INFORMATIQUE THÉORIQUE ET APPLICATIONS

R. BOONYAVATANA
G. SLUTZKI

Ogden's lemma for nonterminal bounded languages

Informatique théorique et applications, tome 20, n^o 4 (1986), p. 457-471
<http://www.numdam.org/item?id=ITA_1986__20_4_457_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 20, n° 4, 1986, p. 457 à 471)

OGDEN'S LEMMA FOR NONTERMINAL BOUNDED LANGUAGES (*)

by R. BOONYAVATANA ([1]) and G. SLUTZKI ([1])

Communicated by Robert CORI

Abstract. – We present an Ogden-type pumping lemma for nonterminal bounded languages. It is shown that these Ogden-type conditions are stronger than the classical-type pumping conditions for nonterminal bounded languages. However, we show that they are not sufficient. In fact, we construct counterexamples at various levels of the Chomsky hierarchy, each of which satisfies the conditions of our Ogden-type lemma.

Résumé. – *Une grammaire est dite bornée pour les non terminaux si tout mot qui dérive de l'axiome contient un nombre de non terminaux borné par un entier K.* On démontre ici un lemme d'itération du type de celui d'Ogden pour les langages engendrés par ces grammaires; ce lemme améliore ceux déjà connus pour ces langages. Nous montrons toutefois qu'il ne constitue pas une condition suffisante en construisant des exemples de langages satisfaisant ce lemme, et pris dans chacune des classes de la hiérarchie de Chomsky.

1. INTRODUCTION

Like the pumping lemma for context-free languages, Ogden's lemma is useful to prove that a given language is not context-free. It is known that Ogden's lemma is more powerful than the pumping lemma for context-free languages [7, 9]. The additional power of Ogden's lemma seems to stem from the use of marked positions which helps to cut down the number of factorizations considered in the pumping lemma. Boasson and Horváth [3] showed that there exist non-context-free languages satisfying Ogden's lemma and hence, Ogden's conditions are not sufficient.

A pumping lemma for nonterminal bounded languages is proved in [4]. The pumping conditions as presented in [4] were shown not to be sufficient.

(*) Received March 1985, revised January 1986.
([1]) Department of Computer Science, Iowa State University. Ames, IA 50010, U.S.A..

In this paper, we prove Ogden-type lemma for nonterminal bounded languages which is a generalization of a similar lemma for linear context-free languages, see section 2.

The paper consists of four sections. The second section gives preliminary definitions and the proof of the linear Ogden's lemma. The third section contains our main result, the Ogden's lemma for nonterminal bounded languages together with examples of its application. In the fourth section we provide counterexamples to show that although Ogden's lemma is an improvement of the pumping lemma, it still does not provide sufficient conditions.

2. PRELIMINARIES

In this section basic definitions are introduced. We mainly follow [1, 7, 8] for our terminology and notation. We will discuss Ogden's lemma for linear context-free grammars at the end of this section with an eye towards a generalization to be discussed in subsequent sections.

A *context-free grammar* (*cfg*) is a construct $G=(N, T, P, S)$. N and T are two disjoint sets of *nonterminals and terminals* respectively; P is a finite set of *productions* each of the form $A \to \alpha$ with A in N and α in $(N \cup T)^*$; the *start symbol* S is in N.

V is used to denote $N \cup T$, the *vocabulary* of G. G_A, where A is a nonterminal of G, will be the grammar resulting from G by making A the start symbol; thus $G=G_S$. Let X be a subset of V and let w be in V^*. $\#_X(w)$ is the number of occurrences of symbols of X in w and $\#_V(w)$ is the length of w. The language generated by G is denoted by $L(G)$; language L is a *context-free language* (*cfl*) if it is generated by some cfg.

For a word in a language L we may regard some positions of the word as *distinguished*; we refer to them as *marked positions*. For z in T^*, $m(z)$ is the number of marked positions of z.

Let t be a derivation tree for some word in $L=L(G)$. We define a node n of t to be a *branch node* if n has at least 2 direct descendants both of which have marked descendants. Let q be a leaf of a particular root-to-leaf path π of t. Then a branch node n on π is a *left branch node* (relative to π) if a direct descendant of n not on the path π has a marked descendant to the left of q; otherwise, n is a *right branch node*. It should be noted that, contrary to the usual definition (cf. Ogden's original proof [9]) the notions of left and right branch nodes are not symmetrical i. e. here, a branch node cannot be simultaneously left and right.

A production $A \to \alpha$, in a cfg G, is *linear* if $\#_N(\alpha) \leqq 1$. A cfg is *linear* (*lcfg*) if all its productions are linear. A cfl is *linear* (*lcfl*) if it is generated by some lcfg.

We now define the rank functions. Let $G=(N, T, P, S)$ be a cfg and let α in V^* be a sentential form. If the set $\{ \#_N(\beta) \mid \alpha \overset{*}{\underset{G}{\Rightarrow}} \beta \}$ is finite then we let $\text{rank}(\alpha) = \max \{ \#_N(\beta) \mid \alpha \overset{*}{\underset{G}{\Rightarrow}} \beta \}$ otherwise $\text{rank}(\alpha)$ is undefined. A cfg G for which $\text{rank}(A)$ is defined for every nonterminal A is called *nonterminal bounded* (*ntbg*). The *rank of* G, $\text{rank}(G)$, is $\max[\text{rank}(A)]$ where A is in N. G is *k-nonterminal bounded* (*k-ntbg*) if $\text{rank}(G)=k$. L is k-nonterminal bounded (k-ntbl) if it is generated by some k-ntbg. Note that $\text{rank}(w)=0$ for w in T^* and for $\alpha=\alpha_1 \alpha_2 \ldots \alpha_k$ in V^*, $\text{rank}(\alpha) = \sum_{i=1}^{k} \text{rank}(\alpha_i)$; moreover, every ntbg has nonterminals of rank 1.

We will use p to denote the maximum number of occurrences of terminals in the productions of a grammar, i. e. $p = \max \{ \#_T(\alpha) \mid A \to \alpha \text{ is a production in } G \}$. ε will denote the empty word. Unit productions are productions of the form $A \to B$ where A, B are nonterminals and an ε-production is a production of the form $A \to \varepsilon$.

Throughout this paper, capital letters will be used for families of languages. Thus CFL and LIN are respectively the class of all cfl's and lcfl's; k-NTBL is the class of all k-ntbl's. In the rest of this section we will present Ogden's lemma for linear languages. It will provide a good background for the generalization which will be discussed in subsequent sections.

LEMMA 2.1 (Ogden's lemma for LIN): *If L is a lcfl then there is a constant n (depending only on L) such that if z is in L and $m(z) \geqq n$ then z can be written as $z = uvwxy$ such that:*

(1) *$m(uvxy) \leqq n$;*

(2) *either each of u, v, w or each of w, x, y contains a marked position;*

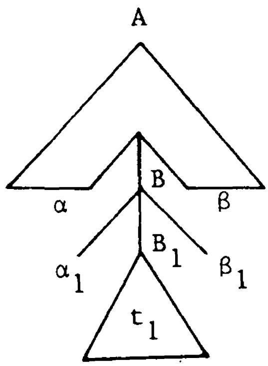
(3) *for every $i \geqq 0$, $uv^i wx^i y$ is in L.*

It is perhaps appropriate to point out here that Ogden's lemmas for CFL and LIN differ precisely in condition (1) of the above lemma where in the CFL-case the requirements is $m(vwx) \leqq n$ rather than $m(uvxy) \leqq n$. We should also mention that in the classical pumping lemma for linear languages condition (1) reduces to $|uvxy| \leqq n$ ([2], proposition 6.6).

We will first prove a claim which relates the number of marked positions of a word with its derivation tree in a grammar.

CLAIM 2.2: *Let G be a lcfg without ε or unit productions. Let t be the derivation tree for the derivation $A \overset{*}{\underset{G}{\Rightarrow}} z$, z in T^*, and C a root-to-leaf path in t with maximum number of branch nodes. If C has $\leqq b$ branch nodes, then $m(\text{yield}(t)) \leqq bp+1$.*

Proof of claim 2.2: By induction on b. For $b=0$, it is clear by the definition of branch node that m (yield$(t)) \leqq 1$. Now let t be as described in the claim (with $b \geqq 1$). Let B be the label of the first branch node on the path C. Since G is a lcfg, the first step in the derivation from B is $B \Rightarrow \alpha_1 B_1 \beta_1$ where α_1, β_1 are in T^* and hence t looks like:



The part of the path C which is in t_1 has $\leqq b-1$ branch nodes and $m(\alpha)=m(\beta)=0$. By induction hypothesis, m (yield$(t_1)) \leqq (b-1)p+1$ and so

$$m(\text{yield}(t)) = m(\alpha_1) + m(\beta_1) + m(\text{yield}(t_1)) \leqq p + (b-1)p + 1 = bp + 1.$$

We have proved the claim. $\square$

Proof of lemma 2.1: Without loss of generality we may assume that ε is not in L and $L=L(G)$ where $G=(N, T, P, S)$ has no ε or unit productions. Put $k=|N|$ (cardinality of N), and $n=2(k+1)p+2$. Let z be in L with

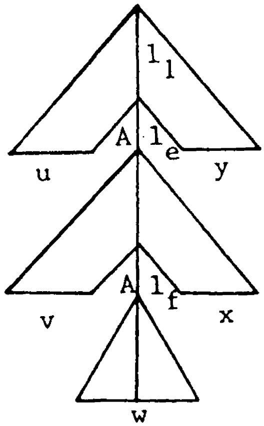


Figure 1.

$m(z) \geqq n$ and let t be the derivation tree for z in G; let C be the root-to-leaf path in t with maximum number of branch nodes.

By claim 2.2, C has at least $2k+3$ branch nodes. Let $b_1, b_2, \ldots, b_{2k+3}$ be the first $2k+3$ branch nodes in the path C. We may assume that at least $k+2$ of $b_1, \ldots, b_{2k+3}$ are left branch nodes. The other case can be treated analogously. Let $l_1, \ldots, l_{k+2}$ be the first $k+2$ left branch nodes in the sequence $b_1, \ldots, b_{2k+3}$. Since there are k nonterminals we can find two nodes among $l_2, \ldots, l_{k+2}$, say l_e and l_f such that (1) l_e and l_f are labeled by the same nonterminal, say A, and (2) l_e is an ancestor of l_f. This situation is shown in figure 1.

Since l_f is an ancestor of b_{2k+3}, the path along C from the root of t down to but excluding l_f has at most $2(k+1)$ branch nodes. By claim 2.2, $m(uvxy) \leqq 2(k+1)p+1 < n$. Since l_1, l_e and l_f are left branch nodes, each of u, v and w has at least one marked position. Hence, condition (2) of lemma 2.1 is satisfied. Finally, we have $S \overset{*}{\underset{G}{\Rightarrow}} uAy$, $A \overset{*}{\underset{G}{\Rightarrow}} vAx$ and $A \overset{*}{\underset{G}{\Rightarrow}} w$. Therefore, $S \overset{*}{\underset{G}{\Rightarrow}} uv^i wx^i y$ for all $i \geqq 0$. The proof of the lemma is now complete. □

3. NONTERMINAL BOUNDED LANGUAGES

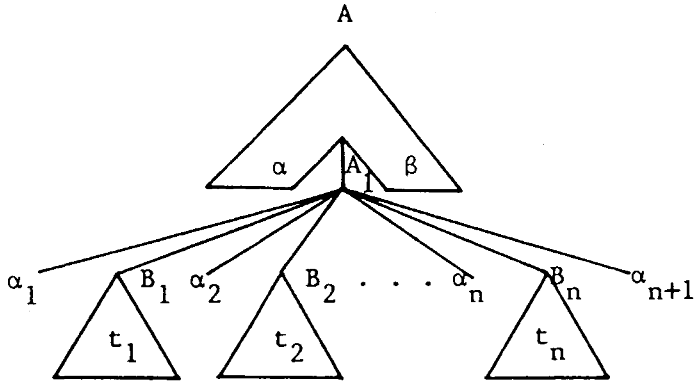
Ogden's lemma for nonterminal bounded languages which is a generalization of lemma 2.1 from previous section will be proved in this section. We will first prove an auxiliary claim analogous to claim 2.2 of section 2, then present the proof of the main lemma.

CLAIM 3.1: *Let G be a ntbg and t be the derivation tree for a derivation* $A \overset{*}{\underset{G}{\Rightarrow}} z$, *z in* T^*, *where* $rank(A)=r$. *If C, a root-to-leaf path in t with maximum number of branch nodes, has at most b branch nodes, then* $m(z) \leqq (rb-r+1)p+r$.

Proof: By induction on r. For $r=1$ we have the linear case which has been treated in claim 2.2. Now let $r \geqq 2$ and suppose the claim holds for nonterminals of rank $<r$. Let A, G, C, z, t and b be as stated. Starting from the root of t, let n_1 (labeled by A_1) be the first branch node on the path C. The first step in the derivation from A_1 is

$$A_1 \quad \Rightarrow \quad \alpha_1 B_1 \alpha_2 B_2 \dots \alpha_n B_n \alpha_{n+1} \qquad (n \geqq 1)$$

and we have the following situation:



where for all i, α_i is in T^*,

$$\operatorname{rank}(B_i)=r_i \qquad , \qquad \sum_{j=1}^{n} r_j \leqq \operatorname{rank}(A_1) \leqq \operatorname{rank}(A)=r,$$

t_i has at most $b-1$ branch nodes and $z=\alpha\alpha_1 \operatorname{yield}(t_1) \dots \operatorname{yield}(t_n)\alpha_{n+1}\beta$. Note that α and β have no marked positions. We want to show that $m(z) \leqq (rb-r+1)p+r$. There are two cases to be considered.

Case 1: $n>1$. We have $r_i<r$ $(i=1, \ldots, n)$. By induction hypothesis $m(\text{yield}(t_i)) \leqq (r_i(b-1)-r_i+1)p+r_i$. Thus:

$$m(z)=\sum_{i=1}^{n+1} m(\alpha_i)+\sum_{i=1}^{n} m(\text{yield}(t_i)) \leqq p+\sum_{i=1}^{n}(r_i(b-1)-r_i+1)p+\sum_{i=1}^{n} r_i$$

$$\leqq p+\sum_{i=1}^{n} r_i(b-1)p+r \leqq p+r(b-1)p+r=(rb-r+1)p+r.$$

Case 2: $n=1$. We argue inductively on b. If $b=1$, A_1 is the only branch node on C, $m(\text{yield}(t_1)) \leqq 1$ and so $m(z) \leqq p+1 \leqq (rb-r+1)p+r$. Now assume that $b \geqq 2$ and the claim holds for C with less than b branch nodes. By induction hypothesis,

$$m(\text{yield}(t_1)) \leqq (r_1(b-1)-r_1+1)p+r_1.$$

Thus,

$$m(z) \leqq p+m(\text{yield}(t_1)) \leqq (r_1 b-r_1+1)p+r_1 \leqq (rb-r+1)p+r.$$

This completes the proof of the claim. □

THEOREM 3.2 (Ogden's lemma for NTBL): *If L is an $r-$ntbl and generated by an $r-$ntbg $G=(N, T, P, S)$ then there exists a constant n (depending on L) such that if z is in L and $m(z) \geqq n$ then z can be written as $z=z_1 z_2 \ldots z_s$, $1 \leqq s \leqq r$, where each z_i can be written as $z_i=u_i v_i w_i x_i y_i$ such that:*

(1) $\sum_{i=1}^{s} m(u_i v_i x_i y_i) \leqq n$;

(2) *either each of u_i, v_i, w_i or each of w_i, x_i, y_i contains a marked position;*

(3) *for all natural numbers $a_i \geqq 0$ $(1 \leqq i \leqq s)$ $z_1^{(a_1)} z_2^{(a_2)} \ldots z_s^{(a_s)}$ is in L where $z_i^{(b)}=u_i v_i^b w_i x_i^b y_i$.*

Proof: By induction on r. For $r=1$, L is lcfl and we have proved the result in lemma 2.1. Let L be r-ntbl and $G=(N, T, P, S)$ an r-ntbg such that

$$L=L(G), \quad r \geqq 2, \qquad k=|N|=\sum_{i=1}^{r} k_i$$

where k_i is the number of nonterminals of rank i in G. Put $n=(2rk+r+1)p+r+1$. Consider a derivation tree t for z in $L(G)$ such that $m(z) \geqq n$. Let C be root-to-leaf path in t with maximum number of branch nodes. By claim 3.1, the path C has at least $2k+3$ branch nodes on it. Let

$b_1, b_2, \ldots, b_{2k+3}$ be the first (topmost) $2k+3$ branch nodes on C. Among these there are at least $k+2$ left branch nodes or at least $k+2$ right branch nodes. Since there are k nonterminals, there exist branch nodes b_f, b_i and b_j, $1 \leqq f < i < j$, each of which is to be as high as possible in C and such that:

(1) b_f, b_i and b_j are of the same type, i. e. either all are left or all are right branch nodes (note that all branch nodes above b_f must be of type different from b_f);

(2) b_i and b_j are labeled by the same nonterminal, say A;

(3) there is at most one nonterminal $B \neq A$, for which there are four (or three) proper ancestor branch nodes of b_j labeled by B (two of each type), and;

(4) for each nonterminal E distinct from A and the nonterminal B of (3), there can be at most two ancestor branch nodes of b_j labeled by E (one of each type).

It is left to the reader to convince himself that such branch nodes can in fact be found. We will consider two cases depending on the form of the derivation tree from the root to b_j.

Case 1: The derivation from S (the root) to $A(b_j)$ uses only linear productions. Since b_j is an ancestor of b_{2k+3}, the path along C from S down to $A(b_j)$ excluding b_j contains at most $2(k+1)$ branch nodes. By lemma 2.1, we can write $z = uvwxy$ such that (1) $m(uvxy) \leqq 2(k+1)p+2 \leqq n$, (2) either each of u, v and w or each of w, x and y contains a marked position, and (3) for every $i \geqq 0$, $uv^i wx^i y$ is in L. Thus, the theorem is obtained with $s=1$.

Case 2: The nodes b_i, b_j occur after a nonlinear production:

$$B \to \alpha_1 B_1 \alpha_2 B_2 \ldots \alpha_a B_a \alpha_{a+1}, \qquad 2 \leqq a \leqq r',$$

where $\operatorname{rank}(B) = r' \leqq r$, $\operatorname{rank}(B_i) = r_i$, α_i in T^* and $\sum_{i=1}^{a} r_i \leqq r'$.

This is illustrated in figure 2. The case where B is at the root of the tree (i. e., $S = B$) is easy and left to the reader; in what follows we assume $S \neq B$. From the way b_i and b_j are chosen it follows that the number of branch nodes on the path from S to B is $\leqq 2 \sum_{i=r'}^{r} k_i + 2$.

By linearity of the derivation from S to S' each branch node on the path from S to S' can contribute at most p marked positions. Thus, we have

$$m(\gamma_1 \gamma_2) + \sum_{i=1}^{a+1} m(\alpha_i) \leqq 2 \sum_{i=r'}^{r} k_i p + 2p.$$

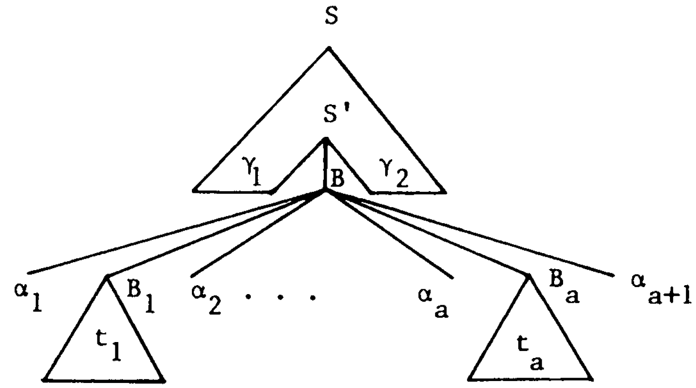


Figure 2.

Since $r \geqq 2$ and $\sum_{i=r'}^{r} k_i \geqq 2$ it is easy to check that

$$2 \sum_{i=r'}^{r} k_i p + 2p \leqq 2 \sum_{i=r'}^{r} k_i rp - 2rp + 2p.$$

We now consider $m(\text{yield}(t_i))$ for each subtree t_i. Put

$$n_i = (2r_i \sum_{j=1}^{r_i} k_j + r_i + 1)p + r_i + 1 \qquad \text{for} \quad 1 \leqq i \leqq a.$$

CLAIM. *There exists* $1 \leqq i \leqq a$ *such that* $m(\textit{yield}(t_i)) \geqq n_i$.

Proof of Claim: Suppose that $m(\text{yield}(t_i)) < n_i$ for all i. Then

$$m(z) = m(\gamma_1 \gamma_2) + \sum_{i=1}^{a+1} m(\alpha_i) + \sum_{i=1}^{a} m(\text{yield}(t_i))$$

$$< m(\gamma_1 \gamma_2) + \sum_{i=1}^{a+1} m(\alpha_i) + \sum_{i=1}^{a} n_i$$

$$\leqq 2 \sum_{i=r'}^{r} k_i rp - 2rp + 2p + \sum_{i=1}^{a} [(2r_i \sum_{j=1}^{r_i} k_j + r_i + 1)p + r_i + 1]$$

$$\leqq 2 \sum_{i=r'}^{r} k_i rp + 2p - rp + 2p \sum_{i=1}^{a} r_i \sum_{j=1}^{r'-1} k_j + ap + r + a.$$

$$\leqslant 2 \sum_{i=r'}^{r} k_i rp + 2p + 2 \sum_{i=1}^{r'-1} k_i rp + 2r$$

$$\leqslant 2krp + p + rp + r + 1 = n.$$

This contradicts the assumption that $m(z) \geqq n$. Hence, the claim is proved.

Let $J = \{ j \mid m(\text{yield}(t_j)) \geqq n_j \}$; by the above claim, $1 \leqq |J| \leqq r$. By induction hypothesis, for each j in J, we can write $\text{yield}(t_j) = z_{j1} z_{j2} \ldots z_{js_j}$ where $1 \leqq s_j \leqq r_j$ and each z_{ji} can be written as $z_{ji} = u_{ji} w_{ji} x_{ji} y_{ji}$ with

(1) $\sum_{i=1}^{s_j} m(u_{ji} v_{ji} x_{ji} y_{ji}) \leqq n_j$

(2) either each of u_{ji}, v_{ji} and w_{ji} or each of w_{ji}, x_{ji} and y_{ji} contains a marked position for all $1 \leqq i \leqq s_j$;

(3) $z_{j1}^{(a_1)} z_{j2}^{(a_2)} \ldots z_{js_j}^{(a_{s_j})}$ is in $L(G_{B_j})$.

Now we will write z (*fig.* 2) in the form $z = z_1 z_2 \ldots z_s$, $1 \leqq s \leqq r$ where z_i's satisfy the conditions stated in the theorem. We will argue in three cases:

Case 1: The first z_1 is defined to be

$$u_1 v_1 w_1 x_1 y_1 \quad \text{where} \quad u_1 = \gamma_1 \alpha_1 \text{yield}(t_1) \ldots \alpha_h u_{h1};$$
$$v_1 = v_{h1};\ w_1 = w_{h1};\ x_1 = x_{h1};\ y_1 = y_{h1} \quad \text{where} \quad h = \min J.$$

Case 2: The last z_s is defined to be $u_s v_s w_s x_s y_s$ where $u_s = u_{ls_l}$; $v_s = v_{ls_l}$; $w_s = w_{ls_l}$; $x_s = x_{ls_l} y_s = y_{ls_l} \alpha_{l+1}$ yield $(t_{l+1}) \ldots \alpha_{a+1} \gamma_2$ where $1 = \max J$.

Case 3: For the interior z_g, $1 < g < s$, there are two possible cases. The first is when $z_g = z_{ji}$ for j in J and $1 < i < s_j$. For the second case, let h and q be consecutive elements of J, i. e. $h < q$ and for $h < i < q$, i is not in J. We then define

$$z_g = u_g v_g w_g x_g y_g \quad \text{where} \quad u_g = u_{hs_h};$$
$$v_g = v_{hs_h}; \qquad w_g = w_{hs_h}; \qquad x_g = x_{hs_h}; \qquad y_g = y_{hs_h} \alpha_{h+1} \text{yield}(t_{h+1}) \ldots \alpha_q.$$

We have obtained $z = z_1 z_2 \ldots z_s (s \geqq 1)$. Moreover,

$$s = \sum_{j \in J} s_j \leqq \sum_{j \in J} r_j \leqq \sum_{j=1}^{a} r_j \leqq r.$$

Thus $1 \leqq s \leqq r$. It remains to show that all the conditions of the theorem hold. By induction hypothesis and the definition of J we have:

$$\sum_{i=1}^{s} m(u_i v_i x_i y_i) = m(\gamma_1 \gamma_2) + \sum_{i=1}^{a+1} m(\alpha_i) + \sum_{j \notin J} m(\text{yield}(t_j)) + \sum_{j \in J} \sum_{i=1}^{s_j} m(u_{ji} v_{ji} x_{ji} y_{ji})$$

$$< m(\gamma_1 \gamma_2) + \sum_{i=1}^{a+1} m(\alpha_i) + \sum_{j \notin J} n_j + \sum_{j \in J} n_j$$

$$= m(\gamma_1 \gamma_2) + \sum_{i=1}^{a+1} m(\alpha_i) + \sum_{j=1}^{a} n_j.$$

But this sum is $<n$ by the proof of the claim shown before. Thus, condition (1) is satisfied. By construction of the z_i's conditions (2) and (3) are obtained immediately. Hence the proof of theorem 3.2 is complete. □

We will now give two examples in which theorem 3.2 is applied. In particular we show that the Ogden's conditions (lemma 2.1 and theorem 3.2) are stronger than pumping conditions. The pumping conditions are obtained from the Ogden's conditions by replacing any mention of the mapping m (and marked positions) by the mapping $|\,.\,|$ (and the concept of length).

Example 1: Let $\Sigma = \{a_1, b_1, c_1\}$. We will use lemma 2.1 to show that

$$L = b\Sigma^* \cup \{a^n b a_1^k b_1^{k+m} c_1^m c^n \mid k, m, n \geqq 1\}$$

is not lcfl but L satisfies a pumping lemma for lcfl [2, 4]. We will first show that L satisfies the linear pumping conditions with constant 7. Let z be in L, $|z| \geqq 7$. Then either

$$z = a^i b a_1^j b_1^{j+k} c_1^k c^i \quad \text{for some } i, j, k \geqq 1$$

or $z = bz_1$ with z_1 in Σ^+. In the second case we may factor $z = uvwxy$ where $u = b$, $v =$ the first symbol of z_1, $w =$ the rest of z and $x = y = \varepsilon$. In the first case let $u = y = \varepsilon$, $v =$ the first a of z, $x =$ the last c of z, $w =$ the rest of z. In both cases, it is easy to see that $|uvxy| \leqq 7$, $|vx| \geqq 1$ and $uv^i wx^i y$ is in L for all $i \geqq 0$. We will now show that L is not lcfl. Suppose L is lcfl and let n be the Ogden's lemma constant for L. Consider $z = a^n b a_1^n b_1^{2n} c_1^n c^n$ where all positions of $a_1^n b_1^{2n} c_1^n$ are marked. Clearly, $m(z) > n$ and thus, we can factor $z = uvwxy$ such that the three conditions of lemma 2.1 hold. It is obvious that v and x can contain only one type of letter and, to satisfy condition (2), at least one of these is within $a_1^n b_1^{2n} c_1^n$. Furthermore, because of condition (1), neither v nor x can contain occurrences of b_1 which implies immediately that pumping of z will yield strings out of L. Thus condition (3) of lemma 2.1 fails, showing that L is not lcfl.

Example 2: Let $\Sigma=\{a_1, b_1, a_2, b_2, \ldots, a_{r+1}, b_{r+1}\}$. We will use theorem 3.2 to show that

$$L=\{a^n b a_1^{n_1} b_1^{n_1} a_2^{n_2} b_2^{n_2} \ldots a_{r+1}^{n_{r+1}} b_{r+1}^{n_{r+1}} c^n \mid n, n_i \geqq 1, i=1, \ldots, r+1\} \cup b\Sigma^*,$$

is not r-ntbl but L satisfies the pumping lemma for r-ntbl, *cf.* [4]. Similarly to example 1, we can prove that L satisfies the pumping lemma for LIN. Clearly, any language that satisfies the linear pumping conditions also satisfies the pumping conditions for NTBL at any rank. Now we will use theorem 3.2 to show that L is not r-ntbl. Suppose L is r-ntbl and let n be the constant corresponding to L in theorem 3.2. Consider $z=a^n b a_1^n b_1^n a_2^n b_2^n \ldots a_{r+1}^n b_{r+1}^n c^n$ where all positions of the subword $z'=a_1^n b_1^n a_2^n b_2^n \ldots a_{r+1}^n b_{r+1}^n$ are marked, and thus $m(z)>n$. By theorem 3.2, z may be written as $z_1 z_2 \ldots z_s$ with $1 \leqq s \leqq r$, where $z_i=u_i v_i w_i x_i y_i$ such that the three conditions of that theorem hold. For each i, each of v_i and x_i can contain only one type of letter and at least one of them is contained in z' to satisfy condition (2). Now z' consists of $2r+2$ distinct letters whereas only at most $2r$ pieces of v_i's and x_i's get pumped. This implies that there exist at least 2 letters that do not get pumped and they must come in pairs of a_i and b_i to keep the correct balance. These pairs of $a_i^n b_i^n$ can be placed either (1) within $y_j u_{j+1}$ for some $1 \leqq j < s$ or (2) within w_j for some $1 \leqq j \leqq s$. In the first case we have $m(y_j u_{j+1}) \geqq 2n$ which violates condition (1) of the theorem. In the second case, $a_i^n b_i^n$ is within z_j and then v_j must consist of a_e's or b_e's for some $e<i$ while x_j must consist of a_f's or b_f's for some $i<f$. Clearly, this violates condition (3) of the theorem because v_j and x_j are pumped together. Thus L is not r-ntbl.

4. LANGUAGES SATISFYING THE OGDEN'S LEMMA

We have seen that lemma 2.1 and theorem 3.2 are necessary conditions for linear and nonterminal bounded languages, respectively. In this section, we will present counterexamples, inspired by [5, 3], to show that they are not sufficient conditions.

Define

$$L_G=\{a^p b^p c^r d^r \mid p, r \geqq 1\} \cup \{a^p b^q c^r d^s \mid 1 \leqq p < q \text{ and } r, s \geqq 1\}$$
$$\cup \{a^p b^q c^r d^s \mid 0 < p-q \leqq r+s \text{ and } p, q, r, s \geqq 1\}.$$

It is easy to see that L_G is cfl and by [5, lemma 2], L_G is not lcfl. Moreover, as we will argue below, L_G satisfies Ogden's condition of lemma 2.1 with $n=8$. Consider z in L_G with at least 8 marked positions. $z=w_1 x_1 w_2 x_2 \ldots w_k x_k w_{k+1}$ where the x_i's are all the symbols at the marked positions and w_i's are in Σ^*, $k \geqq 8$. We only need to argue the case where z is of form $z=a^p b^p c^r d^r$ where p, $r \geqq 1$. There are 4 possibilities.

(1) $x_2=a$. If $r>1$ we choose $u=w_1 x_1 w_2$, $v=x_2$, $x=$last d, $y=\varepsilon$, and $w=$the rest of string z. Clearly, $m(uvxy) \leqq 3$. By pumping down we will get a word with the number of a's less than the number of b's. Pumping v and x up will give a word with the number of a's greater than the number of b's and their difference is bounded by the sum of the number of c's and d's. On the other hand, if $r=1$ then since $m(z) \geqq 8$, $p \geqq 3$ and so we can pump a and b. Let $v=x_2$ and $x=$last b. In this case $m(uvxy) \leqq 5$. By pumping we will get a word of the same form.

(2) $x_2=b$. If $p>1$ we choose $u=w_1 x_1 w_2$, $v=x_2$, $x=y=\varepsilon$, and $w=$the rest of string z. Otherwise, $p=1$ and thus $r \geqq 3$ in which case we put $v=x_3$, $x=$last d, and u, w, and y are defined accordingly. By pumping v and x in both cases, we still get a word in L_G.

(3) $x_2=c$. This implies that $2r \geqq m(c^r d^r) \geqq 7$ which gives $r \geqq 4$. Thus, we can pump down c and d. We pick $v=x_2$, $x=$last d, and u, w, and y are defined accordingly. Pumping v and x will give a word of the same form and thus in L_G.

(4) $x_2=d$. In this case, we pick $v=$first c, $x=x_{k-1}$, u, w, and y are defined accordingly. Clearly, pumping will give a word of the same form and thus in L_G.

Hence, L_G satisfies all the three conditions of lemma 2.1 and gives an example of a context-free language that is not linear but satisfies the linear Ogden's conditions.

Let H be a subset of natural numbers, $\Sigma=\{a, b, c, d\}$ and define

$$A_H=\{a^n b^n c^m d^m \mid n, m \text{ in } H\},$$
$$L_H=A_H \cup \{a^p b^q c^r d^s \mid p \neq q \text{ or } r \neq s\}.$$

By [6, theorem 1 and corollary 1] for appropriate H, A_H is properly cfl (i. e., cfl but not lcfl). Moreover, L_H is properly context-sensitive (respectively: recursive, recursively enumerable, not recursively enumerable) if A_H is properly context-sensitive (respectively: recursive, recursively enumerable, not recursively enumerable). We will now prove that L_H satisfies Ogden's conditions of lemma 2.1 with $n=4$. This will show that there are (uncountably many) languages that satisfy linear Ogden's lemma at various levels of Chomsky hierarchy. Consider z in L_H with at least 4 marked positions. $z = w_1 x_1 w_2 x_2 \ldots w_k x_k w_{k+1}$ where the x_i's are all the symbols at the marked positions and w_i's are in Σ^*, $k \geqq 4$. If z is in A_H, then we can choose $u = w_1 x_1 w_2$, $v = x_2$ and $w =$ the rest of string of z. By pumping v, we will get a new word with either unequal number of a's and b's or c's and d's. Thus, the three conditions of lemma 2.1 are satisfied.

Now suppose z is not in A_H, i.e. z is of form $z = a^p b^q c^r d^s$ where $p \neq q$. The case for $r \neq s$ can be handled similarly. There are two possible subcases; the first one is when $r \neq s$. We choose $u = w_1 x_1 w_2$, $v = x_2$ and $w =$ the rest of string z. If v is a or b, we will still have $r \neq s$ in any new word obtained by pumping v. On the other hand, if $v = c$ or d, pumping v will still give a word with $p \neq q$. Hence, in this case, the three conditions of lemma 2.1 are satisfied. In the second subcase, we have $r = s$. Without loss of generality, we assume $p > q$. There are four possibilities to be considered in this subcase: (1) when x_2 and x_3 are both a's (2) when $x_2 = a$ and $x_3 = b$ (3) when x_2 and x_3 are both b's and (4) when x_2, x_3 is c or d. We will present the argument for (1) and leave it to the reader to convince himself of the other cases. Consider when x_2 and x_3 are both a's (note $x_1 = a$). If $p - q > 1$ then we put $u = w_1 x_1 w_2$, $v = x_2$ and $w =$ the rest of string z. Obviously, a new word obtained by pumping v will still have $p \neq q$. However, if $p - q = 1$, we choose $u = w_1 x_1 w_2$, $v = x_2 w_3 x_3$ and $w =$ the rest of string z. By pumping v down, we have $p < q$ whereas if v is pumped up we have $p > q$. In both cases, $p \neq q$. We conclude that L_H satisfies all the three conditions of lemma 2.1.

It is obvious that any language that satisfies the linear Ogden's conditions of lemma 2.1 also satisfies the Ogden's condition of theorem 3.2. Hence, by changing H, we obtain counterexamples at various levels of the Chomsky hierarchy, each of which satisfies the Ogden's conditions of theorem 3.2.

ACKNOWLEDGMENTS

We thank the referee whose very insightful (and timely) comments helped to remove some vague points and make the exposition clearer.

REFERENCES

1. A. V. Aho and J. D. Ullman, *The Theory of Parsing, Translation, and Compiling*, Vol. 1, Prentice-Hall, 1972.
2. J. Berstel, *Transductions and Context-Free Languages*, Teubner, Stuttgart, 1979.

3. L. Boasson and S. Horváth, *On languages satisfying Ogden's lemma*, RAIRO Informatique Théorique, Vol. 12 (3), 1978, pp. 201-202.
4. R. Boonyavatana and G. Slutzki, *A pumping lemma for Nonterminal Bounded Languages*, manuscript, 1984.
5. S. A. Greibach, *Linearity is Polynomially Decidable for Realtime Pushdown Store Automata*, Inform. Contr., Vol. 42, 1979, p. 27-37.
6. S. A. Greibach, *The Unsolvability of the Recognition of Linear Context-Free Languages*, JACM, Vol. 13, (4), 1966, pp. 582-587.
7. M. A. Harrison, *Introduction to Formal Language Theory*, Addison-Wesley, Reading, Mass., 1978.
8. J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, Reading, Mass, 1979.
9. W. Ogden, *A Helpful Result for Proving Inherent Ambiguity*. Math. Syst. Theory, Vol. 2 (3), pp. 191-194.