RAIRO. INFORMATIQUE THÉORIQUE

H. A. MAURER
G. ROZENBERG
A. SALOMAA
D. WOOD

Pure interpretations of EOL forms

RAIRO. Informatique théorique, tome 13, nº 4 (1979), p. 347-362
<http://www.numdam.org/item?id=ITA_1979__13_4_347_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 13, n° 4, 1979, p. 347 à 362)

PURE INTERPRETATIONS OF EOL FORMS (*)

by H. A. MAURER (1), G. ROZENBERG (2),
A. SALOMAA (3) and D. WOOD (4)

Communicated by M. NIVAT

Abstract. — *The notion of an EOL form combined with pure interpretations is studied. This approach is compared and contrasted with the original approach of an EOL form with its interpretations. In particular it is shown there are no pure very complete EOL forms.*

Résumé. — *On étudie la notion de forme EOL avec interprétation pure. Cette approche se distingue de l'approche originale d'une forme EOL avec ses interprétations. On montre en particulier qu'il n'existe pas de formes EOL pures très complètes.*

1. INTRODUCTION

In [6] the notion of an EOL form was introduced, with its attendant notions of the interpretation of a form, the family of EOL systems derived from a form and the family of EOL languages generated by a form. The approach taken in [6] should be compared with the pioneering paper of Cremers and Ginsburg [2] for (context-free) grammar forms. Apart from the underlying rewriting system being an EOL system rather than a (context-free) grammar, the major distinction is the restriction of interpretations of EOL forms to be strict interpretations in the grammar form framework, *see* [1] and [4]. In a recent paper [5], strict interpretations for non-context-free grammar forms have been investigated.

However, in the present paper we present the results of an investigation into the notion of a pure interpretation. Pure forms and interpretations were first defined in section 7 of [6], where some preliminary results were given. Our present emphasis is upon the pure interpretation of an EOL form rather than of a pure form. This distinction will be clarified in section 2 which includes a brief review of the relevant notions. In section 3 we turn to normal form or reduction results as well as pure completeness. In contrast to the results in [6] on EOL

Received September 1978, Revised, March 1979.

(*) Work carried out partially under the auspices of the Universities of Antwerpen, Karlsruhe and Turku and partially under a Natural Sciences and Engineering Research Council of Canada Grant No. A-7700.

(1) Institut für Informationsverarbeitung, Graz, Austria.
(2) Institute of Applied Mathematics and Computer Science, University of Leiden, Leiden, The Netherlands.
(3) Department of Mathematics, University of Turku, Turku, Finland.
(4) Unit for Computer Science, McMaster University, Hamilton, Canada.

R.A.I.R.O. Informatique théorique/Theoretical Informatics, 0399-0540/1979/347/$ 4.00

forms we show that forms cannot be shortened under pure interpretation, which in turn implies that there are no pure very complete EOL forms. Section 4 investigates which EOL forms give rise to the same language family under both interpretation methods. Section 5 considers the speed-up of a form, showing that the language family is preserved under speed-up when an EOL form fulfills a simple condition. Again the contrary result is obtained under the usual interpretations. Finally, section 6 briefly considers goodness and strong goodness.

2. EOL FORMS AND PURE INTERPRETATIONS

We first review some basic definitions including the definition of an EOL form and its interpretations, before introducing pure interpretations.

An *EOL scheme* T is a triple $T=(V, \Sigma, P)$ where V is a finite set of symbols, $\Sigma \subseteq V$ is called the set of terminals, $V-\Sigma$ the set of nonterminals and P is a finite set of pairs (A, α) with A in V and α in V^* such that for each A in V at least one such pair is in P. The elements $p=(A, \alpha)$ of P are called productions and are usually written as $A \to \alpha$. T is a propagating EOL scheme, abbreviated as an EPOL scheme if in each production $A \to \alpha$ the right hand side differs from ε.

Let $T=(V, \Sigma, P)$ be an EOL scheme. For words $x=A_1 A_2 \ldots A_n$ with A_i in V and $y=y_1 y_2 \ldots y_n$ with y_i in V^* we write $x \Rightarrow_T y$ if $A_i \to y_i$ is a production of P for every i. We write $x \Rightarrow_T^0 x$ for every x in V^* and write $x \Rightarrow_T^n y$ if for some z in V^* $x \Rightarrow_T z \Rightarrow_T^{n-1} y$ holds. By $x \Rightarrow_T^* y$ we mean $x \Rightarrow_T^n y$ for some $n \geqq 0$, and by $x \Rightarrow_T^+ y$ we mean $x \Rightarrow_T^n y$ for some $n \geqq 1$.

For convenience, the EOL scheme will often not be indicated below the arrow $\Rightarrow$ if it is understood by the context.

A sequence of words $x_0, x_1, x_2, \ldots, x_n$ with

$$x_0 \Rightarrow x_1 \Rightarrow x_2 \Rightarrow \ldots \Rightarrow x_{n-1} \Rightarrow x_n$$

is called a derivation (of length n leading from x_0 to x_n).

EOL and EPOL schemes (V, Σ, P) where $V=\Sigma$ are called OL and POL schemes, respectively, and are written as pairs (Σ, P).

An *EOL system* G is a quadruple $G=(V, \Sigma, P, S)$ where (V, Σ, P, S) where (V, Σ, P) is an EOL scheme and S in $V-\Sigma$ is called the start symbol. The notions introduced for EOL schemes are carried over to EOL systems in the obvious manner. The *language generated by* G is denoted by $L(G)$ and defined as $L(G)=\{x \text{ is in } \Sigma^* : S \Rightarrow^* x\}$.

In the same way as adding a start symbol to an EOL scheme yields an EOL system, adding an arbitrary word w, called the axiom, to an OL scheme

$T=(\Sigma, P)$ gives an OL system $G=(\Sigma, P, w)$, where $L(G)$ is defined by $L(G)=\{x : w \Rightarrow^* x\}$.

For convenience, languages which differ by at most the empty word will be considered equal (modulo ε). Classes of languages will be considered equal if for any nonempty language in one class there is an equal (modulo ε) language in the other class, and conversely. The class of EOL languages is denoted by $\mathcal{L}$(EOL), i.e. $\mathcal{L}(\mathrm{EOL})=\{L(G) : G$ is an EOL system$\}$. Similarly, $\mathcal{L}$(FIN), $\mathcal{L}$(REG), $\mathcal{L}$(CF), $\mathcal{L}$(CS) and $\mathcal{L}$(RE) will denote the classes of finite, regular, context-free, context-sensitive and recursively enumerable languages, respectively.

For a word x, *alph*(x) is the set of all symbols occurring in x. For a language L, $LS(L)=\{|x| : x \text{ is in } L\}$ is *the length set of* L. For a set M of symbols and a set N of words $M \to N$ denotes the set of productions $\{A \to \alpha : A \text{ in } M, \alpha \text{ in } N\}$.

Let $G=(V, \Sigma, P, S)$ be an EOL system. A symbol A in V is called *reachable* (from S) if $S \Rightarrow^* x A y$ holds for some x, y.

G is called *reduced* if each A in V is reachable. G is called *looping* if $A \Rightarrow^+ A$ holds for some reachable A in V. G is called *expansive* if $A \Rightarrow^* x A y A z$ holds for some reachable A in V and some x, y, z in V^*. We say $x_0 \Rightarrow^l x_l$ is *nonterminal* [*total nonterminal*] and write $x_0 \Rightarrow^l_{nt} x_l$ [$x_0 \Rightarrow^l_{tnt} x_l$], if for any y_0, z_0 such that $S \Rightarrow^* y_0 x_0 z_0$ and for some [any] sequence of words

$$x_1, x_2, \ldots, x_{l-1} \quad \text{with} \quad x_i \Rightarrow x_{i+1} \quad \text{for } i=0, \ldots, l-1,$$
$$S \Rightarrow^* y_0 x_0 z_0 \Rightarrow y_1 x_1 z_1 \Rightarrow \ldots \Rightarrow y_{l-1} x_{l-1} z_{l-1} \Rightarrow y_l x_l z_l$$

implies $y_i x_i z_i$ contains at least one nonterminal for each i with $1 \leqq i \leqq l-1$.

We now introduce the notions of EOL forms and their interpretations.

DEFINITION: An *EOL form* F is an EOL system, $F=(V, \Sigma, P, S)$. An EOL system $F'=(V', \Sigma', P', S')$ is called an *interpretation of* F (*modulo* μ), symbolically $F' \lhd F(\mu)$, if μ is a substitution defined on V and (i)-(v) hold:

(i) $\mu(A) \subseteqq V'-\Sigma'$ for each A in $V-\Sigma$;

(ii) $\mu(a) \subseteqq \Sigma'$ for each a in Σ';

(iii) $\mu(A) \cap \mu(B)=\emptyset$ for all $A \neq B$ in V;

(iv) $P' \subseteqq \mu(P)$ where $\mu(P)=\bigcup_{A \to \alpha \text{ in } P} \mu(A) \to \mu(\alpha)$;

(iv) S' is in $\mu(S)$.

$\mathcal{G}(F)=\{F' : F' \lhd F\}$ is the *family of EOL forms generated by* F, and $\mathcal{L}(F)=\{\mathrm{L}(F') : F' \lhd F\}$ is called the *family of languages generated by* F.

DEFINITION: Two EOL forms F_1 and F_2 are called *equivalent* if $L(F_1)=L(F_2)$. They are called *form equivalent* if $\mathcal{L}(F_1)=\mathcal{L}(F_2)$.

We now introduce the central concept of this paper, namely, pure interpretations.

DEFINITION: An EOL system $F'=(V', \Sigma', P', S')$ is called a *pure interpretation of an EOL form* $F=(V, \Sigma, P, S)$ (*modulo* μ), $F' \underset{p}{\lhd} F(\mu)$ if μ is a substitution defined on V and (i)-(iv) hold:

(i) $\mu(V) \nsubseteq V'$;

(ii) $\mu(A) \cap \mu(B) = \emptyset$ for all $A \neq B$ in V;

(iii) $P' \subseteq \mu(P)$;

(iv) S' is in $\mu(S)$ and S' is in $V' - \Sigma'$.

As for usual interpretations we introduce $\mathscr{G}_p(F)$ and $\mathscr{L}_p(F)$, the families of EOL forms derived from F under pure interpretations and languages generated by F under pure interpretations, respectively. We say two forms F_1 and F_2 are *pure form equivalent* if $\mathscr{L}_p(F_1) = \mathscr{L}_p(F_2)$.

It should be observed that pure interpretations are more general than the usual interpretation so that $\mathscr{G}(F) \subseteq \mathscr{G}_p(F)$ and $\mathscr{L}(F) \subseteq \mathscr{L}_p(F)$.

Secondly, observe that symbols in F can yield both nonterminal and terminal symbols in F', although the disjointness condition (ii) still holds. Because of this we have added the condition that S' must be nonterminal in F', that is, condition (iv). Since F' is to be an EOL system this, we feel is a reasonable restriction. If S' is allowed to be terminal some of the results of the following sections would be invalidated. Since no distinction is made between terminals and nonterminals in F, F is essentially an OL form with a single symbol axiom, that is, $F=(\Sigma, P, S)$. Contrast this with the notion of a *pure form* in [6, section 7]. A *pure form* F is a pair $F=(\Sigma, P)$ such that (Σ, Σ, P) is an EOL scheme in other words, the nonterminal alphabet is empty. An EOL system $F'=(V', \Sigma', P', S')$ is an interpretation of F modulo μ if (i) $\mu(\Sigma) \subseteq V'$; (ii) $\mu(a) \cap \mu(b) = \emptyset$, for all $a \neq b$; (iii) $p' \subseteq \mu(P)$, and (iv) S' is in $\mu(\Sigma)$. The only distinction is that S' is the interpretation of a terminal in the pure form case whereas S' is an interpretation of a nonterminal in the pure interpretation case.

We make precise the relationship between pure forms and pure interpretations in the following theorem, which should be compared with theorem 7.2 in [6].

THEOREM 2.1: *For every pure form* $F=(\Sigma, P)$,

$$\mathscr{L}(F) = \bigcup_{S \text{ in } \Sigma} \mathscr{L}_p(F_S),$$

where $F_S = (\Sigma, \emptyset, P, S)$, *for all* S *in* Σ.

Proof: Observe that $F_S \lhd F$, for all S in Σ, therefore

$$\bigcup_{S \text{ in } \Sigma} \mathscr{L}_p(F_S) \subseteq \mathscr{L}(F).$$

Conversely, if $F' \lhd F(\mu)$, where $F' = (V', \Sigma', P', S')$, then clearly $F' \underset{p}{\lhd} F_S$, where S' is in $\mu(S)$. Hence we have the reverse inclusion. $\square$

We feel that pure interpretations of an EOL form are preferable to pure interpretations of EOL schemes (or interpretations of pure forms) since we avoid carrying around the union of language families and, more importantly, the sentence symbol S' of a pure interpretation $F' \underset{p}{\lhd} F(\mu)$ is obtained from the sentence symbol S of F. This is in agreement with our previous work on both EOL and ETOL forms [6, 7].

We close this section by summarizing those results of interest which carry over from the EOL form theory [6].

THEOREM 2.2: (1) *the relation* $\underset{p}{\lhd}$ *is decidable and transitive;*

(2) *Let* F *and* F' *be EOL forms*

$$\mathscr{G}_p(F') \subseteq \mathscr{G}_p(F) \quad \text{iff} \quad F' \underset{p}{\lhd} F;$$

(3) *it is decidable for arbitrary* EOL *forms* F *and* F' *whether* $\mathscr{G}_p(F) = \mathscr{G}_p(F')$;

(4) *let* $F = (V, \Sigma, P, S)$, $\overline{F} = (\overline{V}, \overline{\Sigma}, \overline{P}, \overline{S})$ *be* EOL *forms and let* $l \geqq 1$ *be an integer. If* $X \to \alpha$ *in* P *implies* $X \overset{l}{\underset{\overline{F}}{\Rightarrow}} \alpha$ *then* $\mathscr{L}_p(F) \subseteq \mathscr{L}_p(\overline{F})$;

(5) *let* F *and* F' *be EOL forms, where* $F' \underset{p}{\lhd} F$. *If* F' *is looping (expansive) then* F *is looping (expansive).*

3. REDUCTION AND *P*-COMPLETENESS

In [6] a number of reduction results are proved, which are analogous to the usual reduction results for EOL systems. The chief exception being that an EOL form does not necessarily have a form equivalent synchronized EOL form. It is clear that each EOL form under pure interpretations trivially has such a p-form equivalent synchronized EOL form since we can consider all symbols to be nonterminal. However, in contradistinction to the usual EOL form interpretation, we show that reduction to short normal form does not in general preserve the language family under pure interpretations.

NOTATION: We use the prefix p in the following to denote the pure interpretation variant of the usual interpretation terminology, for example, p-interpretation, p-form equivalent, p-complete, etc.

DEFINITION: Consider an EOL form $F=(V,\Sigma,P,S)$. F is *separated* if $A\to\alpha$ in P implies (i) α is in $\Sigma\cup(V-\Sigma)^*$ and (ii) A is in Σ implies α is not in Σ. F is *synchronized* if for each a in Σ, $a\Rightarrow^+\alpha$ implies α is not in Σ^*. F is *short* if $A\to\alpha$ is in P implies $|\alpha|\leqq 2$.

It is clear that the following result holds:

LEMMA 3.1: *For every* EOL *form F a p-form equivalent reduced* EOL *form F can be constructed.*

Since we can consider all members of V in an EOL form $F=(V,\Sigma,P,S)$ to be nonterminal, we can trivially obtain a p-form equivalent separated form from F.

Although an EOL form can be trivially synchronized under p-interpretations, we also have a much stronger synchronization result which tells us that a p-family is similar to the family of EOL systems in this respect.

LEMMA 3.2: *For every propagating* EOL *form $F=(V,\Sigma,P,S)$ and for every L in $\mathscr{L}_p(F)$ there exists an $F'\underset{p}{\lhd} F$ such that F' is synchronized and $L=L(F')$.*

Proof: We need the following synchronization transformation.

DEFINITION: Let $F=(V,\Sigma,P,S)$ be an EOL form. Construct $\mathrm{SYNCH}(F)=\bar{F}=(\bar{V},\Sigma,\bar{P},\bar{S})$ as follows:

Let $\bar{V}=V\cup V_1\cup V_2$, where

$$V_1=\{N_A : A \text{ in } V\},\qquad V_2=\{\bar{A} : A \text{ in } V\},$$

and letting

$$\mu(A)=\{A,N_A,\bar{A}\}\quad\text{for all } A \text{ in } V,$$

$$\bar{P}=\{A\to\beta : A\to\alpha \text{ in } P \text{ and } \beta \text{ is in } \mu(\alpha)\cap V_1^*\}$$

$$\cup\{\bar{A}\to\beta : A\to\alpha \text{ in } P \text{ and } \beta \text{ is in } \mu(\alpha)\cap(V^*\cup V_2^*)\}$$

$$\cup\{N_A\to\beta : A\to\alpha \text{ in } P \text{ and } \beta \text{ is in } \mu(\alpha)\cap V_1^*\}.$$

In the above definition it is clear that $\bar{F}\underset{p}{\lhd} F(\mu)$. Essentially it is the standard synchronization technique, where the symbols N_A play a blocking role.

Notice however, that we also have $F \underset{p}{\lhd} \bar{F}(\bar{\mu})$, that is $F \underset{p}{\lhd} \mathrm{SYNCH}(F)(\bar{\mu})$, where $\bar{\mu}$ is defined by:

for all A in V,

$$\bar{\mu}(A) = \emptyset = \bar{\mu}(N_A),$$

and for all $\bar{A}$ in V_2,

$$\bar{\mu}(\bar{A}) = A.$$

Immediately $F \underset{p}{\lhd} \bar{F}(\bar{\mu})$.

Returning to the lemma in hand, for each L in $\mathscr{L}_p(F)$ there exists $F' \underset{p}{\lhd} F$ with $L = L(F')$. Applying SYNCH to F' we obtain the result since p-interpretation is transitive. □

We need the propagating condition since if $a \to \varepsilon$ is in F then it is also present in SYNCH(F). In fact the lemma is not true, in general, if F is not propagating, consider $F : S \to a;\ a \to \varepsilon$; then F is not synchronized, and for $L = L(F) = \{ a \}$ in $\mathscr{L}_p(F)$, there is no synchronized $F' \underset{p}{\lhd} F$ with $L(F') = L$.

If we can synchronize in the above sense then can we also carry out the propagating transformation ? Or more generally, for all forms F does there exist a propagating form $\bar{F}$ with $\mathscr{L}_p(\bar{F} = \mathscr{L}_p(F)$? This difficult question has opened up a new area of investigation, namely, the family of length sets of a form under pure interpretation, which we hope to return to in the near future. Note that synchronized EOL forms and systems can be reduced to propagating forms or systems which are form equivalent and equivalent, respectively. On the other hand, OL systems cannot be so reduced.

We have the following:

THEOREM 3.3: *Let F be defined by $S \to c_1 ab_1 c_2 ab_2 c_3 ab_3$; $a \to \varepsilon$; $b_i \to c_i ab_i$; $c_i \to c_i$; $1 \leqq i \leqq 3$, then there is no propagating $\bar{F}$ with $\mathscr{L}_p(\bar{F} = \mathscr{L}_p(F)$.*

Proof: Consider an $\bar{F}' \underset{p}{\lhd} \bar{F}$ with $L(\bar{F}) = L(F)$, $\bar{F}'$ is an EPOL system. Now consider $\bar{F}'$ as the POL system $\bar{F}''$ [i.e., Letting $\bar{F}' = (\bar{V}', \bar{\Sigma}', \bar{P}', \bar{S}')$ then $\bar{F}'' = (\bar{V}', \bar{V}', \bar{P}', \bar{S}')$]. Thus

$$L(\bar{F}'') \supseteq L(F) \qquad \text{and} \qquad L(F) = \{ c_1^n ab_1 c_2^n ab_2 c_3^n ab_3 : n \geqq 1 \}.$$

Following Ehrenfeucht and Rozenberg [3], if a POL system $\bar{F}''$ generates a language containing $L(F)$ then $L(\bar{F}'')$ also contains a word $x = x_1 z_1 x_2 z_2 x_3 z_3$

with $|x_1|=|x_2|=|x_3|$ and $|z_1|=|z_2|=|z_3|\geqq 2$ such that $\mathrm{alph}(x_i)\cap \mathrm{alph}(x_j)\neq \emptyset$ for some i, j in $\{1, 2, 3\}$ and $i\neq j$.

Evidently no such word can occur in any language $L'=L(F')$ where $F' \underset{p}{\lhd} F$.

Hence $L(\overline{F}'')$ is not in $\mathscr{L}_p(F)$. Since $\overline{F}'' \underset{p}{\lhd} \overline{F}$ we have $\mathscr{L}_p(\overline{F})\neq \mathscr{L}_p(F)$, as desired. □

We now demonstrate that we cannot shorten EOL forms under pure interpretation. For an EOL form F let *min* $r(F)$ and *max* $r(F)$ denote the length of the shortest and longest, respectively, right hand side of the productions in F. The following technical result will prove useful.

LEMMA 3.4: *For every reduced* EOL *form* $F=(V, \Sigma, P, S)$ *there exists a language* L *in* $\mathscr{L}_p(F)$ *with* $L=\{x\}$ *and* $\min r(F)\leqq |x|\leqq \max r(F)$.

Proof: Let $S\to\alpha$ be in P. Consider the isolating p-interpretation $F'=(V\cup\Sigma'\cup\{X\}, \Sigma', P', X)\underset{p}{\lhd} F(\mu)$, where the only derivation is:

$$X \Rightarrow \alpha' \Rightarrow^{+} \text{blocking},$$

where α' is in Σ'^*. [We can always isolate in this manner. Consider $\Sigma=\emptyset$ without loss of generality and define $\mu(A)=\{A', A\}$ for all A in $V-\{S\}$, and $\mu(S)=\{X, S, S'\}$. Choose $X\to\alpha'$ in $\mu(S\to\alpha)$ such that α' consists of primed symbols, let $A'\to\beta$ for all A in V, β in V^* such that $A\to\beta$ is in P and finally include all the rules in P. Letting $\Sigma'=\{A' : A \text{ in } V\}$, then clearly $X\Rightarrow x'\Rightarrow^{+}$ blocking.]

Immediately, $L(F')=\{\alpha'\}$ and $\min r(F)\leqq |\alpha'|\leqq \max r(F)$. □

We also have the weaker result that every p-family contains a singleton language.

THEOREM 3.5: *There exist* EOL *forms which cannot be shortened under p-interpretation.*

Proof: Let $F=(\{S\}, \emptyset, \{S\to SSS\}, S)$ be an EOL form, then for every short EOL form $\overline{F}$, $\mathscr{L}_p(\overline{F})\neq\mathscr{L}_p(F)$. Consider any short form $\overline{F}$. Then $0\leqq \min r(\overline{F})\leqq \max r(\overline{F})\leqq 2$, hence there is a language L in $\mathscr{L}_p(\overline{F})$ with $L=\{x\}$ and $0\leqq |x|\leqq 2$. However since $\min r(F)=\max r(F)=3$, L is not in $\mathscr{L}_p(F)$. □

We can apply lemma 3.4 more generally.

DEFINITION: An EOL form F is *p-complete* if $\mathscr{L}_p(F)=\mathscr{L}(\mathrm{EOL})$. F is said to be *p-vomplete* (p-very complete) if for all EOL forms $\overline{F}$ there exists $F' \underset{p}{\lhd} F$ with $\mathscr{L}_p(F')=\mathscr{L}_p(\overline{F})$.

In [8] the notion of vompleteness is investigated and it shown that vomplete EOL forms exist. A result to the contrary is now proved for p-interpretations.

THEOREM 3.6: *There are no p-vomplete* EOL *forms.*

Proof: Assume that F is a p-vomplete EOL form. We argue by contradiction. Let $m = \max r(F)$. Consider $\overline{F} = (\{S\}, \emptyset, \{S \rightarrow S^{m+1}\}, S)$. Now $\min r(\overline{F} = \max r(\overline{F} = m+1$, hence for all L in $\mathscr{L}_p(\overline{F})$ the length of the smallest word in L is at least $m+1$. However, for all $F' \lhd_p F$, since $\min r(F') \geqq \min r(F)$ and $\max r(F') \leqq \max r(F) = m$, there exists a singleton language $\{x\}$ in $\mathscr{L}_p(F')$ with $0 \leqq |x| \leqq m$. Hence $\mathscr{L}_p(\overline{F}) \neq \mathscr{L}_p(F')$ for any $F' \underset{p}{\lhd} F$. Therefore F is not p-complete. □

For p-completeness we have more positive results, namely:

THEOREM 3.7:

$$F_1: \quad S \rightarrow S \mid SS \qquad \textit{and} \qquad F_2: \quad S \rightarrow \varepsilon \mid S \mid SS;$$

are p-complete.

Proof: F_1 is shown to be complete and hence p-complete in [6] and since $F_1 \underset{p}{\lhd} F_2$, $\mathscr{L}_p(F_1) \subseteqq \mathscr{L}_p(F_2)$, therefore F_2 is also p-complete. □

An EOL form is a *one-letter form* if its only symbol is S. We can characterize p-complete one-letter forms as follows:

THEOREM 3.8: *Let F be a one-letter form:*

(i) *if F is propagating then F is p-complete iff P contains the rules $S \rightarrow S$* and *$S \rightarrow SS$;*

(ii) *If F is not propagating then F is p-complete iff P contains the rules $S \rightarrow S$ and $S \rightarrow S^m$, for some $m \geqq 2$.*

Proof: Part (i) has been shown in [6], it remains to prove part (ii). Note that a one-letter form is always a pure form in the [6] sense.

If F contains $S \rightarrow \varepsilon$, $S \rightarrow S$ and $S \rightarrow S^m$, for some $m \geqq 2$, then there is a p-interpretation $F' = (\{S, S_1, \ldots, S_m\}, \emptyset, P', S)$ where P' contains $S \rightarrow S_1 \ldots S_m$, $S \rightarrow S_1$, $S_1 \rightarrow S$, $S_2 \rightarrow S$ and $S_i \rightarrow \varepsilon$, $3 \leqq i \leqq m$. Then $S \rightarrow_{F'}^2 S$, $S \Rightarrow_{F'}^2 SS$, therefore F_1 of theorem 3.7 is simulated by F', hence $\mathscr{L}_p(F_1) \subseteqq \mathscr{L}_p(F') \subseteqq \mathscr{L}_p(F)$, by theorem 2.2, hence F is p-complete.

If F is p-complete, F must be looping, that is, $S \Rightarrow^+ S$ must hold since F is a one-letter form we must have $S \rightarrow S$ in P. If F has no rule $S \rightarrow S^m$ with $m \geqq 2$, then no word of length $\geqq 2$ can be generated by any interpretation of F. Therefore $S \rightarrow S^m$ is in P for some $m \geqq 2$. □

In closing this section we need the following:

DEFINITION: Let F be an EOL form. We say F is *regular-p-complete* if $\mathscr{L}_p(F)=\mathscr{L}(\mathrm{REG})$.

We now show that we can obtain the regular languages with pure interpretations as well as with the usual interpretations [6]. We first prove a more general result.

THEOREM 3.9: *Let $F=(V, \Sigma, P, S)$ be an EOL form for which the following three conditions hold:*

(i) $V=V_1 \dot{\cup} V_2$ *and S is in V_1 (to avoid triviality);*

(ii) *For all A in V_1, $A \Rightarrow^* X_1 \ldots X_m Y$ implies X_i is in V_2, $1 \leqq i \leqq m$, $m \geqq 0$ and Y is in V, and*

(ii) *for all A in V_2, $A \Rightarrow^* \alpha$ implies α is in V_2.*

Then $\mathscr{L}_p(F) \subseteq \mathscr{L}(\mathrm{REG})$.

Proof: Note that each $F' \underset{p}{\lhd} F$ will fulfill the same three conditions, hence it suffices to prove that $L(F)$ is in $\mathscr{L}(\mathrm{REG})$.

For all A in V_2 and all $i>0$, define

$$M_i(A)=\{B : A \Rightarrow^i B\} \cap \Sigma.$$

Clearly there exist integers t and p such that $M_t(A)=M_{t+\lambda p}(A)$, for all A in V_2, and all $\lambda \geqq 0$, where $t>0$ and $p>0$. Define a new alphabet $\bar{V}=\{A^{(i)} : A \text{ is in } V \text{ and } 1 \leqq i \leqq t+p\} \cup V$ and a right linear grammar $G=(\bar{V}, \Sigma, \bar{P}, S)$ where $\bar{P}$ contains:

(i) for all i, $0<i \leqq p$, $S \rightarrow S^{(i+t)}$;

(ii) for all i, $1<i \leqq t+p$,

$$A^{(i)} \rightarrow M_{i-1}(X_1) \ldots M_{i-1}(X_m) Y^{(i-1)} \qquad \text{if } A \rightarrow X_1 \ldots X_m Y \text{ is in } P$$

(iii) $A^{(1)} \rightarrow X_1 \ldots X_m Y$ and

$$A^{(1)} \rightarrow M_{p-1}(X_1) \ldots M_{p-1}(X_m) Y^{(p)} \qquad \text{if } A \rightarrow X_1 \ldots X_m Y \text{ is in } P.$$

It should be clear that $L(G) \subseteq L(F)$, and further that $L(F)-L(G)$ is finite; the "initial mess".

Hence $L(F)$ is regular and $\mathscr{L}_p(F) \subseteq \mathscr{L}(\mathrm{REG})$. □

COROLLARY 3.10: *Let F be $S \rightarrow aS \mid a$; $a \rightarrow a$, then F is regular p-complete.*

Proof: Since F fulfills the conditions of the theorem. □

4. *P*-STABILITY

When comparing pure interpretations with the usual interpretations an immediate question is raised, namely, when is $\mathscr{L}_p(F)=\mathscr{L}(F)$? We say an EOL form is *p-stable* when this holds. A straightforward observation yields:

LEMMA 4.1: *Let F be a complete* EOL *form. Then F is p-stable.*

Proof: Clearly $\mathscr{L}(H)\subseteq\mathscr{L}_p(F)$ and since F is complete $\mathscr{L}(F)=\mathscr{L}_p(F)$. □

The reverse is, of course, not necessarily so, since $F: S\to a;\ a\to a;$ is not complete but $\mathscr{L}_p(F)=\mathscr{L}(F)$, the family of all single letter languages. Similarly if F is p-complete F is not necessarily p-stable. Consider $F: S\to S\mid SS$; F is p-complete but $\mathscr{L}(F)=\{\emptyset\}$.

A related question is: can we always transform an EOL form into a p-form equivalent p-stable form? In this case the answer is positive. We first need:

DEFINITION: Let $F=(V,\ \Sigma,\ P,\ S)$ be an EOL form. Construct $\mathrm{SPLIT}(F)=\overline{F}=(\overline{V},\ \overline{\Sigma},\ \overline{P},\ S)\mathrel{\underset{p}{\lhd}} F(\mu)$ with $\overline{V}=V\cup\{\overline{A} : A \text{ in } V\}$ as follows:

Let $\mu(A)=\{A,\ \overline{A}\}$, for all A in V, $\overline{\Sigma}=\Sigma\cup\{\overline{A} : A \text{ in } V-\Sigma\}$, and $\overline{P}=\mu(P)$.

THEOREM 4.2: *Let* $F=(V,\ \Sigma,\ P,\ S)$ *be an* EOL *form. Then* $\mathscr{G}_p(F)=\mathscr{G}(\mathrm{SPLIT}(F))$.

Proof: Since the combination of a p-interpretation and an interpretation is a p-interpretation, $\mathscr{G}(\mathrm{SPLIT}(F))\subseteq\mathscr{G}_p(F)$. Consider $F'=(V',\ \Sigma',\ P',\ S')\mathrel{\underset{p}{\lhd}} F(\mu)$. Define an interpretation $\overline{\mu}$ such that $F'\lhd\mathrm{SPLIT}(F)(\overline{\mu})$.

First define two substitutions μ_t and μ_n on V by:

For all A in V:

(i) $\mu_t(A)\subseteq\Sigma'$ and $\mu_n(A)\subseteq V'-\Sigma'$ and

(ii) $\mu(A)=\mu_t(A)\cup\mu_n(A)$.

Secondly, define $\overline{\mu}$ on $\overline{F}=\mathrm{SPLIT}(F)$ by:

(i) for all A in $V-\Sigma$, $\overline{\mu}(A)=\mu_n(A)$ and $\overline{\mu}(\overline{A})=\mu_t(A)$;

(ii) for all A in Σ, $\overline{\mu}(A)=\mu_t(A)$ and $\overline{\mu}(\overline{A})=\mu_n(\overline{A})$.

Clearly $P'\subseteq\overline{\mu}(\overline{P})$, therefore $F'\lhd\overline{F}(\mu)$, which gives the required result. □

We now prove the main result:

THEOREM 4.3: *Let F be an* EOL *form.*

Then SPLIT(F) *is p-stable.*

Proof: We need to show that $\mathscr{L}_p(\mathrm{SPLIT}(F))=\mathscr{L}(\mathrm{SPLIT}(F))$. We prove a stronger result, namely, $\mathscr{G}_p(\mathrm{SPLIT}(F))=\mathscr{G}(\mathrm{SPLIT}(F))$. To obtain this strong equivalence we only need to show that $\mathcal{G}_p(\mathrm{SPLIT}(F)) \subseteqq \mathcal{G}(\mathrm{SPLIT}(F))$, since an interpretation is always a p-interpretation. Now since $\mathrm{SPLIT}(F)$ is a p-interpretation of F, $\mathcal{G}_p(\mathrm{SPLIT}(F)) \subseteqq \mathcal{G}_p(F) = \mathcal{G}(\mathrm{SPLIT}(F))$ by theorem 4.2. □

Letting $\mathcal{L}$ denote the class $\{\mathcal{L}(F) : F \text{ is an EOL form}\}$, $\mathcal{L}_p$ denote the class $\{\mathcal{L}_p(F) : F \text{ is an EOL form}\}$ and $\mathcal{L}_{\text{synch}}$ denote the class $\{\mathcal{L}(F) : F \text{ is a synchronized EOL form}\}$, theorem 4.2 also leads to:

THEOREM 4.4:

$$\mathcal{L}_p \subsetneqq \mathcal{L} \qquad \textit{and} \qquad \mathcal{L}_p \subseteqq \mathcal{L}_{\text{synch}}.$$

Proof: Since every EOL form F can be transformed into a p-stable EOL form $\mathrm{SPLIT}(F)$, such that $\mathcal{L}_p(F) = \mathcal{L}(\mathrm{SPLIT}(F))$ then $\mathcal{L}_p \subseteqq \mathcal{L}$. We obtain proper inclusion by considering the EOL form $F : S \to a;\ a \to b;\ b \to b$. For all $F' \lhd F$, $L(F')$ contains at least two words. However from lemma 3.4 we know that every p-family contains singleton languages, therefore there is no EOL form $\overline{F}$ with $\mathcal{L}_p(\overline{F}) = \mathcal{L}(F)$.

Using the SYNCH transformation detailed in lemma 3.2 we can synchronize a p-stable form F to give $\overline{F} = \mathrm{SYNCH}(F)$. Now since SYNCH is a p-interpretation we have $\mathcal{G}(\mathrm{SYNCH}(F)) \subseteqq \mathcal{G}_p(F)$, hence $\mathcal{L}(\mathrm{SYNCH}(F)) \subseteqq \mathcal{L}_p(F) = \mathcal{L}(F)$. We observed previously in lemma 3.2 that SYNCH has the property, $F \underset{p}{\lhd} \mathrm{SYNCH}(F)$, that is, $\mathcal{L}_p(F) \subseteqq \mathcal{L}(\mathrm{SYNCH}(F))$.

Hence we have shown that for every EOL form F there exists an EOL form $\overline{F}$ such that $\mathcal{L}_p(F) = \mathcal{L}(\overline{F})$ and $\overline{F}$ is synchronized. Therefore $\mathcal{L}_p \subseteqq \mathcal{L}_{\text{synch}}$. □

We leave as an open problem whether $\mathcal{L}_p \subsetneqq \mathcal{L}_{\text{synch}}$. Notice that we have implicitly introduced a stronger version of p-stability. We say an EOL form F is *strong p-stable* if $\mathcal{G}_p(F) = \mathcal{G}(F)$. Theorem 4.3 states that for every EOL form F there is a strong p-stable EOL form $\overline{F}$ which is p-form equivalent to F. Clearly if F is strong p-stable it is p-stable, is the converse true? Similarly we have shown that SPLIT (F) is always (strong) p-stable. This raises the question whether given a (strong) p-stable EOL form F, there always exists an F such that $F = \mathrm{SPLIT}(F)$?

5. SPEED-UP

We show that the language family of an EOL form under p-interpretations is preserved under speed-up in certain cases. This is in contradistinction to the usual mode of interpretation.

DEFINITION: Let $F=(V, \Sigma, P, S)$ be an EOL form and $n>1$ an integer. Let $\overline{F}=(V, \Sigma, \overline{P}, S)=n\text{-}SPEED(F)$, where

$$\overline{P}=\{S \rightarrow \alpha : S \Rightarrow_F^k \alpha \text{ and } 1 \leqq k \leqq n\}$$
$$\{A \rightarrow \alpha : A \Rightarrow_F^n \alpha,\ A \text{ in } V \text{ and } \alpha \text{ in } V^*\}.$$

We say $\overline{F}$ is a *speed-up of* F.

Consider $F : S \rightarrow a;\ a \rightarrow b;\ b \rightarrow b$; then the 2-speed-up of F is $\overline{F} : S \rightarrow a \mid b;\ a \rightarrow b;\ b \rightarrow b$. Now $\mathscr{L}(\overline{F}) \neq \mathscr{L}(F)$ since each L in $\mathscr{L}(F)$ contains at least two words whereas letting F' be $S \rightarrow b;\ b \rightarrow b$; where $\overline{F}' \lhd \overline{F}$, we have $L(\overline{F}')=\{b\}$.

We say an EOL form $F=(V, \Sigma, P, S)$ is *distinguished* if $P \subseteq V \times (V-\{S\})^*$. We now have:

THEOREM 5.1: *Let $F=(V, \Sigma, P, S)$ be a distinguished* EOL *form. Then for all* $n>0$, $\mathscr{L}_p(F)=\mathscr{L}_p(n\text{-SPEED}(F))$.

Proof: First observe that $L(F)=L(\overline{F})$ where $\overline{F}=n$-SPEED(F) and $\overline{F}=(V, \Sigma, \overline{P}, S)$. Clearly any $\overline{F}$-derivation $S \Rightarrow^m \alpha$ can be expanded as an F-derivation $S \Rightarrow^q \alpha$, where $q=k+ln$, $0 \leqq k<n$ and $l \geqq 0$. Conversely given an F-derivation $S \Rightarrow^m \alpha$, if $m \leqq n$ then $S \rightarrow \alpha$ is in $\overline{P}$, that is, $S \Rightarrow \alpha$ in $\overline{F}$, otherwise $m=in+j$, where $i>0$, and $0 \leqq j<n$. Letting

$$S=\alpha_0 \Rightarrow \alpha_1 \Rightarrow \ldots \Rightarrow \alpha_m=\alpha \text{ in } F$$

we obtain

$$S \Rightarrow \alpha_j \Rightarrow \alpha_{j+n} \Rightarrow \ldots \Rightarrow \alpha_{j+in}=\alpha \text{ in } \overline{F}.$$

Now let $F' \underset{p}{\lhd} F$ and construct n-SPEED(F'). Clearly $L(F')=L(n$-SPEED$(F'))$ and n-SPEED$(F') \underset{p}{\lhd} \overline{F}$. Hence $\mathscr{L}_p(F) \subseteq \mathscr{L}_p(n\text{-SPEED}(F))$.

Conversely, consider $\overline{F}' \underset{p}{\lhd} \overline{F}\ (\mu)$. Construct $F' \underset{p}{\lhd} F$ such that for each rule $A' \rightarrow \alpha'$ in $\overline{F}'$ with A' not in $\overline{\mu}(S)$, there is a unique nt-derivation $A' \Rightarrow^n a'$ in F'. By unique we mean that each symbol appearing in an intermediate word is nonterminal and only appears in that position. Similarly for each rule $S' \rightarrow \alpha'$ in $\overline{F}'$ construct a unique nt-derivation $S' \Rightarrow^k \alpha'$ in F', where $\mu^{-1}(S') \Rightarrow^k \mu^{-1}(\alpha')$ in F. Essentially F' is a tnt-simulation of $\overline{F}'$ apart from the initial starting sequences from S'. Hence $L(F')=L(\overline{F}')$. Clearly $F' \underset{p}{\lhd} F$ so we have demonstrated the required result. □

The assumption that F is distinguished is crucial. Consider

$$F: \quad S \to SS \mid aa; \; a \to NN; \qquad N \to NN,$$

and the 2-speed-up of F,

$$\overline{F}: \quad S \to SS \mid SSS \mid aa \mid aaaa \mid NNN;$$
$$a \to NNNN; \qquad N \to NNNN.$$

Now $L(F) = \{ a^m : m = 2^n, \text{ where } n \geqq 1 \}$ and $S \Rightarrow SS \Rightarrow aaaaaa$ in $\overline{F}$, which is not in $L(F)$.

Similarly, $\mathscr{L}_p(F) \neq \mathscr{L}_p(\overline{F})$. Since each rule in F is a "doubling" rule then for all α such that $S \Rightarrow^+ \alpha$ in F, we have $|\alpha| = 2^n$ for some $n \geqq 1$. This is also true in any interpretation $F' \underset{p}{\lhd} F$. Hence $LS(L(F')) \subseteq \{ 2^n : n \geqq 1 \}$. We have already demonstrated that each $L(\overline{F})$ in $\mathscr{L}_p(\overline{F})$ contains a word of length 6, which is not a power of 2. Hence the result.

However, we can change a form such that it is distinguished using the obvious transformation. That is, if $F = (V, \Sigma, P, S)$, let $\overline{F} = (V \cup \{ \overline{S} \}, \Sigma, P \cup \{ \overline{S} \to \alpha : S \to \alpha \text{ in } P \}, \overline{S})$, then $\mathscr{L}_p(F) = \mathscr{L}_p(\overline{F})$. Hence we have:

THEOREM 5.2: *Let $F = (V, \Sigma, P, S)$ be an* EOL *form. There exists a distinguished* EOL *form $\overline{F}$ such that:*

(i) $\mathscr{L}_p(\overline{F}) = \mathscr{L}_p(F)$, *and*

(ii) *for all* $n > 1$, $\mathscr{L}_p(n\text{-SPEED}(\overline{F})) = \mathscr{L}_p(F)$.

6. P-GOODNESS AND STRONG P-GOODNESS

An EOL form F is said to be *p-good* if for every EOL form G with $\mathscr{L}_p(G) \subseteq \mathscr{L}_p(F)$ there exists $F' \underset{p}{\lhd} F$ such that $\mathscr{L}_p(F') = \mathscr{L}_p(G)$. Clearly F is p-vomplete if F is p-good and p-complete. We have already shown that no p-vomplete forms exist (th. 3.5), since a form cannot be shortened. We would expect from this result that p-goodness is rare. This is indeed the case. But first the positive results.

THEOREM 6.1: $F_1 : S \to S$, $F_2 : S \to S \mid \varepsilon$ *and* $F_3 : S \to a^m$; $a \to a$, $m \geqq 1$ *are p-good.*

Proof: F_1, F_2 and F_3 are p-good by arguments similar to those given in [8] for the goodness of $S \to a$; $a \to a$. The important observation is that each F_i generates a family of finite sets. □

When F is "infinite", however, F cannot be p-good.

THEOREM 6.2: *Let* $F=(V, \Sigma, P, S)$ *be a propagating* EOL *form such that* $\{\alpha : S \Rightarrow_F^* \alpha\} = SF(F)$ *is infinite. Then* F *is not p-good.*

Proof: Assume F is p-good. Let $m = \max r(F)$, and consider α in $SF(F)$ with $|\alpha| > m$ [such an α must exist if $SF(F)$ is infinite]. Let $\overline{F} = (\{\overline{S}\} \cup V, \Sigma, \overline{P}, \overline{S})$ where $\overline{P} = P \cup \{\overline{S} \to \alpha\}$. Clearly $\max r(\overline{F}) > m$ and $\mathscr{L}_p(\overline{F}) \subseteq \mathscr{L}_p(F)$. However, for all $F' \underset{p}{\lhd} F$ there are languages in $\mathscr{L}_p(F')$ which contain words whose lengths are at most m. Whereas the smallest length word in each language in $\mathscr{L}_p(\overline{F})$ has length greater than m. A contradiction. □

COROLLARY 6.3: $F : S \to SS$ *is not p-good.*

Notice that $F_1 : S \to S$ has the unusual property that for all propagating F with $\mathscr{L}_p(F) \subseteq \mathscr{L}_p(F_1)$, not only does there exist $F'_1 \underset{p}{\lhd} F_1$ with $\mathscr{L}_p(F'_1) = \mathscr{L}_p(F)$ but also $F \underset{p}{\lhd} F_1$. In this case we say F_1 is *strong p-good.*

By the technique displayed in the proof of theorem 6.2 we can construct from an arbitrary infinite $F=(V, \Sigma, P, S)$ a particular $\overline{F} = (V \cup \{\overline{S}\}, \Sigma, P \cup \{\overline{S} \to \alpha\}, \overline{S})$, where $S \Rightarrow_F^+ \alpha$ and $|\alpha| > \max r(F)$. Clearly $\mathscr{L}_p(\overline{F}) \subseteq \mathscr{L}_p(F)$ and $\overline{F}$ cannot be a p-i nterpretation of F. Similarly choosing α in Σ^* with $|\alpha| > \max r(F)$ we can obtain an analogous result for the usual interpretations [note that it is sufficient that there is an α (in Σ^*) with $S \Rightarrow_F^+ \alpha$ and $\alpha > \max r(F)$ to give the contradiction in either case].

We have just shown:

THEOREM 6.4: *Let* $F=(V, \Sigma, P, S)$ *be an* EOL *form for which* $S \Rightarrow_F^+ \alpha$ *with* $|\alpha| > \max r(F)$, *for some* α (*some* α *in* Σ^*). *Then* F *is not strong p-good* (*not strong good*).

In other words strong p-goodness is very rare.

REFERENCES

1. E. BERTSCH, *An Observation on Relative Parsing Time, J. of the A.C.M.* Vol. 22, 1975, pp. 493-498.
2. A. B. CREMERS and S. GINSBURG, *Context-Free Grammar Forms,* Vol. 11, 1975, pp. 86-117.
3. A. EHRENFEUCHT and G. ROZENBERG, *EPOL Systems Generating Counting Languages,* Manuscript, 1977.

4. S. GINSBURG and O. MAYER, *On Strict Interpretations of Grammar Forms*, Lecture Notes in Computer Science n° 45, New York, Springer-Verlag, 1976.
5. H. A. MAURER M. PENTTONEN, A. SALOMAA and D. WOOD, *Non-Context-Free Grammar Forms*, Math. Syst. Th., Vol. 12, 1979, pp. 297-324.
6. H. A. MAURER A. SALOMAA and D. WOOD, *EOL Forms*, Acta Informatica, Vol. 8, 1977, pp. 75-96.
7. H. A. MAURER, A. SALOMAA and D. WOOD, *ETOL Forms*, J. Comp. Syst. Sc., Vol. 16, 1978, pp. 345-361.
8. H. A. MAURER A. SALOMAA and D. WOOD, *On Good EOL Forms*, *S.I.A.M. J. Comp.*, Vol. 7, 1978, pp. 158-166.