mathrm{Dp}(a, a', \alpha, A) = \big((i,j), (i+1, \bar{j})\big) .$$

Ceci implique $P(\alpha, A_{n-1})^{-1} b_n = (i,j)$ et, par conséquent, l'existence de $x \in A_{n-1}$ tel que αx ait déplacé b_n de (i',j') (qui est éventuellement $(0,0)$) à (i,j).

En outre il doit exister $y \in A_{n-1}$ tel que $\alpha a_n \leqq \alpha y < b_n$ et que αa_n déplace αy de (i'',j'') à (i,j). De fait αy est le plus grand des éléments de B_{n-1} qui soit déplacé par αa_n. Appliquant l'hypothèse d'induction à $A \setminus \{a^*\} = a^{-1} B_{n-1}$, on en conclut que y est le dernier élément de B_{n-1} déplaçant $a = a_n$ dans $P(\alpha^{-1}, B_{n-1})$ et que par conséquent,

$$P(\alpha^{-1}, B_{n-1})^{-1} a = (i,j) .$$

De façon analogue, l'hypothèse d'induction appliquée à A_{n-1} montre que $a' = \alpha^{-1} b_n$ déplace x de (i',j') à (i,j) dans $P(\alpha^{-1}, A_{n-1})$.

Comme $x < a_n$ il en résulte que l'operation $\leftarrow a'$ déplace $a = a_n$ de (i,j) en $(i+1, \bar{\bar{j}})$ ce qui achève la vérification dans ce cas puisque, trivialement, $\bar{j} = \bar{\bar{j}}$, ces deux nombres ne dépendant que des formes

$$|P(\alpha, A_{n-1} \setminus \{a^*\})| \qquad \text{et} \qquad |P(\alpha^{-1}, B_{n-1} \setminus \{\alpha a_n\})|$$

qui sont identiques d'après l'hypothèse d'induction.

En raison de la symétrie, on a établi du même coup que $\mathrm{Dp}(a_n, a^*, \alpha, A) = 0$ si et seulement si $\mathrm{Dp}(b_n, \alpha a_n, \alpha^{-1}, B) = 0$ ce qui termine la vérification de la remarque.

Observons maintenant que la construction qui vient d'être discutée donne

$$P(\alpha^{-1}, \alpha A_n) = P(\alpha^{-1}, \alpha A_{n-1}) + (a_n)_{i,j}$$

où $(i,j) = |P(\alpha, A_n)| \setminus |P(\alpha, A_{n-1})|$. Donc, supposant déjà établi que $P(\alpha^{-1}, \alpha A_{n-1}) = Q(\alpha, A_{n-1})$, on a encore

$$P(\alpha^{-1}, \alpha A_n) = Q(\alpha, A_n)$$

et, par induction, dans tous les cas,

$$P(\alpha^{-1}, B) = Q(\alpha, A)$$

ce qui est la formule cherchée.

Donnons une application de cette remarque au cas particulier de $A = B$.

Propriété 2. *Une condition nécessaire et suffisante pour que* $\alpha : A \to A$ *soit une involution est que* $P(\alpha, A) = Q(\alpha, A)$.

Démonstration. Il est trivial que $A = B$ et $\alpha = \alpha^{-1}$ entraînent $P(\alpha^{-1}, B) = P(\alpha, A)$, c'est-à-dire $P(\alpha, A) = Q(\alpha, A)$ d'après la formule vérifiée dans cette section.

Réciproquement, supposons $P(\alpha, A) = Q(\alpha, A)$ et montrons qu'il en résulte $a_1 = b_1$, $\alpha a_1 = \alpha b_1$,

$$P(\alpha, A \setminus \{a_1, \alpha a_1\}) = Q(\alpha, A \setminus \{a_1, \alpha a_1\})$$

ce qui, par induction, établit la propriété.

Revenant aux notations de la fin de la section 2, nous distinguons deux cas selon que $i^* > 1$ ou $i^* = 1$.

1° $i^* > 1$. On a $\Delta Q(\alpha, A) = Q(\alpha, A \setminus \{a_1\})$ et l'on sait en déduire αa_1 et $P(\alpha, A \setminus \{a_1\})$. D'après la formule de la présente section

$$P(\alpha, A \setminus \{a_1\}) = Q(\alpha^{-1}, B \setminus \{\alpha a_1\}) .$$

Donc par la remarque 1 :

$$\Delta P(\alpha, A \setminus \{a\}) = \Delta Q(\alpha^{-1}, B \setminus \{\alpha a_1\}) = Q(\alpha^{-1}, B \setminus \{\alpha a_1, b_1\}) .$$

Partons maintenant de $Q(\alpha^{-1}, B)$ qui, toujours d'après la même formule, est égal à $P(\alpha, A)$. Répétant le même calcul que plus haut, on en déduit $Q(\alpha, A \setminus \{a^{-1} b_1, a_1\})$ qui est donc égal à $Q(\alpha^{-1}, B \setminus \{\alpha a_1, b_1\})$ d'après l'hypothèse $P(\alpha, A) = Q(\alpha, A)$. Une troisième application de la formule donne

$$Q(\alpha^{-1}, B \setminus \{\alpha a_1, b_1\}) = P(\alpha, A \setminus \{\alpha_1, \alpha^{-1} b_1\})$$

et le résultat est vrai dans ce cas.

2° $i^* = 1$. Dans ce cas les observations faites à la fin de la section 2 et la formule de la présente section donnent directement

$$\alpha a_1 = a_1 \quad \text{et} \quad P(\alpha, A \setminus \{a_1\}) = Q(\alpha, A \setminus \{a_1\}) .$$

La propriété est donc vérifiée dans tous les cas.

Examinons plus en détail le cas où $A = B \neq \emptyset$ est fini et $P(\alpha,A) = Q(\alpha,A)$ et, pour tout tableau standard S, dénotons par $\mathrm{Imp}\,|S|$ le nombre des $j \in N$ tels qu'il existe un nombre impair de $i \in N$ pour lesquels $(i,j) \in |S|$. Il résulte des définitions que, dans le cas 2° discuté plus haut,

$$|Q(\alpha,A)| \setminus |\Delta Q(\alpha,A)| = (i,j^*)$$

et qu'il n'existe pas d'autre $i \in N$ tels que $(i,j^*) \in |Q(\alpha,A)|$. Donc

$$\mathrm{Imp}\,|Q(\alpha,A \setminus \{a_1\})| = \mathrm{Imp}\,|Q(\alpha,A)| - 1 .$$

Dans le cas 1° soit $(i^*,j^*) = |Q(\alpha,A)| \setminus |\Delta Q(\alpha,A)|$ où par hypothèse $i^* > 1$. Soit $|U'|$ la séquence relative à l'opération Δ dans $P(\alpha,A \setminus \{a_1\})$. Il est facile de voir qu'il existe un entier k tel que $(i_{k'},j_{k'}) \in |U|$ entraîne $(i_{k'},j_{k'}) \in |U'|$ si $k' < k$ et $(i_{k'-1},j_{k'}) \in |U'|$ si $k' \geqq k$. Il s'en déduit que

$$|P(\alpha,A \setminus \{a_1\})| \setminus |\Delta P(\alpha,A \setminus \{a_1\})| = (i^* - i, j^*)$$

et par conséquent, d'après nos remarques antérieures

$$\mathrm{Imp}\,|Q(\alpha,A \setminus \{a_1,\alpha^{-1}b_1\})| = \mathrm{Imp}\,|Q(\alpha,A)| .$$

Par induction, l'on conclut de ces deux relations que si α est une involution sur l'ensemble *fini* A, le nombre des éléments laissés invariants par α est précisément égal à $\mathrm{Imp}\,|Q(\alpha,A)|$.

5. L'opération *I*.

Soit Q un tableau standard tel que $0 < \mathrm{Card}\,\{Q\} = n < \infty$. On définit $Q^I \in \mathscr{T}$ par l'équation

$$Q^I = \sum \{(a_k)_{i_k,j_k} \colon k \in [1,n]\}$$

où a_k désigne le k-ième élément de $\{Q\}$ par ordre croissant et où, ici $(i_k,j_k) = |\Delta^{n-k+1}Q| \setminus |\Delta^{n-k}Q|$ (avec $\Delta^0 Q = Q$). Trivialement, $\{Q\} = \{Q^I\}$, $|Q| = |Q^I|$ et $Q^{IT} = Q^{TI}$ où T indique la transposition. De plus si la bijection $\sigma \colon \{Q\} \setminus \{a_1\} \to \{Q\} \setminus \{a_n\}$ définie par $\sigma a_{k+1} = a_k$ pour $k \in [1,n-1]$ est étendue de façon naturelle à $\mathscr{T}$, on vérifie sans peine que

$$Q^I = \sigma(\Delta Q)^I + (a_n)_{i_n,j_n} .$$

On verra plus bas que Q^I est standard et $Q^{II} = Q$. Par exemple, pour $Q = \begin{smallmatrix} 248 \\ 679 \end{smallmatrix}$ comme plus haut,

$$Q^I = \begin{smallmatrix} 267 \\ 489 \end{smallmatrix}, \quad (\Delta Q)^I = \begin{smallmatrix} 478 \\ 69 \end{smallmatrix}, \quad \sigma(\Delta Q)^I = \begin{smallmatrix} 267 \\ 48 \end{smallmatrix} .$$

Soit maintenant $Q = Q(\alpha,A)$ où $0 < \mathrm{Card}\,A = n < \infty$. La bijection $\bar{\alpha} \colon A \to B$ étant définie par $\bar{\alpha} a_k = \alpha a_{n-k+1}$ pour $k \in [1,n]$, il a été prouvé par Schensted que

$$P(\bar{\alpha}, A) = P(\alpha, A)^T$$

[1, lemme 7, p. 186]. Nous vérifions par induction sur n que $Q(\bar{\alpha}, A) = Q(\alpha, A)^{IT}$ en observant que le résultat est vrai pour $n = 1$ et en supposant qu'il est déjà établi pour $A' = A \setminus \{a_1\}$.

Compte tenu de la relation $\bar{\alpha}\sigma A' = A'$, et écrivant comme d'habitude A_{n-1} pour $A \setminus \{a_n\}$, ceci revient à supposer

$$Q(\bar{\alpha}, A_{n-1}) = \sigma Q(\alpha, A')^{IT} .$$

Maintenant, $Q(\bar{\alpha}, A) = Q(\bar{\alpha}, A_{n+1}) + (a_n)_{i,j}$ et, comme on l'a noté plus haut,

$$Q(\alpha, A)^{IT} = \sigma\big(\Delta Q(\alpha, A)\big)^{IT} + (a_n)_{j_n, i_n} .$$

D'après la remarque 1, $\Delta Q(\alpha, A) = Q(\alpha, A')$ et par conséquent:

$$Q(\bar{\alpha}, A) = Q(\alpha, A)^{IT} - (a_n)_{j_n, i_n} + (a_n)_{i,j} .$$

Il suffit donc de vérifier $|Q(\bar{\alpha}, A)| = |Q(\alpha, A)^{IT}|$. Or ceci résulte immédiatement de $|Q(\alpha, A)^I| = |Q(\alpha, A)|$, de l'égalité de forme des P-symboles et des Q-symboles et de l'égalité $|P(\bar{\alpha}, A)| = |P(\alpha, A)^T|$ impliquée par l'identité de Schensted. La formule est donc établie.

RÉFÉRENCES

1. C. Schensted, *Longest increasing and decreasing subsequences*, Canad. J. Math. 13 (1961), 179–192.

FACULTÉ DES SCIENCES, POITIERS, FRANCE

STUDIES IN LOGIC

AND THE

FOUNDATIONS OF MATHEMATICS

L. E. J. BROUWER / E. W. BETH / A. HEYTING

EDITORS

Computer Programming and Formal Systems

Editors

P. BRAFFORT

and

D. HIRSCHBERG

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM

THE ALGEBRAIC THEORY OF CONTEXT-FREE LANGUAGES*

N. CHOMSKY
Massachusetts Institute of Technology

AND

M. P. SCHÜTZENBERGER
Harvard University

1. Linguistic Motivation

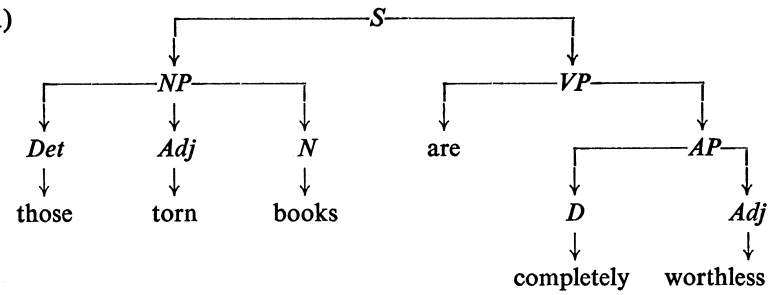
We will be concerned here with several classes of sentence-generating devices that are closely related, in various ways, to the grammars of both natural languages and artificial languages of various kinds. By a *language* we will mean simply a set of strings in some finite set V of symbols called the *vocabulary* of the language. By a *grammar* we mean a set of rules that give a recursive enumeration of the strings belonging to the language. We will say that the grammar *generates* these strings. (Thinking of natural languages, we would call the generated strings *sentences*; in algebraic parlance they would ordinarily be called *words* and the vocabulary would be called an *alphabet*; regarding a grammar as specifying a programming language, the strings would be called *programs*; we will generally use the neutral term *strings*).

For a class of grammars to have linguistic interest, there must be a procedure that assigns to any pair (σ, G), where σ is a string and G a grammar of this class, a satisfactory *structural description* of the string σ with respect to the grammar G. In particular, the structural description should indicate that the string σ is a well-formed sentence of the language $L(G)$ generated by G, where this is the case. If it is, the structural description should contain grammatical information that provides the basis for explaining how σ is understood by speakers who have internalized the grammar G; if it is not, the structural description might indicate in what respects σ deviates from well-formedness.

* This work was supported in part by the U. S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research; and in part by the National Science Foundation; and in part by a grant from the Commonwealth Fund.

We will be concerned with only one aspect of the structural description of a sentence, namely, its subdivision into phrases belonging to various categories. Thus, for example, a structural description of the English sentence "those torn books are completely worthless" should indicate that *those* is a *Determiner*, *torn* and *worthless* are *Adjectives*, *books* is a *Noun*, *completely* is an *Adverb*, *those torn books* is a *Noun Phrase*, *completely worthless* is an *Adjective Phrase*, *are completely worthless* is a *Verb Phrase*, the whole string is a *Sentence*, as well as additional details regarding subclassification. This information can be represented by a diagram such as (1):

(1)



or, equivalently, by a labelled bracketing of the string, as in (2):

(2) $[_S\,[_{NP}\,[_{Det}$ those$][_{Adj}$ torn$][_N$ books$]][_{VP}$ are $[_{AP}\,[_D$ completely$]$ $[_{Adj}$ worthless$]\,]\,]$.

A major concern of the general theory of natural languages is to define the class of possible strings (by fixing a universal phonetic alphabet); the class of possible grammars; the class of possible structural descriptions; a procedure for assigning structural descriptions to sentences, given a grammar; and to do all of this in such a way that the structural description assigned to a sentence by the grammar of a natural language will provide the basis for explaining how a speaker of this language would understand this sentence (assuming no limitations of memory, attention, etc.). The grammar, then will represent certain aspects of the linguistic competence of the speaker of the language.

We will not be concerned here with the empirical question of adequacy of the structural descriptions or the grammars that we will investigate. In fact, the classes of grammars that we will consider, and the kinds of

structural descriptions that they generate, are undoubtedly too narrow to do justice to real human linguistic competence. Nevertheless, the systems we consider (which, in effect, formalize traditional notions of parsing and immediate constituent analysis) bear certain relations to the kinds of systems that seem empirically adequate, and that are, for the time being, too complex to permit abstract study.[1])

In the representation (2), we have, aside from brackets, two kinds of symbols: (i) symbols of the generated string (i.e., the six symbols *those*, *torn*, *books*, *are*, *completely*, *worthless*)[2]); (ii) the symbols *S*, *NP*, *Det*, *Adj*, *N*, *VP*, *AP*, *D* representing phrase-categories. Symbols of type (i) we will call *terminals;* symbols of type (ii), *non-terminals.*

We will assume, below, a fixed stock of terminal and non-terminal symbols from which the grammars of all languages are constructed. The set of terminals can be regarded as constituting a potential common vocabulary for all languages. Thinking of spoken language, we can regard the set of terminals as defined by a universal phonetic alphabet (assuming, as is natural, an upper bound on the length of morphemes[2]). Thinking again of natural language, we can regard the fixed set of non-terminals as a universal set of categories from which the phrase types of all languages are drawn. An important and traditional question of general linguistics has to do with the possibility of giving a concrete interpretation of the non-terminals that constitute the categories in terms of which grammars are constructed — is it possible, in other words, to find a general definition, independent of any particular language, of such categories as Noun, Verb, etc., in terms of semantic content or formal properties of grammars? The problem of giving a concrete interpretation to the set of terminals and non-terminals is, of course, like the problem of empirical adequacy of certain categories of grammars, a crucial issue in the science of language; but it is beyond the range of our immediate interests here.

We can generate the sentence "those torn books are completely worthless", with the structural description (2), by the set of *rewriting rules*:

---

[1]) For further discussion of these questions, see Chomsky [10].

[2]) In a linguistically adequate grammar, we would generate not these symbols, but rather a more abstract representation using the symbols *the*, *demonstrative*, *plural*, *tear*, *participle*, *book*, *plural*, *be*, *plural*, *complete*, *ly*, *worth*, *less*, in this order. Representation in terms of these symbols (called *morphemes*) will be converted to phonetic representation by a set of *phonological rules* which will not concern us at all here. See Chomsky and Miller [15]. We will use actual sentences, such as (2), only for illustrative examples, and will therefore not be concerned with such refinements as this.

(3)
$$\begin{aligned} S &\to NP\ VP \\ NP &\to Det\ Adj\ N \\ Det &\to \text{those} \\ Adj &\to \text{torn} \\ Adj &\to \text{worthless} \\ N &\to \text{books} \\ VP &\to \text{are}\ AP \\ AP &\to D\ Adj \\ D &\to \text{completely} \end{aligned}$$

by a *derivation* that is constructed in the following way. First, write down the *initial symbol* S as the first line of the derivation. Form the $n+1^{st}$ line of the derivation by selecting at will an occurrence of a non-terminal α in the n^{th} line (where this occurrence of α is not labelling a bracket), and replace it by the string: $[_\alpha \varphi]$, where $\alpha \to \varphi$ is one of the rules of (3). Continue until the only non-terminals that appear are those that label brackets, at which point, the derivation is *terminated*. Deleting the brackets of a terminated derivation, with their labels, we have a string containing only terminals. Call this a *terminal string*. Four different terminal strings can be generated by the grammar (3). We can construct a grammar that generates infinitely many terminal strings, each with a structural description, by permitting recursions, e.g., by adding to (3) the rules

(4)
$$\begin{aligned} NP &\to \text{that}\ S \\ VP &\to \text{is}\ AP \\ AP &\to \text{obvious} \end{aligned}$$

in which case we can generate, e.g., "that those torn books are completely worthless is obvious", etc.[1]). Each of the generated sentences will again have a structural description of the appropriate kind.

Grammars of the type (3), (4) we will call *context-free (CF) grammars*. They are characterized by the property that exactly one non-terminal appears on the left-hand side of each rewriting rule. If this restriction is relaxed, we have systems with entirely different formal properties. It

[1]) In this case, infinitely many non-English sentences will also be generated, e.g., "that those torn books is obvious are completely worthless", etc. Hence the grammar ((3), (4)) is unacceptable. The difficulty of avoiding empirical inadequacies of this sort can easily be underestimated. We stress again that this is the key issue for both linguistics and psychology, though it will not concern us directly here. For discussion, see Chomsky [13].

seems that grammars for natural languages must contain at least some rewriting ryles of this more general form, and some rules that are not rewriting rules at all. Cf. Chomsky [8], [10], and [12], Chomsky and Miller [15], for further abstract discussion of systems of these sorts, which we will not consider further here. A set of terminal strings that can be generated by some *CF* grammar we will call a *CF language*.

A *CF* language may generate a terminal *(debracketised)* string φ with several different structural descriptions. In this case, if the grammar is empirically adequate, φ should be structurally ambiguous. Consider, for example, the *CF* grammar with the rules

(5)
$S \to NP\ VP$
$NP \to$ they; $NP \to Adj\ N$; $NP \to N$
$VP \to$ are NP; $VP \to$ Verb NP
Verb $\to$ are flying
$Adj \to$ flying
$N \to$ planes

With this grammar we can generate both (6) and (7):

(6) $[_S[_{NP}$ they$]\ [_{VP}[_{Verb}$ are flying$]\ [_{NP}[_N$ planes$]\ ]\ ]\ ]$.
(7) $[_S[_{NP}$ they$]\ [_{VP}$ are $[_{NP}\ [_{Adj}$ flying$]\ [_N$ planes$]\ ]\ ]\ ]$

Correspondingly, the terminal string "they are flying planes" is structurally ambiguous; it can mean: "my friends, who are pilots, are flying planes"; or: "those spots on the horizon are flying planes". Study of structural ambiguity is one of the most instructive ways to determine the empirical adequacy of a grammar.

We will see below that there are certain *CF* languages that are inherently ambiguous, in the sense that each *CF* grammar that generates them assigns alternative structural descriptions to some of their sentences. Furthermore, we will see that the problem of determining whether a *CF* grammar is ambiguous is recursively unsolvable,[1]) even for extremely simple types of *CF* grammars.

Though *CF* grammars are far from fully sufficient for natural languages, they are certainly adequate for the description of familiar artificial languages, and apparently for the description of certain, perhaps all, programming languages. In particular, a *CF* grammar can be written for

[1]) There is, in other words, no mechanical procedure (algorithm) for determining whether an arbitrary *CF* grammar assigns more than one structural description to some string that it generates.

ALGOL [18], and each program in ALGOL will be one of the terminal strings generated by this grammar. Clearly, a programming language must be unambiguous. Therefore, it is important to determine whether, in fact, a particular programming language meets this condition, or whether a particular infinite set of programs *can* each be unambiguous, given certain techniques for constructing them (e.g., techniques that can be represented as rules for constructing derivations in a *CF* grammar). As indicated in the preceding paragraph, these may be rather difficult questions.

Suppose that G_1 and G_2 are generative systems that specify certain techniques for constructing computer programs; suppose, in fact, that they are grammars that generate the programming languages L_1 and L_2, each of which consists of an infinite number of strings, each string being a possible program. It is often interesting to inquire into the relative power of programming languages. We will see that if G_1 and G_2 are *CF* grammars (as, e.g., in the case of ALGOL), most problems concerning the relation of L_1 to L_2 are recursively unsolvable, in particular, the problem of determining whether L_1 and L_2 have an empty or an infinite intersection, or whether L_1 is contained in L_2 [2], or whether there is a finite transducer (a "compiler") that maps L_1 onto L_2 (Ginsburg and Rose, personal communication). Hence it is possible that general questions concerning the formal properties of *CF* systems and formal relations between them may have a concrete interpretation in the study of data-processing systems, as well as in the study of natural language. This possibility has been pointed out particularly by Ginsburg and Rice [18], Ginsburg and Rose [19].

In considering a grammar as a generative device, we may be concerned with the language (i.e., set of terminal strings) that it generates, or with the set of structural descriptions that it generates (N.B.: each structural description uniquely determines a terminal string, as in (2)). The latter is clearly the much more interesting question. Similarly, in studying generative capacity of a class of grammars (or relative capacity of several such classes, as in evaluating alternative linguistic theories), we may be concerned either with the set of languages that can be generated, or with the set of systems of structural descriptions that can be generated. The latter, again, is a more interesting, but much more difficult question. Investigation of such questions is, altogether, quite recent, and attention has been restricted almost exclusively to generation of languages rather than of systems of structural descriptions. We will consider genera-

tion from a point of view intermediate between the two just mentioned. We will consider a representation of a language not as a set of strings and not as a set of structural descriptions, but as a set of pairs (σ, n), where σ is a string and n expresses its degree of ambiguity; that is, n is the number of different structural descriptions assigned to σ by the grammar G generating the language to which it belongs.

2. Grammars as Generators of Formal Power Series

2.1. Suppose that we are given a finite vocabulary V partitioned into the sets V_T (= terminal vocabulary) and V_N (= non-terminal vocabulary). We consider now languages with the vocabulary V_T, and grammars that take their non-terminals from V_N. Let $F(V_T)$ be the free monoid generated by V_T, i.e., the set of all strings in the vocabulary V_T. A language is, then, a subset of $F(V_T)$.

Consider a mapping r which assigns to each string $f \in F(V_T)$ a certain integer $\langle r, f\rangle$. Such a mapping can be represented by a *formal power series* (denoted also by r) in the non-commutative variables x of V_T. Thus

$$r = \sum_i \langle r, f_i\rangle f_i = \langle r, f_1\rangle f_1 + \langle r, f_2\rangle f_2 + \ldots, \tag{8}$$

where $f_1, f_2, \ldots$ is an enumeration of all strings in V_T. We define the support of r (= Sup(r)) as the set of strings with non-zero coefficients in r. Thus

$$\mathrm{Sup}(r) = \{f_i \in F(V_T) \mid \langle r, f_i\rangle \neq 0\}. \tag{9}$$

We do not insist that the coefficients $\langle r, f_i\rangle$ of the formal power series r in (8) be positive. If, in fact, for each i, $\langle r, f_i\rangle \geq 0$, then we shall say that r is a *positive* formal power series.

If for each $f_i \in F(V_T)$, the coefficient $\langle r, f_i\rangle$ is either zero or one, we say that r is the *characteristic* formal power series of its support.

2.2. If r is a formal power series and n an integer, we define the product nr as the formal power series with coefficients $\langle nr, f\rangle = n\langle r, f\rangle$, where $\langle r, f\rangle$ is the coefficient of f in r. Where r and r' are formal power series, we define $r + r'$ as the formal power series with coefficients $\langle r + r', f\rangle = \langle r, f\rangle + \langle r', f\rangle$, where $\langle r, f\rangle$ and $\langle r', f\rangle$ are, respectively, the coefficients of f in r and r'. We define rr' as the formal

power series with coefficients $\langle rr', f\rangle = \sum_{i,j} \langle r, f_i\rangle \langle r', f_j\rangle$, where $f_i f_j = f$. Thus the set of formal power series form a ring closed under the operations: multiplication by an integer, addition, multiplication.

Notice that where r and r' are positive formal power series the support of $r + r'$ is exactly the set union of the supports of r and r', and the support of rr' is exactly the set product of the supports of r and r' (i.e., the set of all strings $f_i f_j$ such that f_i is in the support of r and f_j in the support of r'). We will discuss the interpretation of other simple set theoretic operations below.

We say that two formal power series r and r' are *equivalent mod degree n* (i.e., $r \equiv r'$ (mod deg n)) if $\langle r,f\rangle = \langle r',f\rangle$ for every string f of length ("degree") $\leq n$. Suppose then that we have an infinite sequence of formal power series $r_1, r_2, \ldots$, such that for each n and each $n' > n$, $r_n \equiv r_{n'}$ (mod deg n). In this case, the limit r of the sequence $r_1, r_2, \ldots$ is well-defined as

$$r = \lim_{n \to \infty} \pi_n r_n \tag{10}$$

where for each n, $\pi_n r_n$ is the polynomial formed from r_n by replacing all coefficients of strings of length $> n$ by zero. Then the ring of the formal power series becomes an *ultrametric*, hence topological, ring.

With these notions defined, we can turn to the problem of relating the representation of languages in terms of formal power series to the representation of languages by generative processes such as *CF* grammars.

2.3. Suppose that G is a generative process generating the language $L(G)$. Each string $f \in F(V_T)$ is assigned a certain number $N(G,f)$ of structural descriptions by G; $N(G,f) > 0$ just in case $f \in L(G)$. $N(G,f)$ expresses the degree of structural ambiguity of f with respect to G. It is natural to associate with G the formal power series $r(G)$ such that $\langle r(G),f\rangle = N(G,f)$, where $\langle r(G),f\rangle$ is the coefficient of f in $r(G)$. Thus $r(G)$ expresses the ambiguity of all terminal strings with respect to the grammar G. The coefficient $\langle r(G),f\rangle$ is *zero* just in case f is not generated by G; it is *one* just in case f is generated unambiguously (in one and only one way) by G; it is *two* just in case there are two different structural descriptions for f, in terms of G; etc.

An $r(G)$ associated with a grammar G will, of course, always be positive; and its support Sup($r(G)$) will be exactly the language $L(G)$ generated by G. We can regard a formal power-series r with both positive and negative coefficients as being associated with *two* generative processes

G_1 and G_2. The coefficient $\langle r,f\rangle$ of f in r can be taken as the difference between the number of times that f is generated by G_1 and by G_2; that is, in this case, $\langle r,f\rangle = N(G_1,f) - N(G_2,f)$.

Suppose that G is a CF grammar with non-terminals $\alpha_1, \ldots, \alpha_n$, where α_1 is the designated initial symbol (i.e., $\alpha_1 = S$, in the example (1), above). We can construct the formal power series $r(G)$ associated with G by a straightforward iterative procedure. To do this, we proceed as follows.

Observe, first of all, that G can be written as a system of equations in the variables $\alpha_1, \ldots, \alpha_n$. Let $\varphi_{i,1}, \ldots, \varphi_{i,m_i}$ be the strings such that $\alpha_i \to \varphi_{i,j}$ $(1 \leq j \leq m_i)$ are rules of G. We then associate with α_i the *polynomial expression* σ_i,

$$\sigma_i = \varphi_{i,1} + \varphi_{i,2} + \ldots + \varphi_{i,m_i} \tag{11}$$

We now associate with the grammar G the set of equations

$$\alpha_1 = \sigma_1; \ldots; \alpha_n = \sigma_n. \tag{12}$$

Let us assume that the grammar G contains no rules of the form

$$\begin{aligned} &\alpha_i \to e \\ &\alpha_i \to \alpha_j. \end{aligned} \tag{13}$$

It is clear that these assumptions do not affect generative capacity [2]. That is, for every CF grammar containing such rules there is another grammar without any such rules, which generates the same language. We will also explicitly require, henceforth, that if G is a CF grammar and α is a non-terminal of G, then there must be terminal strings derivable from α — i.e., if G' contains the rules of G and has α as its initial symbol, then the language generated by G' must be non-null. Again, this requirement obviously does not affect generative capacity.

Returning now to the problem of constructing the power-series that is associated with G and that represents the degree of ambiguity that G assigns to each string, observe that we can regard each equation $\alpha_i = \sigma_i$ of (12) as defining a mapping ψ_i that carries an n-tuple $(r_1, \ldots, r_n)$ of power series into the power series defined by replacing α_j in σ_i by r_j. This is legitimate because of the closure properties of the ring of power series noted above in § 2.2.

Thus the set of equations (12) defines a mapping ψ,

$$\psi(r_1, \ldots, r_n) = (r'_1, \ldots, r'_n), \text{ where } r'_1 = \psi_i(r_1, \ldots, r_n). \tag{14}$$

Consider now the infinite sequence of n-tuples of power series ϱ_0, $\varrho_1, \ldots$, where

$$\begin{aligned} \varrho_0 &= (r_{0,1}, \ldots, r_{0,n}) = (0, \ldots, 0) \\ \varrho_1 &= (r_{1,1}, \ldots, r_{1,n}) \\ \varrho_2 &= (r_{2,1}, \ldots, r_{2,n}) \end{aligned} \tag{15}$$

and where for each i,j $(j > 0)$

$$r_{j,i} = \psi_i (r_{j-1,1}, \ldots, r_{j-1,n}), \tag{16}$$

and where 0 is the power series in which all coefficients are zero. Each $r_{j,i}$ in (15) has only finitely many non-zero coefficients; it is, in other words, a polynomial. Furthermore, we can show that for each i,j,j' such that $j' > j > 0$, $1 \leq i \leq n$, it is the case that

$$r_{j,i} \equiv r_{j',i} \pmod{\deg j}. \tag{17}$$

Consequently, as noted in § 2,2, the limit $r_{\infty,j}$ of the infinite sequence $r_{1,i}, r_{2,i}, \ldots$ is well-defined for each i (it is, of course, in general not a polynomial). We will call the n-tuple $(r_{\infty,1}, \ldots, r_{\infty,n})$, so defined, *the solution* to the set of equations (12). Indeed, the n-tuple $(r_{\infty,1}, \ldots, r_{\infty,n})$ is the only n-tuple within our framework to satisfy the set of equations (12). For this reason we will say that a power series is *algebraic* [42] if it is one of the terms of a solution to a set of equations such as (12), where there is no restriction on the sign of the numerical coefficients. We will call a power series *context-free* if the coefficients in the defining equations are all positive.

In particular, $r_{\infty,1}$, which we will henceforth call *the power series generated by* the grammar G of (12) with initial symbol α_1, is the power series associated with G in the manner described at the outset of § 2.3. Its support is the language $L(G)$ generated by G, and the coefficient $\langle r_{\infty,1}, f \rangle$ of a string $f \in F(V_T)$ determines the ambiguity of f with respect to G, in the way described above.

Notice that if an algebraic power series is context-free, it is positive, but not necessarily conversely. That is, a power series may be a term of the solution to a set of equations and may have only positive coefficients, but may not be a term of the solution to any set of equations with only positive coefficients.[1])

[1]) For example, using notions which will be defined below in § 3.1, the Hadamard square $s \odot s$, for $s \in \lambda_0$, has only positive coefficients (and has the same support as s) but it is not, in general, generated by a set of equations with only positive coefficients.

2.4. As examples of the process described above, consider the two grammars (18) and (19):

(18) $S \to bSS;\ S \to a$

(19) $S \to SbS;\ S \to a.$

Each of these grammars has only a single non-terminal; hence the corresponding set of equations will in each case consist of a single equation. Corresponding to (18) we have (20), and corresponding to (19) we have (21).

(20) $S = a + bSS$

(21) $S = a + SbS.$

The equations (19) and (20) correspond to (12), above, with $n = 1$. Both (19) and (20) meet the condition (13).

Consider first the grammar (18) represented in the form (20). Proceeding in the manner of the preceding section, we regard (20) as defining a mapping ψ such that $\psi(r) = a + brr$, where r is a power series. We then (corresponding to (15)) form the infinite sequence $\varrho_0, \varrho_1, \varrho_2, \ldots$ as follows:

(22)
$$\begin{aligned}
&\varrho_0 = r_0 = 0\\
&\varrho_1 = r_1 = a + br_0r_0 = a + b00 = a\\
&\varrho_2 = r_2 = a + br_1r_1 = a + baa\\
&\varrho_3 = r_3 = a + br_2r_2 = a + b(a + baa)(a + baa)\\
&\qquad\qquad = a + baa + babaa + bbaaa + bbaabaa\\
&\varrho_4 = r_4 = a + br_3r_3\\
&\ldots\ \ldots\\
&\ldots\ \ldots
\end{aligned}$$

Clearly for each j, j' such that $j' > j > 0$, we have $r_j \equiv r_{j'} \pmod{\deg j}$. Consequently the limit r_∞ is well-defined. This power series is the solution to equation (20), and its support is the language generated by the *CF* grammar (18). Notice that the power series r_∞, in this case, is characteristic, and its support is the set of *well-formed formulas* of the "*implicational calculus*" with one variable in parenthesis-free (Polish) notation (with the symbol a playing the role of propositional variable, and b the role of the operator "conditional").

Consider now the grammar (19) represented in the form (21). We regard (21) as defining a mapping ψ such that $\psi(r) = a + rbr$, where r is a power series. We now form the infinite sequence $\varrho_0, \varrho_1, \varrho_2, \ldots$:

(23)
$$\begin{aligned}
&\varrho_0 = r_0 = 0\\
&\varrho_1 = r_1 = a + r_0 b r_0 = a + 0b0 = a\\
&\varrho_2 = r_2 = a + r_1 b r_1 = a + aba\\
&\varrho_3 = r_3 = a + r_2 b r_2 = a + (a + aba)b(a + aba)\\
&\qquad\qquad = a + aba + 2ababa + abababa\\
&\varrho_4 = r_4 = a + r_3 b r_3\\
&\qquad = a + aba + a(ab)^2a + 5(ab)^3a + 6(ab)^4a + 6(ab)^5a +\\
&\qquad\quad 4(ab)^6a + (ab)^7a\\
&\ldots\ \ \ldots\\
&\ldots\ \ \ldots
\end{aligned}$$

Again, for each j, j' such that $j' > j > 0$, we have $r_j = r_{j'}$, (mod deg j), and the limit r_∞ is defined as the power series

$$(24)\quad r_\infty = \sum_n \begin{bmatrix} 2n \\ n \end{bmatrix} \frac{1}{n+1}\ (ab)^n a = a + aba + 2(ab)^2a + 5(ab)^3a + 14(ab)^4a + 42(ab)^5a + \ldots$$

where
$$\begin{bmatrix} 2n \\ n \end{bmatrix} = \frac{2n \times 2n - 1 \times \ldots \times n + 1}{1 \times 2 \times \ldots \times n}$$

The power series r_∞ of (24) is the solution to the equation (21), and its support is the language generated by the grammar (19). It is not, in this case, a characteristic power series. Taking the symbol a again as a propositional variable and b as the sign for "conditional", the grammar (19) is the set of rules for generating the well-formed formulas of the implicational calculus with one variable in ordinary notation, but with the parentheses omitted. The structural descriptions generated by (19) in the manner described in section 1 (cf. (3)) are of course unambiguous, since brackets are preserved, but the terminal strings formed by debracketization are ambiguous, and the degree of ambiguity of each generated terminal string is exactly its coefficient in r_∞ — thus *ababa* can be interpreted in two ways, either as $(ab(aba))$ or $((aba)ba)$, etc. A more general case has been treated by Raney [38] by Lagrange's inversion formula.

In (20) and (21) all coefficients are positive and the solution is therefore a positive power series. Consider, however, the set of equations consisting of the single member

(25)
$$S = a - SbS.$$

In this case we have the sequence

(26)
$$\begin{aligned}
&\varrho_0 = r_0 = 0\\
&\varrho_1 = r_1 = a - r_0 b r_0 = a - 0b0 = a\\
&\varrho_2 = r_2 = a - r_1 b r_1 = a - aba\\
&\varrho_3 = r_3 = a - r_2 b r_2 = a - (a - aba)b(a - aba)\\
&\qquad\qquad = a - aba + 2ababa - abababa\\
&\cdot\\
&\cdot\\
&\cdot
\end{aligned}$$

In fact the coefficients in ϱ_i of (26) are exactly those of ϱ_i of (23) except for sign — the coefficient of f in ϱ_i of (26) is positive just in case f has an even number of b's.

The power series r_∞ which is the solution to (25) is not positive and it is consequently not context-free (though its support happens to be a context-free language, in this case, in fact, the language with (19) as one of its grammars). We can, however, regard r_∞ as the difference between two context-free power-series $r_\infty{}^+$ and $r_\infty{}^-$; and, correspondingly, we can regard its support as the set of strings that are not generated the same number of times by a pair of CF grammars G^+ and G^- which generate $r_\infty{}^+$ and $r_\infty{}^-$, respectively. Suppose we set $S = S^+ - S^-$, so that (25) becomes

(27)
$$\begin{aligned}
S^+ - S^- &= a - (S^+ - S^-)b(S^+ - S^-)\\
&= a - (S^+bS^+ - S^+bS^- - S^-bS^+ + S^-bS^-)\\
&= a + S^+bS^- + S^-bS^+ - (S^+bS^+ - S^-bS^-).
\end{aligned}$$

Consider now the set of equations

(28) (i) $S^+ = a + S^+bS^- + S^-bS^+$

(ii) $S^- = S^+bS^+ + S^-bS^-.$

This is a set of positive equations with two variables S^+ and S^-, and it will have as solution the pair $(r_\infty{}^+, r_\infty{}^-)$, where $r_\infty{}^+$ is the limit of the sequence $r_0{}^+, r_1{}^+, \ldots$ and $r_\infty{}^-$ the limit of the sequence $r_0{}^-, r_1{}^-, \ldots$ of (29):

(29)
$$\begin{aligned}
&\varrho_0 = (r_0{}^+, r_0{}^-) = (0, 0)\\
&\varrho_1 = (r_1{}^+, r_1{}^-) = (a, 0)\\
&\varrho_2 = (r_2{}^+, r_2{}^-) = (a, aba).
\end{aligned}$$

It is clear that where r_∞ is the solution to (25), $r_\infty = r_\infty{}^+ - r_\infty{}^-$. But, furthermore, $r_\infty{}^+$ is the power series generated by the CF grammar G^+ with the initial symbol S^+ and the grammar (28i); and $r_\infty{}^-$ is the power

series generated by the *CF* grammar G^- with the initial symbol S^- and the grammar (28ii).

In a similar manner, any algebraic power series can be represented (in infinitely many different ways) as the difference of two context-free power series, and its support can be regarded, therefore, as the set of strings which are not generated the same number of times by two *CF* grammars. This is as close as we can come to a concrete interpretation for the general notion of algebraic power series.

More generally, the same construction could be carried out for an arbitrary ring of coefficients instead of the ring of natural numbers used above. This is a still unexplored domain. For instance, if the coefficients are taken modulo a prime p (i.e., if we consider as "non-produced" the strings produced a multiple of p times), the formal power series $\Sigma_{n>0} z^{p^n}$ in the single terminal z is algebraic [27], although its support cannot be the support of any of the power series introduced above.

3. Further Operations on Formal Power Series

3.1. In § 2.2 we observed that the set of power series is closed under the operations of addition, product, and multiplication by an integer. We pointed out that the support of $r + r'$ is the union of the supports of r and r', and that the support of rr' is the set product of the supports of r and r', provided that the coefficients are non-negative. We will now turn to two other operations under which the set of power series is closed, and consider the corresponding set-theoretic interpretation for the supports.

It is standard terminology to say that r is *quasi-regular* if $\langle r, e\rangle = 0$. Then $r^{n'} \equiv 0 \pmod{\deg n}$ for $0 < n < n'$ and the element $r^* = \lim_{n\to\infty} \Sigma_{0<n'<n} r^{n'}$ is well-defined. Furthermore, r^* satisfies the identity

$$r + r^*r = r + rr^* = r^*, \tag{30}$$

which determines it uniquely. Thus r^* is usually called the *quasi-inverse* of r. This notion relates directly to the more familiar notion of an inverse by the remark that if $r' = e - r$ and $r'' = e + r^*$, then $r'r'' = (e - r)(e + r^*) = e - r + r^* - rr^* = e = r''r'$, that is, $r'' = r'^{-1}$. Conversely, given r' such that $\langle r', e\rangle = 1$, we can write it as $r' = e - r$, where r is quasi-regular, so that $e + r^*$ is the inverse of r''.

Note that by the very definition of r^*, this power series has only non-

negative coefficients if r does, and that Sup $r^* = (\text{Sup } r)^*$, where on the right side of the equation the star denotes Kleene's star operation [21].

In particular if V is an arbitrary set of letters and if the power series v is defined by $\langle v, x\rangle = 1$ if $x \in V = 0$ if $x \notin V$ (i.e., if v is the characteristic function of V), $e + V^*$ (in Kleene's sense) is the set of all words generated by the letters of V and $e + v^* = (e - v)^{-1}$ is the characteristic function of this set. This follows from the fact that any word $f \in V^*$ appears once and only once in the infinite sum $\Sigma_{n>0} V^n$. Consequently, when we know the characteristic function r of a set of strings, we are able to write also the characteristic function $(1 - V_T)^{-1} - r$ of its complement.

It is worth mentioning that in this case the latter has non-negative coefficients and although it is algebraic in the sense defined above, it is not necessarily context-free.

The second operation that we define is the *Hadamard product*, thus generalizing in one of the possible ways the usual notion of classical analysis. The definition that we give differs from the various extensions to the case of several variables that occur in the literature, but it seems to be most natural extension for non-commutative power series.

Where r and r' are two power series, their *Hadamard product* $r \odot r'$ will be the power series with coefficients

$$\langle r \odot r', f\rangle = \langle r, f\rangle \langle r', f\rangle \tag{31}$$

identically for all strings f. Hence Sup $(r \odot r') = (\text{Sup } r) \cap (\text{Sup } r')$, and $r \odot r'$ is a characteristic function if r and r' are.

Finally we introduce the following notation: given a string $x_{i_1} x_{i_2} \ldots, x_{i_{n-1}} x_{i_n} = f (x_{i_j} \in V)$ we define $\tilde{f}$ (the mirror image of f) to be the string

$$\tilde{f} = x_{i_n} x_{i_{n-1}} \cdots x_{i_2} x_{i_1} \tag{32}$$

Clearly $\tilde{\tilde{f}} = f$ and the relation $ff' = f''$ implies $\tilde{f}'' = \tilde{f}'\tilde{f}$. Formally this mapping is an *involutory anti-automorphism* of the ring of power series and it can be proved to be uniquely characterized by this property (up to a permutation of the elements of V).

3.2. The notation just introduced will be used later on for simplifying the description of grammars in the following way. Suppose that a grammar G contains the rules

$$\begin{aligned} \alpha_1 &= \pi_1 \alpha_2 \pi_2 + \pi_1 \pi_2 + \pi_3 \\ \alpha_2 &= \alpha_2 \pi_4 + \pi_4 \end{aligned} \tag{33}$$

where the π_j's are polynomial expressions in $V - \{\alpha_1\}$. Then the second rule implies

$$\alpha_2 = \left(\sum_{n>0} \pi_4{}^n\right) \tag{34}$$

and the rules (33) can be replaced by the simpler rule

$$\alpha_1 = \pi_1(1 - \pi_4)^{-1}\pi_2 + \pi_3. \tag{35}$$

We can, in fact, give a linguistic interpretation to this simplified form of description. Thus, for example, a pair of rules of the form $\alpha_1 \to f_1\alpha_2 f_2$, $\alpha_2 \to \alpha_2\alpha_2$ (that is, a pair which can now be given in the form: $\alpha_1 \to f_1(1 - \alpha_2)^{-1}f_2$) can be regarded as constituting, in effect, a rule schema: $\alpha_1 \to f_1\alpha_2{}^n f_2$ ($n = (1, 2, \ldots)$. With this reinterpretation, the grammar, though still finitely specified by rule schemata, consists of an infinite number of rules. But now recall the manner in which a structural description (a labelled bracketing) is assigned to a terminal string generated by a *CF* grammar (see above, § 1). A grammar specified by the rule schema given above can generate a structural description of the form

$$\text{--- } [_{\alpha_1} f_1 [_{\alpha_2} p_1][_{\alpha_2} p_2] \ldots [_{\alpha_2} p_n] f_2] \text{ ---} \tag{36}$$

for each n, where each p_k is derived from α_2. In the sentence (terminal string) with this structural description, each $\tilde{p}_k$ is a phrase of type α_2, where $\tilde{p}_k$ is formed by debracketization of p_k. The successive phrases $\tilde{p}_1, \ldots, \tilde{p}_n$ form a "coordinate construction", which, taken together with the strings formed ultimately from f_1 and f_2, is a construction of the type α_1. This is the natural way to extend *CF* grammars to accommodate true coordination, as, e.g., where a string of adjectives of arbitrary length may appear in predicate position with no internal structure defined on them. Cf. Chomsky [10].

3.3. Let us try to relate what we have done so far to classical analysis, writing $\varphi f = \varphi f'$ for any two strings f and f' if they contain exactly the same number of each of the letters (terminal or not).

Clearly φ extends to a mapping of our non-commutative power series onto the ring of the ordinary (commutative) *formal* power series with integral coefficients, and it is easily seen that φ is a homomorphism. For example, if $\alpha = a + b\alpha\alpha$, we have $\varphi\alpha = \varphi a + \varphi b\varphi\alpha\varphi\alpha$, and $\varphi\alpha$ is the ordinary power series

$$\varphi\alpha = (\varphi a)^{n+1}(\varphi b)^n \begin{bmatrix} 2n \\ n \end{bmatrix} \frac{1}{n+1} \tag{37}$$

in the ordinary variables φa, φb. (Here if $\alpha' = a + \alpha' b \alpha'$, we would also have $\varphi\alpha' = \varphi\alpha$).

Furthermore, it can be shown directly from the way our power series are obtained that the coefficients do not grow faster than an exponential function of the degree (length) of the strings. Thus the image φ of any one of our power-series is in fact an ordinary convergent Taylor series expansion of an algebraic function.

Reciprocally, if we are given (ordinary) variables $\bar{x}_1, \ldots, \bar{x}_n$, an (ordinary) algebraic function of this quantity $\bar{y}$ is defined by a polynomial in $\bar{y}$ and the $\bar{x}_i$; and in case $\bar{y}$ admits a development in Taylor series (with integral coefficients) around zero in the $\bar{x}_i$'s, we can associate with it infinitely many formal power series β such that $\varphi\beta = \bar{y}$ and β is defined by formal equations. For instance: starting from the algebraic function $\bar{y}$ of $\bar{a}$ and $\bar{b}$ defined by $\bar{y}^2\bar{b} - \bar{y} + a = 0$, we obtain the two examples given above, and also formal power series

$$\alpha = a + b\alpha\alpha + \pi\alpha - \alpha\pi \tag{38}$$

where π is an arbitrary polynomial in a and b. Thus, e.g., take $\pi = b$. Then

$$\begin{aligned} &\alpha_0 = a \\ &\alpha_1 = a + baa + ba - ab \\ &\ldots\ldots \\ &\text{etc.} \end{aligned} \tag{39}$$

3.4. Let us conclude by indicating some connections between our considerations and Lyndon's theory of equations in a free group (Lyndon, 1960). Let $\{x_i\}$ $(1 \leq i \leq n)$ be a terminal vocabulary, ξ a non-terminal letter and let w be a product of terms of the form $1 - x_i$, $(1 - x_i)^{-1}$, $1 - \xi$, $(1 - \xi)^{-1}$. We define $\deg(w) = d_+ - d_-$ where d_+ and d_- are the number of factors $1 - \xi$ and $(1 - \xi)^{-1}$ in w. Thus, for instance, for $w = (1 - x_2)(1 - x_1)(1 - \xi)(1 - x_i)^{-1}(1 - \xi_i)^{-1}(1 - x_2)^{-1}$, one has $\deg(w) = 1 - 1 = 0$.

As is well-known, the elements $1 - x_i$ generate (by multiplication) a free group G. The relation $w = 1$ may be considered as an equation in the unknown ξ. In our terminology a *solution* of $w = 1$ would be a power series ξ_0 in the x_i's such that $w = 1$ identically when ξ_0 is substituted for ξ in w; ξ_0 will be a *group solution* if, furthermore, $1 - \xi_0 \in G$; i.e., if $1 - \xi_0$ is itself expressible as a product of terms $(i - x_i)^{\pm 1}$. R.C. Lyndon

has proven the very remarkable result that the *totality* of the *group solutions* can be obtained algorithmically.

Let us relate part of this question to our remarks in § 2.3. For this we introduce the new symbols $\xi_i (1 \leq i \leq n)$, η, and equations

$$\xi_i = x_i + \xi_i x_i; \quad y = \xi^2 + \xi\eta \tag{1}$$

so that $(1 - x_i)^{-1} = 1 + \xi_i$ and $(1 - \xi)^{-1} = 1 + \xi + \xi^2 + \xi\eta$

Substituting these expressions in $w = 1$ and simplifying, we obtain a relation

$$(\deg (w))\, \xi = p' \tag{2}$$

where p' is a polynomial in the variables x_i, ξ_i, η having no term of degree less than 2.

Hence if deg $(w) \neq 0$ the system (1), (2) has one and only one solution in power series (the fact that the coefficients are eventually rational instead of integral numbers is irrelevant to the proof in § 2.3) and since the group solutions are a subset of the power series solutions, we have verified directly that if deg $w \neq 0$, the free group equation $w = 1$ has at most one solution.

On the contrary, if deg $w = 0$ (as for instance for the equation $w = (1 - \xi)(1 - x_i)(1 - \xi)^{-1}(1 - x_i)^{-1} = 1$) our approach entirely collapses and says nothing even about the unrestricted solutions of $w = 1$.

For instance $(1 - x_i)(1 - \xi)(1 - x_i)^{\varepsilon}(1 - \xi)^{-1} = 1$ has no solution if $\varepsilon \neq -1$ and has an infinity of *group* solutions if $\varepsilon = -1$, viz. $1 - \xi = (1 - x_i)^{\pm n}$ $(n > 0)$. Indeed, then, the equation can equivalently be written $\xi x_1 = x_1 \xi$ (which has as solutions, in our sense, all the power series in x_1).

Of course, the case deg $w = 0$ is precisely that in which, the unknown $1 - \xi$ disappears when taking the commutative image as in § 3.3 and it is the non-trivial case from a group theoretic point of view.

4. Types of *CF* Grammars and Their Generative Properties

4.1. In terms of conditions on the rules that constitute them, we can define several categories of *CF* grammars that are of particular interest. In the following we will use $\alpha, \beta, \ldots$ for non-terminal symbols; $f, g, \ldots$ for terminal strings (possibly null); and φ, ψ for arbitrary strings. Recall that we have

excluded the possibility of rules of the form $\alpha \to e$ or $\alpha \to \beta$, remarking that this restriction does not affect generative capacity. We will describe *CF* grammars in terms of rules or equations, whichever is more convenient.

If the grammar G contains no non-terminal α from which it is possible to derive both a string f' and a string $f\alpha g$, then the terminal language $L(G)$ generated by G will be finite. In this case, G will be called a *polynomial grammar*.

Consider now grammatical rules of the following kinds:

(40)
(i)	$\alpha \to f\beta$	(right-linear)
(ii)	$\alpha \to \beta f$	(left-linear)
(iii)	$\alpha \to f\beta g$	(linear)
(iv)	$\alpha \to f$	(terminating)

A grammar containing only right-linear and terminating rules or only left-linear and terminating rules will be called a *one-sided linear* grammar.

A grammar containing only rules of the type (40) will be called *linear*. Suppose that G contains only rules of the type (40) and of the type $\alpha_1 \to \varphi$, where α_1 is the initial symbol of G; and that, furthermore, it contains no rule $\beta \to \varphi\alpha_1\psi$. Thus the defining equation for α_1 is $\alpha_1 = \pi_1$, where π_1 is a polynomial not involving α_1. Such a grammar will be called *meta-linear*.

Given a grammar G (i.e., a set of positive equations) which is polynomial, one-sided linear, linear, meta-linear or context-free, we will say that the power series r which is the principle term of its solution (i.e., which it *generates*, in the sense defined in § 2.3) and the language Sup r which it generates are, respectively, polynomial, one-sided linear, linear, meta-linear or context-free. These families of power-series will be designated, respectively, $\mathscr{P}^+$, $\mathscr{L}_0{}^+$, $\mathscr{L}^+$, $\mathscr{L}_m{}^+$, $\mathscr{I}^+$; and for each family $\mathscr{F}$ the family of supports of $\mathscr{F}$ will be designated Sup $(\mathscr{F})$.

Notice that Sup $(\mathscr{P}^+)$ is just the family of finite sets, and that Sup $(\mathscr{L}_0{}^+)$ is the family of regular events, in the sense of Kleene [21] (cf. Chomsky, [7] — note that the class of regular events is closed under reflection).

We consider now certain elementary properties of these families of languages.

It is, first of all, immediate that the following inclusion relations hold among these families:

(41) $\mathrm{Sup}(\mathscr{P}^+) \subset \mathrm{Sup}(\mathscr{L}_0{}^+) < \mathrm{Sup}(\mathscr{L}^+) < \mathrm{Sup}(\mathscr{L}_m{}^+) < \mathrm{Sup}(\mathscr{I}^+)$.

Furthermore, in each of these cases inclusion can be strengthened to proper inclusion. Thus we have:

PROPERTY 1.

$$\mathrm{Sup}(\mathscr{P}^+) \subsetneq \mathrm{Sup}(\mathscr{L}_0{}^+) \subsetneq \mathrm{Sup}(\mathscr{L}^+) \subsetneq \mathrm{Sup}(\mathscr{L}_m{}^+) \subsetneq \mathrm{Sup}(\mathscr{I}^+).$$

The simplest example of a language in $\mathrm{Sup}(\mathscr{L}^+)$ but not in $\mathrm{Sup}(\mathscr{L}_0{}^+)$ is the set of all strings $\{a^n b a^n\}$ $(a, b \in V_T)$. This is generated by the grammar: $\alpha = a\alpha a + b$, and is easily shown not to be a regular event. The product of languages in $\mathrm{Sup}(\mathscr{L}^+)$ is always in $\mathrm{Sup}(\mathscr{L}_m{}^+)$, but not in general in $\mathrm{Sup}(\mathscr{L}^+)$. The language L_{IC} of our example (18) above with the grammar:

$$\alpha = a + b\alpha\alpha \tag{42}$$

and consisting of the set of well-formed formulas of the implicational calculus with one free variable in Polish notation is in $\mathrm{Sup}(\mathscr{I}^+)$ but not in $\mathrm{Sup}(\mathscr{L}_m{}^+)$. This follows from the fact that L_{IC} contains all strings of the form

$$b^{m_1} a^{m_1} b^{m_2} a^{m_2} \ldots b^{m_k} a^{m_k} a, \tag{43}$$

for each $k \geq 1$, $m_i \geq 1$. But each string in L_{IC} contains n occurrences of b and $n + 1$ occurrences of a, for some $n \geq 1$. Consequently, for a fixed integer k, to generate all strings of the form (43), it must be possible to derive from the initial symbol of the grammar of L_{IC} a string φ containing k occurrences of non-terminals. Consequently, this grammar cannot be metalinear.

For empirical interpretation of the theory of *CF* grammars, the relation between $\mathrm{Sup}(\mathscr{I}^+)$ and $\mathrm{Sup}(\mathscr{L}_0{}^+)$ is of particular importance, since a finite device incorporating the instructions of a *CF* grammar G generating $L(G)$ as a representation of its intrinsic competence, will be able to interpret only the sentences of some fixedsubset $R \in \mathrm{Sup}\,(\mathscr{L}_0{}^+)$ of $L(G) \in \mathrm{Sup}(\mathscr{I}^+)$ (with fixed supplementary aids). This relation can be described precisely in terms of certain formal features of the structural descriptions (labelled bracketings) generated by *CF* grammars — cf. § 1. Let us say that G is a *self-embedding grammar* if it generates a structural description of the form

$$\ldots [_\alpha \varphi [_\alpha \psi] \chi] \ldots, \tag{44}$$

where φ and χ contain non-null terminals, and where ψ is a properly bracketed expression. Then we have the following result:

THEOREM 1a.
$L \notin \mathscr{L}_0^+$ if and only if every CF grammar that generates L is self-embedding.
Chomsky [9]. This result can be extended in the following way. Define the *degree of self-embedding* of a structural description D as the largest N such that D contains a subconfiguration:
$[_\alpha\varphi_1[_\alpha\varphi_2[_\alpha \ldots [_\alpha\varphi_{N+1}]\varphi_{N+2}] \ldots]\varphi_{2N+1}]$ where each φ_i contains non-null terminals. Then there is a one-one effective mapping Φ of $\{(G,n) : G$ a *CF* grammar, $n \geq 1\}$ into the set of one-sided linear grammars and a one-one effective mapping Ψ of the set Δ of structural descriptions into Δ such that:

THEOREM 1b.
For each $L \in Sup(\mathscr{I}^+)$, there is a CF grammar G generating L such that for each N, Φ (G,N) generates f with the structural description D if and only if G generates the terminal string f with the structural description $\Psi(D)$, where $\Psi(D)$ has degree of self-embedding $\leq N$.
Chomsky [8]. Thus, intuitively, we can, given G, construct a finite device $\Phi(G,N)$ that will recognize the structure of a string f generated by G just insofar as the degree of self-embedding of a particular structural description of f does not exceed N. This fact suggests certain empirical consequences. For discussion, cf. Chomsky [10], Miller and Chomsky [29].

4.2. We consider now various closure properties of these families of languages.

The families of power series defined above can be given the following algebraic characterization. $\mathscr{P}^+$ is a *semi-ring*.[1]) $\mathscr{L}_0^+$ is the smallest semi-ring containing $\mathscr{P}^+$ and closed by quasi-inversion of quasi-regular elements. $\mathscr{L}^+$ is a *module*, and $\mathscr{L}_m^+$ is the smallest semi-ring containing it. The full set $\mathscr{I}^+$ is a semi-ring closed by quasi-inversion of quasi-regular elements.

Correspondingly, we have the following properties of the supports: $Sup(\mathscr{P})$ is closed under set union and set product; $Sup(\mathscr{L}_0^+)$ is the smallest set containing the finite sets and closed under the operations of set union, set product, and the star operation described in § 3.1 [21]; $Sup(\mathscr{L}^+)$ is closed under set union, but not set product; $Sup(\mathscr{L}_m^+)$ is the smallest set

[1]) The notion of semi-ring generalizes to that of ring in that the additive structure is only a monoid (not necessarily group) structure. A typical semi-ring is the so-called "Boolean ring" with two elements 0 and 1 and the rules
$(0 = 0 + 0 = 00 = 01 = 10;\ 1 = 0 + 1 = 1 + 0 = 1 + 1 = 11)$.

containing the sets of Sup($\mathscr{L}^+$) and closed under set product as well (this is, of course, the motivation behind the construction of $\mathscr{L}_m{}^+$); The full set Sup($\mathscr{I}^+$) is closed by union, product and the star operation.

These properties are immediate, and it is natural to inquire into closure under the other elementary operations on sets, namely, intersection and complementation. It is obvious that Sup($\mathscr{P}^+$) is closed under intersection, and it is well-known that the class Sup($\mathscr{L}_0{}^+$) of regular events is closed under intersection and complementation.

For the other families, we have the following results. The family Sup($\mathscr{I}^+$) of all *CF* languages is not closed under intersection and hence (since it is closed under union) not closed under complementation [40], [2]. The example given, in each of these references, consists of a pair of meta-linear languages whose intersection is not context-free. Hence it follows that Sup ($\mathscr{L}_m{}^+$) is also not closed under intersection or, consequently, complementation. This result can be strengthened to cover linear grammars, in fact (for intersection) even linear grammars with a single non-terminal.

To see this, consider the grammars G_1 and G_2 defined as in (45) and (46) respectively:

(45) $$\alpha = aa\alpha c + b\alpha c + bc$$

(46) $$\alpha = a\alpha cc + a\alpha b + ab.$$

G_1 and G_2 are each linear with a single non-terminal. But the intersection of the languages that they generate is the set of strings $[a^{2n}b^n a^{2n}]$, which is not context-free. This example (along with the fact that these families are closed under union) establishes that

PROPERTY 2.
The families Sup($\mathscr{L}^+$), Sup($\mathscr{L}_m{}^+$), Sup($\mathscr{I}^+$) are not closed under either intersection or complementation; the intersection of two sets in one of these families may not even be in Sup($\mathscr{I}^+$), even when the sets in question are generated by grammars with a single non-terminal.

Presumably the complement of a language of Sup($\mathscr{L}^+$) or of Sup($\mathscr{L}_m{}^+$) is not context-free (i.e., is not a member of Sup($\mathscr{I}^+$)). However, we have no examples to show this.

Thus of the classes of languages discussed above, only the regular events (and the finite sets) are closed under formation of intersections. However, the intersection of a regular event and a context-free grammar

is again a context-free language [2]. We have in fact, the following stronger result which extends a well-known theorem of classical analysis due to R. Jungen [20].

THEOREM 2.
Suppose that $r_1 \in \lambda_0^+$. *Let* U^+ *be one of the families* $\mathscr{P}^+$, λ_0^+, λ^+, λ_m^+, $\mathscr{I}^+$. *Let* $r_1 \odot r_2$ *be the Hadamard product of* r_1, r_2 (*cf.* § 3.1). *Then* $r_1 \odot r_2 \in U^+$, *for every* $r_2 \in U^+$. *Furthermore, if* r_2, $r_3 \in \lambda_0^+$, *then* $r_2 \odot r_3 \in \lambda_0^+$.

Cf. Schützenberger [46]. It follows that the intersection of a language of $\mathrm{Sup}(U^+)$ with a regular event is in $\mathrm{Sup}(U^+)$, for each U^+. The proof of this result, which is related to a similar result concerning closure under transduction, will be outlined in § 8, below.

4.3. The category of linear grammars is of particular interest, as we will see directly, and we will now make a few preliminary observations concerning it. Notice that if L is a language generated by a linear grammar, we can find a vocabulary V' disjoint from V_T, two homomorphic mappings α, α' of $F(V')$ into $F(V_T)$, a regular event R in V', and a finite set $C \subset F(V_T)$ such that L consists of exactly the strings $f = \alpha(g)c\,\alpha'(\tilde{g})$, where $g \in R$, $\tilde{g}$ is the reflection of g, and $c \in C$. Thus a finite process dealing with a collection of pairs of strings or a pair of coordinated finite processes can, in general, be correlated to a linear grammar and studied in this way.

Equivalently, we can characterize a linear language in the following, slightly different way. Let $V' = V^+ \cup V^-$ ($V^+ = \{v_i : 0 \leq i \leq n\}$; $V^- = \{v_i : -n \leq i \leq -1\}$. Where $f \in F(V^+)$, let us define $\tilde{f}$ as the result of substituting v_{-i} for v_i in f, throughout. Then a linear language L is determined by choice of a homomorphic mapping β of $F(V')$ into $F(V_T)$, a regular event R in V^+, and a finite set $C \subset F(V_T)$. L is now the set of strings $\beta(f)c\beta(\tilde{f})$, where $f \in R$ and $c \in C$. We will use this alternative characterization below.

We can now determine special classes of linear languages by imposing further conditions on the underlying regular event R, the mappings α, α', and the class C. In particular, in applications below we will be concerned with the case in which R is simply a free monoid (a regular event defined by a single-state automaton) and where C contains just $c \in V_T$, where $\alpha(f) \neq \varphi c \psi \neq \alpha'(f)$. We will call grammars defined by this condition *minimal linear grammars.*

A minimal linear grammar contains a single non-terminal symbol S and a single terminating rule $S \to c$, and no non-terminating rule $S \to \varphi c \psi$. Thus each string of the language it generates has the designated "central marker" c. This is the simplest set of languages in our framework beyond the regular events, and we will see that they differ markedly from regular events in many formal properties.

For later reference, we give now one particular result concerning minimal linear grammars. Let us take V', $V_T = W \cup \{c\}$ $(c \notin W)$, α and α' as above. Let G be the minimal linear grammar defined by α, α' and generating $L(G)$. Thus G has the defining equation

$$\beta = c + \Sigma\{\alpha(v)\beta\alpha'(v) : v \in V'\} \tag{47}$$

where α, α' are mappings of $F(V')$ into $F(W)$. Then we have:

THEOREM 3.
If α is a monomorphism (isomorphism into), then the complement $F(V_T) \setminus L(G)$ of $L(G)$ with respect to $F(V_T)$ is generated by an unambiguous linear grammar.

Proof: Let $A = \alpha(V')$, $F(A) = \alpha F(V')$, and for any set $F \subset F(W)$, let $F^+ = \{f \in F : f \neq e\}$.

Clearly there is a partition: $F(V_T) \setminus L(G) = L \cup L'$, such that

(48)
$$L = fcf' : f \in F^+(A), f' \in F(W), fcf' \notin L(G);$$
$$L' = F(W) \cup cF(W) \cup ((F(W) \setminus F(A))\, cF(W) \cup F(V_T)cF(V_T)cF(V_T).$$

But L' is a regular event. Hence it suffices to show that L is generated by an unambiguous linear grammar.

Since α is a monomorphism, there exists an isomorphism $\tilde{\alpha} : F(A) \to F(V')$. We extend α' to $F^+(A)$ by defining $\alpha' a = \alpha'(\tilde{\alpha} a)$, for $a \in F^+(A)$.

Suppose that $acf' \in L$. Thus $\in F^+(A)$, $f' \in F(W)$, and $f' \neq \alpha' a$. By definition there are just three mutually exclusive possibilities for acf'.

(49) (i) $f' \in F^+(W)\alpha' a$
(ii) $\alpha' a \in F^+(W) f'$
(iii) $a = a_1 a_2 a_3$ and $f' = hwg\alpha' a_1$ (where $a_1, a_3 \in F(A)$; $a_2 \varepsilon A$; $w \in W$; $h,g \in F(W)$; $\alpha' a_2 \in F^+(W)g$; $\alpha' a_2 \in F(W)wg$).

(49i) is the case in which f' has $\alpha' a$ as a proper right factor;
(49ii) is the case in which $\alpha' a$ has f' as a proper right factor;

(49iii) is the case in which $\alpha' a$ and f' have as their common maximal right factor the string $g\alpha' a_1$, which is a proper substring of both $\alpha' a$ and f'. Thus the three cases are mutually exclusive and exhaustive, and we have a partitioning of L into the three subsets L_1, L_2, L_3, consisting of the strings meeting (i)–(iii), respectively. What we now have to show is that each of L_1, L_2, L_3 is generated by an unambiguous linear grammar.

In the case of L_1 and L_2 this fact is obvious. Let $\bar{A} = \Sigma\{a : a \in A\}$ and $\bar{W} = \Sigma\{w : w \in W\}$. Then L_1 is generated by the grammar (50) and L_2 by the grammar (51) (cf. § 3.2).

$$\beta = \Sigma\{a\beta\alpha' a : a \in A\} + c(1 - \bar{W})^{-1} \tag{50}$$

$$\beta = \Sigma\{a\beta\alpha' a : a \in A\} + (1 - \bar{A})^{-1} c \tag{51}$$

Consider now the case of L_3. For each $a \in A$, let us denote by $B(a)$ the set of all strings wg ($w \in W$, $g \in F(W)$) such that $\alpha' a \in F^+(W)g$ and $\alpha' a \notin F(W)wg$. Clearly $B(a)$ is always a finite set, since g is shorter than $\alpha' a$. We can now generate L_3 by the unambiguous linear grammar with the equations:

$$\begin{aligned} \beta_1 &= \Sigma\{a\beta_1\alpha' a : a \in A\} + \Sigma\{a\beta_2 b : a \in A, b \in B(a)\} \\ \beta_2 &= c + c(1 - \bar{W})^{-1} + (1 - \bar{A})^{-1} c + \Sigma\{a\beta_2 w : a \in A, w \in W\}. \end{aligned} \tag{52}$$

Verification is straightforward. But now we have given $F(V_T) \setminus L(G)$ as the union of the four disjoint sets L_1, L_2, L_3, L', each of which has an unambiguous linear grammar. Consequently, $F(V_T) \setminus L(G)$ itself has an unambiguous linear grammar, as was to be proven.

Notice that if we had taken α originally as an "information-lossless transduction" [43] instead of as a monomorphism, we could prove a result differing from *Theorem 3* only in that the linear grammar constructed would have bounded ambiguity, rather than no ambiguity.

4.4. We have considered several subfamilies of the class of *CF* grammars, classifying on the basis of structural properties of the defining rules. There are other principles of classification that might be considered. Thus, for example, it might be worthwhile to isolate the class of the *star grammars* (*languages*) characterized as follows: G is a star grammar if associated with each non-terminal α_t of G there is a set Σ_t of non-terminals and three terminal strings f_t, f'_t, f''_t, and G contains all and only the rules: $\alpha_t \to f''_t$, $\alpha_t \to f_t \alpha_j f'_t$ ($\alpha_j \in \Sigma_t$), $\alpha_j \to \alpha_k \alpha_l$ ($\alpha_j, \alpha_k, \alpha_l \in \Sigma_t$). These are, in a sense, the most "structureless" *CF* grammars. The interest of

these languages lies in the fact that the equations defining the associated power series are expressible using in an essential manner only the quasi-inverse and addition, as we have observed in § 3.2. Notice, in particular, that the non-metalinear language L_{IC} defined by (42) is a star language. We have suggested a linguistic interpretation for the notion "star language" in § 3.2.

Another principle of classification might be in terms of the number of non-terminals in the minimal defining grammar of a certain power series. However, it does not seem likely that interesting properties of language can correlate with a measure so insensitive to structural features of grammars as this (except for the special case of the languages defined by grammars with only *one* non-terminal), because for monoids, as distinct from groups, the gross numerical parameters do not relate in an interesting way to the fine structure. Notice, incidentally, that for any finite N we can construct a regular event which cannot be generated by a *CF* grammar with less then N non-terminal symbols.

Another principle of classification is suggested by consideration of dependencies among subparts of the grammar. Let us call a *CF* grammar *irreducible* if no proper subset of the set of defining equations constitutes a *CF* grammar (recall that terminal strings must be derivable from each non-initial non-terminal of a *CF* grammar); otherwise, *reducible.* If a *CF* grammar is reducible, in this sense, there must be proper subsets Σ_1 of its rules and Σ_2 of its non-terminals, such that only rules of Σ_1 are involved in extending derivations to terminated derivations at points where symbols of Σ_2 appear in lines of derivations.

One particular extreme form of reducibility has been studied by Ginsburg and Rice (18). Following them, let us call a *CF* grammar G *sequential* if its non-terminals can be ordered as $\alpha_1, \ldots, \alpha_n$ (where α_1 is the initial symbol) in such a way that there is no rule $\alpha_i \to \varphi\alpha_j\psi$ for $j < i$. The solution to a sequential grammar is particularly easy to determine by the iterative procedure described in § 2.3 by successive elimination of variables.

Concerning the family $\mathscr{S}^+$ of sequential grammars and the family $\mathrm{Sup}(\mathscr{S}^+)$ of their supports, Ginsburg and Rice establish the following results, paralleling those mentioned above. First, it is clear that $\mathscr{S}^+$, like $\mathscr{I}^+$ is a semi-ring closed by quasi-inversion of quasi-regular elements. Correspondingly, $\mathrm{Sup}(\mathscr{S}^+)$ is closed by union, product, and the star operation. From this fact, and the fact that $\mathscr{P}^+ \subset \mathscr{S}^+$, it follows that $\mathrm{Sup}(\mathscr{L}_0^+) \subset \mathrm{Sup}(\mathscr{S}^+)$. Furthermore, the inclusion is proper, as we can

see from the grammar (42), which, since it contains a single terminal, is sequential. In fact, we have

$$\text{Sup}(\mathscr{L}_0^+) \subsetneq \text{Sup}(\mathscr{S}^+) \subsetneq \text{Sup}(\mathscr{I}^+). \tag{53}$$

Ginsburg and Rice show that there is no sequential grammar for the language with the vocabulary $\{a, b, c, d\}$ and containing the string

$$a^{n_{2k-1}} d \dots db^{n_2} da^{n_1} ca^{n_1} db^{n_2} d \dots b^{n_{2k-2}} da^{n_{2k-1}} \tag{54}$$

(which is symmetrical about c) for each sequence $(k, n_1, \dots, n_{2k-1})$ of positive integers, although this language is generated by the grammar.

$$\begin{aligned} \alpha &= ad\beta da + a\alpha a + aca \\ \beta &= b\beta b + bd\alpha db. \end{aligned} \tag{55}$$

There is no stronger relation than (53) between $\text{Sup}(\mathscr{S}^+)$ and the families of *Property* 1, § 4.1, however. The grammar (55) is in fact linear, though not sequential, so that $\text{Sup}(\mathscr{L}^+) \not\subset \text{Sup}(\mathscr{S}^+)$; and the grammar (42) is sequential but not meta-linear, so that $\text{Sup}(\mathscr{S}^+) \not\subset \text{Sup}(\mathscr{L}_m^+)$.

Since the grammars (45) and (46) are sequential, we see that *Property 2* (but not *Theorem 2*) can be extended to Sup $(\mathscr{S}^+)$. For further results on sequential languages, see Ginsburg and Rose [19], Shamir [51].

5. An Alternative Characterization of Families of *CF* Languages

In this section we will present a rather different approach to the definition of families of languages, and we will show how it interrelates with the classification presented above. We rely here on the two fundamental notions: *standard regular event* and *Dyck language*, which we now define.

A *standard regular event* A is given by a finite alphabet X, two subsets J_1 and J_2 of (X, X), and the rule that $f \in A$ if and only if

(56) (i) $f \in xF(X) \cap F(X)x'$, where $(x, x') \in J_1$

(ii) $f \notin F(X)xx'F(X)$, where $(x, x') \in J_2$.

Thus A is the set of all strings that begin and end with prescribed letters, and that contain no pair of consecutive letters belonging to J_2. It is, more technically, the intersection of the quasi-ideal determined by J_1 with the complement of the two-sided ideal generated by all products

xx' ($(x, x') \in J_2$). A is, in particular, what is sometimes called a "1-definite event" [21], [35].

We define the *Dyck language* D_{2n} on the $2n$ letters $x_{\pm i}$ ($1 \leq i \leq n$) as the set of all strings f which can be reduced to the empty string by repeated cancellation of consecutive pairs of letters $x_j x_{-j}$ ($-n \leq j \leq n$). The Dyck language is a very familiar mathematical object: if φ is the homomorphism of the free monoid generated by $\{x_{\pm i}\}$ onto the free group generated by the subset $\{x_i : i > 0\}$ that satisfies identically $(\varphi x_i)^{-1} = \varphi x_{-i}$, then D_{2n} is the kernel of φ, that is, the set of strings f such that $\varphi f = 1$.

Concerning these notions, we have the following results.

PROPOSITION 1.
For any regular event $B \subset F(Z)$, we can find a standard regular event A and a homomorphism $\alpha : F(X) \to F(Z)$, such that $B = \alpha A$.

It is worth mentioning that this representation can be chosen in such a way that not only $B = \alpha A$, but, furthermore, each string $f \in A$ has the same degree of ambiguity as the corresponding string $\alpha f \in B$. That is, if $B = \mathrm{Sup}(\beta)$, we can find γ such that $A = \mathrm{Sup}(\gamma)$ and for each f, $\langle \gamma, f \rangle = \langle \beta, \alpha f \rangle$.

We can generalize *Proposition 1* to *CF* languages, making use of the following property of D_{2n}.

PROPERTY 1. *D_{2n} is generated by an unambiguous CF grammar.*
To obtain an unambiguous grammar of D_{2n}, we introduce $2n + 1$ nonterminals $\alpha_{\pm i}$ ($1 \leq i \leq n$) and β. Consider now the $2n + 1$ equations

$$(57)\quad \text{(i)} \qquad \alpha_i = x_i \Big(1 - \sum_{j \neq -i} \alpha_j\Big)^{-1} x_{-i}$$

$$\text{(ii)} \qquad \beta = (1 - \Sigma \alpha_i)^{-1}.$$

(Cf. § 3.2, for notation).

Intuitively, β can be interpreted as the sum of all strings that can be reduced to the empty string by successive cancellation of two consecutive letters $x_i x_{-i}$. Each α_i is the sum of all words in $\mathrm{Sup}(\beta)$ that begin by x_i and have no proper left (or right) factor in $\mathrm{Sup}(\beta)$. The equation (57i) implies that each $f \in \mathrm{Sup}(\alpha_i)$ has one and only one factorization

$$(58) \qquad f = x_i f_1 f_2 \ldots f_m x_{-i}$$

where each f_j belongs to a well-defined set $\mathrm{Sup}(\alpha_j)$ (where j is not $-i$ because we want the initial letter x_i to cancel only with the final letter x_{-i}).

Similarly, each $f \in \mathrm{Sup}(\beta)$ has one and only one factorization $f = f_1 \ldots f_m$, where the f_j's belong to $\cup_i \mathrm{Sup}(\alpha_i)$.

We now have the following result, analogous to *Proposition 1*.

PROPOSITION 2.
Any CF language $L \subset F(Z)$ is given by an integer n, a standard regular event A on $X_{2n} = \{x_{\underline{+}i} : 1 \leq i \leq n\}$, a homomorphism $\varphi : F(X_{2n}) \to F(Z)$, and the rule $L = \varphi(A \cap D_{2n})$.
[48], [49], [11], [12].

Again, as above, this statement implies that the strings are produced with the appropriate ambiguity. Furthermore, it is possible to choose J_1 such that $(x,x') \in J_1$ if x belongs to a certain subset of X (cf.[48]).

Special subfamilies of languages such as those considered above can be defined by imposition of conditions on the underlying standard regular event A and the homomorphism φ. Thus suppose that we take the standard regular event A on the alphabet $X \cup Y$ (where $X = \{x_{\underline{+}i} : 1 \leq i \leq n\}$, $Y = \{y_{\underline{+}i} : 1 \leq i \leq m\}$ defined by the following conditions on J_1 and J_2:

$$(59)\quad J_1 = \{(x_i, x_j) : i > 0\}$$
$$J_2 = \{(x_i, x_j) : \mathrm{sign}\,(i) \neq \mathrm{sign}\,(j)\} \cup \{(y_i, y_j) : i < 0 \text{ or } j > 0\} \cup \{(x_i, y_j) : i < 0 \text{ or } j < 0\} \cup \{(y_i, x_j) : i > 0 \text{ or } j > 0\}.$$

Thus every string has the form $fgg'f'$, where $f, f' \in F(X)$; $g, g' \in F(Y)$; f, g (respectively, f', g') contain only letters with positive (respectively, negative) indices. If we designate by X^+ and X^- the subsets of X consisting of letters with positive indices and negative indices, respectively (similarly, Y^+ and Y^-), we can describe the permitted and excluded transitions by the matrix (60), where the entry 1 (0) indicates that transition is (is not) permitted from the element labelling to row to that labelling the column, and where U is the matrix with all one's and 0 the matrix with all zeroes.

(60)

	Y^+	Y^-	X^+	X^-
Y^+	0	U	0	0
Y^-	0	0	0	U
X^+	U	0	U	0
X^-	0	0	0	U

But consider now the set $A \cap D_{XY}$ (where D_{XY} is the Dyck language

on the alphabet $X \cup Y$). If $fg \in A$ (where $f \in F(X^+ \cup Y^+)$, $g \in F(X^- \cup Y^-)$) meets the additional condition that $fg \in D_{XY}$, then g must be the mirror-image of f (up to a change of the sign of indices). That is, in the notation of the second paragraph of § 4.3, it must be the case that $g = \tilde{f}$. Clearly, if α is a homomorphic mapping of $F(X \cup Y)$ into $F(V_T)$, then $\alpha(A \cap D_{XY})$ is a linear language. Furthermore, if we add the further condition that $y_i = e$ for $i < 0$ and $y_i = c$ for each $i > 0$, where $\alpha x_i \notin F(V_T)cF(V_T)$ for any i, then $L = \alpha(A \cap D_{XY})$ is a minimal linear language with c as designated center symbol, and every minimal linear language is given by such a choice of α. This gives an independent characterization of minimal linear languages.

Furthermore, by adding additional pairs to J_2 we can delimit the defined canonical language A in such a way that $\{f : f \in F(X^+)$ and for some g, $fg \in A$ and $g \in F(Y)F(X)\}$ is an arbitrary regular event (instead of simply the free monoid on X^+, as above), so that $L = \alpha(A \cap D_{XY})$ will be an arbitrary linear language. Thus we have an independent definition of the notion "linear language". (Notice that these further restrictions on J_2 affect only the permitted transitions in the matrices along the main diagonal of (60).

In much the same way, we can give a general definition of "metalinear language". Thus, for example, consider the particular metalinear language generated by the grammar with the equations

(61)
$$\begin{aligned} \xi_i &= \xi_{i1}\xi_2 \\ \xi_1 &= e + \Sigma\{a\xi_1 b : a,b \in V_T\} \\ \xi_2 &= e + \Sigma\{a\xi_2 b : a,b \in V_T\}. \end{aligned}$$

In this case, the matrix for the underlying standard regular event would be

(62)

	X_1^+	X_1^-	X_2^+	X_2^-
X_1^+	U	U	0	0
X_1^-	0	U	U	0
X_2^+	0	0	U	U
X_2^-	0	0	0	U

Any metalinear language, and only these, will be based on a standard event with a matrix of essentially this kind (with, perhaps, additional restrictions along the main diagonal).

Propositions 1 and 2 thus provide for the possibility of very natural definitions of the full class of *CF* languages, and various subfamilies of this class, independently of the approach taken in preceding sections.

6. Undecidability

6.1. In Post [36] it is shown that the following problem, known as the *Correspondence problem*, is recursively unsolvable. Where $\Sigma = \{(f_1,g_1), \ldots, (f_n,g_n)\}$ is a sequence of pairs of strings, let us say that a sequence $I = (i_1, \ldots, i_m)$ of integers $(1 \leq i_j \leq n)$ *satisfies* Σ if

$$f_{i_1} \ldots f_{i_m} = g_{i_1} \ldots g_{i_m}. \tag{63}$$

The correspondence problem is the problem of determining whether, given Σ, there is an index sequence that satisfies Σ. Notice that either Σ is satisfied by no index sequence or else by infinitely many, since if $(i_1, \ldots, i_m)$ satisfies Σ, then so does $(i_1, \ldots, i_m, i_1, \ldots, i_m)$. Post showed that there is no algorithm for determining, for arbitrary Σ, whether there is no index sequence satisfying Σ, or whether there are infinitely many, these being the only alternatives.

We can reformulate the correspondence problem directly in terms of minimal linear grammars. Given $\Sigma = \{(f_1,g_1), \ldots, (f_n,g_n)\}$, form $G(\Sigma)$ with the single non-terminal S and the defining equation:

$$S = a + f_1 S g_1 + \ldots + f_n S g_n, \tag{64}$$

where a is a symbol not in any of the f_i's or g_i's. Clearly there is an index sequence satisfying Σ just in case $G(\Sigma)$ generates a string $faf^{\sim}$. Or, to put it differently, let L_m be the "mirror-image" language consisting of all strings $faf^{\sim}$, $f \in F(V_T)$, and let $L(G(\Sigma))$ be the language generated by G. Then either there is no index sequence satisfying Σ, in which case $L_m \cap L(G(\Sigma))$ is empty; or there are infinitely many index sequences satisfying Σ, in which case $L_m \cap L(G(\Sigma))$ is infinite. From the unsolvability of the correspondence problem and the fact that L_m is generated by a linear grammar with one non-terminal, we conclude directly that:

Undecidability Theorem 1.
There is no algorithm for determining, given two CF grammars G_1 and G_2 generating L_1 and L_2 respectively, whether $L_1 \cap L_2$ is empty or infinite. This is true even where G_1 and G_2 are minimal linear grammars and where G_1 is a fixed particular grammar of L_m.

The problems of emptiness or finiteness of intersection are easily seen to be solvable for one-sided linear grammars, but we see that for the simplest grammars in our framework that go beyond regular events in generative capacity, these problems are no longer solvable.

This observation is generalized in Bar-Hillel, Perles, Shamir [2], where many problems concerning *CF* grammars are shown to be recursively unsolvable. In brief, their method is as follows. Let us limit V_T to the set $\{a, 0, 1\}$. Where $\Sigma = \{(f_1,g_1), \ldots, (f_n,g_n)\}$ is a set of pairs of strings in the vocabulary $\{0, 1\}$ (i.e., $f_i,g_i \in F\{0, 1\}$), let $L(\Sigma)$ be the set of all strings

$$10^{i_k} \ldots 10^{i_1} a f_{i_1} \ldots f_{i_k} a \tilde{g}_{j_l} \ldots \tilde{g}_{j_1} a 0^{j_1} 1 \ldots 0^{j_l} 1, \tag{65}$$

where $1 \leq i_1, \ldots, i_k, j_1, \ldots, j_l \leq n$.

More perspicuously, let us use $\bar{\imath} = 01^i$ as a code for the number i. Then a string of $L(\Sigma)$ is formed by selecting index sequences $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_l)$ and forming

$$\bar{i}_k \ldots \bar{i}_1 a f_{i_1} \ldots f_{i_k} a g_{j_l} \ldots g_{j_1} \bar{a} \bar{j}_1 \ldots \bar{j}_l. \tag{66}$$

$L(\Sigma)$ now plays the same role as the language generated by (64) in the foregoing proof of Undecidability Theorem 1. It is clearly a *CF* language (generated, in fact, by a meta-linear grammar which is an obvious modification of (64). But from *Theorem* 3, § 4.3, above, it follows directly that the complement $F(V_T) \setminus L(\Sigma)$ of $L(\Sigma)$ with respect to the vocabulary V_T is a *CF* language, and that we can construct its grammar given the grammar of $L(\Sigma)$. (Notice, in fact, that we could have used any code, in place of the particular choice $\bar{\imath} = 01^i$, for defining $L(\Sigma)$).

In place of the mirror-image language L_m used in the proof of *Undecidability Theorem 1*, let us consider the "double-mirror-image" language L_{dm} consisting of all strings

$$x_1 a x_2 a \tilde{x}_2 a \tilde{x}_1, \text{ where } x_1 \text{ and } x_2 \text{ are strings in } \{0, 1\}. \tag{67}$$

It is not hard to show that L_{dm} and its complement with respect to V_T are both *CF* languages.

Observe that

$$L(\Sigma) \cap L_{dm} = \{\bar{i}_k \ldots \bar{i}_1 a f_{i_1} \ldots f_{i_k} a g_{i_k} \ldots g_{i_1} a \bar{i}_1 \ldots \bar{i}_k\} \tag{68}$$

where $(i_1, \ldots, i_k)$ satisfies Σ (that is, where $f_{i_1} \ldots f_{i_k} = g_{i_1} \ldots g_{i_k}$).

Observe also that an infinite set of strings of the form of (68) cannot constitute a *CF* language (nor, a fortiori, a regular event).

Suppose now that there is a positive solution to the correspondence

problem for Σ; that is, there is an index sequence satisfying Σ. Then, as we have observed, there are infinitely many such sequences. Consequently $L(\Sigma) \cap L_{dm}$ is infinite. It is therefore neither a regular event nor a *CF* language.

Suppose, on the other hand, that there is no index sequence satisfying Σ. Then $L(\Sigma) \cap L_{dm}$ is empty, and is therefore both a regular event and a context-free language. But $L(\Sigma)$ and L_{dm} are *CF* languages; and, with Σ fixed, we can construct their *CF* grammars $G(\Sigma)$ and G_{dm} (which are, in fact, meta-linear). Thus if there were an algorithm for determining whether the intersection of the languages generated by two *CF* grammars G_1 and G_2 is empty, finite, a regular event, or a *CF* language, this algorithm would also provide a solution to the general correspondence problem. We conclude, then:

UNDECIDABILITY THEOREM 2.
There is no algorithm for determining, given CF grammars G_1 and G_2, whether the intersection of the languages that they generate is empty, finite, a regular event, or a CF language — in particular, this remains true when both are meta-linear and G_2 is a fixed grammar of L_{dm}.

Let $\overline{G}_{dm}$ be the *CF* grammar that generates the complement $\overline{L_{dm}}$ (all complements now are with respect to V_T) of L_{dm}. And, given Σ, let $\overline{G}(\Sigma)$ be the *CF* grammar that generates the complement $\overline{L(\Sigma)}$ of $L(\Sigma)$, as guaranteed by *Theorem 3*, § 4.3. Consider now the grammar G generating the language $L(G) = \overline{L_{dm}} \cup \overline{L(\Sigma)}$. Clearly G is *CF* and can be constructed from G_{dm} and $\overline{G}(\Sigma)$. But the complement $\overline{L(G)}$ of $L(G)$ is just the set $\overline{L_{dm} \cup L(\Sigma)} = L_{dm} \cap L(\Sigma)$, and we know by *Undecidability Theorem 2* that there is no algorithm for determining, given Σ, whether this set is empty, finite, a regular event, or a *CF* language. But given Σ, G is determined as a *CF* grammar. Therefore we have:

UNDECIDABILITY THEOREM 3.
There is no algorithm for determining, given the CF grammar G, whether the complement of the language generated by G is empty, finite, a regular event, or a CF language.

There is, in particular, no general procedure for determining whether the *CF* grammar G generates the universal language $F(V_T)$, or whether G generates a regular event (since the complement of a regular event is a

regular event). Consequently, there is no algorithm for determining, given *CF* languages L_1 and L_2, whether there is a transducer mapping L_1 onto L_2 since all and only regular languages can be obtained by transduction from the *CF* language $F(V_T)$ (Ginsburg and Rose, personal communication). There is, furthermore, no general method for determining whether two *CF* grammars are equivalent, i.e., generate the same language, since if there were such a method, it could be used to determine whether a *CF* grammar G is equivalent to the grammar G_U generating $F(V_T)$. It also follows immediately that there is no algorithm for determining, given *CF* grammars, whether the language generated by one includes the language generated by the other, since this would give a solution for the equivalence problem.

These results have been outlined for languages constructed from a three-element vocabulary V_T, but it is clear that by appropriate recoding, they still apply to languages in a vocabulary of two or more letters. This is worked out in detail in Bar-Hillel, Perles, Shamir [2].

6.2. We observed in § 4 that finite processes involving pairs of strings receive a natural formulation in terms of linear grammars. In particular, as we have just seen, the correspondence problem can be described directly as a problem concerning minimal linear grammars. The same is true of a second combinatorial problem, also due to Post, called the "Tag problem".

We can state a generalized form of the Tag problem in the following way. Let W be the set of strings (the free monoid) in some finite vocabulary, and let P be a finite subset of non-null strings of W meeting the condition that no string of W has more than one left factor in P. That is, there are no p_1, p_2, w_1, w_2, w_3 ($p_i \in P$, $w_i \in W$) such that $p_1 \neq p_2$ and $w_1 = p_1 w_2 = p_2 w_3$. Let V be the set of strings of W that have no left factor in P — that is, $v \in V$ if and only if there is no $p \in P$ such that $v = pw$, for $w \in W$. Clearly V is a recursive, in fact regular, set. Let α be a mapping of P into W (thus α defines a set of pairs of strings (p, w), where $w = \alpha p$, $p \in P$, $w \in W$). Define a mapping T on W, where

(69) $$\begin{aligned} Tf &= f'\alpha p, \text{ if } f = pf' \\ Tf &= H, \text{ if } f \in V\ (H \notin W). \end{aligned}$$

Consider the problem:

(70) given a string f, is there an integer n such that $T^n f = H$?

Regarding T as defining the computation of a Turing machine, (70) is the *halting problem* for this Turing machine. It has been shown by Minsky [30 that (70) is a recursively unsolvable problem.

The Tag problem as formulated by Post is the special case of (70), above, where T meets the following additional conditions: P is the set of all strings of length k, for some fixed $k \geq 2$; αp depends only on the left-most symbol of p. Even with this restriction, the problem (70) is unsolvable, as Minsky has shown. This is a somewhat surprising result, because of the determinacy (*monogenicity*) of the generative procedure T.

As a step towards reformulating the generalized Tag problem in terms of minimal linear grammars, we observe that it can be stated in the following way. Given W, P, V, α, T, as above, the question (70) has a positive answer just in case

(71) there are strings $p_1, \ldots, p_n \in P$ and $v \in V$ such that: $p_1 \ldots p_n v = f\alpha p_1 \ldots \alpha p_n$.

But we can now restate the generalized Tag problem as the following problem concerning linear grammars. Given W, P, V, α, T, let us define the grammar G generating $L(G)$ with the single equation

$$S = \Sigma_i v_i c + \Sigma_i (p_i S \widetilde{\alpha p_i}) \tag{72}$$

where $v_i \in V$, $p_i \in P$, and $c \notin W$ is the distinguished central marker. Let us define the language $M(f) = \{fgc\tilde{g} : g \in W\}$ (thus $M(f) = fL_m$, where L_m is the "mirror-image language" defined above). Then the answer to (71) (equivalently, (70)) is positive if and only if the intersection of $L(G)$ with $M(f)$ is non-empty. Thus we see that there is no algorithm for determining whether, for fixed f, the language $M(f)$ has a non-empty intersection with a language with a grammar meeting (72) (even for the special case in which P is the set of all strings of length k, for fixed $k \geq 2$, and αp depends only on the left-most letter of P).

Notice that *Undecidability Theorem 1*, above, also follows directly from unsolvability of the Tag problem. In fact, the Correspondence and Tag problems both concern the cardinality of the intersection of a minimal linear language L with the languages $M(f)$, where $f = e$ and L is arbitrary, for the case of the Correspondence problem, while f is arbitrary and L meets the condition (72), above, for the case of the Tag problem.

7. Ambiguity

7.1. We have defined the power series r to be *characteristic* just in case each coefficient $\langle r,f\rangle$ is either zero or one. We say that a *CF* grammar is *unambiguous* if the principal term of its solution is a characteristic power series. In this case, each sentence that it generates is provided with a single structural description, and "debracketization" introduces no ambiguities. Let us call a *CF* language *inherently ambiguous* if each of its *CF* grammars is ambiguous.

It is well-known that no regular event is inherently ambiguous — that is, each regular event is the support of a characteristic power series which is the principal term of the solution of a one-sided linear grammar [14], [37]. However, this remark does not carry over to the full class of *CF* grammars. It has been shown by Parikh [34] that there are *CF* languages that are inherently ambiguous.

An example of an inherently ambiguous language is the set

$$\{a^n b^m c^p : n = m \text{ or } m = p\}. \tag{73}$$

In this case, the strings of the form $a^n b^n c^n$ must have ambiguity at least two in any *CF* grammar generating (73) (and there is a *CF* grammar generating (73) in which they have ambiguity exactly two).

We do not have examples illustrating the extent of inherent ambiguity in *CF* languages, or special types of *CF* languages.

Notice that it is an immediate consequence of *Undecidability Theorem 1* of § 6 that there can be no algorithm for determining whether a *CF* grammar, or even a linear grammar, is ambiguous. Suppose in fact that, as above, $\Sigma = \{(f_1, g_1), \ldots, (f_n, g_n)\}$ is a sequence of pairs of strings. Select $n+1$ new symbols $x_0, \ldots, x_n$ and construct the grammars G_f with the rules $S_f \to x_0$, $S_f \to x_i S_f \tilde{f}_i$ $(1 \leq i \leq n)$ and G_g with the rules $S_g \to x_0$, $S_g \to x_i S_g \tilde{g}_i$ $(1 \leq i \leq n)$. Clearly G_f and G_g are unambiguous, and the Correspondence problem for Σ has a positive solution if and only if there is a string generated by both G_f and G_g, that is, if and only if the grammar G_{fg} is ambiguous, where G_{fg} contains the rules of G_f, the rules of G_g, and the rules $S \to S_f$, $S \to S_g$, where S is the initial symbol of G_{fg}. Consequently, there can be no procedure for determining, for arbitrary Σ, whether the grammar G_{fg} associated with Σ in this way is unambiguous.

The grammar G_{fg} is linear with three non-terminals and a designated central marker, and we see that for this class of grammars the ambiguity

problem is unsolvable. Presumably, this remark can be generalized to grammars with two non-terminals. It is an interesting open question, however, whether the ambiguity problem remains unsolvable for minimal linear grammars.

Summarizing the matter of ambiguity, as it stands at present, we have the following results:

AMBIGUITY THEOREM 1. *There are inherently ambiguous CF languages.*

AMBIGUITY THEOREM 2.
There is no algorithm for determining whether a CF grammar (which may even be linear with a designated central marker) is ambiguous.

8. FINITE TRANSDUCTION

We want to describe a particularly simple family of transformations from language to language. The first and most essential one is a *homomorphism.*

Let L be any language on a terminal vocabulary Z and assume that for each $z \in Z$ we are given a language L_z on a second vocabulary X. We denote by θL the set of all strings (in X) which can be obtained by taking a word $g = z_{i_1} z_{i_2} \dots z_{i_m} \in L$, and replacing each z_{i_j} by an arbitrary word from $L_{z_{i_j}}$. The name "homomorphism" is self-explanatory. In fact, if we consider the rings $A(Z)$ and $A(X)$ of formal power series in the variables $z \in Z$ and $x \in X$, and if we denote by θ the homomorphism of $A(Z)$ into $A(X)$ that is induced by the mapping $\theta_z =$ the formal power series associated with L_z, then θL is the support of the image by θ of the formal power series associated with L.

An interpretation within our previous framework can be given if L and the L_z's are CF languages. In this case, suppose that L is produced by the CF grammar G (with non-terminal vocabulary Y) and that each L_z is produced by the CF grammar G_z (with the set of non-terminals Y_z and the initial letter $y_{z,0}$). We assume that the sets Y_z are disjoint and we consider a CF grammar $\bar{G}$ with non-terminals $Y \cup Z \cup \bigcup_{z \in Z} Y_z$ consisting of the rules of G and of the G_z's and the rules $z \to y_{z,0}$ ($z \in Z$). (More simply we identify each z with $y_{z,0}$). It is clear that $\bar{G}$ produces exactly θL.

We now generalize this construction to the following type of context

dependency: Let R_i $(i \in I)$ and $R_{i'}$ $(i' \in I')$ be two finite families of regular events such that every $g \in F(Z)$ belongs to one and only one member of each family. Suppose also that for each triple $(z \in Z, i \in I, i' \in I')$, we have a language $L_{z,i,i'}$ in the vocabulary X.

Then for any $y = z_{j_1} z_{j_2} \dots z_{j_k}$ we replace each z_{j_k} by an arbitrary string from the language $L(z_{j_k}, i, i')$ where i and i' are determined by the condition that the string $z_{j_1} z_{j_2} \dots z_{j_{k-1}}$ is in R_i and the string $z_{j_{k+1}} \dots z_{j_{k+1}}$ in $R_{i'}$. It is easily proven that without loss of generality it may be assumed that for any string g belonging to some set R_{i_1}, and for $z \in Z$, the set R_{i_2} which contains gz depends only upon the index i_1 and the letter z. In other words we may assume that we are given a set of states I, a transition mapping $I \times Z \to I$ and an initial state $i_0 \in I$ such that $z_{j_1} z_{j_2} \dots z_{j_{k-1}} \in R_i$ if and only if i is the state reached from i_0 after reading $z_{j_1} z_{j_2} \dots z_{j_{k-1}}$.

A similar construction applies to R_i, and for the sake of clarity we write the corresponding mapping as a *left* multiplication. Given the two mappings $I \times Z \to I$ and $Z \times I'$, we denote by σg, for each $g = z_{j_1} z_{j_2} \dots z_{j_m} \in F(X)$, the sequence of triples

$$(74) \qquad (i_1, z_{j_1}, i'_m)\,(i_2, z_{j_2}, i'_{m-1}) \dots (i_k, z_{j_k}, i'_{m-k+1}) \dots (i_m, z_{j_m}, i'_1)$$

where inductively

$$(75) \qquad i_2 = i_1 z_{j_1},\ i_3 = i_2 z_{j_2},\ \dots,\ i_m = i_{m-1} z_{j_m},\ i_k = i_{k-1} z_{j_{k-1}} \text{ and}$$
$$i'_2 = z_{j_m} i'_1,\ i'_3 = z_{j_{m-1}} i'_2,\ \dots,\ i'_m = z_{j_2} i'_{m-1}.$$

With these notations the transformation we have been describing can be considered as consisting of two steps:

(76) (i) replacement of every $g \in L$ by the string $\sigma g = (i_1, z_{j_1}, i_m) \dots (i_m, z_{j_m}, i')$ in an alphabet U consisting of triples (i, z, i');

(ii) replacement in σg of every triple $(i_k, z_{j_k}, i'_{m-k+1})$ by an arbitrary string from the language $L(z_{j_k}, i_k, i'_{m-k+1})$.

Since step 2 is only a homomorphism, it is sufficient to discuss step 1. For this let U denote the set of all triples (i, z, i') and consider the language L' obtained from L by adding to its grammar all the rules $z_j \to (i, z_j, i')$ with $i \in I$, $i' \in I'$ arbitrary).

Clearly a string of L' belongs to the set $\{\sigma g : g \in L\}$ if and only if it satisfies the condition (75) above, or, in other words, if it belongs to the

regular event $\bar{R}$ determined by the condition (75) on the set $F(U)$ of all strings in the alphabet U.

Hence step 1 consists only of a homomorphism from L in to the set of all strings on U (which gives L') followed by the intersection of L' with a regular event.

Let us now give a final interpretation of what we have done: For each $z \notin Z$, let μz denote a matrix whose rows and columns are indexed by pairs $(i \notin I, i' \notin I')$ and whose entries are as follows

$$(77) \qquad \mu z_{(i,\, i')(i''\, i''')} = \text{the triple } (i, z, i''') \text{ if } i'' = iz \text{ and } i' = zi''' \\ = 0 \text{ otherwise.}$$

Then if we compute
$\mu z_{j_1} \mu z_{j_2} \dots \mu z_{j_m} = \mu g$, it is easily verified that the entry $(i, i'_m)(i_m, i'_1)$ of μg is precisely σg. From this it follows easily that $\{\sigma g : g \notin L\} = L' \cap \bar{R}$ is also a context-free language. Indeed, μ is a homomorphism — we replace every non-terminal y by a matrix μy whose entries are new non-terminals and we verify that μ commutes with the substitutions used for defining the language as the solution of a system of equations. On the other hand identifying the entries one by one in the image μ of our equations gives a new set of equations of the usual type that exactly defines $L' \cap \bar{R}$ [46]. More simply still we can define μ' as above except that for each non-zero entry we take the formal power series associated with $L(z_j, i, i')$, instead of the triple (i_1, z_j, i'). Then the two steps of the construction are telescoped in a single one and the power series associated with the language (on X) obtained by our transformation is simply an entry of

$$(78) \qquad \Sigma\{\mu' g : g \in L\}.$$

This is the basis for the proof of Theorem 2, § 4, above.

9. Connections with the Theory of Automata

We have so far been studying generative processes, the languages and systems of structural descriptions that they define, and finitary mappings on these languages from a completely abstract point of view. To relate these remarks to the theory of automata, it is convenient to introduce a temporal asymmetry into consideration.

An automaton M can be regarded as a device consisting of a set of states Σ (the *memory* of M) that accepts (equivalently, produces) a

sequence of symbols from a vocabulary (alphabet) V in accordance with fixed, finitely statable instructions (which can be given by associating with each $v \in V$ a mapping φ_v of Σ into itself (or into the set of subsets of Σ, in the case of a "non-deterministic" automaton). If we designate an initial state and a set of final states, we define a language $M(L)$ consisting of the strings that can be accepted by M as it proceeds in accordance with its instructions from the initial state to a final state, proceeding from $S \in \Sigma$ to $S' \in \Sigma$ on accepting v just in case $\varphi_v(S) = S'$ (or $S' \in \varphi_v(S)$, in the non-deterministic case). The size of memory of M, or its rate of growth in the course of computation, provides a certain index of the richness of the language $L(M)$ in terms of which we can compare various families of languages of the kinds we have considered.

Given a set of strings L, let us write $f \sim f'$ just in case for all g, $fg \in L$ if and only if $f'g \in L$. Clearly $\sim$ is an equivalence. Furthermore, it is clear that we can take the equivalence classes defined by $\sim$ as states of an automaton $M(L)$ that accepts L, since all of the information about f relevant to the further computation of $M(L)$, once it has read f, is given by the equivalence class to which f belongs. Notice that L is the union of certain of these equivalence classes, and that $f \sim f'$ implies that $fg \sim f'g$, for all g.

Secondly, given L let us write $f \equiv f'$ if and only if for all g, $gf \sim gf'$. Clearly $f \equiv f'$ if and only if for all g,g', gfg' if and only if $gf'g'$. Thus $\equiv$ is symmetrical, and it is easy to show that $f_1 \equiv f_2$ and $f_3 \equiv f_4$. Thus $\equiv$ is a congruence relation, and the $\equiv$-classes in the set $F(V)$ can be multiplied together giving a quotient monoid of $F(V)$. This quotient monoid $F'(V) = \varphi F(V)$ is such that $L = \varphi^{-1}\varphi L$ — and is canonically associated with L_0 [41].

This observation relates the present theory to the theory of monoids. The interest of this is that in certain cases, the $\sim$-classes (and the quotient monoid) have a simple interpretation that can be translated into the language of automata, and, conversely, that certain algebraic notions (in particular, that of *extension*), receive a simple interpretation.

Returning now to the problem of characterizing families of languages in terms of automata, it is well-known that the sub-family $\mathrm{Sup}(\mathscr{L}_0^+)$ of *CF* languages is uniquely characterized by the fact that for each language $L \in \mathrm{Sup}(\mathscr{L}_0^+)$, there is an automaton $M(L)$ with bounded memory that accepts L.

Consider now the family $\mathrm{Sup}(\mathscr{L}_0)$, that is, the set of supports of power series that are the solutions to systems of "one-sided linear" equations

with positive or negative integral coefficients. As we have observed, $L \in \mathrm{Sup}(\mathscr{L}_0)$ if and only if $L = \mathrm{Sup}(r_1 - r_2)$, where $r_1, r_2 \notin (L_0{}^+)$. It can now be shown that the following statements are equivalent:

(79) (i) $L \in \mathrm{Sup}(\mathscr{L}_0{}^+)$;
(ii) there is a one-one correspondence between the $\sim$-classes for L and a finite dimensional space of integral vectors $v(f)$ such that for each $x \in V$, $v(fx) = v(f)\mu x$, where μx is a matrix;
(iii) $F/\equiv$ is isomorphic to a monoid of finite dimensional integral matrices (i.e., the matrices μ of ii);
(iv) L is accepted by an automaton $M(L)$ with a finite dimensional space of vectors with integral coordinates as memory and transitions as above in ii.

(Schützenberger, [44] — let the class $\mathscr{A}$ of automata be those defined by (79iv)).

Consider now the following two restrictions on the class $\mathscr{A}$ of automata.

(80) (i) there is an N such that, for all $f \in F(V)$, $||v(f)|| < N$;
(ii) for all $f, f', f'' \in F(V)$ and $\varepsilon > 0$,
$$\lim_{n \to \infty} e^{-\varepsilon n} ||v(f'f^n f'')|| = 0$$

where $||v||$ is the length of the vector v, in the usual sense.

Clearly (80i) implies (80ii). Furthermore, it is clear that L is a regular event (that is, $L \in \mathrm{Sup}(\mathscr{L}_0{}^+)$) just in case (80i) is met by an automaton of class $\mathscr{A}$ that accepts L. An automaton of class $\mathscr{A}$ that meets condition (80ii) is called a *finite counting automaton* in Schützenberger [47], where such devices are studied. It can be proved that in a loose way (80ii) means that the amount of information (in bits) stored in the memory does not grow faster than a linear function of the logarithm of the length of one input word.

It is interesting to observe that (just as in the case of the full class of *CF* grammars), there is no algorithm for determining, given $M \in \mathscr{A}$, whether there is an f not accepted by M [28]. Furthermore, the same problem for finite counting automata is easily shown to be unsolvable, if Hilbert's tenth problem (the problem of the existence of an integral solution for an arbitrary diophantine equation) is unsolvable [47].

Consider now an automaton M with a structure of the following kind: the states of M (the $\sim$-classes in the input language of M) are identified with strings in a certain new ("internal") alphabet, and for

each $v \in V$, the "computing instruction" φ_v mapping [$\sim$-class of f] $\to$ [$\sim$-class of fv] consists of addition or deletion of letters at the right-hand end of the internal string associated with [$\sim$-class of f]. Such an automaton we can call (in accordance with usual terminology) a *pushdown storage* (*PDS*) *automaton*. PDS automata constitute a restricted subclass of the class of linear bounded automata studied by Myhill [319], Ritchie [39].

Where M is a PDS automaton, the language L that it accepts is a *CF* language, and each *CF* language can be obtained by a homomorphism from a language accepted by a PDS automaton [48], [49]. In particular, where D is a Dyck language and A a standard regular event (cf. § 5), $D \cap A$ is accepted by a PDS automaton.

A *non-deterministic PDS automaton* is an automaton of the type described above, except for the fact that φ_v maps a state into a set of states. We can now prove directly that *CF* languages (languages of the class Sup($\mathscr{J}^+$)) are exactly those that are accepted by non-deterministic PDS automata [11], [12].

REFERENCES

[1] Ajdukiewicz, K., Die Syntaktische Konnexität. *Studia Philosophica*, (1935), 1, 1–27.

[2] Bar-Hillel, Y., Perles, M. and Shamir, E., On formal properties of simple phrase structure grammars. *Tech. Report* 4, July 1960.

[3] —, Applied Logic Branch. The Hebrew University of Jerusalem. Now published in *Zeit. für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, Band 14, Heft 2, (1961), 143–172.

[4] — and Shamir, E., Finite state languages, *Bull. Research Council Israel*, 8F (1960), 155–166.

[5] Birkeland, R., Sur la convergence de développement, qui expriment le racins de l'équation algebrique générale. *C. R. Acad. Sciences* 171 (1920) 1370–1372; 172, (1921) 309–311.

[6] Culik, K., Some notes on finite state languages. *Časopis pro pěstovani Mat.*, (1961), 86, 43–55.

[7] Chomsky, N., Three models for the description of language. *I.R.E. Trans.* PGIT 2, (1956), 113–124.

[8] —, On certain formal properties of grammars. *Information and Control*, (1959), 2, 137–167.

[9] —, A note on phrase structure grammars. *Information and Control*, (1959), 2, 393–395.

[10] —, On the notion "Rule of Grammar". *Proc. Symp. Applied Math. 12, Am. Math. Soc.*, (1961).

[11] —, Context-free grammars and pushdown storage. Quarterly Progress Reports no. 65, Research Laboratory of Electronics, M.I.T., (1962).

[12] —, Formal properties of grammars 1962. To appear in Bush, Galanter, Luce (eds.), *Handbook of Mathematical Psychology*, vol. 2. Wiley.

[13] —, The logical basis for linguistic theory. *Proc IXth Int. Cong. Linguists*, Cambridge, Mass. (1962).

[14] — and Miller, G. A., Finite state languages. *Information and Control*, 1 (1958), 91–112.

[15] — and —, Introduction to the formal analysis of natural languages 1962. To appear in Bush, Galanter, Luce (eds.), *Handbook of mathematical psychology*, vol. 2, Wiley.

[16] Davis, M., *Computability and Unsolvability*. New York, McGraw-Hill, (1958).

[17] Elgot, C. C. Decision problems of finite automata design and related arithmetics. *Trans. Am. Math. Soc.*, 98 (1961), 21–51.

[18] Ginsburg, S. and Rice, H. G., Two families of languages related to Algol. *Technical Memorandum.* Systems Development Corporation; Santa Monica, California, (1961).

[19] — and Rose, G. F., Operations which preserve definability in languages. *Technical Memorandum.* Systems Development Corporation. Santa Monica, California, (1961).

[20] Jungen, R., Sur les séries de Taylor n'ayant que des singularités algebrico-logarithmiques sur leur cercle de convergence. *Comm. Math. Helvetici*, 3 (1931), 286–306.

[21] Kleene, S. C., Representation of events in nerve nets and finite automata. *Automata Studies*, Princeton University Press, (1956), 3–41.

[22] Kulagina, O., Ob odnom sposobe opredelenija grammatičeskix ponjatij. *Problemy Kibernetiki*, 1, Moscow, (1958).

[23] Lambek, J., The mathematics of sentence structure. *Am. J. Math.*, 65 (1958), 153–170.

—, On the calculus of syntactic types. *Proc. Symposium Applied Math.* 12, Am. Math. Soc., (1961).

[25] Lyndon, R. C., Equations in free groups, *Trans. Am. Math. Soc.* 96 (1960), 445–457.

[26] McNaughton, R., The theory of automata. To appear in *Advances in Computers*, vol II (Academic Press).

[27] Mahler, K., On a theorem of Liouville in fields of positive characteristic. *Canadian J. of Math.* 1, (1949), 397–400.

[28] Markov, A. A., Ob odnoi nevazrešimoi probleme, *Doklady Akad. Nauk:* n.s. 78, (1951), 1089–1092.

[29] Miller, G. A. and Chomsky, N., Finitary models of language users. 1962. To appear in Bush, Galanter, Luce (eds.), *Handbook of Mathematical Psychology*, vol. 2, Wiley.

[30] Minsky, M. L., Recursive unsolvability of Post's problem of Tag. *Ann. of Math.*, 74, (1961), 437–455.

[31] Myhill, J., Linear bounded automata. WADD *Tech. Note* 60–165. Wright Air Dvpt. Division. Wright Patterson Air Force Base Ohio, (1960).

[32] NEWELL, A. and SHAW, J. C., Programming the logic theory machine. *Proc. Western Joint Computer Conference*, (1957), 230.

[33] OETTINGER, A. G., Automatic syntactic analysis and the pushdown store. *Proc. of Symposia in Applied Math.*, 12, Am. Math. Soc., (1961).

[34] PARIKH, R. J., Language generating devices. *Quarterly Progress Report* no. 60. Research Laboratory of Electronics, M.I.T. January (1961), pp. 199–212.

[35] PERLES, M., RABIN, M. O. and SHAMIR, E., The theory of definite automata. *Tech. Report* no. 6, O.N.R., (1961).

[36] POST, E., A variant of a recursively unsolvable problem. *Bull. Amer. Math. Soc.*, (1946), 52, 264–268.

[37] RABIN, M. O. and SCOTT, D. Finite automata and their decision problems. *I.B.M. Journal of Research*, 3, (1959), 115–125.

[38] RANEY, G. N., Functional composition patterns and power-series reversion, *Trans. Am. Math. Soc.*, 94 (1960), 441–451.

[39] RITCHIE R. W., *Classes of recursive functions of predictable complexity*. Doctoral Diss, Dept. of Math, Princeton U, (1960).

[40] SCHEINBERG, S., Note on the Boolean properties of context-free languages. *Information and Control*, 3 (1960), 372–375.

[41] SCHÜTZENBERGER, M. P., On an application of semi-group methods. *I.R.E. Trans.*, IT2, (1956), 47–60.

[42] —, Un problème de la théorie de automates. *Séminaire Dubreil-Pisot* (Paris) Dec., (1959).

[43] —, A remark on finite transducers. *Information and Control*, 4 (1961), 185–196.

[44] —, On the definition of a family of automata. *Information and Control*, 4 (1961), 245–270.

[45] —, Some remarks on Chomsky's context-free languages. *Quarterly Progress Report* no. 68, Research Laboratory of Electronics, M.I.T., Oct. (1961).

[46] —, On a theorem of R. Jungen, 1962. To appear in *Proc. Am. Math. Soc.*

[47] —, Finite counting automata, 1962. To appear in *Information and Control.*

[48] —, On Context-free languages and pushdown storage. To appear in I.B.M. *Journal of Research.*

[49] —, Certain elementary families of automata. *Symp. on mathematical theory of automata*, Polytechnic Institute of Brooklyn, 1962.

[50] SHEPHERDSON, J. C., The reduction of two-way automata to one-way automata. I.B.M. *Journal of Research*, (1959), 198–200.

[51] SHAMIR, E., On sequential languages. *Tech. Report* no. 7, O.N.R., (1961).

[52] YAMADA, A., Counting by a class of growing automata. *Doctoral Diss.*, Univ. of Penna., Philadelphia, (1960).

Table des matières

Tome V

Marcel-Paul Schützenberger

ŒUVRES COMPLÈTES

éditées par Jean Berstel, Alain Lascoux et Dominique Perrin

Les treize tomes de cette édition contiennent l'ensemble des œuvres de Marcel-Paul Schützenberger qui ont fait l'objet d'une publication dans une revue scientifique ou un livre. Ses travaux couvrent une période de plus de 50 ans, depuis sa première note aux Comptes Rendus en 1943 jusqu'à son dernier article, paru en 1997.
Les publications sont présentées dans l'ordre chronologique. Chaque tome est précédé d'une courte introduction qui essaie d'éclairer certains des travaux, tant pour leur intérêt scientifique intrinsèque que pour l'écho qu'ils ont rencontré et les développements qu'ils ont suscités.

Tome 5 : 1961 – 1963

C'est une période extrêmement féconde de l'œuvre de Schützenberger. Elle voit la parution de nombreux articles dans plusieurs de ses domaines principaux de recherche, à savoir les langages algébriques (ou context-free), les séries rationnelles, la combinatoire, la théorie des codes, les automates et les transductions.
Ce tome contient notamment l'article « The algebraic theory of context-free languages » écrit avec Noam Chomsky, et qui est le plus connu de cette période. On y trouve la définition des langages de Dyck, des langages rationnels locaux, et le célèbre théorème de Chomsky-Schützenberger.
Citons aussi les articles « On a special class of recurrent events » et « On a family of submonoids » qui contiennent les principaux résultats de la théorie des codes bifixes.
L'article « On the definition of a family of automata » est le document fondateur pour la théorie des séries rationnelles en variables non commutatives.
L'article « The equation $a^m = b^n c^p$ in a free group » contient ce que l'on appelle le théorème de Lyndon et Schützenberger.
L'article « Quelques remarques sur une construction de Schensted », est le début d'une longue série d'articles concernant les tableaux de Young.