JOURNAL OF COMPUTER AND SYSTEM SCIENCES 20, 285–309 (1980)

Adherences of Languages

LUC BOASSON AND MAURICE NIVAT

UER de Mathématiques, Université Paris 7, France, and L.A. 248 Informatique Théorique et Programmation

Received October 15, 1979

This paper studies context-free sets of finite and infinite words. In particular, it gives a natural way of associating to a language a set of infinite words. It then becomes possible to begin a study of families of sets of infinite words rather similar to the classical studies of families of languages.

I. INTRODUCTION

The motivation of this work has to be found in the fact that algebraic grammars are special kinds and good examples of nondeterministic programs (cf. [1]). We can extend the computation domain, which is traditionally the free monoïd X^* over a finite alphabet by adding the set of infinite words X^ω, limits of increasing sequences of elements of X^* ordered by the relation "is a left factor of." The set of finite and infinite words $X^\infty = X^* \cup X^\omega$ is an algebraic computation domain in the sense of Scott [11]. The definition of the 0-ary multivalued function computed by an algebraic grammar G, considered as a nondeterministic program leads to the definition of the ∞-language $L^\infty(G, \xi)$ generated by G from nonterminal $\xi : L^\infty(G, \xi)$ is the union of $L(G, \xi)$ the ordinary language generated by G from ξ, containing only finite words, and $L^\omega(G, \xi)$ the set of infinite words generated by G from ξ whose definition was given by Nivat [8]. As one could expect there is a link between $L(G, \xi)$ and $L^\omega(G, \xi)$, an infinite word generated by G being in some sense a "limit" of finite words generated by G: to be precise if G is a Greibach reduced grammar, $L^\omega(G, \xi)$ is the adherence of $L(G, \xi)$ where the adherence of a language L is defined by

$$\mathrm{Adh}(L) = \{u \in X^\omega \mid \forall v < u \ \ w \in X^* : vw \in L\}.$$

In other words the adherence is exactly the set of infinite words all the finite left factors of which are left factors of words in L.

This theorem is established in Nivat [9].

It so happens that $\mathrm{Adh}(L)$ can be given an other definition if one considers on X^∞ the natural ultrametric topology associated with the distance

$$d(f, g) = 2^{-s(f,g)},$$

0022-0000/80/030285-25$02.00/0

where $s(f, g)$ is the length of the longest common left factor of f and g. (which may be infinite in which case f and g are the same infinite word and the distance $d(f, g) = 2^{-\infty} = 0$). The alternative definition of Adh(L) is then

Adh(L) is the set of cluster points of L in this d-topology.

The equivalence of the two definitions is proved below. A number of properties can be derived from both definitions: we retrieve, in the special case of algebraic grammars, general results on nondeterministic programs computing in complete metric spaces established by Arnold and Nivat [1]. But the notion of adherence has its own interest in the field of language theory: at the end of this paper we focus our attention on such aspects of adherences and raise the general question of defining families of adherences corresponding to the classical families of languages (rational cones and AFL's). The main result is that the adherence of any algebraic language can be obtained as the image in a continuous sequential mapping of the adherence of the Dyck set. The reader should be warned that our definition of $L^{\omega}(G, \xi)$, the set of infinite words generated by our algebraic grammar G from a nonterminal ξ is different from the definition used by Cohen and Gold [3] and Linna [7]: this difference is sufficient to explain apparently contradictory results. In the first part of this paper we present a summary of [8, 9] since the definitions and results therein contained are constantly used in the sequel.

II. Infinite Words and Sets of Infinite Words Generated by Algebraic Grammars

Let X be a finite alphabet. We denote by X^* the free monoïd generated by X, i.e., the set of finite words written with X as an alphabet, including the empty word denoted ϵ. If $f \in X^*$ is a word, we denote by $|f|$ its length, i.e., the number of occurrences of letters in f ($|\epsilon| = 0$) and, assuming that Ω is a letter which does not belong to X we define

$$\begin{aligned} f(n) &= \text{the } n\text{th letter of } f && \text{if} \quad n \leqslant |f|, \\ f(n) &= \Omega && \text{if} \quad n > |f|. \end{aligned}$$

The relation $\leqslant$ on X^* is defined by

$$f \leqslant g \Leftrightarrow \forall n \in P\colon n \leqslant |f| \Rightarrow f(n) = g(n).$$

In this writing P denotes $N \backslash \{0\}$, the set of strictly positive integers. If $f \leqslant g$ we say that f is a left factor of g. We say that f is a proper left factor of g if and only if

$f \leqslant g$ and $f \neq g$. We then write $f < g$.

An infinite word u on X is a mapping $u\colon P \to X$. The nth letter of u is $u(n)$. We shall denote $u[n]$ the finite word $u[n] = u(1)\, u(2) \cdots u(n)$.

The set of infinite words on X is denoted X^ω and the set of all finite and infinite words on X is denoted $X^\infty = X^* \cup X^\omega$. The relation $\leqslant$ is extended to X^∞ in the following obvious way:

$$\forall \alpha, \beta \in X^\infty \qquad \alpha \leqslant \beta \Leftrightarrow \forall n \in P \qquad n \leqslant |\alpha| \Rightarrow \alpha(n) = \beta(n)$$

with the convention that $u \in X^\omega \Rightarrow |u| = \infty$ and

$$\forall n \in P : n < \infty$$

we get that

$$\forall u, v \in X^\omega \qquad u \leqslant v \Leftrightarrow u = v$$

and

$$\forall f \in X^*, u \in X^\omega \qquad f \leqslant u \Leftrightarrow f = u[|f|].$$

We denote by $FG(\alpha)$ the set of finite left factors of $\alpha \in X^\infty$, i.e.,

$$\begin{aligned} FG(f) &= \{g \mid g \leqslant f\} \qquad && \text{for all} \quad f \in X^*, \\ FG(u) &= \{u[n] \mid n \in P\} \qquad && \text{for all} \quad u \in X^\omega. \end{aligned}$$

We call ∞-language any subset L of X^∞, language any subset L of X^*, ω-language any subset L of X^ω. And we denote in all three cases by $FG(L)$ the set

$$FG(L) = \{FG(\alpha) \mid \alpha \in L\}.$$

The manipulation of infinite words is made possible by three lemmas which will be in constant (implicit) use in the sequel:

LEMMA 1. *If $u_1 \leqslant u_2 \leqslant \cdots \leqslant u_n \leqslant \cdots$ is an increasing sequence of finite words in X^* ordered by $\leqslant$ and $|u_n|_{n\to\infty} \to \infty$ then there exists a unique $u \in X^\omega$ such that $\forall n \in P$ $u_n \in FG(u)$.*

This unique u is called the least upper bound (lub) *of the sequence $\{u_n\}$ and denoted by $u = \mathrm{Sup}\{u_n\}$.*

LEMMA 2. *For all $u \in X^\omega$, $L \subset X^\infty$*

$$\mathrm{card}(FG(u) \cap FG(L)) = \infty \Rightarrow FG(u) \subset FG(L).$$

LEMMA 3. (This is the celebrated Koenig's lemma). *Let, for all $n \in P$, E_n be a finite nonempty subset of a set E and $R \subset E \times E$ be a relation on E such that*

—$\mathrm{card}(\bigcup_{n\in P} E_n) = \infty$,

—$\forall n \in P, y \in E_{n+1}\ \exists x \in E_n : (x, y) \in R$.

Then there exists an infinite sequence $x_1, x_2, \ldots, x_n, \ldots$ of elements of E such that

$$\forall n \in P \quad x_n \in E_n \qquad \textit{and} \qquad (x_n, x_{n+1}) \in R.$$

We consider now the monoïd structure of X^* which we shall extend into a monoïd structure on X^∞.

The product fg of two words in X^* is defined by

$$\begin{array}{lll} \forall n \in P & n \leqslant |f| & fg(n) = f(n), \\ \forall n \in P & |f| < n \leqslant |f| + |g| & fg(n) = g(n - |f|), \\ \forall n \in P & |f| + |g| < n & fg(n) = \Omega. \end{array}$$

It is quite natural to define the product in X^ω by the same rules which give us $\forall f \in X^*$, $u \in X^\omega : fu$ is the infinite word given by

$$\begin{array}{lll} \forall n \in P & n \leqslant |f| & fu(n) = f(n), \\ \forall n \in P & n > |f| & fu(n) = u(n - |f|), \\ \multicolumn{2}{c}{\forall \alpha \in X^\infty, \quad u \in X^\omega} & u\alpha = u. \end{array}$$

This definition leads to the

LEMMA 4. $\forall \alpha, \beta \in X^\infty$

$$\alpha \geqslant \beta \Leftrightarrow \exists \gamma \in X^\infty : \alpha\gamma = \beta.$$

We retrieve the standard definition of "is a left factor of" extended to X^∞. We define now the product of two ∞-languages L and L' as $LL' = \{\alpha\beta \mid \alpha \in L \ \beta \in L'\}$ and we denote, for all $L \subset X^\infty$

$$L^{\mathrm{fin}} = L \cap X^*,$$
$$L^{\mathrm{inf}} = L \cap X^\omega.$$

We can state

LEMMA 5. $\forall L_1, L_2 \subset X^\infty$

$$\begin{aligned} (L_1 \cup L_2)^{\mathrm{fin}} &= L_1^{\mathrm{fin}} \cup L_2^{\mathrm{fin}}, \\ (L_1 \cup L_1)^{\mathrm{inf}} &= L_1^{\mathrm{inf}} \cup L_2^{\mathrm{inf}}, \\ (L_1L_2)^{\mathrm{fin}} &= L_1^{\mathrm{fin}} L_2^{\mathrm{fin}}, \\ (L_1L_2)^{\mathrm{inf}} &= L_1^{\mathrm{inf}} \cup L_1^{\mathrm{fin}} L_2^{\mathrm{inf}}, \\ (L_1^*)^{\mathrm{fin}} &= (L_1^{\mathrm{fin}})^*, \\ (L_1^*)^{\mathrm{inf}} &= (L_1^{\mathrm{fin}})^* L_1^{\mathrm{inf}}. \end{aligned}$$

A new operation will be used: to any $L \subset X^*$ we shall make correspond the ω-language L^ω given by

$$L^\omega = \{u \in X^\omega \mid \forall n \in P\ \exists p_n \in P : u[p_n] \in L^n \text{ and } p_{n_{n\to\infty}} \to \infty\}.$$

We also write

$$L^\omega = \{u \in X^\omega \mid f_1, f_2, \dots, f_n, \dots \in L\backslash\{\epsilon\}, u = f_1 f_2 \cdots f_n \cdots\}.$$

We can say that L^ω is the set of infinite products of nonempty words in L. It is coherent with our previous notation to define then

$$L^\infty = L^* \cup L^\omega.$$

We call L^∞ the infinite power of L.

Remark. The notion of least upper bound in X^∞ ordered by $\leqslant$ can be extended to directed subsets.

Call $L \subset X^\infty$ directed iff for all $\alpha, \beta \in L$ there exists $\gamma \in L$ such that $\alpha \leqslant \gamma$ and $\beta \leqslant \gamma$.

Then every directed subset L of X^∞ has a least upper bound $\mathrm{Sup}(L)$ such that for all $\alpha \in L$ $\alpha \leqslant \mathrm{Sup}(L)$ and for all $\gamma \in X^\infty$ the condition $\forall \alpha \in L$ $\alpha \leqslant \gamma$ implies $\mathrm{Sup}(L) \leqslant \gamma$.

To prove that let us remark that a directed subset L contains at most one infinite word for if $\alpha, \beta, \gamma \in X^\omega$

$$\alpha \leqslant \gamma \quad \text{and} \quad \beta \leqslant \gamma \quad \text{implies} \quad \alpha = \gamma = \beta$$

Thus if $L \cap X^\omega \neq \varnothing$ $L \cap X^\omega = \{u\}$ and u is the least upper bound. Otherwise $L \cap X^\omega = \varnothing$: the fact $L \subset X^*$ implies that L is countable. Then order the elements of L

$$l_1, l_2, \dots, l_n, \dots$$

and build the following increasing sequence of elements of L

$$f_1 = l_1$$

for all n, take $f_{n+1} \in L$ such that $f_n \leqslant f_{n+1}$ and $l_{n+1} \leqslant f_{n+1}$. We know that this is possible since L is directed.

If L is finite the process will stop leading to an element f_{n+1} which is a greatest element of L and a fortiori the least upper bound. If L is infinite then $\{f_n\}$ is an infinite sequence such that $|f_n| \to \infty$: there exists then $u = \mathrm{Sup}\{f_n\} \in X^\omega$ which is the least upper bound we look for. Indeed for all $n \in P$ one has $l_n \leqslant f_n < u$ and thus u is an upper bound of L. Suppose γ is an other upper bound : clearly γ cannot be a finite word.

The condition $\forall n \in P\ f_n < \gamma$ implies $\gamma = u$. ▮

A partially ordered set in which every directed subset has a least upper bound is called a complete partial order, abbreviated cpo. The structure of cpo plays a major role in the semantics of programming languages (Scott [11], Vuillemin [12], and Courcelle and Nivat [4]).

We now give definitions concerning the infinite words generated by an algebraic grammar and recall results from [8, 9].

Let G be the grammar on the terminal alphabet X and nonterminal alphabet Ξ.

$$\xi_i = P_i, \qquad i = 1,..., N,$$

where P_i is, for all $i \in [N]$, a finite subset of $(X \cup \Xi)^*$

An infinite derivation in G is a sequence

$$t_1, t_2, ..., t_n, ...$$

of words in $(X \cup \Xi)^*$ satisfying,

$$\forall i \in P, \qquad t_i \xrightarrow[G]{} t_{i+1},$$

where the relation $t \rightarrow_G t'$ is defined as usual.

We say that the infinite derivation $\{t_n\}$ is successful if and only if the length of the longest terminal left factor of t_n (denote it by $a(t_n)$) tends to infinity with n.

If $\{t_n\}$ is a successful infinite derivation, there exists according to Lemma 1, a unique infinite word u such that, $\forall n \in P$, $a(t_n) \in FG(u)$ and we say that u is produced by the derivation $\{t_n\}$ and write

$$t_1 \xrightarrow[G]{\omega} u.$$

We shall keep the standard notation $L(G, \xi_i)$ to denote the set of finite words which can be derived from ξ_i in G in finitely many steps

$$L(G, \xi_i) = \{g \in X^* \mid \xi_i \xrightarrow[G]{*} g\}.$$

We shall denote

$$L^\omega(G, \xi_i) = \{u \in X^\omega \mid \xi_i \xrightarrow[G]{\omega} u\},$$

and

$$L^\infty(G, \xi_i) = L(G, \xi_i) \cup L^\omega(G, \xi_i).$$

We call any subset of X^∞ which is equal to $L^\infty(G, \xi_i)$ for some algebraic grammar G and nonterminal ξ_i a ∞-algebraic language.

In the same way we shall talk about ω-algebraic language.

In order to state the principal results we need to extend to infinite words the notion of substitution.

Let $\mathbf{Q} = \langle Q_1, ..., Q_N \rangle$ be a vector of subsets of $(X \cup \Xi)^\infty$ and $f = g_0 \xi_{i_1} g_1 \xi_{i_2} \cdots g_{k-1} \xi_{i_k} g_k \in (X \cup \Xi)^*$

Then the result of the substitution of Q_i to ξ_i for all $i \in [N]$ is the set

$$f[\mathbf{Q}/\boldsymbol{\xi}] = \{g_0 h_1 g_1 h_2 \cdots g_{k-1} h_k g_k \mid h_j \in Q_{i_j}^{\text{fin}} \text{ for all } j \in [k]\}$$

$$\cup \{g_0 h_1 \cdots g_{l-1} h_l g_l u \mid l \in [k], h_j \in Q_{i_j}^{\text{fin}} \text{ for all } j \in [l] \text{ and } u \in Q_{i_{l+1}}^{\text{inf}}\}.$$

If u is an infinite word in $(X \cup \Xi)^\omega$ which we can write

$$u = g_0 \xi_{i_1} g_1 \xi_{i_2} \cdots g_{l-1} \xi_{i_l} g_l \cdots$$

the set $u[\mathbf{Q}/\boldsymbol{\xi}]$ is the following subset of $(X \cup \Xi]^\omega$

$$u[\mathbf{Q}/\boldsymbol{\xi}] = \{g_0 h_1 g_1 h_2 \cdots g_{l-1} h_l g_l \cdots \mid h_j \in Q_{i_j}^{\text{fin}} \text{ for all } j \in P \text{ and } \operatorname{card}\{j \mid h_j = \epsilon\} < \infty\}$$

$$\cup \{g_0 h_1 \cdots g_{l-1} h_l g_l v \mid h_j \in Q_{i_j}^{\text{fin}} \text{ for all } j \in [l], v \in Q_{i_{l+1}}^{\text{inf}}\}.$$

Note that we require that $u[\mathbf{Q}/\boldsymbol{\xi}]$ contain only infinite words.
In a trivial manner we then pose for all $L \subset (X \cup \Xi)^*$

$$L[\mathbf{Q}/\boldsymbol{\xi}] = \bigcup \{\alpha[\mathbf{Q}/\boldsymbol{\xi}] \mid \alpha \in L\}.$$

To the grammar $G : \xi_i = P_i$, $i \in [N]$, is now associated the mapping

$$\hat{G}(\mathbf{Q}) = \langle P_1[\mathbf{Q}/\boldsymbol{\xi}], \ldots, P_N[\mathbf{Q}/\boldsymbol{\xi}] \rangle = \mathbf{P}[\mathbf{Q}/\boldsymbol{\xi}].$$

If $((X \cup \Xi)^\infty)^N$ is ordered by inclusion componentwise it forms a complete lattice. We have results concerning fixed points of $\hat{G}$ which we now state after recalling the classical result of Schützenberger.

THEOREM 1 (Schützenberger [10]). *The mapping $\hat{G}$ associated with the algebraic grammar G has a smallest fixed point in $((X \cup \Xi)^*)^N$ given by*

$$Y(\hat{G}) = \bigcup_{n \in P} \hat{G}^n(\boldsymbol{\varnothing}) = \langle L(G, \xi_1), \ldots, L(G, \xi_N) \rangle.$$

The vector $\boldsymbol{\varnothing}$ is the N-vector all of whose components are equal to $\varnothing$.

THEOREM 2 (Nivat [8]). *The mapping $\hat{G}$ associated with the algebraic grammar G has a smallest fixed point in $((X \cup \Xi)^\infty)^N$ which is the same $Y(\hat{G})$ as in Schützenberger's theorem.*

If the grammar G is weakly Greibach, i.e., satisfies, $\forall i \in [N]$, $P_i \subset (X \cup \Xi)^* \times X \times (X \cup \Xi)^*$ then $\hat{G}$ has also a greatest fixed point given by

$$Z(\hat{G}) = \bigcap_{n \in P} \hat{G}^n(\mathbf{X}^\infty) = \langle L^\infty(G, \xi_1), \ldots, L^\infty(G, \xi_N) \rangle.$$

In this writing the vector $\mathbf{X}^\infty$ denotes the N-vector all of whose components are equal to X^∞.

Note that if G is weakly Greibach $\hat{G}$ has a unique fixed point in $((X \cup \Xi)^*)^N$ which is $Y(\hat{G})$.

Another extremely useful theorem concerning substitution is the following

THEOREM 3 (Nivat [9]). *Let $G : \xi_i = P_i$, $i \in [N]$, be an algebraic grammar and $\bar{G}$ be the following grammar constructed from G : the set of nonterminal symbols of $\bar{G}$ is the set $\bar{\Xi} = \{\bar{\xi}_1, ..., \bar{\xi}_N\}$ of barred letters corresponding with the nonterminals in Ξ.*

For every $\bar{\xi}_i$ we write the equation

$$\bar{\xi}_i = \{s\bar{\xi}_j \mid s \in (X \cup \Xi)^*,\ t \in (X \cup \Xi)^*\ s\xi_j t \in P_i\}$$

Then for all $i \in [N]$ we have

$$L^\omega(G, \xi_i) = L^\omega(G, \bar{\xi}_i)[\mathbf{L}(\mathbf{G}, \boldsymbol{\xi})/\boldsymbol{\xi}].$$

In this writing $\mathbf{L}(\mathbf{G}, \boldsymbol{\xi})$ is the N-vector $\langle L(G, \xi_1), ..., L(G, \xi_N)\rangle$

As a corollary we get, considering that $\bar{G}$ is a right-linear grammar generating only infinite words,

COROLLARY 1 (Nivat [9], Cohen and Gold [3]). *For any algebraic grammar G, and nonterminal ξ_i the ω-language $L^\omega(G, \xi_i)$ can be written as*

$$L^\omega(G, \xi_i) = \bigcup_{l=1}^{l=p} L_l(L'_l)^\omega,$$

where all the L_l, L'_l are ordinary algebraic languages in X^*.

Indeed a right-linear grammar like $\bar{G}$ generating only infinite words can be solved in a way analogous to the way one can solve ordinary right-linear grammars. Each component of the solution is then an element of what Cohen and Gold [3] call the ω-Kleene closure of Rat, i.e., a finite union of products

$$\bigcup_{l=1}^{e=p} R_l(R'_l)^\omega,$$

where R_l, R'_l are rational languages.

It suffices to substitute $\mathbf{L}(\mathbf{G}, \boldsymbol{\xi})/\boldsymbol{\xi}$ to obtain the corollary. With our definition the reverse property is not true: Cohen and Gold obtain it thanks to a more elaborate definition of ω-languages generated by algebraic grammars.

The standard example of the properties in this paragraph is the following. Let $G : \xi = a\xi\xi + b$ be the grammar generating the Lukasiewicz language

$$L = \{f \in \{a, b\}^* \mid |f|_a = |f|_b - 1 \text{ and } \forall f' < f\ |f'|_a \geqslant |f'|_b\}.$$

Clearly $FG(L) = \{f' \in \{a, b\}^* \mid |f'|_a \geqslant |f'|_b\}$ for if $|f'|_a - |f'|_b = n \geqslant 0$ then $f'b^n \in L$.

Since G is Greibach and reduced

$$L^\omega(G, \xi) = \mathrm{Adh}(L) = \{u \in \{a, b\}^\omega \mid \forall n \mid u[n]|_a \geqslant |u[n]|_b\}$$

Applying Theorem 3 we construct the grammar

$$\bar{G} : \bar{\xi} = a\bar{\xi} + a\xi\bar{\xi}$$

with the obvious solution $L^\omega(\bar{G}, \bar{\xi}) = (a + a\xi)^\omega$ whence

$$L^\omega(G, \xi) = (a + a\xi)^\omega[L/\xi] = (a + aL)^\omega.$$

III. Adherences of Languages

The definition was given in [9]:

Definition 1. Let L be a language, $L \subset X^*$: the adherence $\mathrm{Adh}(L)$ of L is the set $\mathrm{Adh}(L) = \{u \in X^\omega \mid \forall n \in P\ \exists v \in X^* : u[n]v \in L\}$. Note that $\mathrm{Adh}(L)$ is the closure of $FG(L)$ in the sense of Eilenberg [6].

We can immediately make a few useful remarks.

The adherence is also defined by

$$\mathrm{Adh}(L) = \{u \in X^\omega \mid FG(u) \subseteq FG(L)\}.$$

Clearly if L is finite $\mathrm{Adh}(L)$ is empty. Also $\mathrm{Adh}(L) = \mathrm{Adh}(FG(L))$ and $L_1 \subseteq L_2 \Rightarrow \mathrm{Adh}(L_1) \subseteq \mathrm{Adh}(L_2)$.

Definition 2. We say that the set $A \subseteq X^\omega$ is an adherence iff there exists $L \subseteq X^*$ such that

$$A = \mathrm{Adh}(L).$$

A is a rational (resp. algebraic) adherence iff there exists a rational (resp. algebraic) language L such that $L = \mathrm{Adh}(L)$.

Examples. (1) $A = \{a^\omega\}$ is a rational adherence since

$$A = \mathrm{Adh}(a^*).$$

(2) $B = \{u \in \{a, b\}^\omega \mid \forall v < u \mid v|_a \geqslant |v|_b\}$

is an algebraic adherence since B is the adherence of the Lukaziewicz language

$$L = \{v \in \{a, b\}^* \mid |v|_a = |v|_b - 1 \text{ and } \forall w < v \mid a|_a \geqslant |w|_b\}.$$

One can remark that B is also the adherence of the Dyck language on one letter

$$D_1' = \{v \in \{a, b\}^* \mid |v|_a = |v|_b \text{ and } \forall w \leqslant v \mid w|_a \geqslant |w|_b\}.$$

(3) $C = a^* b^\omega$ is not an adherence.

For if we had $C = \mathrm{Adh}(L)$ we would have $FG(C) \subseteq FG(L)$ whence $a^* \subseteq FG(L)$ which implies $a^\omega \in \mathrm{Adh}(L)$.

We establish now the first properties:

PROPERTY 1. *If* $A = \mathrm{Adh}(L)$ *we have*

$$FG(A) = \{v \in X^* \mid \mathrm{card}(vX^* \cap L) = \infty\}$$

and $A = \mathrm{Adh}(FG(A))$.

Proof. If $v \in FG(A)$, there exists $u \in A$, $n \in P$ such that $v = u[n]$. The fact that $u \in \mathrm{Adh}(L)$ implies $\forall p \in P$ there exists $w_p \in X^*$ such that $u[p]\, w_p \in L$. Thus $vu(n+1) \cdots u(p)\, w_p \in vX^* \cap L$ for all $p > n$ and $\mathrm{card}(vX^* \cap L)$ is infinite word $u \in \mathrm{Adh}(L)$ such that $v < u$.

Reversely suppose $\mathrm{card}(vX^* \cap L) = \infty$. We apply Koenig's lemma to build an infinite word $u \in \mathrm{Adh}(L)$ such that $v < u$.

Take $E_n = vX^n \cap FG(L)$ which is clearly finite and nonempty for all n. The union $E = vX^* \cap FG(L)$ is infinite by hypothesis. For all $n \in N$, $w \in E_{n+1}$ there exists $w' \in E_n$ such that $w \in w'X$. Whence there exists an infinite sequence $w_n \in E_n$ such that $w_{n+1} \in w_n X$. The sequence w_n is thus increasing for $\leqslant$ and $|w_n|_{n\to\infty} \to \infty$. The least upper bound of this sequence belongs to $\mathrm{Adh}(L)$.

The identity $A = \mathrm{Adh}(FG(A))$ comes easily from $FG(A) \subseteq FG(L)$ which implies

$$\mathrm{Adh}(FG(A)) \subseteq \mathrm{Adh}(FG(L)) = \mathrm{Adh}(L) = A$$

and, for all $u \in A$, $FG(u) \subseteq FG(A)$ which implies $A \subseteq \mathrm{Adh}(FG(A))$. ∎

Property 1 leads to the definition

DEFINITION 3. If L is a language $L \subset X^*$ we call L^c, the center of L defined by

$$L^c = FG(\mathrm{Adh}(L)) = \{v \in X^* \mid vX^* \cap L \text{ is infinite}\}.$$

We state as a lemma a number of properties of the center which are immediately deducible from the definition and Property 1.

LEMMA 6. *For all* $L \subseteq X^*$ *one has*

$$L^c \subseteq FG(L),$$
$$L^c = FG(L^c),$$
$$\epsilon \in L^c,$$
$$(L^c)^c = L^c,$$
$$L^c \text{ is empty if and only if } L \text{ is finite},$$
$$\mathrm{Adh}(L) = \mathrm{Adh}(L^c).$$

For all $L_1, L_2 \subseteq X^*$

$$\mathrm{Adh}(L_1) = \mathrm{Adh}(L_2) \Leftrightarrow L_1{}^c = L_2{}^c,$$
$$L_1 \subseteq L_2 \Rightarrow L_1{}^c \subseteq L_2{}^c.$$

DEFINITION 4. A language $L \subseteq X^*$ is said to be central iff it is equal to its center, i.e., $L = L^c$.

We shall establish some closure properties of the family of central languages. We first need

PROPERTY 2. *Let* L_1 *and* L_2 *be two languages. One has*

$$\mathrm{Adh}(L_1 \cup L_2) = \mathrm{Adh}(L_1) \cup \mathrm{Adh}(L_2),$$
$$\mathrm{Adh}(L_1 L_2) = \mathrm{Adh}(L_1) \cup L_1 \, \mathrm{Adh}(L_2),$$
$$\mathrm{Adh}(L_1^*) = L_1^* \, \mathrm{Adh}(L_1) \cup L_1{}^\omega.$$

Proof. (1) $\mathrm{Adh}(L_1) \cup \mathrm{Adh}(L_2) \subseteq \mathrm{Adh}(L_1 \cup L_2)$ is obvious since $FG(L_1)$ and $FG(L_2)$ are contained in $FG(L_1 \cup L_2)$.

Suppose now $u \in \mathrm{Adh}(L_1 \cup L_2)$. One has $FG(u) \subseteq FG(L_1 \cup L_2) = FG(L_1) \cup FG(L_2)$. Since $FG(u)$ is infinite its intersection with $FG(L_1)$ or $FG(L_2)$ is infinite. Whence $FG(u) \subseteq FG(L_1)$ or $FG(u) \subseteq FG(L_2)$ and $u \in \mathrm{Adh}(L_1) \cup \mathrm{Adh}(L_2)$.

(2) $FG(L_1 L_2) = FG(L_1) \cup L_1 \, FG(L_2)$.

If $FG(u) \subseteq FG(L_1) \cup L_1 \, FG(L_2)$ for the same reason as above $FG(u) \subseteq FG(L_1)$ or $FG(u) \subseteq L_1 \, FG(L_2)$.

Suppose $FG(u) \not\subseteq FG(L_1)$: there exists a maximal n such that $u[n] \in FG(L_1)$. Now $FG(u) \subseteq L_1 \, FG(L_2)$ implies that for all $p > n$, $u[p]$ can be factorized in

$$u[p] = vw, \qquad v \in L_1, \; w \in FG(L_2), \qquad |v| \leqslant n.$$

This means that for an infinite number of p's $u[p]$ can be factorized in vw_p for the same $v \in L_1$.

Thus $u = vu'$, where $u' \in X^\omega$ is such that

$$u'[p] = w_p \in FG(L_2) \qquad \text{for all} \quad p.$$

We can conclude that either $u \in \mathrm{Adh}(L_1)$ or $u \in L_1 \,\mathrm{Adh}(L_2)$.
The reverse inclusion is obvious.

(3) $FG(L_1^*) = L_1^* FG(L_1)$.

Let $u \in \mathrm{Adh}(L_1^*)$. Either for arbitrary large n there exists p_n such that $u[p_n] \in L_1{}^n$ and p_n tends to infinity with n, or there exists a maximal p such that $u[p] \in L_1^*$. In the first case we have by definition $u \in L_1{}^\omega$.

In the second case $FG(u) \subseteq L_1^* FG(L_1)$ implies that, for all $n > p$, $u[n]$ can be factorized in $u[n] = vw$, where $v \in L_1^*$ $w \in FG(L_1)$ and $|v| \leqslant p$.
By the same reasoning as above we conclude in that case that $u \in L_1^* \,\mathrm{Adh}(L_1)$.
The reverse inclusion is obvious. ▮

As a corollary to Property 2 we can establish the

LEMMA 7. *Let L_1 and L_2 be two languages*

$$(L_1 \cup L_2)^c = L_1{}^c \cup L_2{}^c,$$

$$\begin{aligned} (L_1L_2)^c &= FG(L_1) \cup L_1L_2{}^c && \text{if } L_2 \text{ is infinite} \\ &= L_1{}^c && \text{if } L_2 \text{ is finite,} \end{aligned}$$

$$\begin{aligned} (L_1^*)^c &= \varnothing \quad \text{if } L_1 = \varnothing \text{ or } L_1 = \{\epsilon\} \\ &= L_1^* FG(L_1) \text{ in the other cases.} \end{aligned}$$

Proof.

$$\begin{aligned} (L_1 \cup L_2)^c &= FG(\mathrm{Adh}(L_1 \cup L_2)) \\ &= FG(\mathrm{Adh}(L_1) \cup \mathrm{Adh}(L_2)) \\ &= L_1{}^c \cup L_2{}^c \\ (L_1L_2)^c &= FG(\mathrm{Adh}(L_1L_2)) \\ &= FG(\mathrm{Adh}(L_1) \cup L_1 \,\mathrm{Adh}(L_2)) \end{aligned}$$

Thus, if $\mathrm{Adh}(L_2) = \varnothing$, $(L_1L_2)^c = L_1{}^c$, if $\mathrm{Adh}(L_2) \neq \varnothing$,

$$(L_1L_2)^c = L_1{}^c \cup FG(L_1)\, L_2{}^c = FG(L_1)\, L_2{}^c;$$

since $L_1{}^c \subseteq FG(L_1)$ and $\epsilon \in L_2{}^c$,

$$(L_1^*)^c = FG(\mathrm{Adh}(L_1^*)) = FG(L_1^* \,\mathrm{Adh}(L_1) \cup L_1^\omega).$$

If $L_1 = \varnothing$ or $L_1 = \{\epsilon\}$ both $\mathrm{Adh}(L_1)$ and $L_1{}^\omega$ are empty.
Otherwise $(L^*)^c = L^* FG(L_1)$. ▮

We can now state

PROPERTY 3. *The family of infinite central languages is closed under union product and star.*

Proof. Immediate from Lemma 4. ▮

To end this paragraph we state a number of results concerning rational and algebraic adherences. The first theorem below is one of the principal motivations for the present study of adherences.

THEOREM 4. *If G is a Greibach grammar, i.e., satisfies,*

$$\forall i \in [N],\ P_i \subset X(X \cup \Xi)^*$$

then,

$$\forall i \in [N],\ L^{\omega}(G, \xi_i) \supset \mathrm{Adh}(L(G, \xi_i)).$$

If G is moreover reduced, i.e., satisfies, $\forall i \in [N], L(G, \xi_i) \neq \varnothing$ then one has equality.

For any ∞-algebraic language $L \subset X^{\infty}$, there exists a Greibach grammar G such that $L = L^{\infty}(G, \xi_i)$ for some nonterminal ξ_i if and only if

$$L^{\mathrm{inf}} \supset \mathrm{Adh}(FG(L))$$

(*such a language L will be said to be closed later on*).

This theorem is a slight modification of Theorems 5 and 6 of Nivat [9] and needs no new proof.

PROPERTY 4. *The center L^c of a rational* (*resp. algebraic*) *language L is rational* (*resp. algebraic*).

Proof. This results stems from Theorems 4 and 3. Take a Greibach grammar G generating $L = L(G, \xi_i)$ (G will be taken right-linear in case L is rational). From Theorem 4 one has $\mathrm{Adh}(L) = L^{\omega}(G, \xi_i)$. Then from Theorem 3 one has $L^{\omega}(G, \xi_i) = L^{\omega}(\bar{G}, \xi_i) \times [\mathbf{L}(\mathbf{G}, \boldsymbol{\xi})/\boldsymbol{\xi}]$. But $L^c = FG(\mathrm{Adh}(L)) = FG(L^{\omega}(\bar{G}, \xi_i)[\mathbf{L}(\mathbf{G}, \boldsymbol{\xi})/\boldsymbol{\xi}]$.

Very standard constructions from language theory then prove the result. ▮

PROPERTY 5. *The subset L of X^{ω} is a rational* (*resp. algebraic*) *adherence iff L is an adherence and $FG(L)$ is a rational* (*resp. algebraic*) *language in X^*.*

Proof. L is an adherence iff $L = \mathrm{Adh}(FG(L))$. One implication is thus obvious. Reversely suppose $L = \mathrm{Adh}(L')$, $L' \subset X^*$, L' rational (resp. algebraic). Then $L = \mathrm{Adh}(L'^c)$ and since $FG(L)$ and L'^c are both central we have $L'^c = FG(L)$. The result follows from Property 4. ▮

IV. Topology X^ω

DEFINITION 5. The ∞-language $L \subset X^\infty$ is said to be closed if and only if $L^{\inf} \supset \mathrm{Adh}(FG(L))$.

PROPERTY 6. *The family of closed ∞-languages is closed under union, product and infinite power.*

Proof. We have the following sequences of inclusions

$$\begin{aligned}
(L_1 \cup L_2)^{\inf} &= L_1^{\inf} \cup L_2^{\inf} \\
&\supset \mathrm{Adh}(FG(L_1)) \cup \mathrm{Adh}(FG(L_2)) \\
&= \mathrm{Adh}(FG(L_1) \cup FG(L_2)) \qquad \text{by Property 2} \\
&= \mathrm{Adh}(FG(L_1 \cup L_2)) \\
(L_1 L_2)^{\inf} &= L_1^{\inf} \cup L_1^{\mathrm{fin}} L_2^{\inf} \\
&\supset \mathrm{Adh}(FG(L_1)) \cup L_1^{\mathrm{fin}}\, \mathrm{Adh}(FG(L_2))
\end{aligned}$$

But $FG(L_1L_2) = FG(L_1) \cup L_1^{\mathrm{fin}} FG(L_2)$ and

$$\mathrm{Adh}\, FG(L_1L_2) = \mathrm{Adh}(FG(L_1)) \cup L_1^{\mathrm{fin}}\, \mathrm{Adh}(FG(L_2)).$$

To prove the third part of Property 6 we first make precise the definition of L^∞ for $L \subset X^\infty$. From $L^\infty = L^* \cup L^\omega$ we get

$$\begin{aligned}
(L^\infty)^{\mathrm{fin}} &= (L^{\mathrm{fin}})^*, \\
(L^\infty)^{\inf} &= (L^{\mathrm{fin}})^* L^{\inf} \cup (L^{\mathrm{fin}})^\omega.
\end{aligned}$$

Thus assuming $L^{\inf} \supset \mathrm{Adh}\, FG(L)$ we get

$$(L^\infty)^{\inf} \supset (L^{\mathrm{fin}})^*\, \mathrm{Adh}\, FG(L) \cup (L^{\mathrm{fin}})^\omega.$$

But $FG(L^\infty) = (L^{\mathrm{fin}})\, FG(L)^*$ whence

$$\begin{aligned}
\mathrm{Adh}(FG(L^\infty)) &= \mathrm{Adh}(L^{\mathrm{fin}})^* \cup (L^{\mathrm{fin}})^*\, \mathrm{Adh}(FG(L)) \\
&= (L^{\mathrm{fin}})^*\, \mathrm{Adh}(L^{\mathrm{fin}}) \cup (L^{\mathrm{fin}})^\omega \cup (L^{\mathrm{fin}})^*\, \mathrm{Adh}(FG(L))
\end{aligned}$$

and since $\mathrm{Adh}(L^{\mathrm{fin}}) = \mathrm{Adh}(FG(L^{\mathrm{fin}})) \subset \mathrm{Adh}(FG(L))$

we have $(L^\infty)^{\inf} \supset \mathrm{Adh}(FG(L^\infty))$. ∎

We now justify the terminology by considering on X^∞ the topology whose closed sets are precisely the ∞-languages satisfying $L^{\inf} \supset \mathrm{Adh}(FG(L))$.

Consider the mapping cl of 2^{X^∞} into itself given by

$$\mathrm{cl}(L) = L \cup \mathrm{Adh}\, FG(L)$$

Clearly $\mathrm{cl}(L)$ is the least closed ∞-language containing L:

—$\mathrm{cl}(L)$ is closed since $\mathrm{cl}(L)^{\mathrm{inf}} = L^{\mathrm{inf}} \cup \mathrm{Adh}(FG(L))$ and $\mathrm{Adh}(FG(\mathrm{cl}(L)) = \mathrm{Adh}\, FG(L)$

—Suppose $L \subset L'$ and L' is closed:

$$(L')^{\mathrm{inf}} \supset \mathrm{Adh}\, FG(L') \supset \mathrm{Adh}\, FG(L) \qquad \text{implies} \quad \mathrm{cl}(L) \subset L'.$$

The mappings cl satisfies the 4 following axioms

—$\mathrm{cl}(\varnothing) = \varnothing$,

—$\mathrm{cl}(\mathrm{cl}(L)) = \mathrm{cl}(L)$,

—$L \subset \mathrm{cl}(L)$,

—$\mathrm{cl}(L_1 \cup L_2) = \mathrm{cl}(L_1) \cup \mathrm{cl}(L_2)$.

This is the standard set of axioms which allows to define a topology from a closure mapping by taking as open sets the complements of the closed sets, a closed set being by definition a set which satisfies $L = \mathrm{cl}(L)$ (cf. Dugundji [5]).

The principal result in this section is that this topology is also definable by means of an ultrametric distance on X^∞ and that, with respect to this metric, X^∞ is a complete metric space.

Definition 6. For α, β in X^∞ the distance $d(\alpha, \beta)$ is defined by

$$d(\alpha, \beta) = 2^{-\min\{n \in P \mid \alpha(n) \neq (\beta n)\}} \qquad \text{if} \quad \exists n \in P : \alpha(n) \neq \beta(n)$$
$$d(\alpha, \beta) = 0 \qquad \text{iff} \quad \forall n \in P : \alpha(n) = \beta(n).$$

Property 7. *The distance d is an ultrametric distance, i.e., satisfies for all $\alpha, \beta, \gamma \in X^\infty$*

$$d(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta,$$
$$d(\alpha, \beta) = d(\beta, \alpha),$$
$$d(\alpha, \beta) \leqslant \max\{d(\alpha, \gamma), d(\beta, \gamma)\}.$$

We define as usual the associated topology on X^∞ by taking as a basis of neighborhoods of $\alpha \in X^\infty$ the family of open balls

$$B(\alpha, n) = \{\beta \in X^\infty \mid d(\alpha, \beta) < 2^{-n}\}.$$

Since $d(\alpha, \beta) < 2^{-n}$ iff $\alpha[n] = \beta[n]$ an alternative definition of $B(\alpha, n)$ is

$$B(\alpha, n) = \{\beta \in X^\infty \mid \alpha \text{ and } \beta \text{ have the same left factor of length } n\}.$$

A cluster point of a subset L of X^∞ is then a word $\alpha \in X^\omega$ such that every neighborhood of α contains an element of L distinct from α. In other words α is a cluster point of L iff

$$\forall n \in P \qquad \exists \beta \neq \alpha : \alpha[n] = \beta[n]$$

and the set of cluster points of L, also called the derived set of L, denoted L', is precisely $L' = \mathrm{Adh}(FG(L))\backslash L^{\mathrm{inf}}$.

A set $L \subset X^\infty$ is thus closed iff it contains its derived set and this is equivalent to the condition

$$L^{\mathrm{inf}} \supset \mathrm{Adh}\, FG(L).$$

We retrieve the above definition of closed sets: the topology defined by the closure mapping cl and the topology associated to the distance d are the same. We state

THEOREM 5. *For every $L \subset X^*$ the adherence* Adh(L) *is the set of cluster points of L in the metric topology defined on X^∞.*

Terminological remark. In the standard terminology the adherence of L is $L \cup \mathrm{Adh}(L)$: we prefer to keep the word adherence to designate Adh(L) for we shall have to distinguish constantly between finite and infinite words and the word adherence is very convenient to talk about the ω-part of the topological closure of L which is $\bar{L} = L \cup \mathrm{Adh}(L)$.

Having a metric topology on X^ω we can now use freely all the definitions and results concerning sequences, convergence, continuity, and so on.

A sequence $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ is a d-Cauchy (or simply Cauchy) sequence iff

$$\forall n \in P \qquad \exists N \in P \qquad \forall p, q \in P\backslash[N] : \alpha_p[n] = \alpha_q[n].$$

The squence $\{\alpha_n\}$ converges to a limit β iff

$$\forall n \in P \qquad \exists N \in P \qquad \forall p \in P\backslash[N] : \alpha_p[n] = \beta[n].$$

Clearly in X^ω a sequence $\{\alpha_n\}$ converges iff it is a d-Cauchy sequence.

Thus X^∞ is a complete metric space.

Let us write $\alpha_n \to \beta$ iff the sequence $\{\alpha_n\}$ converges to β. We have the following properties: every increasing sequence $\{\alpha_n\}$ converges.

Suppose $\alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_n \leqslant \cdots$. Then either $|\alpha_n| \to \infty$ and $\alpha_n \to \mathrm{Sup}\, \alpha_n$ or the sequence $|\alpha_n|$ is bounded : in this case the sequence α_n is stationary, i.e., there exists an integer N such that, $\forall n \in P\backslash[N]$, $\alpha_n = \alpha_N$. Then $\alpha_n \to \alpha_N$. Now suppose $\{\alpha_n\}$ is d-Cauchy, i.e.,

$$\forall n \in P \qquad \exists N_n \qquad \text{such that} \quad \forall p, q \in P\backslash[N_n], \qquad \alpha_p[n] = \alpha_q[n].$$

If one takes the smallest possible N_n for all n then the sequence $\alpha_N \mathrm{t}[N_n]$ is an increasing sequence : if $N_n \to \infty$ then $\alpha_N \mathrm{t}[N_n] \to \mathrm{Sup}\, \alpha_N \mathrm{t}[N_n]$ and $\{\alpha_n\}$ is convergent with the same limit. If N_n is bounded then $\{\alpha_n\}$ is stationary and thus also convergent. ∎

Other useful topological notions are continuity of functions and compactness. According to the standard definition the mapping f of X^∞ into Y^∞ is said to be continuous iff the reverse image $f^{-1}(L)$ of any closed set $L \subset Y^\infty$ is closed in X^∞. Since every point in X^∞ and Y^∞ have a countable basis of neighborhoods the mapping f is continuous iff for every sequence $\{\alpha_n\}$, $\alpha_n \to \beta$ implies $f(\alpha_n) \to f(\beta)$.

Using an argument similar to the one above one can show the useful

PROPERTY 8. *The increasing mapping f: $X^\infty \to Y^\infty$ is continuous iff for every increasing sequence $\{\alpha_n\}$ one has $f(\alpha_n) \to f(\mathrm{Sup}\ \alpha_n)$.*

Remark. This property asserts that the two notions of continuity which exist in X^∞ coïncide for increasing mappings: one notion is the notion of continuity according to the d-topology. The other notion is the notion of continuity as defined by Scott [11] in a complete partially ordered set as is X^∞ (ordered by $\leqslant$).

A last remark is that X^∞ is compact. This come from the fact that the metric d is totally bounded, i.e., satisfies: for every $\epsilon > 0$ and covering of X^∞ by open balls of diameter ϵ there exists a finite subcovering (Dugundji [5]).

We shall use the following consequence: every continuous mapping of X^∞ into Y^∞ is closed that is maps any closed subset of X^∞ into a closed subset of Y^∞.

V. SEQUENTIAL MAPPINGS OF ∞-LANGUAGES

Up to now we have given definitions of algebraic ω-languages, algebraic ∞-languages, and closed algebraic ∞-languages. We shall denote these families by ω-Alg, ∞-Alg, and $\bar{\infty}$-Alg, respectively. Our aim in this last section is to look at the structure of these families in a very similar way to that done in the finite case, using rational transductions: we assume the reader is familiar with the theory of rational cones, semi-AFLs and AFLs and the standard classification of subcones of Alg.

Obviously we need first to extend to infinite words the notion of rational transduction, that is, to define $\tau(u)$ when given a rational transduction τ and an infinite word u. This can only be done by continuity, looking at the sequence $\{\tau(u[n])\}$.

In the genercal case we do not have much information on this sequence which is a sequence of sets.

That is why we choose here to deal with the subcase of sequential mappings (or gsm mappings) for which we known that the sequence $\tau(u[u])$ is an increasing one. This will be sufficient to exhibit a "generator" of ω-Alg, namely, the adherence of the Dyck set and raise a few questions.

We recall the definition of a gsm (generalized sequential machine): a gsm is a finite automata with a partial output function.

$$\mathscr{A} = \langle X, Y, Q, q_0, \lambda, \delta \rangle$$

where

X is the input alphabet

Y is the output alphabet

Q is a finite set of states

$q_0 \in Q$ is the initial state

λ maps $Q \times X$ into Q (λ is the transition function)

δ maps $Q \times X$ into $Y^* \cup \{0\}$ (δ is the output function)

Such a machine defines a mapping $\gamma_{\mathscr{A}}$ in the following way:

—First extend λ into a mapping of $Q \times X^*$ into Q by the rules

$$\forall q \in Q, \quad \lambda(q, \epsilon) = q,$$
$$\forall q \in Q, \quad f \in X^*, x \in X, \qquad \lambda(q, xf) = \lambda(\lambda(q, x), f);$$

—then extend δ into a mapping of $Q \times X^*$ into $Y^* \cup \{0\}$:

$$\forall q \in Q, \quad \lambda(q, \epsilon) = \epsilon,$$
$$\forall q \in Q, \quad f \in X^*, \quad x \in X, \qquad \delta(q, xf) = \delta(q, x) \cdot \delta(\lambda(q, x), f)$$

with the convention that $0.g = g.0 = 0.0 = 0$ for all $g \in Y^*$.

The mapping $\gamma_{\mathscr{A}}$ is then given by

$$\gamma_{\mathscr{A}}(f) = \delta(q_0, f) \qquad \text{for all} \quad f.$$

By definition a mapping of X^* into $Y^* \cup \{0\}$ is said to be sequential iff there exists a gsm $\mathscr{A}$ such that $\gamma = \gamma_{\mathscr{A}}$.

Now consider an infinite word $u \in X^\omega$. Three cases have to be considered

—If there exists an $n \in P$ such that $\gamma(u[n]) = 0$ then, for all $n' \geqslant n$, $\gamma(u[n']) = 0$.

—The sequence $\gamma(u[n])$ is stationary.

—The sequence $\gamma(u[n])$ is increasing with $|\gamma(u[n])| \to \infty$.

If we just wish to respect continuity we are lead to define

$$\gamma(u) = 0 \qquad \text{in the first case,}$$
$$\gamma(u) = \operatorname{Sup}\{\gamma(u[n])\}$$

in the two other cases with, in the second case, the fact that $\gamma(u)$ is a finite word.

We shall see later that letting $\gamma(u)$ belong to X^* for some $u \in X^\omega$ prevents some important properties from being true.

The best way to avoid that is to restrict ourselves to the consideration of faithful sequential mappings.

DEFINITION 7. The sequential mapping γ is faithful iff

$$\forall g \in Y^* \ \mathrm{card}\{f \in X^* \mid \gamma(f) = g\} \text{ is finite.}$$

LEMMA 8. *Let γ be a sequential mapping.*
For all $u \in X^\omega$, $\gamma(u) \neq 0 \Rightarrow \gamma(u) \in Y^\omega$ iff γ is faithful.

Proof. Suppose, $\forall n \in P$, $\gamma(u[n]) \neq 0$ and $\mathrm{Sup}\,\gamma(u[n]) \in Y^*$. Then there exists an $N \in P$ such that, $\forall n \geqslant N$, $\gamma(u[n]) = \mathrm{Sup}\,\gamma(u[n])$ and thus $\mathrm{card}\{f \in X^* \mid \gamma(f) = \mathrm{Sup}\,\gamma(u[n])\}$ is infinite.

Conversely suppose there exists some $g \in Y^*$ such that $\mathrm{card}\{f \in X^* \mid \gamma(f) = g\}$ is infinite. A straightforward use of Koenig's lemma gives an infinite word u such that $\gamma(u[n]) = g$ for all sufficiently large n where $\gamma(u) = g$ which contradicts the condition $\gamma(u) \in Y^\omega$. ▮

Let us now prove the important properties of sequential mappings thus extended.

PROPERTY 9. *If γ is a faithful sequential mapping of X^* into Y^* then for all $L \subset X^*$*

$$\gamma(\mathrm{Adh}(L)) = \mathrm{Adh}(\gamma(L)).$$

Proof. Suppose $u \in \gamma(\mathrm{Adh}(L))$. For some $v \in \mathrm{Adh}(L)$ $u = \gamma(v)$ and thus

$$u = \lim\{\gamma(v[n])\}.$$

Then $\forall p \in P\ \exists n \in P : u[p] \leqslant \gamma(v[n])$ which implies $FG(u) \subset \gamma(FG(v))$. But $v \in \mathrm{Adh}(L) \Rightarrow FG(v) \subset FG(L)$. Since γ is sequential $\gamma(FG(L)) \subset FG(\gamma(L))$ whence $FG(u) \subset \gamma(FG(v)) \subset \gamma(FG(L)) \subset FG(\gamma(L))$ and this implies $u \in \mathrm{Adh}(\gamma(L))$. Conversely let u belong to $\mathrm{Adh}(\gamma(L))$: then

$$\forall n \in P \qquad \exists v_n \in L : u[n] \leqslant \gamma(v_n).$$

By an obvious remark on sequential mappings

$$\forall n \in P \quad \exists f_n \in X^*,\ x_n \in X : f_n x_n \leqslant v_n\,,\quad u[n] \leqslant \gamma(f_n x_n) \qquad \text{and} \qquad \gamma(f_n) < u[n]$$

Thus for all $n \in P$ the set

$$E_n = \{fx \in FG(L) \mid f \in X^*,\ x \in X,\ u[n] \leqslant \gamma(fx),\ \gamma(f) < u[n]\}$$

is nonempty.

The faithfulness of γ implies that E_n is finite and that $\bigcup_{n \in P} E_n$ is infinite. Clearly for all $n \in P$, $g \in E_{n+1}$ there exists $f \in E_n$ such that $f \leqslant g$.

Applying Koenig's lemma we obtain an infinite increasing sequence $\{g_n\}$ such that $g_n \in E_n$ and $|g_n|_{n\to\infty} \to \infty$. The limit $v = \lim\{g_n\}$ is such that $\gamma(v) = u$ and $v \in \mathrm{Adh}(L)$. ▮

Remarks. (1) The following example shows that Property 9 does not hold if γ is not faithful:

Take $L = \{a^n b^n c^p \mid n, p \in P\}$. Consider the morphism $\varphi: \{a, b, c\}^* \to \{b, c\}^*$ defined by

$$\varphi(a) = \epsilon, \qquad \varphi(b) = b, \qquad \varphi(c) = c.$$

We have $\mathrm{Adh}(L) = \{a^\omega\} \cup \{a^n b^n c^\omega \mid n \in P\}$

$$\varphi(\mathrm{Adh}(L)) = \{\epsilon\} \cup \{b^n c^\omega \mid n \in P\},$$
$$\mathrm{Adh}(\varphi(L)) = \mathrm{Adh}\{b^n c^p \mid n, p \in P\} = \{b^\omega\} \cup \{b^n c^\omega \mid n \in P\}.$$

(2) Let us say that the mapping $\gamma: X^* \to Y^* \cup \{0\}$ is faithful on L, for $L \subset X^*$ iff

$$\forall g \in Y^* \ \mathrm{card}\{f \in L \mid \gamma(f) = g\} \text{ is finite.}$$

The proof we gave of Property 9 also proves the stronger result:

PROPERTY 10. *If the sequential mapping $\gamma: X^* \to Y^* \cup \{0\}$ is faithful on $FG(L)$ then*

$$\gamma(\mathrm{Adh}(L)) = \mathrm{Adh}(\gamma(L)).$$

Examples of Faithful Sequential Mappings

The two following examples will play an important role in the sequel

(1) Let $K \subset X^*$ be a rational language such that $K = FG(K)$. Such a language can be recognized by a finite automaton in which all states but the sink states are final states.

Define the mapping $I_K: X^* \to X^* \cup \{0\}$ by

$$I_K(f) = f \quad \text{if } f \in K,$$
$$I_K(f) = 0 \quad \text{if } f \notin K.$$

According to our definition, $I_K(u) = u$ iff $u[n] \in K$ for all $n \in P$. That is

$$I_K(u) = u \quad \text{if } u \in \mathrm{Adh}(K),$$
$$I_K(u) = 0 \quad \text{if } u \notin \mathrm{Adh}(K).$$

In other words

LEMMA 9. *For every rational language $K \subset X^*$ such that $K = FG(K)$ the mapping $I_K: X^\infty \to X^\infty \cup \{0\}$ defined by*

$$I_K(\alpha) = \alpha \quad \textit{if } \alpha \in cl(K),$$
$$I_K(\alpha) = 0 \quad \textit{if } \alpha \notin cl(K).$$

is a faithful sequential mapping.

We have a reverse property.

LEMMA 10. *Let* $B = \mathrm{Adh}(K)$ *be a rational adherence. Then there exists a faithful sequential mapping* γ *such that, for all* $A \subset X^\omega$, $A \cap B = \gamma(A)$.

Proof. It suffices to take $\gamma = I_H$ where $H = FG(K)$. ▮

(2) An other important faithful sequential mapping is the following. Denote D_n the Dyck set on n letters $x_1, x_2, \dots, x_n$ a grammar of which is

$$\xi = \sum_{i=1}^{i-n} x_i \xi \overline{x_i} \xi + \epsilon.$$

Consider D_2 on the alphabet $\{a, b\}$ a grammar of which is

$$\xi = a\xi\,\bar{a}\,\xi + b\xi\,\bar{b}\,\xi + \epsilon.$$

It is well known that $D_n = \varphi^{-1}(D_2)$, where φ is the morphism of $\{x_1, \dots, x_n\}^*$ into $\{a, b\}^*$ given by

$$\varphi(x_i) = ab^{i-1}a,$$
$$\varphi(\overline{x_i}) = \bar{a}\bar{b}^{i-1}\bar{a}.$$

The special from of the morphism φ allows us to describe φ^{-1} as a sequential mapping.

We take a set of $2n + 2$ states $\{q_0, q_1, \dots, q_n, q_s, \bar{q}_1, \dots, \bar{q}_n\}$ and define the transition and output functions λ and δ by

$$\lambda(q_0, a) = q_1, \qquad \lambda(q_0, \bar{a}) = \bar{q}_1,$$
$$\delta(q_0, a) = \epsilon, \qquad \delta(q_0, \bar{a}) = \epsilon.$$

For all $i = 1, \dots, n - 1$,

$$\lambda(q_i, b) = q_{i+1}, \qquad \lambda(\bar{q}_i, \bar{b}) = \bar{q}_{i+1},$$
$$\delta(q_i, b) = \epsilon, \qquad \delta(\bar{q}_i, b) = \epsilon,$$
$$\lambda(q_i, a) = q_0, \qquad \lambda(\bar{q}_i, a) = q_0,$$
$$\delta(q_i, a) = x_i, \qquad \delta(\bar{q}_i, a) = \bar{x}_i,$$
$$\lambda(q_n, b) = q_s, \qquad \lambda(\bar{q}_n, b) = q_s,$$
$$\delta(q_n, b) = 0, \qquad \delta(\bar{q}_n, b) = 0,$$
$$\lambda(q_n, a) = q_0, \qquad \lambda(\bar{q}_n, a) = q_0,$$
$$\delta(q_n, a) = x_n, \qquad \delta(\bar{q}_n, a) = \bar{x}_n.$$

λ maps all the pairs which do not appear in this list on q_s and δ maps all the pairs which do not appear in this list on 0. The proof that φ^{-1} is the sequential mapping defined by the above gsm is straightforward. ▮

Let us now define inverse sequential mappings in the usual way: if $\gamma\colon X^\infty \to Y^\infty$ then for all $\beta \in Y^\infty$

$$\gamma^{-1}(\beta) = \{\alpha \in X^\infty \mid \gamma(\alpha) = \beta\}.$$

PROPERTY 11. *If γ is a faithful sequential mapping and $A = \mathrm{Adh}(L)$ then $\gamma^{-1}(A) = \mathrm{Adh}(\gamma^{-1}(FG(L)))$*

Proof. Suppose $u = \gamma(v)$, $u \in \mathrm{Adh}(L)$. Then, for all $n \in P$, $\gamma(v[n]) < u$. Thus $FG(v) \subset \gamma^{-1}(FG(u)) \subset \gamma^{-1}(FG(L))$ whence $v \in \mathrm{Adh}(\gamma^{-1}(FG(L)))$.

Conversely suppose for all $n \in P$, $v[n] \in \gamma^{-1}(FG(L))$: then $\gamma(v[n])$ is an increasing of words in $FG(L)$ which has a limit $u \in \mathrm{Adh}(FG(L))$ since γ is faithful. ∎

Remark. The equality $\gamma^{-1}(A) = \mathrm{Adh}(\gamma^{-1}(L))$ is false as shown by the example $L = a^*b$, φ is the morphism of $\{a\}^*$ in $\{a, b\}^*$ given by $\varphi(a) = a$. Then $\varphi^{-1}(\mathrm{Adh}(L)) = \varphi^{-1}(a^\omega) = a^\omega$ and $\mathrm{Adh}(\varphi^{-1}(L)) = \varnothing$.

We now prove the main result of this section:

THEOREM 6. *For every algebraic adherence A these exists a faithful sequential mapping γ such*

$$\gamma(\mathrm{Adh}(D_2)) = A.$$

Proof. The proof uses a slight modification of the Chomsky–Schützenberger theorem [2].

Consider $A = \mathrm{Adh}(L)$ and suppose, this is no restriction, that $\epsilon \notin L$. Then take a grammar G generating L which is in Greibach quadratic from with all rules of one of the following three forms:

(1) $\xi_i \to x\xi_j\xi_k$,

(2) $\xi_i \to x\xi_j$,

(3) $\xi_i \to x$.

We number the rules from 1 to n in such a way that the rules $r_1, \ldots, r_{n_1}$ are of the first form, the rules $r_{n_1+1}, \ldots, r_{n_2}$ are of the second form, the rules $r_{n_2+1}, \ldots, r_n$ are of the third form.

We consider the alphabet $Z = \{z_1, \ldots, z_n, \bar{z}_1, \ldots, \bar{z}_n\}$ and the grammar $\bar{G}$ thus built: to the rule r_1, $1 \leqslant l \leqslant n$, we associate the rule

$$\bar{r}_l = \xi_i \to z_l\xi_j\bar{z}_l\xi_K ;$$

to the rule r_1, $n_1 + 1 \leqslant l \leqslant n_2$, we associate the rule

$$\bar{r}_l\colon \xi_i \to z_l\bar{z}_l\xi_j ;$$

to the rule r_l, $n_2 + 2 \leqslant l \leqslant n$, we associate the rule

$$\bar{r}i\colon \xi_i \to z_l\bar{z}_l$$

The grammar $\bar{G}$ is the collection of the rules $\bar{r}_1, \ldots, \bar{r}_n$.

Define the morphism $\varphi: Z^* \to X^*$ by

$$\varphi(z_1) = \text{the first letter of the right hand member of the rule } r_1,$$
$$\varphi(\bar{z}_1) = \epsilon.$$

We certainly have $\varphi(L(\bar{G}, \xi_i)) = L(G, \xi_i)$.
Define then the collection of rational languages K_i for $i \in [N]$.

$$K_i = Z_i Z^* \backslash Z^* V Z^*,$$

where V is a set of forbidden transitions, the same for all K_i, and Z_i is a set of initial letters depending on i.

We prefer to describe $Z^r \backslash V$ as the set of all words of length r in Z

$z_l \bar{z}_l$, $\quad n_1 < l \leqslant n,$

$z_l \bar{z}_m$, $\quad 1 \leqslant l \leqslant n_1$ and the left hand member of r_m is ξ_j,

$\bar{z}_l \bar{z}_m$, $\quad n_2 < l \leqslant n$ and $1 \leqslant m \leqslant n_1$ and the left hand member of r_l is ξ_j,

$\bar{z}_l z_m$, $\quad 1 \leqslant l \leqslant n_1$ and the left hand member of r_m is ξ_k,

$\bar{z}_l z_m$ $\quad n_1 < l \leqslant n_2$ and the left hand member of r_m is ξ_j.

The sets Z_i are given by

$$Z_i = \{1 \leqslant l \leqslant n \mid \text{the left hand member of } r_l \text{ is } \xi_i\}.$$

Note that in any word in K_i there can be at most two consecutive barred letters.

Call D_n the Dyck set on the n letters $z_1, ..., z_n$. The proof that for all $i \in [N]$

$$L(\bar{G}, \xi_i) = D_n \cap K_i$$

is extremely similar to the original proof of the Chomsky–Schützenberger theorem and we leave it to the reader. The main difference with the original proof is that $K_i = FG(K_i)$ and φ is faithful on K_i. These are the two properties we now use.

The mapping $I_{K_1} \circ \varphi$ is a faithful sequential mapping of D_n onto $L(G, \xi_1)$ since $L(G, \xi_1) = \varphi(L(\bar{G}, \xi_1)) = \varphi(D_n \cap K_1)$. According to property 10 if we extend $I_{K_1} \circ \varphi$ to infinite words we get $(I_{K_1} \circ \varphi)\ \mathrm{Adh}(D_n) = \mathrm{Adh}(L(G, \xi_1)) = A$ since we have taken $A = \mathrm{Adh}(L), L = L(G, \xi_1)$.

But we have exhibited above a faithful sequential mapping γ of D_2 onto D_n so that

$$\mathrm{Adh}(D_n) = \gamma(\mathrm{Adh}(D_2)) \text{ by the same property 10.}$$

We obtain the result by composing γ and $I_{K_1} \circ \varphi$. ∎

Remark. We have proved in fact the stronger result.

For every closed context-free ∞-language L, there exists a faithful sequential mapping γ such that

$$\gamma(cl(D_2)) = L.$$

VI. Conclusion

Towards a theory of families of adherences:

A way of approach to the study of algebraic languages is their classification according to the partial ordering.

L dominates L' (written $L \rightsquigarrow L'$) iff there exists a rational transduction τ such that $\tau(L) = L'$.

Intuitively (and this is supported by a number of theorems):

L dominates L' means that L has a richer internal structure.

We can extend that to adherences by defining

A dominates A' (we write it the same way) iff there exists a faithful sequential mapping γ such that $\gamma(A) = A'$.

And we may ask the question of whether we retrieve or not the same classification for adherences as for languages.

We believe so and will try to prove it in forthcoming papers.

An interesting conjecture in this respect is the following:

Conjecture. Let $\mathscr{C}(L)$ be the rational cone generated by L, i.e., the family of all languages L' such that $L \rightsquigarrow L'$.

Then Adh(L) dominates all the adherences Adh(L'), $L' \in \mathscr{C}(L)$ iff the language $FG(L)$ faithfully generates $\mathscr{C}(L)$ (this meaning that for all $L' \in \mathscr{C}(L)$ there exists a faithful rational transduction τ such that $\tau(FG(L)) = L'$. And in the light of this conjecture some questions may be raised whose answer does not seem obvious.

Question. Does there exist a central language faithfully generating Oct, the family of one counter languages?

References

1. A. Arnold and M. Nivat, Non deterministic recursive programs, *in* "Fundamentals of Computation Theory," Lecture Notes in Computer Science No. 56, pp. 12–21, Springer-Verlag, Heidelberg, 1977.
2. N. Chomsky and M. P. Schützenberger, The algebraic theory of context-free languages, in Computer Programming and Formal Languages, 1963.
3. R. Cohen and A. Gold, Theory of ω-languages, I, II, *J. Comput. System Sci.* **15** (1977), 169–208.
4. B. Courcelle and M. Nivat, Algebraic families of interpretations, *in* "17th Symposium on Foundations of Computer Science," Houston, 1976.
5. J. Dugundji, "Topology," Allyn & Bacon, Boston, 1976.
6. S. Eilenberg, "Automata, Languages and machines," Vol. A, Academic Press, New York, 1974.
7. M. Linna, On ω-words and ω-computations, *Ann. Univ. Turkuensis*, 1975.

8. M. Nivat, Mots infinis engendrés par une grammaire algébrique, *R.A.I.R.O. Inform. Théor.* **11** (1977), 311–327.
9. M. Nivat, Sur les ensembles de mots infinis engendrés par une grammaire algébrique, *R.A.I.R.O. Inform. Théor.* **12** (1978), 259–278.
10. M. P. Schützenberger, Push-down automata and context-free languages, *Inform. Contr.* **6** (1963), 246–264.
11. D. Scott, Outline of a mathematical theory of computation, Memo, Oxford Programming Research Group, 1972.
12. J. Vuillemin, "Syntaxe, sémantique et axiomatique d'un langage de programmation," thèse de Doctorat, Université Paris VII, 1974.