Abstract

Two closely related types of context-free grammars called X-bar and strong X-bar grammrs are presented along the lines of *Jackendoff* (1977). The family **E** of languages generated by strong X-bar grammars is properly included in the family **K** of languages generated by X-bar grammars, which, in turn, is a proper subset of the context-free family L_2. **E** and **K** cannot be compared to the regular, OL, and TOL families, and both are *anti-AFLs*.

J. Demetrovics, G.O.H. Katona, and A. Salomaa (eds):
COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
42. Algebra, Combinatorics and Logic in Computer Science
Győr (Hungary), 1983.

$\bar{X}$ - GRAMMARS

A. KORNAI

Abstract. Two closely related types of context-free grammars called $\bar{X}$ (read *x*-bar) and strong $\bar{X}$ grammars are presented along the lines of J a c k e n d o f f (1977). The family E of languages generated by strong $\bar{X}$ grammars is properly included in the family K of languages generated by $\bar{X}$ grammars which, in turn, is a proper subset of the context-free family L_2. E and K cannot be compared to the regular, OL and TOL families, and both are *anti-AFLs*.

INTRODUCTION

In this paper we investigate a rewriting system, the so-called '$\bar{X}$-*theory*', which is quite widely used (as the phrase-structure component) in describing natural languages.

This investigation is not devoid of purely mathematical interest because $\bar{X}$ grammars occupy some sort of middle ground between (type-2) phrase-structure grammars *(PSGs)* and Lindenmayer systems (*L*-systems). While there is no terminal/nonterminal distinction in L-systems, and the terminals and nonterminals of PSGs are entirely unrelated, $\bar{X}$ grammars have the property of *lexicality*, namely: each nonerminal is linked to one (and only one) of the terminals. This is expressed by the '*bar notation*': nonterminals

are of the form $X^{(i)}$, $(x \in V_T,\ i \geq 1)$.

The $x^{(i)}$-s are called *projections* of x, and x is the *head* of $x^{(i)}$. By definition, $x^{(0)} = x$, $x^{(1)} = \bar{x}$, $x^{(2)} = \bar{\bar{x}}, \ldots$ and these are taken as single (unanalyzed) symbols of length one.

1. BASIC DEFINITIONS AND LEMMAS

We now give the formal definition of (strong) $\bar{X}$ grammars. We will assume that the reader is familiar with the basic notions and notations of formal language theory. Throughout this paper, all terms and symbolism used follow S a l o m a a's book (1973), unless indicated otherwise. Given an alphabet V, we define a partial ordering D on V^* the following way: $\alpha D \beta$ iff α can be arrived at by the *deletion* of certain (possibly none) letters from β. The *'bar-adding'* and *'bar-deleting'* homomorphisms μ and ν will be frequently used in the sequel:

$$\mu(x^{(n)}) = x^{(n+1)} \ (x \in V_T),$$

$$\nu(x^{(n)}) = x^{(n-1)} \ (x^{(n)} \in V_N).$$

$\mu(\alpha)$ will also be denoted by $\bar{\alpha}$.

Definition 1.1. An $X^{(n)}$ grammar (read x n-bar grammar) is a CF grammar $G = (V_N, V_T, P, S)$ satisfying

(i) $V_N = \{x^{(i)} \mid 1 \leq i \leq n,\ x \in V_T\}$ (lexicality)

(ii) $S = v^{(n)}$, $v \in V_T$ (centrality)

(iii) the rules in P have the form $x^{(i)} \to \alpha\ x^{(i-1)}\beta$, where α and β are (possibly empty) strings over $V_M = \{x^{(n)} \mid x \in V_T\}$, (maximality, uniformity, and succesion).

Remark 1.1. The succession property in the above definition means simply that an $x^{(i-1)}$ must appear in any rule rewriting

$x^{(i)}$. Symbols to the right of this (unique) successor are called *complements*, and symbols to its left are called *specifiers* (of $x^{(i)}$). Uniformity means that the maximal pro ection $x^{(max)}$ is the same for every $x \in V_T$, and maximality means that the specifiers and complements belong inthe subset V_M of V_N containing the maximal projection for each head. Since V_N is finite, lexicality enables us to define V_M as the set of those nonterminals that have no higher projections in V_N. This way, lexicality and sussessivity can be maintained without any reference to uniformity. Strong $\bar{X}$ grammars have the additional property of *optionality:* some (or all) $x^{(i)}$ specifiers and/or complements need not appear in the string resulting from the application of some $x^{(i)}$ rule, but if they do, their order must conform tothe one prescribed by the rule in question. Since ordinary *CF* rules never have shis interpretation, optionality will be indicated by the paranthesis notation traditionally employed in linguistics for precisely this purpose: e.g. $x^{(2)} \to (x^{(3)})x^{(1)}(z^{(3)})$ is equivalent to the *BNF* rule $x^{(2)} := x^{(3)}x^{(1)}z^{(3)} | x^{(3)}x^{(1)} | x^{(1)}z^{(3)} | x^{(1)}$. Instead of modifying the interpretation of rewriting rules, we will simulate the effect of optionally interpreted rules by sets of traditional r les:

Definition 1.2. A strong $X^{(n)}$ grammar is an $X^{(n)}$ grammar $G = (V_N, V_T, P, S)$ with the additional properties that

(i) for every $x^{(i)} \to \alpha\, x^{(i-1)} \beta \in P$, $\alpha' D\, \alpha$ and $\beta' D\, \beta$,

$$x^{(i)} \to \alpha' x^{(i-1)} \beta' \quad \text{is also in } P. \text{ (optionality)}$$

(ii) for every $x^{(i)} \in V_N$, there exists a unique rewriting rule of maximal length.

Remark 1.2. It is obvious from the above definition, that in a strong $X^{(n)}$ grammar every rule can be written in the following *'paranthesis form'*:

$$x^{(i)} \to (S_1)(S_2) \ldots (S_\ell)x^{(i-1)}(C_1)(C_2) \ldots (C_r), \quad (0 \leq \ell, r)$$

where the *specifiers* S_i and *complements* C_j are in V_M. In case $n=3$, this is precisely the *Uniform 3-Level Hypothesis (U3LH)* of J a c k e n d o f f (1977). (Condition (ii) stipulating that every nonterminal has a unique (maximal) expansion in the paranthesis form is only implicit in the *U3LH*. As it will never play a crucial role in our proofs, but sometimes makes their presentation easier, we included it in the definition merely for the sake of convenience). A language that can be generated by a (strong) $X^{(n)}$ grammar will be called a (strong) *X-n-bar language* Our first two lemmas will show that the 'n' can be suppressed.

Lemma 1.1. *Given a (strong) $X^{(n)}$ language, there exists a (strong) $X^{(n+1)}$ grammar generating it.*

Proof. For any (strong) $X^{(n)}$ grammar $G = (V_N, V_T, P, v^{(n)})$, we can construct an equivalent (strong) $X^{(n+1)}$ grammar $\bar{G} = (V_N', V_T, \bar{P} \cup P_O, v^{(n+1)})$ by taking $V_N' = V_N \cup \{x^{(n+1)} | x \in V_T\}$ $\bar{P} = \{\bar{\alpha} \to \bar{\beta} \mid \alpha \to \beta \in P\}$ *and* $P_0 = \{\bar{x} \to x | x \in V_T\}$.
$\bar{G}$ is obviously a (strong) $X^{(n+1)}$ grammar, and for every α generated by G $\bar{\alpha}$ can be derived from $v^{(n+1)}$ by using the rules in $\bar{P}$. This derivation can be finished by applying the appropriate rules in P_0, therefore the language generated by G is included in the language generated by $\bar{G}$. Conversely, given a derivation of some word α in $\bar{G}$, the terminals in α can result only from rules in P_O. By omitting these rules fromthe derivation, $\mu(\alpha)$ can be derived from $y^{(n+1)}$ in $\bar{G}$ utilizing only rules in $\bar{P}$, therefore α can be derived from $v^{(n)}$ according to P, i.e. the language generated by $\bar{G}$ is included the language generated by G.

Lemma 1.2. *Given a (strong) $X^{(n)}$ language, there exists a (strong) $X^{(1)}$ grammar generating it.*

Proof. Suppose that the language in question is generated by a strong $X^{(n)}$ grammar $G = (V_N, V_T, P, v^{(n)})$ utilizing the following $n|V_T|$ paranthesis rules:

$$x^{(i)} \to (S^{(n)}_{i,1})(S^{(n)}_{i,2}) \ldots (S^{(n)}_{i,\ell_i} kx^{(i-1)} (C^{(n)}_{i,1})(C^{(n)}_{i,2}) \ldots (C^{(n)}_{i,r_i})$$

An equivalent strong $X^{(1)}$ grammar $G'=(V'_N, V_T, P', \bar{v})$ can be constructed the following way: $V'_N = \{\bar{x} | x \in V_T\}$ and P' contains the following paranthesis rule for $\bar{x}$:

$$x^{(1)} \to (\bar{S}_{n,1})(\bar{S}_{n,2}) \ldots (\bar{S}_{n,\ell_n})(\bar{S}_{n-1,1})(\bar{S}_{n-1,2}) \ldots (\bar{S}_{n-1,\ell_{n-1}}) \ldots$$

$$\ldots (\bar{S}_{1,1})(\bar{S}_{1,2}) \ldots (\bar{S}_{1,\ell_1} kx (\bar{C}_{1,2}) \ldots (\bar{C}_{1,r_1})(\bar{C}_{2,1}) \ldots$$

$$\ldots (\bar{C}_{n,1})(\bar{C}_{n,2}) \ldots (\bar{C}_{n,r_n}).$$

In effect, we substitute $x^{(n-1)}$ in the rule expanding $x^{(n)}$ by the right side of the rule expanding $x^{(n-1)}$ and therefore eliminate $x^{(n-1)}$ from the grammar. Then we eliminate $x^{(n-2)}$, and so on until only elements of V_M and V_T remain in the rules. As a final step, we rename every $x^{(max)}$ to $\bar{x}$. This process leaves us with a grammar equivalent to G, as can be seen from the fact that the corresponding transformations in the equation system associated with the grammar are simply substitutions and therefore leave the formal power series associated with the $x^{(n)}$-s unchanged (and the last step of renaming is obviously an isomorphy).

Notice, that the process of eliminating intermediate $x^{(i)}$-s from the grammar does not hinge on the uniqueness clause in Definition 1.2 or on the paranthesis form. In general this process can be carried through by substituting iteratively the nonmaximal $x^{(i)}$-s by all their possible expansions in every rule that contains them, and 'renaming' will turn the result into an (equivalent) $X^{(1)}$ grammar.

Remark 1.3. Although the above proof depends crucially on the possibility of eliminating intermediate nonterminals i.e. on maximality, it can be easily seen (cf. Remark 1.1) that it is independent

of uniformity.

In what follows, the family of (strong) $\bar{X}$ languages will be denoted by K (E).

The last lemma in this section describes the (strong) $\bar{X}$ languages over a one-letter alphabet $\{v\}$. The *length set* F associated with such a language L is simply $\{n \mid v^n \in L\}$. $\mathbb{N}$ denotes the set of natural numbers (including zero). Addition and multiplication of subsets of $\mathbb{N}$ is defined in the usual manner, in particular $kA \stackrel{d}{=} \{k\}A$ $(=\{ka \mid a \in A\})$ and $k+A \stackrel{d}{=} \{k\}+A$.

Lemma 1.3. *The length set F of a (nonempty) $\bar{X}$ language L has the form*

$$F = 1 + k_1\mathbb{N} + \ldots + k_s\mathbb{N} \tag{1}$$

where the k_i-s can be determined from the productions of the grammar generating L. Conversely, every set F satisfying (1) for some $s, k_1, \ldots, k_s$ (≥ 0) is he length set of some $\bar{X}$ language (over one letter).

Proof. By force of Lemma 1.2, any $\bar{X}$ language over $\{v\}$ can be generated by some grammar $G = (\{\bar{v}\}, \{v\}, P, \bar{v})$. If P contains, say, $s+1$ productions, these have the form $\bar{v} \to \bar{v}^{\ell_i} v \bar{v}^{r_i}$ $(0 \leq i \leq s)$. Since the language in question is not empty, there has to be a production in P which has fewer nonterminals on its [illegible] side than on its left, therefore we can suppose $\ell_0 = r_0 = 0$, i.e. existence of a trivial production $\bar{v} \to v$. We claim that (1) holds if we take k_i to be $\ell_i + r_i$ (for $1 \leq i \leq s$). Given some $v^n \in L$, the last sentential form in its derivation has to be $v^a \bar{v} v^b$, $a+b = n-1$. Since $\bar{v}$ in this string can be rewritten by an arbitrary rule in P, $v^a \bar{v}^{\ell_i} v \bar{v}^{r_i} v^b$ is also a sentential form of G and (by multiple application of the trivial rule) $v^a v^{\ell_i} v v^{r_i} v^b$ is also in L. This way, F satisfies $F - 1 + k_i + 1 \subset F$, $(1 \leq i \leq s)$, and (by virtue of the trivial rule) $1 \in F$.

Moreover, F is the *minimal* set containing 1 and closed under the operations of adding k_i $(1 \leq i \leq s)$, and such a set obviously has the form prescribed by (1).

Constructing a grammar for arbitrary $s, k_1, \ldots, k_s$ (≥ 0) is now a trivial matter: the above proof makes it clear that the rules $\bar{v} \to v$, $\bar{v} \to v\bar{v}^{k_i}$ $(1 \leq i \leq s)$ will have just the desired effect.

Corollary 1.1. *v and $v^{\pm}$ are the only strong $\bar{X}$ languages over $\{v\}$.*

Proof. A strong $X^{(1)}$ grammar G over $\{v\}$ possessing only the trivial $\bar{v} \to v$ rule will generate v, and nothing else. If G contains any nontrivial rule, than, by virtue of optionality, it also contains some rule with $k_i = 1$, and as $1\mathbb{N}$ equals $\mathbb{N}$, the generated language will have the length set $1 + \mathbb{N}$, i.e. it will be v^+.

Corollary 1.2. *v is the only finite language that can be generated by a (strong) $\bar{X}$ grammar over $\{v\}$.*

Proof. We have already seen that v can be generated by such a grammar, and no other finite language has a linear length set.

2. GEOGRAPHY AND CLOSURE UNDER OPERATIONS

First we investigate the position of the families E and K relative to the Chomsky hierarchy.

Therom 2.1. *E is properly contained in K and K is properly contained in L_2. Neither E nor K can be compared to L_3.*

Proof. It follows from the definitions that $E \subset K$ and $K \subset L_2$. Corollary 1.1 gives that $E \neq K$ (even over a one-letter alphabet). Since finite languages are regular, Corollary 1.2 gives $L_3 \not\subset E, K$, and this in turn shows that K can not equal L_2. What remains to be seen is that neither E nor K is contained in L_3.
The language generated by the strong $\bar{X}$ grammar G_0 that has the productions $\bar{v} \to v(\bar{w}), \bar{w} \to (\bar{v})w$ is obviously

(2) $$L_0 = \{v^i w^{i-1} \mid i > 0\} \cup \{v^i w^i \quad i \geq 1\}.$$

Since $v^r w^r \in L_0$ but $v^r w^{r+k} \notin L_0$, w^r and w^{r+k} are incongruent for any $r,k \geq 0$. Therefore, all the w^i-s belong in different classes of the (syntactic) congruence induced by L_0 and, by virtue of Kleene's theorem, L_0 is not regular. This way $E \not\subset L_3$ (and, a fortiori, $K \not\subset L_3$) and we conclude the proof of incomparability by pointing out that (as a corollary to the fact that CFLs over a one-letter alphabet are regular) neither $K \cap L_3$ nor $E \cap L_3$ is empty.

Since the nonterminal alphabet of an $X^{(n)}$ grammar can be inferred from the terminal alphabet, $\bar{X}$ grammars can be taken as triples much like Lindenmayer systems. This gives rise to the suspicion that they may fit into the hierarchy of L-systems better than into the Chomsky-hierarchy. Unfortunately this is not the case, as can be seen from

Theorem 2.2. *Neither* E *nor* K *can be compared to* $L(OL)$ *or* $L(TOL)$.

Proof. Given the facts $E \subset K \subset L_2 \not\subset L_{(OL)} \subset L_{(TOL)}$, it sufficies to prove $E \not\subset L_{(TOL)}$, and the language L_0 of the above theorem can serve as example here as well. We can suppose indirectly that L_0 is generated by some $TOLS$ $T=(V,P,\omega)$ with $V=\{v,w\}$. No rule rewriting v can have the form $v \to \alpha\, vw\, \beta$ in any table, since such a rule (coupled with any rule $w \to \gamma$) would derive $\alpha vw\beta\alpha vw\beta\gamma\gamma \notin L_0$ from $v^2w^2 \in L_0$. By the same token, no rule can have the form $v \to \alpha w v \beta$ (since w never preceeds v in strings of L_0), therefore any rule rewriting v can only have the form $v \to w^a$ *or* $v \to v^b$. The former possibility is trivially excluded, and the same reasoning as above will show that w rules can only have the form $w \to w^a$. This way, any table of P can contain only the following rules:

$$v \to v^{a_1} | v^{a_2} | \ldots | v^{a_\ell}$$

$$w \to w^{b_1} | w^{b_2} | \ldots | w^{b_r} \qquad (\ell, r \geq 1).$$

string $v^k w^k$ $(k > 1)$ will give us $v^{ka_i} w^{kb_j} \in L_0$ by hypothesis, and this implies $ka_i = kb_j$ or $ka_i + 1 = kb_j$. Since inthe latter the right side is divisible by k but the left side is not, only the former equation can hold, therefore $a_i = b_j$ for arbitrary $1 \leq i \leq \ell$, $1 \leq j \leq r$, and the table in question can only have the form

$$v \to v^a \qquad w \to w^a \qquad (a \geq 0).$$

Applying this table to $v^{k+1} w^k$ $(k > 1)$ will gives us $v^{ak+a} w^{ak}$ which has to belong to L_0, therefore $a=1$, or $a=0$. This way, the only tables possible are

$$v \to v \quad w \to w \quad \text{and} \quad v \to \lambda \quad w \to \lambda.$$

The second table is excluded because $\lambda \notin L_0$, therefore the TOL system in question generates nothing but its axiom, a contradiction.

In spite of this result, L-systems and $\bar{X}$ grammars are rather similar, and the properties characterizing $\bar{X}$ grammars receive very palpable interpretations if applied to L-systems.

Exmaple 2.1. If in the (deterministic) OL system $D = (\{\bar{\bar{r}}, \bar{r}, r\}, \bar{\bar{r}}, \{\bar{\bar{r}} \to \bar{r}\,\bar{r}, \bar{r} \to \bar{\bar{r}}\,\bar{r}, r \to \lambda\})$ the number of bars denotes life expectancy and the r-s are taken as (couples of) rabbits, the classical problem of Fibonacci (1202, see e.g. Archibald, 1918) can be analyzed quite naturally. Let us denote the number of old, middle aged and young rabbits after the n-th mating season (rewriting) by a_n, b_n, and c_n respectively. Obviously $a_{n+1} = b_n$, $b_{n+1} = c_n$ and $c_{n+1} = b_n + c_n$. Therefore, the string (population) after the n-th rule application (mating season) contains f_{n+1} $\bar{\bar{r}}$-s, f_n $\bar{r}$-s, and f_{n-1} r-s, where the f_i-s are the Fibonacci numbers.

In sum lexicality means that bars can serve as timing units in L-systems: succession is simply the passing of time, maximality expresses the fact that each cell comes into being with its own prescribed life expectancy, and uniformity means that this is the

same for every kind of cell. Centrality means that we start with one cell (or pair of animals). Optional rules can be interpreted as 'virility quotas': any cell of a given kind (at a given age) can produce only a certain number of new cells (of various kinds) under optimal circumstances, bug may produce fewer than the maximum expressed by the right side of the paranthesis rule in question.

The most striking indication of the similarity between $\bar{X}$ grammars and L-systems however, is the fact that both manifest the same resistance to operations: the rest of this section is devoted to this topic.

Lemma 2.1. *The families* K *and* E *are not closed under union.*

Proof. This result is trivial if we allow for start symbols with different heads: $\{v,w\}$ cannot be an $\bar{X}$ language since every word in those contains the head of the start symbol at least once. The following example, however, makes it clear that nonclosure under union does not hinge on this property: all the $\bar{X}$ languages here (and in what follows) will be centred around v.

The (strong) $\bar{X}$ grammars given by the production sets $\bar{v} \to \bar{v}(\bar{w})$, $\bar{w} \to w(\bar{w})$ *and* $\bar{v} \to v(\bar{w})$, $\bar{w} \to w(\bar{v})$ generate the languages $L_1 = v(w)^*$ and $L_2 = (vw)^+ \cup v(wv)^*$ respectively. Suppose indirectly that $L_1 \cup L_2$ is an $\bar{X}$ language: than by Lemma 1.2. it can be generated by some $X^{(1)}$ grammar $G = (V_N, V_T, P, S)$. Obviously V_T can be only be $\{v,w\}$ and S has to be $\bar{v}$ because $v \in L_1 \cup L_2$. This means that $\bar{v} \to v \in P$, and $vw \in L_1 \cup L_2$ necessitates $\bar{v} \to v\bar{w} \in P$ and $\bar{w} \to w \in P$. The Parikh-image of L_1 is $\{(1,k) \mid k \in \mathbb{N}\}$ and this, by virtue of Parikh's theorem, is possible only in case $w^a \bar{w} w^b$ $(a+b>0)$ can be derived from $\bar{w}$ according to G. Since every $\bar{X}$ rule introduces precisely one terminal, the derivation of $vwv (\in L_1)$ according to P has to take three steps, the first of which can introduce at most two nonterminals (and at least one). In the former case, either $\bar{v} \to v\bar{w}\bar{v}$ *or* $\bar{v} \to \bar{v}\bar{w}v$ is in P, and these rules would enable us to derive vw^{a+b+1} according to P, a contradiction. This way, the first step in the derivation can introduce at most (and obviously

as least) one nonterminal, therefore it has to be $\bar{v} \to v\bar{w}$, $\bar{v} \to \bar{w}v$. $\bar{v} \to v\bar{v}$ or $\bar{v} \to \bar{v}v$. The last two of these would enable us to generate v^+, and $\bar{v} \to \bar{w}v$ would lead to $wv \notin L_1 \cup L_2$, therefore the first step in the derivation of vwv has to be $\bar{v} \to v\bar{w}$. This way, the second step can only be $\bar{w} \to w\bar{v}$ which therefore has to belong in P. Since $w^a\,\bar{w}w^b$ can be derived from $\bar{w}(a+b>0)$, the same holds for $w^{2a}\,\bar{w}w^{2b}$, and (with the aid of our last rule) also for $w^{2a+1}vw^{2b}$. Since we already know that $\bar{v} \to v\bar{w}$ is a rule in P, $vw^{2a+1}vw^{2b}$ can be derived from $\bar{v}$ according to $P(a+b>0)$, a contradiction.

Lemma 2.2. *The families* E *and* K *are not closed under intersection and complementation.*

Proof. Nonclosure under complementation is a trivial corollary to Lemma 1.3, and nonclosure under intersection will be shown the aid of the (strong) $X^{(1)}$ language L_3 generated by the grammar G_3 that has the productions $\bar{v} \to (\bar{v})v(\bar{w})$ and $\bar{w} \to w$. In order to establish the intersection of L_3 with the language L_0 of Theorem 2.1, first we eliminate the production $\bar{w} \to w$ and break up the paranthesis form the following way: $\pi_1 = \bar{v} \to \bar{v}vw$; $\pi_2 = \bar{v} \to \bar{v}v$; $\pi_3 = \bar{v} \to vw$; $\pi_4 = \bar{v} \to v$. The reader can verify for himself that the grammar $G_3 = (\{\bar{v}\}, \{v,w\}, \{\pi_1, \pi_2, \pi_3, \pi_4\}, \bar{v})$ is indeed equivalent to G_3. Since all the strings in L_0 (bar v) end in the symbol w, no derivation according to G_3' that starts with π_2 can result in a string belonging in L_0. Derivations starting with π_1 can lead to words in L_0 only if the next step is π_4, and in this case the result is v^2w. Derivations starting with π_3 or π_4 cannot be continued, but the result is in L_0, therefore $L_0 \cap L_3 = \{v, vw, v^2w\}$. Again, if we suppose indirectly that this is an $\bar{X}$ language, the grammar G generating it can be taken as one-bar, and the same reasoning as above will show that the start symbol has to be $\bar{v}$, G has to possess the trivial productions $\bar{v} \to v$ and $\bar{w} \to w$, the production $\bar{v} \to v,\bar{w}$ and the derivation of v^2w has to take three steps. We cannot introduce two nonterminals in the first step, since

both $\bar{v} \to \bar{v}v\bar{w}$ and $\bar{v} \to v\bar{v}\bar{w}$ can be iterated to give words not in $L_0 \cap L_3$, so the first step has to be either $\bar{v} \to v\bar{v}$, $\bar{v} \to \bar{v}v$, or $\bar{v} \to v\bar{w}$. The first of these overgenerates, and the second is impossible because the 'barless' v on the right side blocks the introduction of any symbol to the right of it. This way, the first step in the derivation of v^2w has to be $\bar{v} \to v\bar{w}$, and the second can only be $\bar{w} \to \bar{v}w$. In this case, however, $\bar{v} \to v\bar{w} \to v\bar{v}w \to vv\bar{w}w \to vvww$ is a possible derivation according to G, i.e. $v^2w^2 \in L_0 \cap L_3$, a contradiction.

Lemma 2.3. *The families* K *and* E *are not closed under (Kleene) product and (λ-free) closure.*

Proof. Nonclosure under multiplication is a trivial consequence of Lemma 1.3 (or its Corollaries) and the effect of $^+$ will be investigated again on the language L_0 of Theorem 2.1. The reader can easily verify that

$$L_0^+ = \{v^i \mid i \geq 1\} \cup \{v^{a_1} w^{b_1} v^{a_2} w^{b_2} \ldots v^{a_s} w^{b_s} \mid s \geq 1,\ a_i \geq b_i \quad (1 \leq i \leq s)\}$$

Again, the usual indirect hypothesis will give us an $X^{(1)}$ grammar $G = (V_N, V_T, P, S)$ generating L_0^+. $\{V_T = v, w\}$; $S = \bar{v}$; $\bar{v} \to v \in P$; $\bar{v} \to v\bar{w} \in P$ and $\bar{w} \to w \in P$ follows immediately from $v \in L_0^+$ and $vw \in L_0^+$ respectively. Since $v^2 \in L_0^+$, either $\bar{v} \to v\bar{v}$ or $\bar{v} \to \bar{v}v$ is in P, and this time we search for a derivation (of length 4) of v^2w^2. In order to give the proof in a coincise form, we remark that (in the presence of trivial rules) a rule (or derivation) $\bar{v} \to \alpha\, \bar{v}\beta$ implies $v(\alpha)\, L_0^+ v(\beta) \subseteq L_0^+$, therefore the nonexistence of such a rule can be proved by any word $\gamma \in L_0^+$ for which $v(\alpha)\gamma v(\beta) \notin L_0^+$ holds. For instance, if the rule in question is $\bar{v} \to v\bar{v}\bar{w}^2$, $vL_0^+w^2 \subset L_0^+$ would follow, and this can be disproved with the aid of $\gamma = vwvw \in L_0^+$.

Apart from the above rule, the only production that can introduce three nonterminals at the beginning of the derivation v^2w^2 is $\bar{v} \to \bar{v}v\bar{w}^2$, and this leads to a contradiction as can be shown with $\gamma = vw$. If the first rule introduces two nonterminals, than it can be $\bar{v} \to v\bar{v}\bar{w}$, $\bar{v} \to \bar{v}v\bar{w}$, or $\bar{v} \to v\bar{w}\bar{w}$. The first possibility can be

excluded with the aid of $\gamma = vwvw$ L_0^+ and leftmost derivations starting with the other two would imply $\bar{v} \to v^2\overline{vw}^2$ or $\bar{v} \to v\overline{vw}^2$, and these can be excluded withthe aid of $\gamma = vwvw$ and vw respectively. This way, the first step in the derivation can introduce only one nonterminal, therefore it is $\bar{v} \to \bar{v}v$, $\bar{v} \to \bar{w}v$, $\bar{v} \to v\bar{v}$ or $\bar{v} \to v\bar{w}$. The first two obviously cannot lead to v^2w^2 and the third implies a contradiction, namely that $vw^2 \notin L_0^+$ can be derived from $\bar{v}$. Therefore, the first step has to be $\bar{v} \to v\bar{w}$ and the second step can introduce at most two nonterminals. If two are introduced $v\bar{v}w\bar{w}$ or $v\overline{vw}w$ can be derived from $\bar{v}$, and we have already seen that this leads to a contradiction. If only one nonterminal is introduced in the second step of the derivation, the $v\bar{w}w$, $v\bar{v}w$, $w\bar{v}$, or $vw\bar{w}$ can be derived from $\bar{v}$. The last two possibilities can be disregarded, since these can not lead to v^2w^2, and the second is already out. This means that $v\bar{w}w$ can be directly rewriten from $\bar{v}w$, i.e. $\bar{w} \to \bar{w}w$ is a rule in P. This rule, however, can apply iteratively after $\bar{v} \to v\bar{w}$ to generate $vw^2 \notin L_0^+$, a contradiction.

Corollary 2.1. K and E are not closed under substitution.

Proof. This follows from the fact that L_0^+ can taken as the result of substituting the "family" $\{L_0\}$ into the (strong) $\bar{X}$ language v^+.

Lemma 2.4. *The families* K *and* E *are not closed under union, intersection, product, and quotient with regular sets.*

Proof. Nonclosure under union, intersection and product with regular sets is a direct consequence of Corollary 1.2. The quotient of L_0 with the language $\{w\}$ is $L=\{v^i n^{i-1} \mid i \geq 1\} \cup \{v^{i+1} n^{i-1} \mid i \geq 1\}$, and as this contains v and v^2, the grammar generating it has to posses the rules $\bar{v} \to v$ and $\bar{v} \to \bar{v}v$ or $\bar{v} \to v\bar{v}$, but these rules would enables us to generate $v^3 \notin L$, a contradiction.

Lemma 2.5. *The families* E *and* K *are not closed under (λ-free) homomorphism or inverse homomorphism. They are not closed under (inverse) gsm mapping.*

Proof. The second sentence is a consequence of the first. The (λ-free) homomorphism $v \to v^2$ takes $\{v\}$ into $\{v^2\}$ and this is not an $\bar{X}$ language by Corollary 1.2. The inverse image of language L_2 of Lemma 2.1 under the (λ-free) homomorphism $v \to v$, $w \to vw$ contains both v and w, therefore it cannot be an $\bar{X}$ language.

The above lemmas make it clear that the usual operations on *strong* $\bar{X}$ languages lead out not only of the family E but also of K. In fact the only customary operation that does not lead out of these families is that of reversal (mirror image): this is due to the fact that the mirror image of the right side of an $\bar{X}$ rule has the form of $\bar{X}$ rules. We can sum up these results in the following

Theorem 2.2. *The families* K *and* E *are anti-AFLs closed under reversal.*

REFERENCES

[1] A r c h i b a l d , R.C., *Golden section*, Amer. Math. Monthly 25, 1918, 232-238.

[2] J a c k e n d o f f, R., $\bar{X}$ Syntax: A Study of Phrase Structure, *MIT Press*, Cambridge, Mass., 1977.

[3] S a l o m a a, A., Formal languages, *Academic Press*, New York, 1973.

A. KORNAI

Computer and Automation
Institute Hungarian Academy
of Sciences
Budapest,
1132 Victor Hugo u. 18-22.
Hungary