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 series in which every coefficient $<r, f>$ 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 $F'_r = \{f \in F : <r, f> \neq 0\}$.

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 $n<r, f>$

addition: $<r+r', f> = <r, f> + <r', f>$, for all f

multiplication: $<rr', f> = \Sigma <r, f> <r', f''>$, 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 $<r, f> = <r', f>$ for every word $f \in F$ of degree ("length") strictly, less than n and $<r, f> \neq <r', f>$ for at least one $f \in F$ of degree n.

Thus, $\|r-r'\| = 0$ if $<r, e> \neq <r', e>$, 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 $<r, f> \neq 0$ for some f of degree (= length) n. Thus r has a finite norm if and only if $<r, e> \neq 0$.

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 $<r, e> = 1$.

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 $F^*_{r_1} = \bigcup_{n>0} \left(F_{r_1}\right)^n$ 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

$r_{F'} = \sum_{f \in F'} f$ (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 $r, f \in \overline{A}(X)$ 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 $\overline{A}(X)$ by disregarding the order of the letters in the words $f \in F$. Formally, let α be a bijection (one-to-one mapping onto) $X \to \overline{X}$. An ordinary power series $\overline{r}$ in the variates $\overline{x}_i \in \overline{X}$ is an infinite sum $\overline{r} = \Sigma\, a_{n_1 n_2 \dots n_m} \overline{x}_1^{n_1} \overline{x}_2^{n_2} \dots \overline{x}_m^{n_m}$ extended to all the monomials $\overline{x}_1^{n_1} \overline{x}_2^{n_2} \dots \overline{x}_m^{n_m}$.

We can consider that any such r (with integral coefficients $a_{n_1 n_2 \dots n_m}$) is the image by the homomorphism α of at least one $r \in \overline{A}(X)$ by defining $a_{n_1 n_2 \dots n_m}$ 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 $\alpha f = \overline{x}_1^{n_1} \overline{x}_2^{n_2} \dots \overline{n}_m^{n_m}$, where α is the homomorphism sending F(X) onto the free commutative monoid generated by $\overline{X}$. 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 $\overline{A}(X)$

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 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 <u>polynomial</u> if its support is a finite set; an element $r \in \bar{A}(X)$ is <u>rational</u> 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 <u>free ring generated by</u> 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 $\bar{r} = \alpha r$ represents a rational function.

Consider, for instance, the formal power series r with $<r, f> = 1$ if and only if $f = x_2^7 x_1^{3+2n_1} x_2 x_1^{3+2n_2} x_2 \ldots x_2 x_1^{3+2n_m} x_2^5$, and $<r, f> = 0$, otherwise. This series r belongs to $\bar{R}(X)$ because r is equal to $x_2^6\left(e-x_2x_1^3\left(e-x_1^2\right)^{-1}\right)^{-1} x_2^5$, and αr can be reduced to the quotient of two polynomials by writing

$$
\begin{aligned}
\alpha r &= \bar{x}_2^6\left(e-\bar{x}_2\bar{x}_1^3\left(e-\bar{x}_1^2\right)^{-1}\right)^{-1} \bar{x}_2^5 \\
&= \bar{x}_2^6\left(\left(e-\bar{x}_1^2\right)^{-1}\left(e-\bar{x}_1^2-\bar{x}_2\bar{x}_1^3\right)\right)^{-1} \bar{x}_2^5 \\
&= \left(e-x_1^2\right)\left(e-x_1^2-x_2x_1^3\right)^{-1} x_2^{11} \\
&= \frac{x_2^{11}\left(1-x_1^2\right)}{1-x_1^2-x_1^2x_2}
\end{aligned}
$$

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 $\boxdot$ of m new elements ξ_i, and we denote by $\bar{\sigma}$ an m-tuple of polynomials σ_ξ in the (noncommutative) variates $y \in Y = X \cup \boxdot$ that satisfy the condition that $<\sigma_\xi, e> = <\sigma_\xi, \xi'> = 0$ for all $\xi, \xi' \in \boxdot$.

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

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

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

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

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

Indeed, the relation $\|w-w'\| \leq 1/n$ expressed the fact that the coefficients $<w_\xi, f>$ and $<w'_\xi, f>$ are equal for every coordinate $\xi \in \boxdot$ 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 $\overline{\sigma}$, a stronger result can be proved when w and w' satisfy the supplementary condition that $<w_\xi, e> = <w'_\xi, e> = 0$ for all ξ. Then, obviously, this last condition is still verified for $\overline{\sigma}w$ and $\overline{\sigma}w'$ (because $<\sigma_\xi, e> = 0$). Furthermore (because $<\sigma_\xi, \xi'> = 0$), we can conclude from $\|w-w'\| \leq 1/n$ that $\|\overline{\sigma}w-\overline{\sigma}w'\| \leq 1/n+1$. 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 $w, w_1, \dots, w_n, \dots$, where $w_o = (0, 0, \dots, 0)$ and $w_{n+1} = \overline{\sigma}w_n$. 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 $\lim_{n\to\infty} \overline{\sigma}w_n - w_n = 0$. This suggests that we speak of w as of a solution of the system of equations $\xi = \sigma_\xi$ (i.e., $w = \overline{\sigma}w$), 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_ξ.

We shall say, accordingly, that w_ξ is an algebraic element of $\overline{A}(X)$. Because of our definition of $\overline{\sigma}$, any w has a finite norm (i.e., $<w_\xi, e> = 0$). This restriction would be artificial; we shall denote by $\overline{S}(X)$ 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 $\overline{\sigma}$, or, as we prefer to say, by a set of "equations" $\xi = \sigma_\xi$.

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 $\boxdot$ and $\boxdot'$ of m and m' elements, and we consider $\boxdot''$ the union of $\boxdot$, $\boxdot'$ 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

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

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 $\left(x_1^{n_1}x_2^{n_1}\right)\left(x_1^{n_2}x_2^{n_2}\right) \dots \left(x_1^{n_m}x_1^{n_m}\right)$ 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.

Since the equation can also be written in the form $r = (e-x_1rx_2)^{-1}$, 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_1rx_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_1rx_2 satisfies the equation $x_1(e-x_1rx_2)^{-1} x_2 = x_1rx_2$. Taking the homomorphic image as α and writing $\bar{r}' = \alpha(x_1rx_2)$, we get the ordinary equation $\bar{x}_1\bar{x}_2(1-\bar{r}')^{-1} = \bar{r}'$; i. e., $\bar{r}'^2 - \bar{r} + \bar{x}_1\bar{x}_2 = \theta$.

By construction, the ordinary power series r' takes the value θ for $x_1x_2 = 0$ 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 $\langle x_1rx_2, f\rangle = 0$ 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_1rx_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 $\boxdot$ be the union of ζ, η and of ξ_i (i=1, ... 2m) and agree that $\xi_{i+m} = \xi_{i'}$, when $i = i'+m$. Let $X = \{x_i\}$ i = 1, 2, ..., 2m, and consider the 2m equations

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

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

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 $(4m^2-t^2)u^2 - 4m^2u + 2mt^2 = 0$, 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 $\xi \in \boxdot$.

Within $\overline{S}_1(X)$ itself we shall distinguish the special case $S_0(X)$ for which $\sigma = f + \Sigma \xi'f''$; i.e., only one-sided linear equations are considered.

Clearly, after taking the homomorphic image as α, both $\overline{S}_1(X)$ and $\overline{S}_0(X)$ collapse onto the ring of the ordinary rational functions but, at the level of $\overline{A}(X)$, the sets from $\overline{S}_0(X)$ 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 $\overline{S}_0^+(X)$. 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 $\overline{S}_0(X)$, but not even to $\overline{S}^+(X)$.

A priori the inclusion relations shown in Fig. XII-1 hold. Here, $P_0(X)$ and $P_0^+(X)$

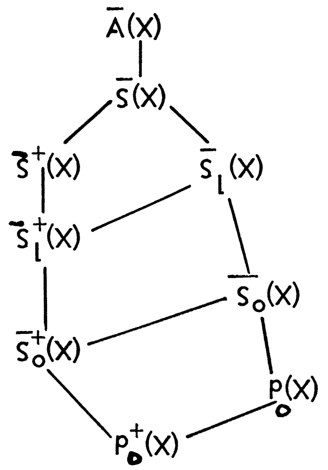

Fig. XII-1.

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.

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 $g_1gg_2 \in D(\xi, G)$. 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

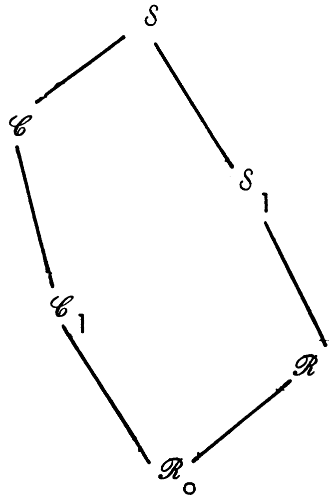

Fig. XII-2.

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

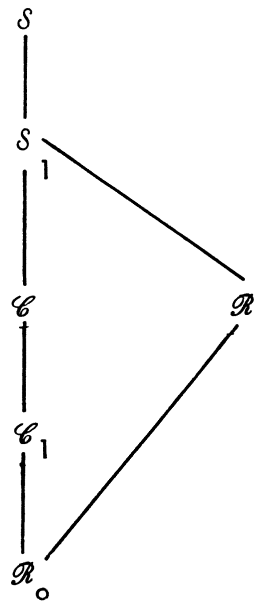

Fig. XII-3a.

or

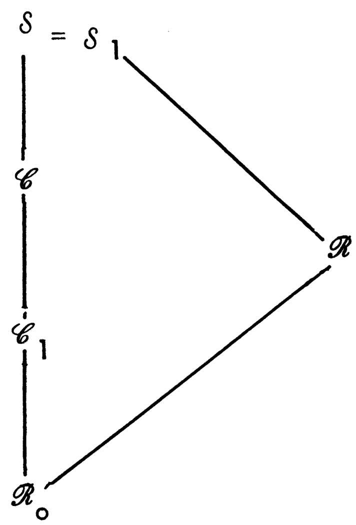

Fig. XII-3b.

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 $h \geq 1$ 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 $0 \leq k < mN$, then $\langle r, x^{n'} \rangle = 0$ for all $n' \geq N$; 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.

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 $\mathcal{C}_1$ is an undecidable one.

Let $\Xi = \{\xi\}$, $X = \{a, b, c\}$ 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 $i_1, i_2, i_3 \dots, i_n$ of indices such that the word $f_{i_1} f_{i_2} \dots f_{i_n}$ is equal to the mirror image of $f'_{i_n} f'_{i_{n-1}} \dots f'_{i_2} f'_{i_1}$. 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 $\mathcal{C}$ contains as a proper subset the family $\mathcal{C}'$ corresponding to the case in which the set of equations $w = \bar{\sigma} w$ has the following property which these authors call the "= sequential property": There exists an indexing $\xi_1, \xi_2 \dots \xi_m$ of the variates $\xi \in \Xi$ 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 $\bar{S}^1(X)$ of those $s \in \bar{S}(X)$ 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 $\mathcal{C}$ (in fact, in the closure of $\mathcal{C}_1$ by finite union and set product) which cannot be the support of an $s \in \bar{S}^1(X)$ having this property.

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

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

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

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

Ginsburg and Rice. Indeed, the last equation can be written $w_4 - x_1w_4 = x_1$. 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 $w_3 = x_2w_3x_2 + \sum_{n>0} x_2x_1^nx_2$, 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:

$$\begin{aligned} r(m,n,m',n') &= 0 \quad \text{if } m \neq m' \text{ and } n \neq n' \\ &= 1 \quad \text{if } m \neq m' \text{ and } n = n' \\ &= 1 \quad \text{if } m = m' \text{ and } n \neq n' \\ &= 2 \quad \text{if } m = m' \text{ and } n = n'. \end{aligned}$$

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 β^*f be the sum $\Sigma\beta(\phi f_1, \phi f_2)$ extended to all factorizations $f = f_1f_2$ of f, and say that f is accepted if and only if β^*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 $\{x_1^n x_2^n x_1^n = n \geq 0\}$) 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 → A the mapping $j : \boxdot \to A$ defined by $j\xi = \Sigma\{g: (\xi, g) \in G\}$.

For any $g \in F(X \cup \boxdot)$, 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 $\xi' \in \boxdot$. 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_f^n, ... 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 $\xi' \in \boxdot$ 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 $\xi \in \boxdot$ at least once at one time or another.

Let us now denote by u the sum $\Sigma\{\xi: \xi \in \boxdot'\}$ for any subset $\boxdot'$ of $\boxdot$. 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 → A. Let δ_o denote the retraction $A(X \cup \boxdot) \to A(X)$ induced by $\delta_o \xi = 0$ for each $\xi \in \boxdot$; (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 $\xi \in \boxdot$.)

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

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

We have

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

$j\beta = b$.

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

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

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

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

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

M. P. Schützenberger

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.