

H. A. MAURER

A. SALOMAA

D. WOOD

Relative goodness of EOL forms

RAIRO. Informatique théorique, tome 12, n° 4 (1978), p. 291-304

http://www.numdam.org/item?id=ITA_1978__12_4_291_0

© AFCET, 1978, tous droits réservés.

L'accès aux archives de la revue « RAIRO. Informatique théorique » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

RELATIVE GOODNESS OF EOL FORMS (*) (1)

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

Communiqué par J. Berstel

Abstract. — The topic of “relative goodness” of “EOL forms” is introduced, which leads to the interesting notion of “mutual goodness” of two “EOL forms”. For example, it is proved that “right linear” and “left linear forms” are not “mutually good”. Finally, a number of open problems are presented.

1. INTRODUCTION

In [3] the notion of an EOL form was introduced. The investigation of complete EOL forms, that is, ones which have as their grammatical family all EOL languages, was begun. This has been followed up in [2] where almost all two-letter forms have been classified with respect to completeness. [3] also introduced *very* complete or *vomplete* EOL forms, that is, ones which can via interpretation generate every EOL grammatical family. This was followed up in [4] where the more general concept of good EOL forms was considered. (Goodness has also been considered by [6] for ETOL forms.) Although goodness was seen to play a central role in EOL form theory results have however been difficult to obtain. One reason for this may be observed in the present paper since it is shown that the consideration of goodness is closely related to the consideration of the classes of families of languages of interpretations of EOL forms, in other words, classes of grammatical families.

Our major theme is the study of relative goodness of forms. In [4] a study was made of goodness of EOL forms relative to the class of all EOL grammatical families. Here we study the goodness of EOL forms relative to the class of all grammatical families obtained from a given EOL form (via interpretation).

(*) Reçu le 28 novembre 1977, et dans la version définitive le 6 avril 1978.

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

(2) Institut für Informationsverarbeitung, Technische Universität Graz, A-8010 Graz, Austria.

(3) Mathematics Department, University of Turku, Finland.

(4) Applied Mathematics Department, McMaster University, Hamilton, Ontario, L 8 S 4 K 1, Canada.

2. BASIC TERMINOLOGY AND PRELIMINARY RESULTS

The notions of a "master grammar" and its structurally related grammars under interpretation was first introduced in [1]. The application of these notions to L systems appeared for the first time in [3]. It is to this latter paper that we refer the reader for further clarification of EOL forms and their basic properties, while we now briefly review the essential notation for the present paper.

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 $A \rightarrow \alpha$. T is called propagating and an EPOL scheme if in each production $A \rightarrow \alpha$ the right hand side differs from ε .

Let $T = (V, \Sigma, P)$ be an EOL scheme. For words $A = A_1 A_2 \dots A_n$ with A_i in V and $y = y_1 y_2 \dots y_n$ with y_i in V^* we write $x \Rightarrow_T y$ if $A_i \rightarrow y_i$ 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 $n \geq 0$, and by $x \Rightarrow_T^+ y$ we mean $x \Rightarrow_T^n y$ for some $n \geq 1$.

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 $x_0, x_1, x_2, \dots, x_n$ with $x_0 \Rightarrow x_1 \Rightarrow x_2 \Rightarrow \dots \Rightarrow x_{n-1} \Rightarrow x_n$ 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) .

If one uses n sets of productions P_1, P_2, \dots, P_n instead of a single set of productions P (in derivations, at each step, productions of one P_i may only be used) one talks of ETOL, EPTOL, TOL, etc. schemes instead of EOL, EPOL, EOL, etc. schemes.

An *EOL system* G is a quadruple $G = (V, \Sigma, P, S)$ 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. A derivation $x \Rightarrow^* y$ ($x \Rightarrow^+ y$ or $x \Rightarrow^n y$) is said to be a sentential derivation if $x = S$. 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 ε will be considered equal. Classes of languages will be considered equal if for any nonempty language in one class a language in the other class, and conversely, exists which differs by at most ε . The class of EOL languages is denoted by $\mathcal{L}(\text{EOL})$, i. e. $\mathcal{L}(\text{EOL}) = \{L(G) : G \text{ is an EOL system}\}$. Similarly, $\mathcal{L}(\text{FIN})$, $\mathcal{L}(\text{REG})$, $\mathcal{L}(\text{CF})$, $\mathcal{L}(\text{CS})$ and $\mathcal{L}(\text{RE})$ will denote the classes of finite, regular, context-free, context-sensitive and recursively enumerable languages, respectively.

For a word x , $\text{alph}(x)$ 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 $M \rightarrow N$ denotes the set of productions $\{A \rightarrow \alpha : A \text{ in } M, \alpha \text{ in } N\}$.

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

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 $A \Rightarrow^* \alpha A \beta A \gamma$ holds for some reachable A in V and some α, β, γ in V^* . We say $x_0 \Rightarrow_G^l x_l$ is *nonterminal* [*total nonterminal*] and write $x_0 \Rightarrow_{tmG}^l x_l$ [$x_0 \Rightarrow_{tmG}^l x_l$], if for some [any] sequence of words x_1, x_2, \dots, x_{l-1} with $x_i \Rightarrow_G x_{i+1}$ for $i = 0, \dots, l-1$,

$$S \Rightarrow_G^* y_0 x_0 z_0 \Rightarrow_G y_1 x_1 z_1 \Rightarrow_G \dots \Rightarrow_G y_{l-1} x_{l-1} z_{l-1} \Rightarrow_G y_l x_l z_l$$

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

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 $F' \triangleleft F(\mu)$, if μ is a substitution defined on V and (i)–(v) hold:

- (i) $\mu(A) \subseteq V' - \Sigma'$ for each A in $V - \Sigma$,
- (ii) $\mu(a) \subseteq \Sigma'$ for each a in Σ ,
- (iii) $\mu(A) \cap \mu(B) = \emptyset$ for all $A \neq B$ in V ,
- (iv) $P' \subseteq \mu(P)$ where $\mu(P) = \bigcup_{A \rightarrow \alpha \text{ in } P} \mu(A) \rightarrow \mu(\alpha)$,
- (v) S' is in $\mu(S)$.

$\mathcal{G}(F) = \{F' : F' \triangleleft F\}$ is the *family of EOL forms generated by F* , and $\mathcal{L}(F) = \{L(F') : F' \triangleleft F\}$ 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 $\mathcal{L}(F_1) = \mathcal{L}(F_2)$.

DEFINITION: Let F be an EOL form. The class of language families generated by F is denoted by $\hat{\mathcal{L}}(F)$, and is defined by:

$$\hat{\mathcal{L}}(F) = \{ \mathcal{L}(F') : F' \triangleleft F \}.$$

We now prove stronger versions of the two simulation lemmas in [3].

We first state without proof the basic result from [3].

LEMMA 2.1: Let $F = (V, \Sigma, P, S)$ and $\bar{F} = (\bar{V}, \bar{\Sigma}, \bar{P}, \bar{S})$ be forms and let $l \geq 1$ be an integer such that $X \Rightarrow_{\bar{F}}^l \alpha$ for each $X \rightarrow \alpha$ in P . Then for each $F' = (V', \Sigma', P', S') \triangleleft F(\mu)$ a form $\bar{F}' = (\bar{V}', \bar{\Sigma}', \bar{P}', \bar{S}') \triangleleft \bar{F}(\bar{\mu})$ can be constructed such that

$$X' \rightarrow \alpha' \text{ is in } P' \text{ iff } X' \text{ is in } \mu(V) \text{ and } X' \Rightarrow_{F'}^l \alpha'.$$

We now have:

LEMMA 2.2. *One-many simulation:* Suppose $F = (V, \Sigma, P, S)$ and $\bar{F} = (\bar{V}, \bar{\Sigma}, \bar{P}, \bar{S})$ are forms and for some integer $l \geq 1$, $X \Rightarrow_{\bar{F}}^l \alpha$ holds for each $X \rightarrow \alpha$ in P . Then $\mathcal{L}(F) \subseteq \mathcal{L}(\bar{F})$ and $\hat{\mathcal{L}}(F) \subseteq \hat{\mathcal{L}}(\bar{F})$.

Proof: Note that we may assume, as in [3], that $X \Rightarrow_{\bar{F}}^l \alpha$, for each $X \rightarrow \alpha$ in P , without any loss of generality. And since $\mathcal{L}(F) \subseteq \mathcal{L}(\bar{F})$ has been proved in [3], we only need to show that $\hat{\mathcal{L}}(F) \subseteq \hat{\mathcal{L}}(\bar{F})$. Consider $F' \triangleleft F(\mu)$, an arbitrary interpretation of F . Let $\bar{F}' \triangleleft \bar{F}(\bar{\mu})$ be the constructed form of Lemma 2.1. Now:

a) $\mathcal{L}(F') \subseteq \mathcal{L}(\bar{F}')$. Notice that F' and \bar{F}' fulfill the conditions of the present lemma;

b) $\mathcal{L}(\bar{F}') \subseteq \mathcal{L}(F')$. F' and \bar{F}' fulfill the conditions of Lemma 2.3, the many-one simulation lemma.

DEFINITION: For any EOL form $F = (V, \Sigma, P, S)$ and integer $l \geq 1$ define $V(l) = \{ X : X \text{ in alph}(\alpha), S \Rightarrow_{\bar{F}}^m \alpha \text{ with } m \text{ a multiple of } l \}$.

LEMMA 2.3. *Many-one simulation:* Let $F = (V, \Sigma, P, S)$ and $\bar{F} = (\bar{V}, \bar{\Sigma}, \bar{P}, \bar{S})$ be forms. Suppose for some integer $l \geq 1$ the conditions X in $V(l)$ and $X \Rightarrow_{\bar{F}}^l \alpha$ imply.

(i) $X \Rightarrow_{\bar{F}}^l \alpha$ and

(ii) $X \rightarrow \alpha$ is in P .

Then $\mathcal{L}(\bar{F}) \subseteq \mathcal{L}(F)$ and $\hat{\mathcal{L}}(\bar{F}) \subseteq \hat{\mathcal{L}}(F)$.

Proof: The first part is proved in [3]. Consider the second part. Let $\bar{F}' = (\bar{V}', \bar{\Sigma}', \bar{P}', S') \triangleleft \bar{F}(\bar{\mu})$, be an arbitrary interpretation of \bar{F} . Construct $F' = (V', \Sigma', P', S)$ from \bar{F} , where $V' \subseteq V \subseteq \Sigma'$, and $P' = \{X \rightarrow \alpha \text{ in } P : X \Rightarrow \frac{1}{F} \alpha \text{ and } X \text{ in } \bar{V}(l)\} \subseteq P$. Now $F' \triangleleft F$ and $\hat{\mathcal{L}}(F') \subseteq \hat{\mathcal{L}}(F)$.

Let $F'' = (\bar{\mu}(V'), \bar{\mu}(\Sigma'), P'', S')$ where $X' \rightarrow \alpha'$ is in P'' iff $X' \Rightarrow \frac{1}{\bar{F}'} \alpha'$ and X' is in $\bar{V}'(l)$. Clearly $F'' \triangleleft F'(\bar{\mu})$, therefore $\mathcal{L}(F'')$ is in $\hat{\mathcal{L}}(F') \subseteq \hat{\mathcal{L}}(F)$. To conclude the proof observe that $\mathcal{L}(\bar{F}') = \mathcal{L}(F'')$, by applying the first parts of Lemma 2.2 and the present lemma.

Consider the following “application”:

Example: Let

$$F : S \rightarrow a \mid S \mid SS; \quad a \rightarrow a \mid S \mid SS;$$

and

$$G : S \rightarrow A; \quad A \rightarrow a \mid S \mid SS; \quad a \rightarrow A;$$

be two forms.

Note that

$$S \Rightarrow_G A \Rightarrow_G SS; \quad S \Rightarrow_G A \Rightarrow_G a; \quad S \Rightarrow_G A \Rightarrow_G S;$$

and

$$a \Rightarrow_G A \Rightarrow_G SS; \quad a \Rightarrow_G A \Rightarrow_G a; \quad a \Rightarrow_G A \Rightarrow_G S;$$

are total nonterminal derivations.

Therefore by Lemmas 2.2 and 2.3, not only do we have $\mathcal{L}(F) = \mathcal{L}(G)$ but also $\hat{\mathcal{L}}(F) = \hat{\mathcal{L}}(G)$.

This example demonstrates the following :

THEOREM 2.4: *Given two EOL forms $F = (V, \Sigma, P, S)$ and $\bar{F} = (\bar{V}, \bar{\Sigma}, \bar{P}, S)$ and an integer $l \geq 1$ such that:*

$$X \rightarrow \alpha \text{ is in } P \text{ iff } X \Rightarrow_{int \bar{F}}^l \alpha \text{ and } X \text{ is in } \bar{V}(l),$$

then $\mathcal{L}(F) = \mathcal{L}(\bar{F})$ and $\hat{\mathcal{L}}(F) = \hat{\mathcal{L}}(\bar{F})$.

We say F and \bar{F} are int -simulation related written $F \vdash_{int} \bar{F}$.

Proof: By Lemmas 2.2 and 2.3.

3. RELATIVE AND MUTUAL GOODNESS

For two EOL forms F and G we say F is *good relative to G* (G -good) if $\mathcal{L}(G)$ is in $\hat{\mathcal{L}}(F)$. In other words there exists $F' \triangleleft F$ with $\mathcal{L}(F') = \mathcal{L}(G)$. Similarly, F is *bad relative to G* (G -bad) if:

- (i) $\mathcal{L}(G) \subseteq \mathcal{L}(F)$, and
- (ii) $\mathcal{L}(G)$ is not in $\hat{\mathcal{L}}(F)$.

Condition (i) excludes trivially bad forms.

Examples:

a) $F_1 : S \rightarrow \varepsilon \mid a \mid S \mid SS; a \rightarrow S$; is by [4] known to be G -good for all EOL forms G . In other words $\hat{\mathcal{L}}(F_1)$ contains every EOL form language family.

b) $F_2 : S \rightarrow a; a \rightarrow a$; is also known to be G -good for all G such that $\mathcal{L}(G) \subseteq \mathcal{L}(F_2)$.

c) $F_3 : S \rightarrow a; a \rightarrow N; N \rightarrow N$; is F_4 -bad, where $F_4 : S \rightarrow a; a \rightarrow b; b \rightarrow b; \triangleleft F_2$.

Essentially any nonempty language in $\mathcal{L}(F_4)$ contains at least two words, whereas for all $F'_3 \triangleleft F_3$, $\mathcal{L}(F'_3)$ contains singleton languages.

d) $F_5 : S \rightarrow aa; a \rightarrow N; N \rightarrow N$;

now F_3 is neither F_5 -good nor F_5 -bad (or vice versa).

We extend the notions of relative goodness and badness to families of forms, since this is our prime area of interest.

DEFINITION: Let \mathcal{F} be a family of EOL forms. We let $\mathcal{L}(\mathcal{F})$ denote $\{ \mathcal{L}(F) : F \text{ is in } \mathcal{F} \}$, so that, in particular, $\mathcal{L}(\mathcal{G}(F)) = \hat{\mathcal{L}}(F)$.

We say a form F is \mathcal{F} -good if $\mathcal{L}(\mathcal{F}) \subseteq \hat{\mathcal{L}}(F)$. Similarly a form is \mathcal{F} -bad if for all G in \mathcal{F} , F is G -bad.

REMARKS:

1) A form F is always good with respect to an interpretation of itself, that is for all $F' \triangleleft F$, F is $\mathcal{G}(F')$ -good.

2) The relative goodness of F is preserved under inverse interpretations. That is, if F' is $\mathcal{G}(G)$ -good, where $F' \triangleleft F$ then F is $\mathcal{G}(G)$ -good.

3) However, it is easy to see that F is $\mathcal{G}(G')$ -good does not necessarily imply that F is $\mathcal{G}(G)$ -good, where $G' \triangleleft G$. The converse is, of course, true. F is $\mathcal{G}(G)$ -good implies F is $\mathcal{G}(G')$ -good. We can replace "good" with "bad".

4) Relative badness is preserved under interpretations that fulfill condition (i) of the Definition.

5) If F is \mathcal{F}_1 -good and \mathcal{F}_2 -good, then it is also $\mathcal{F}_1 \cup \mathcal{F}_2$ -good, $(\mathcal{F}_1 \cap \mathcal{F}_2)$ -good and $(\mathcal{F}_1/\mathcal{F}_2)$ -good. This is also true for badness. In [4] the following weaker notion of badness was used.

DEFINITION: Let F be an EOL form and \mathcal{F} a family of EOL forms, we say F is *weakly \mathcal{F} -bad* if (i) and (ii) hold.

- (i) For all G in \mathcal{F} , $\mathcal{L}(G) \subseteq \mathcal{L}(F)$, and
- (ii) There exists G in \mathcal{F} such that F is G -bad.

Example: Consider

$$F : S \rightarrow a; a \rightarrow N; N \rightarrow N;$$

and

$$G : S \rightarrow a; a \rightarrow a;$$

F is neither $\mathcal{G}(G)$ -bad nor $\mathcal{G}(G)$ -good, but F is weakly $\mathcal{G}(G)$ -bad, since

- (i) for all $G' \triangleleft G$, $\mathcal{L}(G') \subseteq \mathcal{L}(F)$, and
- (ii) for $G' : S \rightarrow a; a \rightarrow b; b \rightarrow b$; F is G' -bad. In fact we know [4] that for any synchronized F such that $\mathcal{L}(G) \subseteq \mathcal{L}(F)$, F is weakly $\mathcal{G}(G)$ -bad.

REMARKS:

1) F is weakly \mathcal{F}_1 -bad implies F is weakly \mathcal{F}_2 -bad for all $\mathcal{F}_2 \supseteq \mathcal{F}_1$ such that for all H in \mathcal{F}_2 , $\mathcal{L}(H) \subseteq \mathcal{L}(F)$. In particular if F is weakly $\mathcal{G}(G')$ -bad for some $G' \triangleleft G$, and for all $G'' \triangleleft G$, $\mathcal{L}(G'') \subseteq \mathcal{L}(F)$ then F is weakly $G(G)$ -bad.

2) F is weakly \mathcal{F}_1 -bad and weakly \mathcal{F}_2 -bad implies F is weakly $(\mathcal{F}_1 \cup \mathcal{F}_2)$ -bad, but, in general, not weakly $(\mathcal{F}_1 \cap \mathcal{F}_2)$ -bad.

The notion of relative goodness leads naturally to the mutual goodness of two EOL forms. Two forms F and G are *mutually good* if F is $\mathcal{G}(G)$ -good and G is $\mathcal{G}(F)$ -good.

REMARKS:

1) Mutual goodness is clearly an equivalence relation, which we denote by \equiv_{mg} . $F \equiv_{mg} G$ is true iff $\hat{\mathcal{L}}(F) = \hat{\mathcal{L}}(G)$. Hence $F \equiv_{mg} G$ implies $\mathcal{L}(F) = \mathcal{L}(G)$, but the converse clearly does not hold.

2) Mutual goodness could be defined at the simpler level, that, is when F is G -good and G is F -good. However this is only a reformulation of form equivalence and is of little interest.

3) Mutual badness, however appealing, would under the natural definition be ill-defined since it would imply $\mathcal{L}(F) = \mathcal{L}(G)$. Mutual weak badness does exist but it isn't transitive or reflexive and seems to be of little interest.

Examples:

a) $F_n : S_i \rightarrow S_{i+1}, 0 \leq i < n; S_n \rightarrow a; a \rightarrow N; N \rightarrow N$; where S_0 is the sentence symbol.

$\mathcal{L}(F_i) = \mathcal{L}(F_j)$, for all $i, j \geq 0$.

Further, for all $i, j \geq 0, i \neq j, \mathcal{G}(F_i) \neq \mathcal{G}(F_j)$, therefore F_i and F_j are not strong form equivalent. However F_i *tnt*-simulates F_0 , for all $i > 0$, therefore from Theorem 2.4 we deduce that $\hat{\mathcal{L}}(F_i) = \hat{\mathcal{L}}(F_0)$, therefore $F_i \equiv_{mg} F_j$, for all $i, j \geq 0$.

b) $F : S \rightarrow a \mid bb; a \rightarrow a; b \rightarrow N; N \rightarrow N$ and $G : S \rightarrow a \mid bb; a \rightarrow N; N \rightarrow N; b \rightarrow b$; are mutually weakly bad.

The following result is important and by the preceding remarks and examples straightforward.

LEMMA 3.1: For EOL forms F and G :

- (i) $\mathcal{G}(F) = \mathcal{G}(G)$ implies $F \equiv_{mg} G$, but not vice versa;
- (ii) $F \equiv_{mg} G$ implies $\mathcal{L}(F) = \mathcal{L}(G)$, but not vice versa, and
- (iii) $F \vdash_{nt} G$ implies $F \equiv_{mg} G$, but not vice versa.

Lemma 2.2 leads to the following results on *nt*-simulation.

LEMMA 3.2: For EOL forms, F, G and H , and a family of forms \mathcal{F} :

- (i) F *nt*-simulates G implies F is both G -good and $\mathcal{G}(G)$ -good;
- (ii) F *nt*-simulates G and G is \mathcal{F} -good implies F is \mathcal{F} -good;
- (iii) G *nt*-simulates H and F is $\mathcal{G}(G)$ -good ($\mathcal{G}(G)$ -bad) implies F is $\mathcal{G}(H)$ -good ($\mathcal{G}(H)$ -bad);
- (iv) H *nt*-simulates G , F is weakly $\mathcal{G}(G)$ -bad and $\mathcal{L}(H) \subseteq \mathcal{L}(F)$ implies F is weakly $\mathcal{G}(H)$ -bad.

Proof: Lemma 2.2 states that F *nt*-simulates G implies $\hat{\mathcal{L}}(G) \subseteq \hat{\mathcal{L}}(F)$, hence these results follow immediately.

DEFINITION: Let \mathcal{F} be a family of EOL forms. For F in \mathcal{F} we say \mathcal{F} is *F-interpretation closed* if $\mathcal{G}(F) \subseteq \mathcal{F}$. Let $\text{Int}(\mathcal{F})$, the *interpretation closure* of \mathcal{F} , be $\bigcup_{F \text{ in } \mathcal{F}} \mathcal{G}(F)$. \mathcal{F} is *interpretation closed* if $\mathcal{F} = \text{Int}(\mathcal{F})$.

We are now able to give a “transitivity” result for goodness.

THEOREM 3.3: For forms F and G , and families of forms \mathcal{F}_1 and \mathcal{F}_2 , if F is \mathcal{F}_1 -good, G is \mathcal{F}_2 -good and \mathcal{F}_1 is G -interpretation closed, then F is \mathcal{F}_2 -good.

Proof:

$$\mathcal{L}(\mathcal{F}_2) \subset \hat{\mathcal{L}}(G), \mathcal{L}(\mathcal{F}_1) \subseteq \hat{\mathcal{L}}(F) \quad \text{and} \quad \mathcal{G}(G) \subseteq \mathcal{F},$$

immediately implies $\mathcal{L}(\mathcal{F}_2) \subseteq \hat{\mathcal{L}}(F)$, that is, F is \mathcal{F}_2 -good.

COROLLARY 3.4: *For EOL forms F, G and H, F is $\mathcal{G}(G)$ -good and G is $\mathcal{G}(H)$ -good implies F is $\mathcal{G}(H)$ -good.*

At the present time we have not settled the decidability of mutual goodness or of *tnt*-simulation relatedness.

We close this section by proving that left and right linear EOL forms are not mutually good, a somewhat surprising result.

THEOREM 3.5: *For $F : S \rightarrow a \mid bS; a \rightarrow a; b \rightarrow c; c \rightarrow a$; there is no left linear form H with $\mathcal{L}(H) = \mathcal{L}(F)$. ($L(F) = a^+ \cup a^*ca$).*

Proof: Assume there is such an H . Then there is an interpretation $H' \triangleleft H$ such that $L(H') = L(F)$. Now let $H' = (V', \Sigma', P', S')$, where $\{a, c\} \subset \Sigma'$. Note first of all that since F is deterministic on terminals, for any interpretation $F' \triangleleft F(\mu)$, for each word y in $L(F')$ such that $\mu^{-1}(y) = a^i$, for some $i \geq 2$, there is a word x in $L(F')$ with $\mu^{-1}(x) = a^{i-2}ca$. Therefore the derivation of $a^i ca$ cannot be "isolated" from the derivation of a^{i+2} in H' , since this would immediately imply $\mathcal{L}(H) \neq \mathcal{L}(F)$.

Secondly, since the derivations of $a^i ca$ and a^{i+2} in H' are tied together, consider the two possibilities:

(i) $a^{i+2} \Rightarrow^+ a^i ca$; this implies $a \Rightarrow^+ c$, in other words since a is in $L(F)$, c is in $L(F)$, a contradiction;

(ii) $a^i ca \Rightarrow^+ a^{i+2}$; this implies $a \Rightarrow^+ a$ and $c \Rightarrow^+ a$, any other possibility would lead to a contradiction. For example $c \Rightarrow^+ \varepsilon$ or $c \Rightarrow^+ a^j, j > 1$ would enable a language to be obtained as an interpretation of H' , which could not be obtained as an interpretation of F .

Thirdly, since H' is left linear and $a^i ca \Rightarrow_{H'}^+ a^{i+2}$, the arrangement to deposit a "c" in the penultimate position of a word $a^i ca$, for any $i \geq 0$, must be made at the second derivation step (since the symbols are deposited in a right-to-left manner with a left linear grammar). This implies :

there is a terminal d such that $d \Rightarrow^+ d$ and $d \Rightarrow^+ c$.

Now if $d \neq a$ and $d \neq c$ then $a^i da$ is in $L(H')$, a contradiction. If $d = a$, then c is in $L(H')$, a contradiction. Hence we are left with $d = c$. However, consider an interpretation $H'' \triangleleft H'$, in which $c \Rightarrow_{H''}^+ a$ is not possible, then $L(H'')$ is not in $\mathcal{L}(F)$.

COROLLARY 3.6: $F_1 : S \rightarrow a \mid aS$; $a \rightarrow a$; and $F_2 : S \rightarrow a \mid Sa$; $a \rightarrow a$; are not mutually good.

Proof: If F_1 and F_2 are mutually good then for all $F'_1 \triangleleft F_1$ there exists $F'_2 \triangleleft F_2$ with $\mathcal{L}(F'_1) = \mathcal{L}(F'_2)$ and vice versa. However F of Theorem 3.5 is an interpretation of F_1 and by Theorem 3.5 there is no $F'_2 \triangleleft F_2$ with $\mathcal{L}(F) = \mathcal{L}(F'_2)$.

4. WHEN ARE FORMS RELATIVELY BAD OR GOOD?

By [4] and the results of the last section we have many examples of relatively bad and good forms. This however leaves open the problem of determining when a form is bad or good. We present some first steps towards a solution to these problems in this section.

DEFINITION: An EOL form $F = (V, \Sigma, P, S)$ is:

- (i) *synchronized* if for all a in Σ , $a \Rightarrow^+ \alpha$ implies α is not in Σ^* ;
- (ii) *weakly synchronized* if for all x in Σ^+ such that $S \Rightarrow^+ x$, then $x \Rightarrow^+ \alpha$ implies α is not in Σ^* ;
- (iii) *k-weakly synchronized*, for some $k > 0$, if for all x in Σ^+ with $S \Rightarrow^+ x$, for all $l \geq k$, $x \Rightarrow^l \alpha$ implies α is not in Σ^* .

Surprisingly, the authors are still unaware whether a weakly synchronized form can always be replaced by a form equivalent synchronized form.

We now have a generalization of a result in [4].

THEOREM 4.1: *Given two forms F and G with $\mathcal{L}(G) \subseteq \mathcal{L}(F)$, F k -weakly synchronized for some $k > 0$, and in G every infinite sentential derivation contains at least $k+1$ distinct terminal words, then F is $\mathcal{G}(G)$ -bad.*

Proof: Let property ⁽⁵⁾ be:

Now since G fulfills property ⁽⁵⁾ then all nontrivial $G' \triangleleft G$ must also fulfill property ⁽⁵⁾, because of:

- (i) the completeness condition for EOL forms;
- (ii) distinct terminal words remain distinct under interpretation;
- (iii) every derivation in G' is the image of a derivation in G .

Immediately, all L in $\mathcal{L}(G')$ contain at least $k+1$ distinct words. By a similar argument, each $F' \triangleleft F$ is k -weakly synchronized, and hence there exists L in $\mathcal{L}(F')$ containing at most k words. Therefore for all $F' \triangleleft F$, for all $G' \triangleleft G$, $\mathcal{L}(F') \neq \mathcal{L}(G')$.

⁽⁵⁾ Every infinite sentential derivation has at least $k+1$ distinct terminal words.

We obtain immediately the following corollaries:

COROLLARY 4.2: *Given two forms F and G with $\mathcal{L}(G) \subseteq \mathcal{L}(F)$, F k -weakly synchronized and in G every infinite sentential derivation contains at least $k+1$ nonempty terminal words, then F is weakly $\mathcal{G}(G)$ -bad, and there exists $G' \triangleleft G$ such that F is $\mathcal{G}(G')$ -bad.*

Proof: By observation that we can obtain a $G' \triangleleft G$ which fulfills property (5) of Theorem 4.1.

COROLLARY 4.3: *Given two forms F and G with $\mathcal{L}(G) \subseteq \mathcal{L}(F)$, F is synchronized (k -weakly synchronized) and G is not weakly synchronized (not l -weakly synchronized, $l \leq k$), then F is weakly $\mathcal{G}(G)$ -bad and there exists $G' \triangleleft G$ such that F is $\mathcal{G}(G')$ -bad.*

We have similar results for propagating forms:

DEFINITION: G is *inherently non-propagating* if $\mathcal{L}(F) \neq \mathcal{L}(G)$ for every propagating EOL form F .

THEOREM 4.4: *There exist inherently non-propagating forms.*

Proof: From [4]: $S \rightarrow aba$; $a \rightarrow c$; $b \rightarrow \varepsilon$; $c \rightarrow c$; and from [6]: $S \rightarrow a$; $a \rightarrow abba$; $b \rightarrow \varepsilon$.

Immediately we obtain:

LEMMA 4.5: *If G is an inherently non-propagating EOL form $\mathcal{L}(G) \subseteq \mathcal{L}(F)$ and F is a propagating EOL form, then F is G -bad.*

COROLLARY 4.6: *Let F be a propagating EOL form and H an EOL form such that $\mathcal{L}(H) \subseteq \mathcal{L}(F)$ and there exists an inherently non-propagating EOL form $H' \triangleleft H$. Then F is weakly $\mathcal{G}(H)$ -bad.*

In [3] it was shown that the "standard" method of transforming a synchronized non-propagating EOL system F to an equivalent synchronized and propagating EOL system H also preserved language families, that is, not only is $L(F) = L(H)$ but also $\mathcal{L}(F) = \mathcal{L}(H)$. However it is open whether this transformation also preserves the class of language families, that is, whether $\hat{\mathcal{L}}(F) = \hat{\mathcal{L}}(H)$.

5. SPOILERS

In the previous section we investigated the relationship of propagation and synchronization with badness. We provided partial answers to the question, given an arbitrary propagating or synchronized form F , what

conditions on a form G are needed to ensure that F is G -bad or $\mathcal{G}(G)$ -bad. Such a G is important since it applies to a whole class of forms rather than a single form, we say G spoils such a class. In the following we are interested in $\mathcal{G}(G)$ -badness.

DEFINITION: Let F and G be EOL forms. If F is $\mathcal{G}(G)$ -bad, we say that G is an F -spoiler. Similarly G is an \mathcal{F} -spoiler for a family of forms \mathcal{F} if, for all F in \mathcal{F} , G is an F -spoiler. And finally G is a weak \mathcal{F} -spoiler if there is an F in \mathcal{F} such that G is an F -spoiler. Let \mathcal{F} (EOL), \mathcal{F} (propagating) and \mathcal{F} (synchronized) denote the families of all forms which are EOL, propagating and synchronized, respectively.

We can unify the results of the previous section as follows:

THEOREM 5.1: *Let F , G and H be EOL forms.*

(i) *If G is a weak \mathcal{F} (synchronized)-spoiler (weak \mathcal{F} (propagating)-spoiler) then $G' \triangleleft G$ is a weak \mathcal{F} (synchronized)-spoiler (weak \mathcal{F} (propagating)-spoiler).*

(ii) *If G is an \mathcal{F} (synchronized)-spoiler (\mathcal{F} -(propagating)-spoiler) and $G \vdash_{int} H$ then H is an \mathcal{F} (synchronized)-spoiler (\mathcal{F} (propagating)-spoiler).*

Proof:

(i) *If F is $\mathcal{G}(G)$ -bad then clearly F is $\mathcal{G}(G')$ -bad for $G' \triangleleft G$.*

(ii) *If F is $\mathcal{G}(G)$ -bad and $G \vdash_{int} H$ then G is $\mathcal{G}(H)$ -bad.*

LEMMA 5.2: *Let \mathcal{F} be a family of EOL forms. If F is in \mathcal{F} then F is not an \mathcal{F} -spoiler.*

Proof: If F is an \mathcal{F} -spoiler and F is in \mathcal{F} then F is $\mathcal{G}(F)$ -bad, a contradiction.

COROLLARY 5.3: *No EOL form is an \mathcal{F} (EOL)-spoiler.*

THEOREM 5.4: *Let G and H be EOL forms. Then G is a $\mathcal{G}(H)$ -spoiler iff H is $\mathcal{G}(G)$ -bad.*

Proof: G is a $\mathcal{G}(H)$ -spoiler implies G is an H -spoiler and therefore H is $\mathcal{G}(G)$ -bad. If H is $\mathcal{G}(G)$ -bad, then $H' \triangleleft H$ is $\mathcal{G}(G)$ -bad, hence G is an H' -spoiler, that is, G is a $\mathcal{G}(H)$ -spoiler.

And finally:

LEMMA 5.5: *There exists a good EOL form which is also a spoiler.*

Proof: $F : S \rightarrow a; a \rightarrow b; b \rightarrow b$ is good and it is an \mathcal{F} (synchronized)-spoiler.

6. GOOD AND NO GOOD LANGUAGES

In this penultimate section we consider the goodness (and its absence) of languages, rather than forms.

DEFINITION: Let $L \subseteq \Sigma^*$. We say L is a *good language* if $L = L(F)$ for some good form F . L is *no good* if there is no good form F with $L = L(F)$.

THEOREM 6.1: Let $L \subseteq \Sigma^*$ and $\{a\}^* \subseteq L$, for some a in Σ . Then L is good.

Proof: Let $G = (V, \Sigma, P, S)$ be a synchronized EOL system with $L = L(G)$.

Construct an EOL form $F = (V_1, \Sigma, P_1, S_1)$ from G as follows:

$$V_1 = V \cup \{S_1, S_2\}$$

and

$$P_1 = P \cup \{S_1 \rightarrow S; S_1 \rightarrow S_2; S_2 \rightarrow \varepsilon \mid a \mid S_2 \mid S_2 S_2; a \rightarrow S_2\}.$$

Now clearly $L \subseteq L(F)$. $L(F) \subseteq L$ since whenever a word a^i is generated from either S or S_2 its only terminal successors are words in $\{a\}^*$ (from rules not in P). On the other hand any word in $L(F) - \{a\}^*$ has no successor terminal words.

Finally, since $F' : S \rightarrow \varepsilon \mid a \mid S \mid SS; a \rightarrow S$; is complete and $F' \triangleleft F$ we have the result.

THEOREM 6.2: $L = \{a, abba\}$ is no good.

Proof: Assume L is good, then there exists a good EOL form F with $L(F) = L$. Now it is clear that F must not be synchronized, since it is shown in [4] that no synchronized form is good. Letting $F = (V, \{a, b\} \cup \Sigma, P, S)$ we must have:

$$\text{either } S \Rightarrow^+ a \Rightarrow^+ abba,$$

or

$$S \Rightarrow^+ abba \Rightarrow^+ a.$$

The latter alternative leads to an immediate contradiction. In the former situation $abba \Rightarrow^+ \alpha$ implies α is not in $(\Sigma \cup \{a, b\})^*$. Otherwise we have either $abba \Rightarrow^+ a$, which again is impossible or $abba \Rightarrow^+ abba$, from which a contradiction ensues. (For since $a \Rightarrow^+ abba \Rightarrow^+ abba$, if $a \Rightarrow^+ a$ and $b \Rightarrow^+ b$ then we can obtain $abbabbabba$, if $a \Rightarrow^+ \varepsilon$ and $b \Rightarrow^+ abba$ then we can obtain $(abba)^m$ for some $m > 1$.)

Hence $abba \Rightarrow_F^+ \alpha$ implies α is not terminal. Thus F is 2-weakly synchronized. This provides our contradiction, since a 2-weakly synchronized form cannot be good.

7. OPEN PROBLEMS

The basic open problem is whether mutual goodness is decidable. However the question of the decidability of *tnt*-simulation relatedness although a weaker result is still of interest.

Consider the following definition:

Let $\mathcal{L}^{(0)}(F) = L(F)$,

and for all $i > 0$, let $\mathcal{L}^{(i)}(F) = \{ \mathcal{L}^{(i-1)}(F') : F' \triangleleft F \}$. It is clear that $\mathcal{L}^{(1)}(F) = \mathcal{L}(F)$ and $\mathcal{L}^{(2)}(F) = \hat{\mathcal{L}}(F)$. It is easy to construct examples of EOL forms F_1 and F_2 such that:

- 1) $\mathcal{L}^{(0)}(F_1) = \mathcal{L}^{(0)}(F_2)$ but $\mathcal{L}^{(1)}(F_1) \neq \mathcal{L}^{(1)}(F_2)$. Weak equivalence;
- 2) $\mathcal{L}^{(i)}(F_1) = \mathcal{L}^{(i)}(F_2)$, $i = 0, 1$ but $\mathcal{L}^{(2)}(F_1) \neq \mathcal{L}^{(2)}(F_2)$. Weakly equivalent and form equivalent, but not mutually good.

However does there exist F_1 and F_2 such that;

- 3) $\mathcal{L}^{(i)}(F_1) = \mathcal{L}^{(i)}(F_2)$, $i = 0, 1, 2$ but $\mathcal{L}^{(3)}(F_1) \neq \mathcal{L}^{(3)}(F_2)$?

And can this question be answered in general?

Finally it is worth noting that Lemmas 2.2 and 2.3 can be generalized for $\mathcal{L}^{(i)}(F)$ and $\mathcal{L}^{(i)}(\bar{F})$ by an inductive argument.

REFERENCES

1. A. B. CREMERS and S. GINSBURG, *Context-free grammar forms*, Journal of Computer and Systems Sciences, Vol. 11, 1975, pp. 86-117.
2. M. K. CULIK, H. A. MAURER and Th. OTTMANN, *Two letter complete EOL forms*, Theoretical Computer Sciences, Vol. 6, 1978, pp. 69-92.
3. H. A. MAURER, A. SALOMAA and D. WOOD, *EOL forms*, Acta Informatica, Vol. 8, 1977, pp. 75-96.
4. H. A. MAURER, A. SALOMAA and D. WOOD, *On good EOL forms*, S.A.I.M. Journal of Computing, Vol. 7, 1978, pp. 158-166.
5. H. A. MAURER, A. SALOMAA and D. WOOD, *ETOL forms*. Journal of Computer and System Sciences, Vol. 16, 1978, pp. 345-361.
6. SKYUM, *On good ETOL forms*, Theoretical Computer Science (to appear), 1978.