Ogden's lemma for coupled-context-free languages; the set of Lyndon words is not coupled-context-free

Peeter Laud
Institute of Computer Science
Tartu University
Liivi 2, 50409 TARTU, ESTONIA
e-mail peeter_l@ut.ee

Abstract

Coupled-context-free grammars are a natural generalization of context-free grammars by combining non-terminals to parenthesizes which can only be substituted together. The generalized context-free grammars defined here further generalize that notion by dropping the requirement that sentential forms of derivation are correctly bracketed, but retaining the requirement that the non-terminals can be substituted together only if they were introduced during the same derivation step. We prove a generalization of Ogden's lemma for generalized context-free grammars, that strictly covers the classical case. As its application, we generalize a result by Berstel and by proving that the set of Lyndon words over two-letter alphabet is not generalized context-free.

1 Introduction

The coupled-context-free grammars, a generalization of context-free grammars, have been introduced by G. Hotz and investigated by Y. Guan [5]. This class has many good properties of context-free grammars, for example being closed under union, concatenation, homomorphisms, being semilinear, etc. Also, there exist polynomial parsing algorithms for this class [9, 6].

There exist several well-known pumping lemmas for context-free languages (see e.g. [1]) which state some necessary conditions for a language to be context-free. One of these lemmas has been proven by W. Ogden [7]. In the current paper we prove that coupled-context-free languages must satisfy some conditions that are similar to those in Ogden's lemma, actually Ogden's result is a simple consequence of ours. Y. Guan [5] has proved a simpler pumping lemma for coupled-context-free languages, similar to Bar-Hillel's lemma [2].

It has been conjectured in [4] that the set of all primitive words (words which are not proper powers of some other word) is not context-free. A weaker result has been obtained by J. Berstel and L. Boasson [3] — they have shown that the set of Lyndon words (see sec. 4) is not context-free. In the current paper, as an application of our pumping lemma, we show that the set of Lyndon words is not coupled-context-free, either. In the light of this result we also conjecture that the set of all primitive words is not coupled-context-free.

We introduce a new class of grammars to emphasize these aspects of coupled-context-free derivations, which seem to be the most relevant for our proofs. This class of grammars that is bigger than the class of coupled-context-free grammars (i.e. some grammars from this class are not coupled-context-free). Still, we don't have any definite results concerning the generative power of the newly defined formalism — it is an open question, whether these grammars are more powerful than coupled-context-free grammars.

2 Definitions

Let $\mathbb{N}$ denote the set of positive integers, let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Let $[i..j]$, where $i \leq j$ denote the set $\{i, i+1, \ldots, j\}$. We denote the set inclusion by $\subseteq$ and strict inclusion by $\subset$. Let V be a finite set called *alphabet. A word over the alphabet* V is a finite sequence of the elements — letters — of V. The set of all words over V is denoted by V^*. The set V^* equipped with concatenation is a monoid which unit element is the empty word ϵ. We denote the length of a word w by $|w|$ and its i-th letter by $w(i)$.

2.1 Coupled-context-free grammars

Definition 2.1. Let K be a finite set and let $g : K \longrightarrow \mathbb{N}$ be a mapping. The set

$$[K, g] := \{(x, n) \in K \times \mathbb{N} \,:\, 1 \leq n \leq g(x)\}$$

is called *bracket alphabet* and g its *arity function*.

We denote the elements (x, n) of $[K, g]$ also (mainly) by x_n. If the arity of some $x \in K$ is 1, then we do not distinguish x and its only component $(x, 1)$.

Definition 2.2. The *bracket language* $\mathrm{D}(K, g)$ over bracket alphabet $[K, g]$ is the smallest subset of $[K, g]^*$ satisfying the following conditions:

1. $\epsilon \in \mathrm{D}(K, g)$;
2. $u, v \in \mathrm{D}(K, g)$ implies $uv \in \mathrm{D}(K, g)$;
3. $x \in K$, $g(x) = k$, $w_1, \ldots, w_{k-1} \in \mathrm{D}(K, g)$ implies $x_1 w_1 x_2 w_2 \cdots x_{k-1} w_{k-1} x_k \in \mathrm{D}(K, g)$.

Example 2.1. Let $K = \{x, y, z\}$ and $g(x) = 1, g(y) = 2, g(z) = 3$, then the following words are elements of $\mathrm{D}(K, g)$:

$$\epsilon, \quad y_1 y_1 y_2 y_2, \quad y_1 z_1 x z_2 y_1 y_2 x z_3 y_2.$$

Example 2.2. Let $K = \{x^{(1)}, \ldots, x^{(n)}\}$. If for all $i \in [1..n]$ holds $g(x^{(i)}) = 2$ then the set $\mathrm{D}(K, g)$ is just the semi-Dyck language over bracket pairs

$$\{(x^{(1)}, 1), (x^{(1)}, 2), (x^{(2)}, 1), (x^{(2)}, 2), \ldots, (x^{(n)}, 1), (x^{(n)}, 2)\}.$$

Let $X \subseteq \mathbb{N}_0$, $k \in \mathbb{N}_0$. Denote $X + k := \{x + k \,:\, x \in X\}$.

Definition 2.3. Let $[K, g]$ be a bracket alphabet and let $w \in \mathrm{D}(K, g)$. The *w-partition* is a partition $\mathcal{S}(w)$ of the set $[1..|w|]$ defined inductively by

1. $\mathcal{S}(\epsilon) := \emptyset$;
2. $\mathcal{S}(uv) := \mathcal{S}(u) \cup \{X + |u| \,:\, X \in \mathcal{S}(v)\}$ for $u, v \in \mathrm{D}(K, g)$;
3. Let $x \in K, g(x) = k, w_1, \ldots, w_{k-1} \in \mathrm{D}(K, g)$. Then

$$\mathcal{S}(x_1 w_1 x_2 w_2 \cdots x_{k-1} w_{k-1} x_k) :=$$
$$\{\{i + \sum_{j=1}^{i-1} |w_j| \,:\, i \in [1..k]\}\} \cup \bigcup_{i=1}^{k-1} \{X + (i + \sum_{j=1}^{i-1} |w_j|) \,:\, X \in \mathcal{S}(w_i)\}.$$

The w-partition describes, which bracket components in the word w are corresponding components.

Example 2.3. For $[K, g]$ and words from example 2.1 the partitions defined by those words are respectively

$$\emptyset, \quad \{\{1, 4\}, \{2, 3\}\}, \quad \{\{1, 9\}, \{2, 4, 8\}, \{3\}, \{5, 6\}, \{7\}\}.$$

Definition 2.4. *Coupled-context-free grammar* (CCFG) is quadruple $G = ([V, g], \Sigma, P, s)$ where

- $[V, g]$ is a bracket alphabet with $V = N \dot{\cup} \Sigma$, $N, \Sigma \neq \emptyset$ and the arity of all elements of Σ equal to 1.
- P is a finite set of *coupled-context-free productions* $(x_1, \ldots, x_k) \to (y_1, \ldots, y_k)$, where $x \in N$, $g(x) = k$, $y_1, \ldots, y_k \in [V, g]^*$, $y_1 \cdots y_k \in \mathrm{D}(V, g)$.
- $s \in N$, $g(s) = 1$, is the *axiom* of the grammar.

Definition 2.5. Let $G = ([V,g],\Sigma,P,s)$ be a CCFG and $w_1,w_2 \in \mathrm{D}(V,g)$. w_2 is *directly derivable* from w_1 by G (denoted $w_1 \Rightarrow w_2$) if there exist such $u_1,u_2 \in [V,g]^*$, $v_1,\dots,v_{k-1} \in \mathrm{D}(V,g)$ and $(x_1,\dots,x_k)\to(y_1,\dots,y_k) \in P$, that $w_1 = u_1x_1v_1x_2v_2\cdots x_{k-1}v_{k-1}x_ku_2$ and $w_2 = u_1y_1v_1y_2v_2\cdots y_{k-1}v_{k-1}y_ku_2$.

In this way the nonterminals $x_1,\dots,x_{g(x)}$ in w_1 can be substituted together iff they are the corresponding bracket components.

Let $\overset{*}{\Rightarrow}$ be the reflexive transitive closure of $\Rightarrow$. The language generated by G is

$$\mathcal{L}(G) := \{w \in \Sigma^* \,:\, s \overset{*}{\Rightarrow} w\}.$$

Thus CCFGs are a generalization of context-free grammars (context-free grammars are just CCFGs where the maximum arity of nonterminals is 1). For a CCFG $G = ([V,g],\Sigma,P,s)$ the maximum arity of the elements of V is called the *degree* of G and denoted by $\mathbf{grad}\,G$.

Example 2.4. The coupled-context-free grammar G,

$$G = ([\{s,A,a,b,c,d\},g],\{a,b,c,d\},\\ \{s\to A_1A_2,(A_1,A_2)\to(aA_1b,cA_2d),(A_1,A_2)\to(\epsilon,\epsilon)\},s),$$

where $g(A)=2$, generates the language $\{a^nb^nc^nd^n \,:\, n \in \mathbb{N}_0\}$.

2.2 Generalized context-free grammars

By sentential forms of coupled-context-free derivations, the property of being correctly bracketed enables us to define which letters are corresponding. By proving our results we do not use the fact that the sentential forms of derivation are correctly bracketed, we only need the well-defined correspondence of sentential forms' positions. Therefore we define a new class of grammars where this correspondence is introduced explicitly. We call the new class of grammars generalized context-free, for the lack of better name.

The sentential forms of derivations (and their corresponding bracket components) are defined as follows:

Let $X \subset \mathbb{N}$ be a finite subset and let $n \in \mathbb{N}$. We define $\mathcal{O}rd_X(n) := |X \cap [1..n]|$. Suppose $n \in [1..|X|]$. We define $\mathcal{E}l_X(n) := \min\{k \in \mathbb{N} \,:\, |X\cap[1..k]| = n\}$ (i.e. $\mathcal{E}l_X(n)$ is the n-th element of X as of a subset of $(\mathbb{N}_0,\leq)$).

Definition 2.6. Let $[K,g]$ be a bracket alphabet, $n \in \mathbb{N}_0$ and $W = \{W_1,\dots,W_k\}$ some partition of the set $[1..n]$. Denote by l_w a mapping $W \longrightarrow K$ such that for every $i \in [1..k]$ it holds $g(l_w(W_i)) = |W_i|$. W and l_w define a word $w \in [K,g]^*$ in the following way. Namely, let $|w| := n$ and for every $i \in [1..n]$ put $w(i) := (l_w(W_j),\mathcal{O}rd_{W_j}(i))$, where j is chosen such that $i \in W_j$. Thus, W and w uniquely determine the mapping l_w. We call the pair $\langle w;W\rangle$ a *bracketing* over the bracket alphabet $[K,g]$ and denote the set of all bracketings over $[K,g]$ by K_g^*.

Example 2.5. Let K and g be defined as in example 2.1. Following bracketings are elements of K_g^*.

$$\langle\epsilon;\emptyset\rangle,\quad \langle y_1y_1y_2y_2;\{\{1,4\},\{2,3\}\}\rangle,\quad \langle y_1y_1y_2y_2;\{\{1,3\},\{2,4\}\}\rangle\\ \langle y_1z_1xz_2y_1y_2xz_3y_2;\{\{1,6\},\{2,4,8\},\{3\},\{5,9\},\{7\}\}\rangle.$$

For every $w \in \mathrm{D}(K,g)$ holds $\langle w;\mathcal{S}(w)\rangle \in K_g^*$.

Let $\langle w; W\rangle$ be a bracketing. If the word w uniquely determines the partition W (i.e. there exists no partition W' of $[1..|w|]$, such that $W' \neq W$ and $\langle w; W'\rangle$ being a bracketing, too) then we denote this bracketing simply by w.

We should also give an explicit description for the right sides of productions. The right sides of coupled-context-free productions have corresponding bracket components implicitly defined, as they are tuples of words whose concatenation is correctly bracketed. Intuitively the right sides of generalized context-free productions should be tuples of something, whose concatenation is a bracketing. The exact definition is

Definition 2.7. The *bracketed tuple* over $[K, g]$ is a pair $\langle (w_1, \ldots, w_n); W\rangle$, where $n \in \mathbb{N}$, $w_1, \ldots, w_n \in [K, g]^*$ and W is partition of the set $[1..\sum_{i=1}^{n} |w_i|]$, such that $\langle w_1 \cdots w_n; W\rangle \in K_g^*$.

We denote the set of all n-component bracketed tuples over $[K, g]$ by $n{-}K_g^*$ (such notation was introduced in [9]). We do not distinguish the sets K_g^* and $1{-}K_g^*$.

Let $\langle w; W\rangle$ be a bracketing, let $X \in W$, $|X| = n$, let $\mathcal{U} = \langle (u_1, \ldots, u_n); U\rangle$ be a bracketed tuple. By *substituting the symbol $l_w(X)$ in position X in the bracketing $\langle w; W\rangle$ by $\mathcal{U}$* we get a bracketing $\langle w; W\rangle[X \longleftarrow \mathcal{U}] = \langle v; V\rangle$, where for every $i \in [1..n]$ the symbol $(l_w(X), i)$ in the word w in position $\mathcal{E}l_X(i)$ is replaced by u_i and the partition V is such, that those positions, that belong together in $\langle w; W\rangle$ (except X), and those that belong together in $\langle (u_1, \ldots, u_n); U\rangle$, also belong together after the substitution. Formally,

$$V = \big\{\{y + \sum_{i=1}^{\mathcal{O}rd_X(y)} (|u_i| - 1) \,:\, y \in Y\} \,:\, Y \in W\backslash\{X\}\big\}$$
$$\cup \big\{\{y + \mathcal{E}l_X(\mathcal{O}rd_Z(y)) - \mathcal{O}rd_Z(y) \,:\, y \in Y\} \,:\, Y \in U\big\},$$

where $Z = \{1 + \sum_{i=1}^{l} |u_i| \,:\, 0 \leq l \leq n - 1\}$. We can also define in the same way a substitution of a symbol in a position in a bracketed tuple by a bracketed tuple.

Definition 2.8. A *generalized context-free grammar* (GCFG) is a quadruple $G = ([V, g], \Sigma, P, s)$, where $[V, g]$, Σ and s are the same as in the definition of CCFG (def. 2.4) and P is the set of *generalized context-free productions* — pairs $x \to \mathcal{Y}$, where $x \in N$ and $\mathcal{Y} \in g(x){-}V_g^*$.

Definition 2.9. Let $G = ([V, g], \Sigma, P, s)$ be a GCFG and let $\langle (u_1, \ldots, u_k); U\rangle$, $\langle (u'_1, \ldots, u'_k); U'\rangle \in k{-}V_g^*$ where $k \in \mathbb{N}$. The bracketed tuple $\langle (u'_1, \ldots, u'_k); U'\rangle$ is called *directly derivable* from bracketed tuple $\langle (u_1, \ldots, u_k); U\rangle$ if there exist such $X \in U$ and $x \to \mathcal{Y} \in P$ that $l_{u_1 \cdots u_k}(X) = x$ and

$$\langle (u'_1, \ldots, u'_k); U'\rangle = \langle (u_1, \ldots, u_k); U\rangle[X \longleftarrow \mathcal{Y}].$$

Denote it by $\langle (u_1, \ldots, u_k); U\rangle \Rightarrow \langle (u'_1, \ldots, u'_k); U'\rangle$.

Let $\stackrel{*}{\Rightarrow}$ be the reflexive-transitive closure of the relation $\Rightarrow$. The *language generated by G* is given by

$$\mathcal{L}(G) := \{w \in \Sigma^* \,:\, \langle s; \{\{1\}\}\rangle \stackrel{*}{\Rightarrow} \langle w; \{\{1\}, \{2\}, \ldots, \{|w|\}\}\rangle\}.$$

The GCFGs are a generalization of CCFGs. A GCFG $([V,g],\Sigma,P,s)$ is CCFG iff for all its productions $x \to \langle (y_1,\ldots,y_{g(x)});Y\rangle$ hold $y_1\cdots y_{g(x)} \in \mathrm{D}(V,g)$ and $Y = \mathcal{S}(y_1\cdots y_{g(x)})$.

Example 2.6. Let $G' = (V',\Sigma,P',s')$ be a context-free grammar and let $n \in \mathbb{N}$. We can construct a GCFG G (of degree n) generating the language $\{w^n : w \in \mathcal{L}(G')\}$ in a very straightforward way.

Let $G = ([V,g],\Sigma,P,s)$, where $V = V'\dot{\cup}\{s\}$ and for every $x \in V'\backslash\Sigma$ is defined $g(x) = n$. Now we are going to describe the production set P. Define a mapping $\mathcal{C} : V' \times [1..n] \longrightarrow [V,g]$ by $\mathcal{C}(x,i) := x_i$ iff $x \in V'\backslash\Sigma$ and $\mathcal{C}(x,i) := x$ iff $x \in \Sigma$ $(i \in [1..n])$. Extend the mapping $\mathcal{C}$ to $V'^* \times [1..n] \longrightarrow [V,g]^*$ by defining $\mathcal{C}(w,i) := \mathcal{C}(w(1),i)\cdots\mathcal{C}(w(|w|),i)$ for every $w \in V'^*$ and $i \in [1..n]$. For every context-free production $x \to y \in P'$ let $\mathrm{Par}_n(x \to y)$ denote a generalized context-free production representing n context-free productions "in parallel"; formally

$$\begin{aligned}\mathrm{Par}_n(x\to y) := x \to \langle \big(\mathcal{C}(y,1),\mathcal{C}(y,2),\ldots,\mathcal{C}(y,n)\big);\\ \big\{\{i+j|y| \;:\; j\in[0..n-1]\} \;:\; i\in[1..|y|], y(i)\in V'\backslash\Sigma\big\}\\ \cup\big\{\{i+j|y|\} \;:\; i\in[1..|y|], y(i)\in\Sigma, j\in[0..n-1]\big\}\rangle.\end{aligned}$$

The production set is defined as

$$P := \{s \to s'_1 s'_2 \cdots s'_n\} \cup \{\mathrm{Par}_n(p') \;:\; p' \in P'\}.$$

We are not aware of any similar construction for CCFGs. Only if the grammar G' is linear, the resulting grammar G will be coupled-context-free.

2.3 Derivation trees

The derivation trees of a GCFG are quite similar to those of context-free grammars. The difference is that there is a partition defined on the vertices of the derivation tree (similarly to positions of sentential forms) that describes which vertices are "corresponding".

Definition 2.10. Let $G = ([V,g],\Sigma,P,s)$ be a GCFG. A *derivation tree* of G is a tree with a fixed partition of its vertices, that are labeled by elements from the set $[V,g] \cup [P,g]$, together with the arity $g(x \to \mathcal{Y}) := g(x)$ for every $x \to \mathcal{Y} \in P$ given. This tree is defined recursively by:

- A tree with one vertex labeled by s is a derivation tree; the partition is the single partition of the one-element set.
- Let T be a derivation tree, let $\{N_1,N_2,\ldots,N_k\}$ be a class of the partition of its vertices, and let $N_1,\ldots,N_k$ be labeled with $(x,1),\ldots,(x,k)$ respectively; here $x \in V\backslash\Sigma$ and $g(x) = k$. Let $p = x \to \langle (y_1,\ldots,y_k);Y\rangle \in P$, where for every $i \in [1..k]$ it holds $y_i = a_{i1}\cdots a_{il_i}$, with $l_i \in \mathbb{N}_0$ and $a_{ij} \in [V,g]$. If the tree T' is obtained from T by the following procedure 1.–3., then it is also a derivation tree.

  1. relabel the vertices $N_1,\ldots,N_k$ by $(p,1),\ldots,(p,k)$, respectively;

2. for every $i \in [1..k]$ add vertices $N_{i1}, \ldots, N_{il_i}$ to the tree as descendants of N_i; also, for every class X in the partition Y add the class

$$\{N_{ij} \,:\, j + \sum_{r=1}^{i-1} l_r \in X\}$$

to the partition of vertices of the tree;

3. for all $i \in [1..n]$ and $j \in [1..l_i]$ label the vertex N_{ij} by a_{ij},

Let N be a class in the partition of the vertices of the tree T and let $|N| = n$. We denote the vertices in N by $N^{(1)}, \ldots, N^{(n)}$ (traversing the tree T in preorder) and write $N = (N^{(1)}, \ldots, N^{(n)})$.

Definition 2.11. Let $G = ([V, g], \Sigma, P, s)$ be a GCFG. A derivation tree of G whose vertices are labeled by the elements of the set $\Sigma \cup [P, g]$ is called a *parse tree* of G.

Definition 2.12. The *frontier* of a labeled tree is the word that is got by writing down the labels of its leaves, from left to right.

Let T be a derivation tree of some GCFG, let $M = (M^{(1)}, \ldots, M^{(m)})$ and $N = (N^{(1)}, \ldots, N^{(n)})$ be two classes in the partition of its vertices. From definition 2.10 follows that if some vertex from N is a direct descendant of some vertex from M, then every vertex from N is a direct descendant of some vertex from M. Thus we can call that the class N a *direct descendant* of the class M (and call M the direct ancestor of N).

Let X be a vertex of the derivation tree T. The *frontier* of the vertex X is defined as the frontier of the subtree of T with the root X. Let $N = (N^{(1)}, \ldots, N^{(n)})$ be a class of the partition of the vertices of T. The *frontier* of N is defined as the sequence $(w^{(1)}, \ldots, w^{(n)})$, where $w^{(i)}$ is the frontier of $N^{(i)}$.

3 Ogden's lemma

Theorem 3.1 (Ogden's lemma for GCFGs). *Let $G = ([V, g], \Sigma, P, s)$ be a GCFG. There exists an integer $\mathbf{k}$ such that if $z \in \mathcal{L}(G)$ and $|z| \geq \mathbf{k}$ and at least $\mathbf{k}$ positions are* distinguished *in z, then there exist an integer $\bar{g}$, $1 \leq \bar{g} \leq \mathbf{grad}\, G$, integers $k_1, \ldots, k_{\bar{g}} \in \mathbb{N}_0$, $k_1 + \cdots + k_{\bar{g}} = \bar{g}$, such that*

1.
$$z = t_0 \prod_{j=1}^{\bar{g}} (v_{j,0} \prod_{i=1}^{k_j} (w_{l_j+i} v_{j,i}) t_j), \tag{3.1}$$

where $l_j := \sum_{i=1}^{j-1} k_j$ and $t_j, v_{j,i}, w_j \in \Sigma^*$, moreover, if $k_j \neq 0$ then $l_j < j \leq l_j + k_j \; (= l_{j+1})$;

2. *there exists such $j \in [1..\bar{g}]$, that w_j contains distinguished positions;*

3. *there exist such $j \in [0..\bar{g}]$, $j_1 \in [1..\bar{g}]$ and $j_2 \in [0..k_{i_1}]$, that in the decomposition (3.1) of z the subword v_{j_1,j_2} lies between words w_j and w_{j+1} (or before the word w_1, if $j = 0$; or after the word $w_{\bar{g}}$, if $j = \bar{g}$) and the words v_{j_1,j_2} and t_j both contain distinguished positions;*

4. *For every* $j \in [1..\bar{g}]$ *the word* w_j *contains distinguished positions iff the product* $v_{j,0}\prod_{i=1}^{k_j}(w_{l_j+i}v_{j,i})$ *contains distinguished positions; if the word* w_j *contains distinguished positions then the product*

$$t_{l_j}\prod_{r=l_j+1}^{l_{j+1}}(v_{r,0}\prod_{i=1}^{k_r}(w_{l_r+i}v_{r,i})t_r)$$

contains distinguished positions;

5. *There exists a non-terminal* A*, such that* $g(A) = \bar{g}$ *and*

$$\begin{aligned} s &\stackrel{*}{\Rightarrow} t_0A_1t_1A_2\cdots t_{\bar{g}-1}A_{\bar{g}}t_{\bar{g}} \\ (A_j)_{j=1}^{\bar{g}} &\stackrel{*}{\Rightarrow} (v_{j,0}A_{l_j+1}v_{j,1}A_{l_j+2}\cdots v_{j,k_j-1}A_{l_j+k_j}v_{j,k_j})_{j=1}^{\bar{g}} \\ (A_j)_{j=1}^{\bar{g}} &\stackrel{*}{\Rightarrow} (w_j)_{j=1}^{\bar{g}}. \end{aligned}$$

We will need following two definitions and a lemma for giving the proof:

Definition 3.1. Let T be a derivation tree of z (from the statement of Ogden's lemma) and let M be a class of the partition of the vertices of T. We say that M is a *splitting* class, if it has at least two direct descendants whose frontiers contain distinguished positions.

Definition 3.2. Let T be a derivation tree of z, $N_i = (N_i^{(1)}, \ldots, N_i^{(e_i)})$ a splitting class in the partition of the vertices of T, $N_{i+1} = (N_{i+1}^{(1)}, \ldots, N_{i+1}^{(e_{i+1})})$ a fixed direct descendant of N_i whose frontier contains distinguished positions and $M = (M^{(1)}, \ldots, M^{(f)})$ some other direct descendant of N_i whose frontier contains distinguished positions. Further, let h be a word over the alphabet $\{a, \bar{a}, b, \bar{b}, \#\}$ containing exactly e_{i+1} letters from the set $\{a, \bar{a}\}$, exactly f letters from the set $\{b, \bar{b}\}$ and exactly $(e_i - 1)$ letters $\#$. We say that h is a *splitting manner* of N_i, if

- For every $j \in [1..e_{i+1}]$ and $k \in [1..f]$ the j-th letter from the set $\{a, \bar{a}\}$ is to the left of the k-th letter from the set $\{b, \bar{b}\}$ in the word h iff the vertex $N_{i+1}^{(j)}$ appears before the vertex $M^{(k)}$ when traversing the tree T in preorder.

- For every $j \in [1..e_{i+1}]$ [resp. $j \in [1..f]$] and $k \in [0..e_i - 1]$ there are exactly k letters $\#$ to the left of the j-th letter from the set $\{a, \bar{a}\}$ [resp. $\{b, \bar{b}\}$] in the word h iff the vertex $N_{i+1}^{(j)}$ [resp. $M^{(j)}$] is a direct descendant of the vertex $N_i^{(k+1)}$.

- For every $j \in [1..e_{i+1}]$ [resp. $j \in [1..f]$] the j-th letter from the set $\{a, \bar{a}\}$ [resp. $\{b, \bar{b}\}$] in the word h is equal to $\bar{a}$ [resp. $\bar{b}$] iff the frontier of the vertex $N_{i+1}^{(j)}$ [resp. $M^{(j)}$] contains distinguished positions.

Note that if the splitting class N_i has three or more direct descendants whose frontiers contain distinguished positions, then N_i may have more than one splitting manner.

Denote

$$\begin{aligned}\mathcal{H}(\bar{g}) := |\{h \in \{a, \bar{a}, b, \bar{b}, \#\}^* \ :\ &\text{There exist a GCFG } G \text{, a word } z \in \mathcal{L}(G),\\ &\text{a parse tree } T \text{ of } z \text{ and a splitting class } N_i \text{ of } T,\\ &\text{such that } \mathbf{grad}\, G \leq \bar{g} \text{ and } h \text{ is a splitting manner of } N_i\}|.\end{aligned}$$

Note that $\mathcal{H}(\bar{g})$ is well-defined (i.e. is not infinite) because the words h (the elements of the set, whose cardinality is $\mathcal{H}(\bar{g})$) can contain no more than $\bar{g}$ letters from the set $\{a, \bar{a}\}$, no more than $\bar{g}$ letters from the set $\{b, \bar{b}\}$, and no more than $(\bar{g} - 1)$ letters $\#$.

Lemma 3.2. *Let $m, n, k \in \mathbb{N}$ and $x \in [0..m]$. Let $f_1, \ldots, f_n : [1..m] \longrightarrow [1..m]$ be arbitrary functions such that $|\bigcap_{t=1}^n I(f_t)| \geq x$, where $I(f_t)$ denotes the set of all fixed points of f_t. For every $1 \leq i \leq j \leq n$ define $f_{i,j} := f_j \circ f_{j-1} \circ \cdots \circ f_i$. There exists a number $\mathcal{N}(m, x, k)$, such that if $n \geq \mathcal{N}(m, x, k)$ then there exist numbers $0 \leq i_0 < i_1 < \ldots < i_k \leq n$, such that $f_{i_0+1,i_1} = f_{i_1+1,i_2} = \cdots = f_{i_{k-1}+1,i_k}$ and $f_{i_0+1,i_1} = f^2_{i_0+1,i_1}$.*

In the following we do not need full power of this lemma, but we believe that it may be useful for proving even more general pumping lemmas for GCFGs. Also, we are not aware of any shorter proofs for the weaker form (in proving Ogden's lemma for GCFGs we only need that $\mathcal{N}(m, 0, 2)$ exists).

Proof. Let $X \subseteq \bigcap_{t=1}^n I(f_t)$ be an arbitrary fixed subset such that $|X| = x$. We prove the lemma by induction over $m - x$.

Base. $m - x = 0$. In this case $f_1 = \cdots = f_n = \mathrm{id}_{[1..m]}$ and if $n \geq k$ then defining $i_j := j$ ($j \in [0..k]$) satisfies the claims of the lemma. Thus $\mathcal{N}(m, m, k) = k$.

Step. Let $m - x > 0$. By induction assumption the number $\mathcal{N}(m, x + 1, k)$ exists. If

$$n \geq ((k - 1)x^{m-x} + 1)(m - x)\mathcal{N}(m, x + 1, k) + 1$$

then at least one of the following two statements is true:

1. There exist an integer $l \geq (k - 1)x^{m-x} + 1$ and integers $0 = j_0 < j_1 < \cdots < j_l \leq n$, such that $\mathrm{Im} f_{1,j_1} = \mathrm{Im} f_{j_1+1,j_2} = \cdots = \mathrm{Im} f_{j_{l-1}+1,j_l} = X$.

2. There exist integers $1 \leq p \leq q \leq n$, such that $q - p \geq (m - x)\mathcal{N}(m, x + 1, k)$ and $X \subset \mathrm{Im} f_{p,q}$.

Indeed, assume that the first statement does not hold. Let l be the greatest number such that there exist numbers $0 = j_0 < j_1 < \cdots < j_l$, for which $\mathrm{Im} f_{j_0+1,j_1} = \mathrm{Im} f_{j_1+1,j_2} = \cdots = \mathrm{Im} f_{j_{l-1}+1,j_l} = X$. The numbers $j_0, \ldots, j_l$ can be chosen so that for every $s \in [1..l]$ it holds $\mathrm{Im} f_{j_{s-1}+1,j_s-1} \neq X$ (if $j_s - j_{s-1} \geq 2$). Consider $(l + 1)$ differences $j_1 - j_0, j_2 - j_1, \ldots, j_l - j_{l-1}, n - j_l$. The sum of these differences is n, thus at least one of them must be at least $\lceil \frac{n}{l+1} \rceil$. Thus, if we take $j_{l+1} := n$, then there exists $s \in [1..l + 1]$ such that $j_s - j_{s-1} \geq (m - x)\mathcal{N}(m, x + 1, k) + 1$. Put $p := j_{s-1}$ and $q := j_s - 1$. This choice implies that the second statement holds.

Suppose that the first statement holds. The cardinality of the set $\{f : [1..m] \longrightarrow [1..m] \ :\ \mathrm{Im} f = I(f) = X\}$ equals x^{m-x}. By the pigeonhole principle, there exist numbers $1 \leq r_1 < r_2 < \cdots < r_k \leq l$, such that

$f_{j_{r_1-1}+1,j_{r_1}} = f_{j_{r_2-1}+1,j_{r_2}} = \cdots = f_{j_{r_k-1}+1,j_{r_k}}$. For every $s \in [0..k-1]$ define $i_s := j_{r_{s+1}-1}$, also define $i_k := j_{r_k}$. Such choice of $i_0, \ldots, i_k$ satisfies the claims of the lemma as for every $0 \le p < q \le l$ holds $f_{j_p+1,j_{p+1}} = f_{j_p+1,j_q}$.

Suppose that the second statement holds. Then there exists a number r' such that $f_{p,q}(r') \notin X$. For every $q' \in [p..q]$ holds $f_{p,q'}(r') \in [1..m] \backslash X$. By the pigeonhole principle there exist a number $r \in [1..m] \backslash X$, an integer $l \ge \mathcal{N}(m, x+1, k) + 1$ and numbers $p \le j_1 < j_2 < \cdots < j_l \le q$, such that for every $s \in [1..l]$ holds $f_{p,j_s}(r') = r$.

For every $s \in [1..l-1]$ define the function $f'_s := f_{j_s+1,j_{s+1}}$. For every $s \in [1..l-1]$ holds $f'_s(r) = r$ and thus $X \dot{\cup} \{r\} \subseteq I(f'_s)$. The condition $l - 1 \ge \mathcal{N}(m, x+1, k)$ holds, thus by induction assumption there exist numbers $0 \le i'_0 < i'_1 < \cdots < i'_k \le l-1$, such that $f'_{i'_0+1,i'_1} = f'_{i'_1+1,i'_2} = \cdots = f'_{i'_{k-1}+1,i'_k}$ and $f'_{i'_0+1,i'_1} = {f'}^2_{i'_0+1,i'_1}$. For every $s \in [1..k]$ it is true that $f'_{i'_{s-1}+1,i'_s} = f_{j_{i'_{s-1}+1}+1,j_{i'_s+1}}$. Defining $i_s := j_{i'_s+1}$ for every $s \in [0..k]$ we get that the numbers $i_0, \ldots, i_k$ satisfy the claims of the lemma. $\square$

Proof of Ogden's lemma. Let l be the the maximal length of right sides of the productions in P, i.e. $l = \max\{|y| : x \to \langle y; Y \rangle \in P\}$. Let us take

$$\mathbf{k} = l^{\mathcal{N}(\mathbf{grad}\, G, 0, 2)\mathcal{H}(\mathbf{grad}\, G)|V \backslash \Sigma|+1}.$$

The proof begins in the same way as in the context-free case:

Let $z \in \mathcal{L}(G)$, such that $|z| \ge \mathbf{k}$. Distinguish $\mathbf{k}$ or more letters in z. Let T be a fixed parse tree of z.

We construct a sequence $N_1, N_2, \ldots$ of classes of the partition of the vertices of T in the following way:

1. N_1 is the class containing (only) the root of T.

2. Let $N_1, \ldots, N_i$ be fixed. If the class N_i contains a leaf (and in this case it contains nothing else), then the sequence is constructed.

3. Let $N_1, \ldots, N_i$ be fixed. The class N_{i+1} will be defined as a direct descendant of N_i whose frontier contains the largest number of distinguished positions (if the frontiers of several direct descendants contain the same number of distinguished positions, then any of them).

Let r be the length of the constructed sequence.

Claim. *Let $i \in [1..r]$. If there are at most s splitting classes among the classes $N_1, \ldots, N_{i-1}$ then there are at least $\frac{\mathbf{k}}{l^s}$ distinguished positions in the frontier of the class N_i.*

The claim can be proved with mathematical induction over i. Indeed, if $i = 1$ then $s = 0$ and the class N_i contains only the root vertex of the tree T, whose frontier contains at least $\mathbf{k}$ distinguished positions. Assume that the claim holds for i. If N_i happens not to be a splitting class then, by induction assumption, its frontier contains at least $\frac{\mathbf{k}}{l^s}$ distinguished positions and thus the frontier of N_{i+1} also contains at least $\frac{\mathbf{k}}{l^s}$ distinguished positions. If N_i happens to be a splitting class then there are $(s-1)$ splitting classes among $N_1, \ldots, N_{i-1}$ and, by induction assumption, the frontier of N_i contains at least $\frac{\mathbf{k}}{l^{s-1}}$ distinguished positions. The class N_i has no more than l direct descendants. Thus, there exists a direct descendant of N_i whose frontier contains at least $\frac{\mathbf{k}}{l^s}$ distinguished positions. By the construction of the sequence $N_1, N_2, \ldots$, the class N_{i+1} is the direct descendant of N_i whose frontier contains at least as many distinguished positions as the frontier of any other direct descendant of N_i. Thus the frontier of N_{i+1} also contains at least $\frac{\mathbf{k}}{l^s}$ distinguished positions and the claim holds.

Thus there are at least $(\mathcal{N}(\mathbf{grad}\, G, 0, 2)\mathcal{H}(\mathbf{grad}\, G)|V\setminus\Sigma|+1)$ splitting classes in the sequence $N_1, \ldots, N_r$.

The pigeonhole principle gives, that there are at least $(\mathcal{N}(\mathbf{grad}\, G, 0, 2)\mathcal{H}(\mathbf{grad}\, G) + 1)$ splitting classes in the sequence $N_1, \ldots, N_r$ that are labeled with the same non-terminal, denoted by A. Moreover, as these splitting classes can have no more than $\mathcal{H}(\mathbf{grad}\, G)$ splitting manners, there should be at least $(\mathcal{N}(g(A), 0, 2) + 1)$ splitting classes in the sequence $N_1, \ldots, N_r$ that are labeled by the same non-terminal A and have the same splitting manner, denoted by h. Let $X_1, X_2, \ldots, X_{\mathcal{N}(g(A),0,2)+1}$ be the last $(\mathcal{N}(g(A), 0, 2) + 1)$ splitting classes in the sequence $N_1, \ldots, N_r$ that are labeled with the non-terminal A and whose splitting manner is h.

Let $i \in [1..\mathcal{N}(g(A), 0, 2)]$. Define the functions $f_i : [1..g(A)] \longrightarrow [1..g(A)]$ as follows: $f_i(k) := j$, where j is chosen so, that the vertex $X_{i+1}^{(k)}$ is contained in the subtree of T whose root is $X_i^{(j)}$.

Let $1 \le i \le j \le \mathcal{N}(g(A), 0, 2)$, let $f_{i,j} := f_j \circ f_{j-1} \circ \cdots \circ f_i$. It follows from the definition of f_i that the vertex $X_{j+1}^{(k)}$ is contained in the subtree of T whose root is $X_i^{(f_{i,j}(k))}$. According to the lemma 3.2 there exist i_0, i_1, i_2, such that $0 \le i_0 < i_1 < i_2 \le \mathcal{N}(g(a), 0, 2)$ and $f_{i_0+1,i_1} = f_{i_1+1,i_2} = f^2_{i_0+1,i_1}$. Denote $Y_0 := X_{i_0+1}$, $Y_1 := X_{i_1+1}$, $Y_2 := X_{i_2+1}$.

Let $\bar{g} := g(A)$ and $k_j := |f^{-1}_{i_1+1,i_2}(j)|$ for every $j \in [1..\bar{g}]$. Denote the frontier of the class Y_2 by $(w_j)_{j=1}^{\bar{g}}$, the frontier of the class Y_1 by

$$(v_{j,0} \prod_{i=1}^{k_j} (w_{l_j+i} v_{j,i}))_{j=1}^{\bar{g}},$$

(the quantities l_j are defined as in the statement of the lemma) and the frontier of the root vertex (i.e. the entire word z) by

$$t_0 \prod_{j=1}^{\bar{g}} (v_{j,0} \prod_{i=1}^{k_j} (w_{l_j+i} v_{j,i}) t_j).$$

(These were the definitions of w_i, $v_{i,j}$ and t_i.) Such decomposition of z satisfies the claim of the theorem. Indeed,

1. The decomposition of z has the required form. Also, if $k_j > 0$ then $l_j < j \le l_{j+1}$; this follows from $f^2_{i_1+1,i_2} = f_{i_1+1,i_2}$, i.e. the image of f_{i_1+1,i_2} coincides with the set of its fixed points. Indeed, from the definitions of k_j and l_j and the monotonicity of f_{i_1+1,i_2} follows, that for every $q \in [1..\bar{g}]$ holds $f^{-1}_{i_1+1,i_2}(q) = [l_q + 1..l_{q+1}]$. Let $q \in [1..\bar{g}]$ be such, that $k_q > 0$. In this case q belongs to the image of f_{i_1+1,i_2}, thus $f_{i_1+1,i_2}(q) = q$, i.e. $q \in [l_q + 1..l_{q+1}]$.

2. Y_2 is one element of the sequence $N_1, \ldots, N_r$, thus its frontier contains distinguished positions.

3. Y_1 is a splitting class, thus there exist such j_1, j_2 that v_{j_1,j_2} contains distinguished positions. Y_0 is a splitting class, thus there exists such j that t_j contains distinguished positions. The numbers j_1, j_2, j can be chosen as described in the claim of the theorem, because the splitting manners of Y_0 and Y_1 are the same and $f_{i_0+1,i_1} = f_{i_1+1,i_2}$.

4. It follows from the equality of splitting manners of Y_0, Y_1 and Y_2 that for every $j \in [1..\bar{g}]$ the frontier of $Y_0^{(j)}$ contains distinguished positions iff the frontier of $Y_1^{(j)}$ contains distinguished positions iff the frontier $Y_2^{(j)}$ contains distinguished positions.

5. It is obvious that there exist derivations described in the claim of the lemma for the non-terminal A.

□

Observe that the proof does not depend on the choice of the parse tree T. Thus we could have fixed it in the claim of the theorem. We need this observation in the following section.

4 On the set of Lyndon words

Definition 4.1. Let Σ be an alphabet. A word $w \in \Sigma^*$ is called *primitive* if there exist no $v \in \Sigma^*$ and $n \in \mathbb{N}\backslash\{1\}$, such that $w = v^n$. Let Σ be linearly ordered; this defines lexicographical order on Σ^*. A primitive word is called a *Lyndon word* if it is minimal among its cyclic permutations (two words w_1, w_2 are cyclic permutations of each other if there exist words u, v, such that $w_1 = uv$ and $w_2 = vu$).

It has been conjectured in [4] that the set of all primitive words is not context-free. A weaker result has been obtained in [3], where the set of Lyndon words is proven to be non-context-free. Ogden's lemma for context-free grammars is the main tool of that proof. As an application of our pumping lemma, we show that the set of Lyndon words is not generalized context-free, either. We also conjecture that the set of all primitive words is not generalized context-free.

Theorem 4.1. *Let $\Sigma = \{a, b\}$, let $a < b$. The set L of Lyndon words over the alphabet Σ is not generalized context-free.*

We will need the following definitions and a lemma for giving the proof.

Definition 4.2. Let $G = ([V, g], \Sigma, P, s)$ be a GCFG. A production $p = x \to \langle (y_1, \ldots, y_{g(x)}); Y \rangle \in P$ is called a *ϵ-production* if there exists such $i \in [1..g(x)]$, that $y_i = \epsilon$.

Definition 4.3. A GCFG $G = ([V, g], \Sigma, P, s)$ is called *ϵ-free* either if P contains no ϵ-productions or if P contains exactly one ϵ-production — $s \to \epsilon$ — and in this case the right side of no production contains s.

Lemma 4.2. *For every GCFG $G = ([V, g], \Sigma, P, s)$ there exists a ϵ-free GCFG $G' = ([V', g], \Sigma, P', s')$, such that $\mathcal{L}(G) = \mathcal{L}(G')$.*

For coupled-context-free grammars the proof can be found in [9, 6] where the algorithm for transforming coupled-context-free grammars to Chomsky normal form, also defined there, is given. Similar argument works also for GCFG-s.

Proof of theorem 4.1. Assume that there exists a GCFG G such, that $\mathcal{L}(G) = L$. We may assume w.l.o.g. that G is ϵ-free. Let $\mathbf{k}$ be the Ogden's constant of G. Put $\mathbf{c} := 6(\mathbf{k}+2)(2\mathbf{k}+3)(\mathbf{grad}\, G)$ and consider the Lyndon word $z := (a^{\mathbf{k}+1}b)^2(a^{\mathbf{k}}b)^{\mathbf{c}}$. For every $x \in [1..\mathbf{c}]$ let $\mathcal{A}[x]$ denote the set

$$[\mathbf{k}+4+x(\mathbf{k}+1)..2\mathbf{k}+3+x(\mathbf{k}+1)].$$

Thus the x-th subword $a^{\mathbf{k}}$ of z, that is immediately preceded and succeeded by letter b, stands at the positions $\mathcal{A}[x]$. Fix a parse tree T of z.

Let $x \in [1..\mathbf{c}]$ be fixed. Let the positions from the set $\mathcal{A}[x]$ be distinguished in z and let

$$z = t_0 \prod_{j=1}^{\bar{g}} (v_{j,0} \prod_{i=1}^{k_j} (w_{l_j+i} v_{j,i}) t_j)$$

be the decomposition of z in Ogden's lemma.

According to the third claim of Ogden's lemma there exist such $j_0 \in [0..\bar{g}]$, $j_1 \in [1..\bar{g}]$ and $j_2 \in [0..k_{i_1}]$, that subword v_{j_1,j_2} lies between subwords w_{j_0} and w_{j_0+1} (with exceptions mentioned in the third claim of Ogden's lemma for cases $j_0 = 0$ and $j_0 = \bar{g}$) and the subwords v_{j_1,j_2} and t_{j_0} both contain distinguished positions. According to the fourth claim of Ogden's lemma, w_{j_1} contains distinguished positions. Let $j \in [1..\bar{g}]$ be such, that $k_j > 0$ and $l_j + 1 \leq j_1 \leq l_{j+1}$. According to the fourth claim of Ogden's lemma, w_j contains distinguished positions. Observing the decomposition of z we notice that the subwords t_{j_0}, v_{j_1,j_2} and w_j can lie in the word z in two possible orders — $t_{j_0}, v_{j_1,j_2}, w_j$ or $w_j, v_{j_1,j_2}, t_{j_0}$. Moreover, in the first case holds $l_j \leq j_0$, and in the second case holds $j_0 \leq l_{j+1}$.

Suppose that subwords t_{j_0}, v_{j_1,j_2} and w_j lie in the first possible order. Observe the following subword of z:

$$t_{l_j} v_{l_j+1,0} t_{l_j+1} v_{l_j+2,0} t_{l_j+2} \cdots v_{j-1,0} t_{j-1} v_{j,0} w_{l_j+1} v_{j,1} w_{l_j+2} \cdots v_{j,j-l_j-1} w_j. \quad (4.1)$$

We are going to prove that t_{l_j} contains distinguished positions. Again, there are two possibilities:

1. $j = l_j + 1$. In this case the subword (4.1) has the form $t_{l_j} v_{j,0} w_j$, thus $j_1 = j, j_2 = 0$ and $j_0 = l_j$, thus t_{l_j} contains distinguished positions.

2. $j > l_j + 1$. We are going to heavily use the fact that there are no "gaps" between distinguished positions. Subword t_{j_0} contains distinguished positions, thus w_{l_j+1} contains distinguished positions. According to the fourth claim of Ogden's lemma,

$$t_{l_{l_j+1}} \prod_{r=l_{l_j+1}}^{l_{l_j+2}} (v_{r,0} \prod_{i=1}^{k_r} (w_{l_r+i} v_{r,i}) t_r)$$

contains distinguished positions, thus $t_{l_{l_j+2}}$ contains distinguished positions. $l_j + 2 \leq j$, thus $l_{l_j+2} \leq l_j$ and t_{l_j} contains distinguished positions.

If the subwords t_{j_0}, v_{j_1,j_2} and w_j lie in the second possible order, then we can similarly show that $t_{l_{j+1}}$ contains distinguished positions.

We have shown that there exists such $j \in [1..\bar{g}]$ that

1. $k_j > 0$ and w_j contains distinguished positions.
2. One of the following statements is true:
   (a) t_{l_j} contains distinguished positions and thus

$$\prod_{s=l_j+1}^{j-1} (v_{s,0}t_s)v_{j,0} \prod_{s=1}^{j-l_j-1} (w_{l_j+s}v_{j,s})$$

   contains only distinguished positions. Moreover, there exists such $s \in [l_j+1..j-1]$ that $v_{s,0} \neq \epsilon$, or such $s \in [0..j-l_j-1]$ that $v_{j,s} \neq \epsilon$.

   (b) $t_{l_{j+1}}$ contains distinguished positions and thus

$$\prod_{s=j-l_j+1}^{k_j} (v_{j,s-1}w_{l_j+s})v_{j,k_j} \prod_{s=j+1}^{l_{j+1}} (t_{s-1}v_{s,0})$$

   contains only distinguished positions. Moreover, there exists such $s \in [j-l_j..k_j]$ that $v_{j,s} \neq \epsilon$, or such $s \in [j+1..l_{j+1}]$ that $v_{s,0} \neq \epsilon$.

Thus by pumping up the word z accordingly to the fifth claim of Ogden's lemma we get a set of words, where each word contains a subword belonging to $\{a\}^*$ and not standing in the beginning of this word, and such that the length of these subwords is unbounded.

Let $\mathbf{j}$ be the least j such that the above two conditions 1. and 2. are satisfied. Define the numbers $H_{\mathrm{b}}(x)$ and $H_{\mathrm{e}}(x)$ so that the subword $w_{\mathbf{j}}$ of z stands on the positions $[H_{\mathrm{b}}(x)..H_{\mathrm{e}}(x)]$ in the word z. Define $d(x) := \mathbf{j}$.

Therefore, by pumping up the word accordingly to the fifth claim of Ogden's lemma the size of the first group of letters a must increase boundlessly. Thus, there must exist such $j' \in [1..\bar{g}]$, that

1'. $k_{j'} > 0$ and the first position of $w_{j'}$ is no more than the $(\mathbf{k}+2)$-th position of z.

2'. There exists a non-empty word among the words $v_{j',0}, v_{j',1}, \dots, v_{j',k_{j'}}$, $w_{l_{j'}+1}, \dots, w_{j'-1}, w_{j'+1}, \dots, w_{l_{j'+1}}$.

3'. If the first position of $w_{j'}$ is exactly the $(\mathbf{k}+2)$-th position of z, then there exists a non-empty word among the words $v_{j',0}, \dots, v_{j',j'-l_{j'}-1}$, $w_{l_{j'}+1}, \dots, w_{j'-1}$.

Moreover, the last position of the word $w_{j'}$ can be no more than the $(2\mathbf{k}+3)$-th position of the word z. Otherwise, by pumping down the word z accordingly to the fifth claim of Ogden's lemma, the first group of letters a would get smaller, but the second group would still contain $(\mathbf{k}+1)$ letters.

Let $\mathbf{j}'$ be the least j' such that the above three conditions 1'–3' are satisfied. Define the numbers $H'_{\mathrm{b}}(x)$ and $H'_{\mathrm{e}}(x)$ so that the subword $w_{\mathbf{j}'}$ of z stands on the positions $[H'_{\mathrm{b}}(x)..H'_{\mathrm{e}}(x)]$ in the word z. Define $d'(x) := \mathbf{j}'$.

In this way, we have defined functions $H_{\rm b}, H_{\rm e}, d, H'_{\rm b}, H'_{\rm e}, d'$ from $[1..\mathbf{c}]$ to the set of positive integers. For every $x \in [1..\mathbf{c}]$ holds $H_{\rm b}(x) > \mathbf{k} + 4 + x(\mathbf{k}+1)$ or $H_{\rm e}(x) < 2\mathbf{k} + 3 + x(\mathbf{k}+1)$, thus the cardinality of the set $\{(H_{\rm b}(x), H_{\rm e}(x)) : x \in [1..\mathbf{c}]\}$ is at least $3(\mathbf{k}+2)(2\mathbf{k}+3)(\mathbf{grad}\, G)$. Also, the cardinality of the image of $H'_{\rm b}$ is not greater than $(\mathbf{k}+2)$ and the cardinality of the image of $H'_{\rm e}$ is not greater than $(2\mathbf{k}+3)$. Thus there exists $X \subseteq [1..\mathbf{c}]$, such that $|X| \geq 3(\mathbf{grad}\, G)$ and for every $x_1, x_2 \in X$ $(x_1 \neq x_2)$ hold $(H_{\rm b}(x_1), H_{\rm e}(x_1)) \neq (H_{\rm b}(x_2), H_{\rm e}(x_2))$ and $(H'_{\rm b}(x_1), H'_{\rm e}(x_1)) = (H'_{\rm b}(x_2), H'_{\rm e}(x_2))$.

For every $x \in X$ let $Y_2(x)$ [resp. $Y_1(x)$] denote the class of the partition of the vertices of T which is denoted by Y_2 [resp. Y_1] in the proof of Ogden's lemma, for the case where the positions from the set $\mathcal{A}[x]$ are distinguished in the word z. For all $x \in X$ the frontiers of $Y_2(x)^{(d'(x))}$ are the same. Thus for every $x_1, x_2 \in X$ the classes $Y_2(x_1)$ and $Y_2(x_2)$ either coincide or one is a descendant of the other.

Let $x_0 \in X$ be such that the class $Y_2(x_0)$ is the closest to the root of T among the classes $\{Y_2(x) : x \in X\}$. As the frontiers of the vertices $Y_2(x)^{(d(x))}$ $(x \in X)$ are all different, the element x_0 can be chosen in at most $\mathbf{grad}\, G$ different ways. Thus the vertices of $Y_2(x_0)$ have at least $2\,\mathbf{grad}\, G$ descendants which have the form $Y_2(x)^{(d(x))}$, where $x \in X$. The vertex $Y_2(x_0)^{(d'(x_0))}$ has no descendants which have this form. Thus there exist $d_0 \in [1..|Y_2(x_0)|]$, and $x_1, x_2, x_3 \in X$ $(x_1 < x_2 < x_3)$, such that the vertices $Y_2(x_1)^{(d(x_1))}$, $Y_2(x_2)^{(d(x_2))}$ and $Y_2(x_3)^{(d(x_3))}$ are descendants of the vertex $Y_2(x_0)^{(d_0)}$. The classes $Y_1(x_2)$ and $Y_2(x_0)$ must either coincide or one must be a descendant of the other. Suppose that the class $Y_1(x_2)$ is a descendant of $Y_2(x_0)$. In this case the vertex $Y_1(x_2)^{(d'(x_2))}$ is a descendant of the vertex $Y_2(x_0)^{(d'(x_0))}$ and an ancestor of the vertex $Y_2(x_2)^{(d'(x_2))}$. The length of the frontier of $Y_1(x_2)^{(d'(x_2))}$ is greater than the length of the frontier of $Y_2(x_2)^{(d'(x_2))}$. Consequently the frontiers of the vertices $Y_2(x_0)^{(d'(x_0))}$ and $Y_2(x_2)^{(d'(x_2))}$ cannot be equal. It contradicts the choice of the set X. Thus the class $Y_2(x_0)$ must be either a descendant of the class $Y_1(x_2)$ or these classes must coincide. The frontier of the vertex $Y_2(x_0)^{(d_0)}$ contains the entire subword of z that stands on positions from the set $\mathcal{A}[x_2]$, thus the frontier of the class $Y_1(x_2)$ contains this subword, too. It follows that when the positions from the set $\mathcal{A}[x_2]$ are distinguished in the word z, the decomposition of z does not satisfy the claims of Ogden's lemma (there exist no such $\bar{\jmath} \in [0..\bar{g}]$ that $t_{\bar{\jmath}}$ contains distinguished positions). Thus, there is no GCFG G such that $\mathcal{L}(G) = L$. $\square$

References

[1] A. V. Aho, J. D. Ullman. The Theory of Parsing, Translation and Compiling. Vol. 1: Parsing. Prentice-Hall, Inc. Englewood Cliffs, N. J. 1972.

[2] Y. Bar-Hillel, M. Perles, E. Shamir. On formal properties of simple phrase-structure grammars. Z. Phonetik. Sprachwiss. Komm. **14** (1961), pp143-172.

[3] J. Berstel, L. Boasson. The Set of Lyndon Words is not Context-Free. Bulletin of the European Assoc. for Theor. Comp. Sci. **63** (1997), pp139-140.

[4] P Dömösi, S. Horváth, M. Ito, L. Kászonyi, M. Katsura. Some combinatorial properties of words, and the Chomsky-hierarchy. in Proc. 2nd Internat. Coll. Words, Languages and Combinatorics (Kyoto, 1992) (World Scientific, Singapore, 1994), pp105-123.

[5] Y. Guan. Klammergrammatiken, Netzgrammatiken und Interpretationen von Netzen. Dissertation, Universität des Saarlandes, 1992.

[6] G. Hotz, G. Pitsch. On parsing coupled-context-free languages. Theor. Comp. Sci. **161** (1996), pp205-233.

[7] W. Ogden. A helpful result for proving inherent ambiguity, Mathematical Systems Theory **2** (1968), pp191-194.

[8] H. Petersen. On the language of primitive words. Theor. Comp. Sci. **161** (1996), pp141-156.

[9] G. Pitsch. Analyse von Klammergrammatiken. Dissertation, Universität des Saarlandes, 1993.