ON THE MULTIPLICITY EQUIVALENCE PROBLEM FOR CONTEXT-FREE GRAMMARS

WERNER KUICH
Institut für Algebra und Diskrete Mathematik
Technische Universität Wien
Wiedner Hauptstraße 8-10, A-1040 Wien

Meinem Freund Arto zum 60. Geburtstag gewidmet.

Abstract. Two context-free grammars are called multiplicity equivalent iff all words over the common terminal alphabet are generated with the same degree of ambiguity. Generalizing a technique introduced by D. Raz, we show for some classes of context-free grammars that their multiplicity equivalence problem is decidable.

1. Introduction

The result of Harju, Karhumäki [3], which showed the decidability of the equivalence problem for deterministic multitape finite automata by algebraic methods has again raised interest to solve the same problem for deterministic pushdown automata, i. e., for deterministic context-free languages. Raz [8], [9] has, in two versions of a paper, introduced a new technique which allows to decide the multiplicity equivalence of certain context-free grammars. This technique is described in Section 2. In Section 3 we will generalize this technique and in Section 4 we will apply the results of Section 3 to certain classes of context-free grammars.

The decidability of the multiplicity equivalence by the technique of Raz [8], [9] depends on the following result of Kuich, Salomaa [4]:

Corollary 16.19. *It is decidable whether or not two given formal power series in* $\mathbb{Q}^{\mathrm{alg}}\langle\langle c(\Sigma^*)\rangle\rangle$ *are equal.* $\square$

Here $c(\Sigma^*)$ denotes the free commutative monoid generated by the alphabet Σ.

We now assume the reader to be familiar with the basic notions and results concerning context-free grammars and languages (see Salomaa [10]) and algebraic formal power series (see Kuich, Salomaa [4]).

If Σ is an alphabet, then $|w|_a$ denotes the number of occurences of the symbol $a \in \Sigma$ in the word $w \in \Sigma^*$. If $\Sigma = \{a_1, \ldots, a_m\}$, the *Parikh mapping* $\psi : \Sigma^* \to \mathbb{N}^m$ is the morphism defined by $\psi(a_i) = (0, \ldots, 1, \ldots, 0)$, where the 1 stands at position i and the addition of vectors is componentwise. This yields $\psi(w) = (|w|_{a_1}, \ldots, |w|_{a_i}, \ldots, |w|_{a_m})$.

In the sequel we only consider ε-free context-free grammars G with terminal alphabet $\Sigma = \{a_1, \ldots, a_m\}$ such that for all words over Σ the number of different leftmost derivations with respect to G is finite. The number of different leftmost derivations for the word w is called *degree of ambiguity* of w and is denoted by $d(G,w)$. Since $d(G,w)$ is finite for all $w \in \Sigma^*$, the formal power series $\sum_{w \in \Sigma^*} d(G,w)w$ is in $\mathbb{N}^{\mathrm{alg}}\langle\langle \Sigma^* \rangle\rangle$. = multiplicity

Commutative closure

Let G_1 and G_2 be two context-free grammars with terminal alphabet Σ. Then G_1 and G_2 are called *multiplicity letter equivalent*, denoted by $G_1 \sim_{\#} G_2$, iff for all $(n_1, \ldots, n_m) \in \mathbb{N}^m$

$$\sum_{\psi(w) = (n_1, \ldots, n_m)} d(G_1, w) = \sum_{\psi(w) = (n_1, \ldots, n_m)} d(G_2, w).$$

Since $\sum_{w \in \Sigma^*} d(G_l, w)c(w) = \sum_{\alpha \in c(\Sigma^*)} \left(\sum_{c(w) = \alpha} d(G_l, w)\right)\alpha$, $l = 1,2$, where $c(w)$ is the commutative version of w, is in $\mathbb{N}^{\mathrm{alg}}\langle\langle c(\Sigma^*)\rangle\rangle$, this definition allows us to express the above quoted Corollary 16.19 of Kuich, Salomaa [4] in terms of the theory of context-free grammars:

Theorem 1.1. *It is decidable whether or not two given context-free grammars are multiplicity letter equivalent.* $\square$

The next definition is the crucial one. Two context-free grammars G_1 and G_2 are *multiplicity equivalent,* denoted by $G_1 \sim G_2$, iff for all words $w \in \Sigma^*$

$$d(G_1, w) = d(G_2, w).$$

With the notation $\mathrm{diff}(G_1, G_2, w) = d(G_1, w) - d(G_2, w)$ we have

$$G_1 \sim G_2 \quad \text{iff} \quad \mathrm{diff}(G_1, G_2, w) = 0 \text{ for all } w \in \Sigma^*.$$

As usual, G_1 and G_2 are called *equivalent,* denoted by $G_1 \equiv G_2$, iff $L(G_1) = L(G_2)$. Observe that, for unambiguous context-free grammars G_1 and G_2,

$$G_1 \sim G_2 \quad \text{iff} \quad G_1 \equiv G_2.$$

Hence, the decidability of the multiplicity equivalence for the members of a class of context-free grammars would imply the decidability of the equivalence for the unambiguous context-free grammars of this class, i. e., the decidability of the equality of the generated context-free languages. This would also show the decidability of the equivalence for the deterministic context-free grammars of this class.

A language is called *acommutative* iff its Parikh mapping is injective. In this case, two context-free grammars are multiplicity letter equivalent iff they are multiplicity equivalent. Hence, Corollary 16.19 of Kuich, Salomaa [4] yields at once the next result.

Theorem 1.2. *It is decidable whether or not two given context-free grammars which generate acommutative languages are multiplicity equivalent.* □

Example. If $L(G_1), L(G_2) \subseteq a_1^* \dots a_m^*$ then it is decidable whether or not $G_1 \sim G_2$. (See Ginsburg [2] for the theory of bounded context-free languages.) □

A language over a one-letter alphabet is always acommutative. This yields the next result.

Corollary 1.3. *It is dedidable whether or not two given context-free grammars that generate languages over $\Sigma = \{a_1\}$ are multiplicity equivalent.* □

The following remark is in order. A context-free grammar G that generates a language over $\Sigma = \{a_1\}$ always generates a regular language. But in general, there is no regular grammar G' such that $d(G', a_1^n) = d(G, a_1^n)$ for all $n \geq 0$. These facts are due to

$$\mathbb{B}^{rat}\langle\langle a_1^* \rangle\rangle = \mathbb{B}^{alg}\langle\langle a_1^* \rangle\rangle \quad \text{and} \quad \mathbb{N}^{rat}\langle\langle a_1^* \rangle\rangle \subsetneqq \mathbb{N}^{alg}\langle\langle a_1^* \rangle\rangle.$$

Example. Consider the context-free grammar $G = (\{S\}, \{a_1\}, \{S \to a_1SS, S \to a_1\}, S)$. We obtain, for all $n \geq 0$,

$$d(G, a_1^{2n+1}) = \frac{(2n)!}{n!(n+1)!}, \qquad d(G, a_1^{2n}) = 0.$$

The language $L(G) = \{a_1^{2n+1} \mid n \geq 0\}$ is regular. But there exists no regular grammar G' such that $d(G', a_1^n) = d(G, a_1^n)$ for all $n \geq 0$. □

2. Previous results by D. Raz

In this section we cover some of the results of Raz [8], [9]. Most of the constructions and proofs are modified versions of those of Raz [8], [9].

The basic idea of Raz [8] is to construct, for a given context-free grammar G with terminal alphabet $\Sigma = \{a_1, \ldots, a_m\}$, a context-free grammar G' such that, for all $w \in \Sigma^*$,

$$d(G', w) = \sum_{w' \leq w} d(G, w').$$

Here the summation is over all words $w' \in \Sigma^*$ that are less or equal to w, where $\leq$ is the following partial order over Σ^* (total lexicographic order on Σ^n, $n \geq 0$):

?

$$w_1 \leq w_2 \quad \text{iff} \quad |w_1| = |w_2|,\ w_1 = va_iv_1,\ w_2 = va_jv_2,\ i < j, \quad \text{or } w_1 = w_2.$$

length-preserving

This means that a_1 is the least and a_m is the largest element of Σ.

Let G_1 and G_2 be context-free grammars. Let G_1' and G_2' be context-free grammars such that, for all $w \in \Sigma^*$,

$$d(G_l', w) = \sum_{w' \leq w} d(G_l, w'), \quad l = 1,2.$$

Then, for all $w \in \Sigma^*$, we have

$$\mathrm{diff}(G_1', G_2', w) = \sum_{w' \leq w} \mathrm{diff}(G_1, G_2, w')$$

Construction 2.1. Let $G = (\Phi, \Sigma, P, S)$ be a context-free grammar in binary Greibach normal form, i.e., all the productions are of the form

$$A \to aA_1A_2, \; A \to aA_1, \; A \to a, \qquad A, A_1, A_2 \in \Phi, \; a \in \Sigma.$$

Let $\bar{G} = (\bar{\Phi}, \Sigma, \bar{P}, \bar{S})$ be the context-free grammar defined by

(i) $\bar{\Phi} = \Phi \cup \Phi' \cup \Phi'' \cup \{\bar{S}\}$, where Φ' and Φ'' are primed and double-primed versions of Φ, respectively, and $\bar{S}$ is a new start variable;

(ii) $\bar{P} = P \cup P_0 \cup P_1 \cup P_2 \cup \{\bar{S} \to S, \bar{S} \to S''\}$, where
$P_0 = \{A' \to b, A'' \to c \mid A \to a \in P, \; b, c \in \Sigma, \; c > a\}$,
$P_1 = \{A' \to bA_1', A'' \to aA_1'', A'' \to cA_1' \mid A \to aA_1 \in P, \; b, c \in \Sigma, \; c > a\}$,
$P_2 = \{A' \to bA_1'A_2', A'' \to aA_1''A_2', A'' \to aA_1A_2'', A'' \to bA_1'A_2' \mid A \to aA_1A_2 \in P, b, c \in \Sigma, \; c > a\}$.

Then we obtain

$$d(\bar{G}, w) = \sum_{w' \leq w} d(G, w') \qquad \text{for all } w \in \Sigma^*. \qquad \square$$

Theorem 2.2. *For a given context-free grammar G, where $d(G,w)$ is finite for all words w, a context-free grammar G' can be effectively constructed such that, for all w,*

$$d(G', w) = \sum_{w' \leq w} d(G, w').$$

Proof. By Theorems 14.9, 14.31 and 14.6 of Kuich, Salomaa [4] (the basic semiring is $\mathbb{N}$) we can transform the grammar G into a multiplicity equivalent context-free grammar G'' in binary Greibach normal form. Now we apply Construction 2.1 yielding $G' = \bar{G}''$. □

In the sequel, we denote the grammar G' of Theorem 2.2 by $\mathrm{I}(G)$. Iteration of this construction yields the context-free grammars $\mathrm{I}^s(G)$, $s \geq 0$: $\mathrm{I}^0(G) = G$, $\mathrm{I}^{i+1}(G) = \mathrm{I}(\mathrm{I}^i(G))$, $i \geq 0$.

Before studying some of the properties of the context-free grammars $\mathrm{I}^s(G)$, $s \geq 0$, we need two combinatorial lemmas and one more grammatical construction.

Lemma 2.3. *Let $a_{k,i} \in \mathbb{R}$, $k \geq 0$, $i \geq 1$, and assume $a_{k,i} = \sum_{1 \leq j \leq i} a_{k-1,j}$ for all $k, i \geq 1$. Then, for $k, i \geq 1$,*

$$a_{k,i} = \sum_{1 \leq j \leq i} \binom{k+i-j-1}{k-1} a_{0,j}.$$

Proof. Let $\mathrm{B}_k(z) = \sum_{i \geq 1} a_{k,i} z^i = \frac{1}{1-z} \mathrm{B}_{k-1}(z)$, $k \geq 1$, be the generating function of the $a_{k,i}$. The equation $\mathrm{B}_k(z) = \frac{1}{(1-z)^k} \mathrm{B}_0(z)$, $k \geq 1$, implies $a_{k,i} = \sum_{1 \leq j \leq i} a_{0,j} \cdot c_{k,i,j}$, where $c_{k,i,j}$ is the coefficient of z^{i-j} in $(1-z)^{-k}$. The Binomial Expansion yields

$$c_{k,i,j} = \binom{k+i-j-1}{k-1}.$$

□

Lemma 2.4. *Let $a_{k,i} \in \mathbb{R}$, $k \geq 0$, $i \geq 1$, and assume $a_{k,i} = \sum_{1 \leq j \leq i} a_{k-1,j}$ for all $k, i \geq 1$. Let now $k, t \geq 1$ be fixed. Assume $a_{s,t} = 0$ for all $1 \leq s \leq k$ and $|\{j \mid a_{0,j} \neq 0,\ 1 \leq j \leq t\}| \leq k$. Then $a_{0,j} = 0$ for all $1 \leq j \leq t$.*

Proof. Assume that $a_{0,s_1}, a_{0,s_2}, \ldots, a_{0,s_l}$, $l \leq k$, are possibly unequal to 0. Since $a_{s,t} = 0$, we obtain by Lemma 2.3

$$\sum_{1 \leq j \leq l} \binom{s+t-s_j-1}{s-1} a_{0,s_j} = 0, \qquad 1 \leq s \leq k.$$

The determinant of the $l \times l$-matrix with i,j-entry

$$\binom{i+t-s_j-1}{i-1}$$

is unequal to 0. Hence $a_{0,s_j} = 0$ for $1 \leq j \leq l$. □

Theorem 2.5. *Let G_1 be a context-free grammar and G be a regular grammar. Then a context-free grammar G' can be effectively constructed such that*

$$d(G',w) = d(G_1,w)\cdot d(G,w).$$

Proof. By Theorems 14.9, 9.18 and Corollary 13.6 of Kuich, Salomaa [4] (the basic semiring is $\mathbb{N}$), the formal power series

$$\sum_{w\in\Sigma^*} d(G_1,w)\cdot d(G,w)w \quad = \quad \sum_{w\in\Sigma^*} d(G_1,w)w \odot \sum_{w\in\Sigma^*} d(G,w)w$$

is $\mathbb{N}$-algebraic in a constructive sense. (Here, $\odot$ denotes the Hadamard product.) □

In terms of grammars, the construction given in the proof of Theorem 6.7 of Salomaa [10] yields a grammar G' with the properties stated in Theorem 2.5. In the sequel, we denote the grammar G' of Theorem 2.5 by $G_1 \odot G$.

Lemma 2.6. *Let $k,n \geq 0$. Assume that G_1 and G_2 are context-free grammars such that, for all $1\leq s\leq k$, $\mathrm{diff}(\mathrm{I}^s(G_1),\mathrm{I}^s(G_2),a_m^n)=0$. Furthermore, assume that $k_n = |\{w \mid \mathrm{diff}(G_1,G_2,w)\neq 0,\ w\in\Sigma^n\}| \leq k$. Then $k_n = 0$.*

Proof. Denote, for $1\leq i\leq |\Sigma|^n$, the i-th word in the lexicographic order over Σ^n by $w_i^{(n)}$. The equalities

$$d(\mathrm{I}^s(G_l),w_i^{(n)}) = \sum_{1\leq j\leq i} d(\mathrm{I}^{s-1}(G_l),w_j^{(n)}), \qquad l=1,2,$$

imply the equality

$$\mathrm{diff}(\mathrm{I}^s(G_1),\mathrm{I}^s(G_2),w_i^{(n)}) = \sum_{1\leq j\leq i} \mathrm{diff}(\mathrm{I}^{s-1}(G_1),\mathrm{I}^{s-1}(G_2),w_j^{(n)}).$$

We are now in the position to apply Lemma 2.4: Define

$$a_{s,i} = \mathrm{diff}(\mathrm{I}^s(G_1),\mathrm{I}^s(G_2),w_i^{(n)}) \text{ and } t=|\Sigma|^n,$$

i. e., $w_t^{(n)} = a_m^n$. By our assumptions, we have $|\{j \mid a_{0,j}\neq 0,\ 1\leq j\leq t\}|\leq k$ and $a_{s,t} = \mathrm{diff}(\mathrm{I}^s(G_1),\mathrm{I}^s(G_2),a_m^n)=0$ for all $1\leq s\leq k$. Hence, for all $1\leq j\leq |\Sigma|^n$, $a_{0,j} = \mathrm{diff}(G_1,G_2,w_j^{(n)}) = 0$, i. e., $k_n = 0$. □

Theorem 2.7. *Let G be a regular grammar with* $\mathrm{d}(G,w)=1$, $w \in a_m^+$, *and* $\mathrm{d}(G,w)=0$, *otherwise. Let G_1 and G_2 be context-free grammars. Then*

$$G_1 \sim G_2 \qquad \textit{iff} \qquad \mathrm{I}^s(G_1) \odot G \sim \mathrm{I}^s(G_2) \odot G \textit{ for all } s \geq 1.$$

Proof. Assume that $G_1 \sim G_2$ is not valid. Then there exists an $n \geq 1$ such that $k_n > 0$, contradicting Lemma 2.6. □

Let $k \geq 0$. Two context-free grammars G_1 and G_2 are called *k-length close* iff, for all $n \geq 0$, $k_n \leq k$, where k_n is defined in Lemma 2.6. Obviously, $G_1 \sim G_2$ iff they are 0-length close. # words of length n where G_1, G_2 differ.

Theorem 2.8. *Let G be a regular grammar with* $\mathrm{d}(G,w)=1$, $w \in a_m^+$, *and* $\mathrm{d}(G,w)=0$, *otherwise. Let G_1 and G_2 be context-free grammars that are k-length close for some $k \geq 0$. Then*

$$G_1 \sim G_2 \qquad \textit{iff} \qquad \mathrm{I}^s(G_1) \odot G \sim \mathrm{I}^s(G_2) \odot G \textit{ for all } 1 \leq s \leq k.$$

Proof. Our assumptions yield $k_n \leq k$ for all $n \geq 0$, where k_n is from Lemma 2.6. $\mathrm{I}^s(G_1) \odot G \sim \mathrm{I}^s(G_2) \odot G$ implies by Lemma 2.6 $k_n = 0$ for all $n \geq 0$, i. e., $\mathrm{diff}(G_1,G_2,w) = 0$ for all $w \in \Sigma^n$, $n \geq 0$. This means $G_1 \sim G_2$. □

Corollary 2.9. *Let G_1 and G_2 be context-free grammars that are k-length close for some $k \geq 0$. Then it is decidable whether or not G_1 and G_2 are multiplicity equivalent.*

Proof. Apply Corollary 1.3. □

Recently, in a series of papers, Andrasiu, Dassow, Paun, Salomaa [1] and Paun, Salomaa [5], [6], [7] introduced and considered thin and slender languages. For $k \geq 0$ a language $L \subseteq \Sigma^*$ is called *properly k-thin* iff L contains at most k words of length n for all $n \geq 0$, i. e., $|L \cap \Sigma^n| \leq k$. A language is *slender* iff it is properly k-thin for some $k \geq 0$.

Corollary 2.10. *It is decidable whether or not two context-free grammars generating slender languages are multiplicity equivalent.*

There are a bounded number of words in the language of any fixed length.

Proof. Let G_1 and G_2 generate properly k_1- and k_2-thin context-free languages. Then G_1 and G_2 are (k_1+k_2)-length close. Apply Corollary 2.9. □

Corollary 2.11. *Let G_1 and G_2 be unambiguous context-free grammars. Furthermore, let $\mathrm{L}(G_1)$ be slender. Then it is decidable, whether or not $\mathrm{L}(G_1) = \mathrm{L}(G_2)$.*

Proof. By Andrasiu, Dassow, Paun, Salomaa [1], it is decidable for an unambiguous context-free grammar whether or not it is slender. If G_2 is not slender, $\mathrm{L}(G_2)$ is different from $\mathrm{L}(G_1)$. If G_2 is slender, apply Corollary 2.10. Since G_1 and G_2 are unambiguous, $\mathrm{L}(G_1) = \mathrm{L}(G_2)$ iff $G_1 \sim G_2$. □

As mentioned earlier, the results presented in the first part of this section are due to Raz [8]. Theorem 2.2 is Lemma 5 of [8]; Theorem 2.7 is Algorithm A of [8]; Theorem 2.8, Corollary 2.9 are Theorem 9 of [8] (but $\mathrm{l}^s(G_1) \sim \mathrm{l}^s(G_2)$ is replaced by our condition $\mathrm{l}^s(G_1) \odot G \sim \mathrm{l}^s(G_2) \odot G$); Corollary 2.10 is Corollary 10 of [8]; and Corollary 2.11 is Corollary 11 of [8].

We now turn to Raz [9]. The basic idea in this paper is to construct, for a given context- free grammar G with terminal alphabet $\Sigma = \{a_1, \ldots, a_m\}$, a context-free grammar G' such that, for all $w \in \Sigma^*$

$$\mathrm{d}(G',w) = \mathrm{o}(w)\,\mathrm{d}(G,w).$$

Here, $\mathrm{o}(w) = i$ iff w is the i-th word of length n in the lexicographic order on Σ^n, where $|w| = n$.

Let G_1 and G_2 be context-free grammars. Let G_1' and G_2' be context-free grammars such that, for all $w \in \Sigma^*$,

$$\mathrm{d}(G_l',w) = \mathrm{o}(w)\,\mathrm{d}(G_l,w), \qquad l=1,2.$$

Then, for all $w \in \Sigma^*$, we have

$$\mathrm{diff}(G_1',G_2',w) = \mathrm{o}(w)\,\mathrm{diff}(G_1,G_2,w).$$

Construction 2.12. Let $G=(\Phi,\Sigma,P,S)$ be a context-free grammar in binary Greibach normal form. Let $\bar{G}=(\bar{\Phi},\Sigma,\bar{P},\bar{S})$ be the context-free grammar defined by

(i) $\bar{\Phi}=\Phi\cup\Phi'\cup\Phi''\cup\{\bar{S}\}\cup\{T_b^a \mid a,b\in\Sigma\}$, where Φ' and Φ'' are primed and double-primed versions of Φ, respectively, and $\bar{S}$ and T_b^a are new variables;

(ii) $\bar{P}=P\cup P_0\cup P_1\cup P_2\cup\{\bar{S}\to S, \bar{S}\to S'\}\cup\{T_b^a\to a \mid a,b\in\Sigma\}$, where

$$
\begin{aligned}
P_0 &= \{A'\to T_c^a,\ A''\to T_b^a \mid A\to a\in P,\ b,c\in\Sigma,\ c<a\},\\
P_1 &= \{A'\to aA_1',\ A'\to T_c^a A_1'',\ A''\to T_b^a A_1'' \mid A\to aA_1\in P,\ b,c\in\Sigma,\ c<a\},\\
P_2 &= \{A'\to aA_1A_2',\ A'\to aA_1'A_2'',\ A'\to T_c^a A_1''A_2'',\ A''\to T_b^a A_1''A_2'' \mid\\
&\qquad\qquad\qquad\qquad A\to aA_1A_2\in P,\ b,c\in\Sigma,\ c<a\}.
\end{aligned}
$$

Then we obtain

$$\mathrm{d}(\bar{G},w) = \mathrm{o}(w)\,\mathrm{d}(G,w) \qquad \text{for all } w\in\Sigma^*. \qquad \square$$

Theorem 2.13. *For a given context-free grammar G, where $\mathrm{d}(G,w)$ is finite for all words w, a context-free grammar G' can effectively be constructed such that, for all w,*

$$\mathrm{d}(G',w) = \mathrm{o}(w)\,\mathrm{d}(G,w).$$

Proof. Analogous to the proof of Theorem 2.2; use Construction 2.12 instead of Construction 2.1. $\square$

In the sequel, we denote the grammar G' of Theorem 2.13 by $\mathrm{K}(G)$. Iteration of this construction yields the context-free grammars $\mathrm{K}^s(G)$, $s\geq 0$: $\mathrm{K}^0(G)=G$, $\mathrm{K}^{i+1}(G)=\mathrm{K}(\mathrm{K}^i(G))$, $i\geq 0$. We have

$$\mathrm{d}(\mathrm{K}^s(G),w) = \mathrm{o}(w)^s\,\mathrm{d}(G,w)$$

and

$$\mathrm{diff}(\mathrm{K}^s(G_1),\mathrm{K}^s(G_2),w) = \mathrm{o}(w)^s\,\mathrm{diff}(G_1,G_2,w)$$

for context-free grammars G, G_1, G_2 and all $s\geq 0$, $w\in\Sigma^*$.

Lemma 2.14. *Let $k\geq 0$. Assume that G_1 and G_2 are context-free grammars such that, for all $0\leq s\leq k-1$,*

$$\mathrm{K}^s(G_1) \sim_{\#} \mathrm{K}^s(G_2).$$

Furthermore, let $(n_1,\dots,n_m)\in\mathbb{N}^m$ *and assume that*

$$k_{n_1,\dots,n_m} := |\{w \mid \mathrm{diff}(G_1,G_2,w)\neq 0,\ \psi(w)=(n_1,\dots,n_m)\}| \le k.$$

Then $k_{n_1,\dots,n_m}=0$.

Proof. There are at most k different words $w\in\Sigma^*$, $\psi(w)=(n_1,\dots,n_m)$, such that possibly $\mathrm{diff}(G_1,G_2,w)\neq 0$. Denote these words by $w_1,\dots,w_l$, $l\le k$. We have, for all $s\ge 0$,

$$\sum_{1\le i\le l}\mathrm{diff}(\mathrm{K}^s(G_1),\mathrm{K}^s(G_2),w_i) = \sum_{1\le i\le l}\mathrm{o}(w_i)^s\,\mathrm{diff}(G_1,G_2,w_i).$$

Hence, by our assumption $\mathrm{K}^s(G_1)\sim_{\#}\mathrm{K}^s(G_2)$, we obtain, for all $0\le s\le l-1$,

$$\sum_{1\le i\le l}\mathrm{o}(w_i)^s\,\mathrm{diff}(G_1,G_2,w_i) = 0.$$

Since the Vandermonde matrix is nonsingular, we obtain $\mathrm{diff}(G_1,G_2,w_i)=0$ for all $1\le i\le l$, i.e., $k_{n_1,\dots,n_m}=0$. □

Theorem 2.15. *Let* G_1 *and* G_2 *be context-free grammars. Then*

$$G_1\sim G_2 \qquad \textit{iff} \qquad \mathrm{K}^s(G_1)\sim_{\#}\mathrm{K}^s(G_2)\ \textit{for all } s\ge 1.$$

Proof. Assume that $G_1\sim G_2$ is not valid. Then there exists $(n_1,\dots,n_m)\in\mathbb{N}^m$ such that $k_{n_1,\dots,n_m}>0$ contradicting Lemma 2.14. □

Let $k\ge 0$. Two context-free grammars G_1 and G_2 are called *k-letter count close* iff, for all $(n_1,\dots,n_m)\in\mathbb{N}^m$, $k_{n_1,\dots,n_m}\le k$, where $k_{n_1,\dots,n_m}$ is defined in Lemma 2.14. Obviously $G_1\sim G_2$ iff they are 0-letter count close.

Theorem 2.16. *Let* G_1 *and* G_2 *be context-free grammars that are k-letter count close for some* $k\ge 0$. *Then*

$$G_1\sim G_2 \qquad \textit{iff} \qquad \mathrm{K}^s(G_1)\sim_{\#}\mathrm{K}^s(G_2)\ \textit{for all } 0\le s\le k-1.$$

Proof. Our assumptions yield $k_{n_1,\dots,n_m}\le k$ for all $(n_1,\dots,n_m)\in\mathbb{N}^m$, where

$k_{n_1,\ldots,n_m}$ is from Lemma 2.14. $\mathrm{K}^s(G_1) \sim_{\#} \mathrm{K}^s(G_2)$ implies by Lemma 2.14 $k_{n_1,\ldots,n_m} = 0$ for all $(n_1,\ldots,n_m) \in \mathbb{N}^m$, i.e., $\mathrm{diff}(G_1,G_2,w) = 0$ for all $w \in \Sigma^*$. This means $G_1 \sim G_2$. □

Corollary 2.17. *Let G_1 and G_2 be context-free grammars that are k-letter count close for some $k \geq 0$. Then it is decidable whether or not G_1 and G_2 are multiplicity equivalent.*

Proof. Apply Theorem 1.1. □

Clearly, Corollaries 2.10 and 2.11 are at once proved again by Corollary 2.17.

As mentioned earlier, the results presented in the second part of this section are due to Raz [9]. Theorem 2.13 is Lemma 6 of [9]; Theorem 2.16 and Corollary 2.17 are Theorem 8 of [9].

3. Generalizations

In this section we will generalize the results of Raz [8], [9] as presented in Section 2. Before stating one of these generalizations, we have to generalize the combinatorial Lemma 2.4.

Lemma 3.1. *Let $a_{k,i} \in \mathbb{R}$, $k \geq 0$, $i \geq 1$, and assume $a_{k,i} = \sum_{1 \leq j \leq i} a_{k-1,j}$ for all $k, i \geq 1$. Let $k, l > 0$ and $t_0 = 1 \leq t_1 < \ldots < t_l$. Assume $a_{s,t_r} = 0$ for all $1 \leq s \leq k$, $1 \leq r \leq l$ and, for all $0 \leq r \leq l-1$, $|\{j \mid a_{0,j} \neq 0,\ t_r \leq j \leq t_{r+1}\}| \leq k$. Then $a_{0,j} = 0$ for all $1 \leq j \leq t_l$.*

Proof. The proof is by induction on l. For $l=1$, Lemma 3.1 is identical to Lemma 2.4. Hence, assume $l>1$. By definition, let $s_0=1, s_1=t_2, \ldots, s_{l-1}=t_l$. Then we obtain by Lemma 2.4

$$\{j \mid a_{0,j} \neq 0,\ s_0 \leq j \leq s_1\} = \{j \mid a_{0,j} \neq 0,\ t_1 \leq j \leq t_2\}.$$

Hence, for all $0 \leq r \leq l-2$, $|\{j \mid a_{0,j} \neq 0,\ s_r \leq j \leq s_{r+1}\}| \leq k$. We obtain now by our induction hypothesis $a_{0,j}=0$ for all $1 \leq j \leq s_{l-1} = t_l$. □

Theorem 3.2 is a generalization of Theorem 2.8.

Theorem 3.2. *Let $k,l>0$. Let G be an unambiguous regular grammar. Assume that G_1 and G_2 are context-free grammars such that, for all $n\geq 1$, the following condition is satisfied: There exist words $v_0=a_1^n<v_1<\dots$ $\dots<v_{l-1}<v_l=a_m^n$ of length n in $\mathrm{L}(G)$ such that*

$$|\{w \mid \mathrm{diff}(G_1,G_2,w)\neq 0,\ v_r\leq w\leq v_{r+1}\}|\leq k \qquad \text{for all } 0\leq r\leq l-1.$$

Moreover, assume that

$$\mathrm{I}^s(G_1)\odot G\sim \mathrm{I}^s(G_2)\odot G \qquad \text{for all } 1\leq s\leq k.$$

Then $G_1\sim G_2$.

Proof. Let $n\geq 1$. Define $a_{s,i}$ as in the proof of Lemma 2.6:

$$a_{s,i}=\mathrm{diff}(\mathrm{I}^s(G_1),\mathrm{I}^s(G_2),w_i^{(n)}).$$

By $\mathrm{I}^s(G_1)\odot G\sim \mathrm{I}^s(G_2)\odot G$ we obtain

$$a_{s,\mathrm{o}(v_r)}=0 \text{ for } 1\leq s\leq k,\ 0\leq r\leq l.$$

Our condition and Lemma 3.1 yield $\mathrm{diff}(G_1,G_2,w)=0$ for all w with $1\leq \mathrm{o}(w)\leq|\Sigma|^n$, i. e., for all words of length n. □

We now use a construction slightly different from that in Theorem 3.2.

Lemma 3.3. *Let $k,l>0$. Let G be an unambiguous regular grammar. Assume that G_1 and G_2 are context-free grammars such that, for some $n\geq 1$, the following conditions are satisfied:*

(i) There exist words $v_0=a_1^n<v_1<\dots<v_{l-1}<v_l$ of length n in $\mathrm{L}(G)$ such that

$$|\{w \mid \mathrm{diff}(G_1,G_2,w)\neq 0,\ v_r\leq w\leq v_{r+1}\}|\leq k \qquad \text{for all } 0\leq r\leq l-1.$$

(ii) $\mathrm{diff}(\mathrm{I}(\mathrm{K}^s(G_1))\odot G,\mathrm{I}(\mathrm{K}^s(G_2))\odot G,w)=0$ *for all* $0\leq s\leq k-1,\ w\in\Sigma^n$.

Then $\mathrm{diff}(G_1,G_2,w)=0$ *for all $w\in\Sigma^n$, such that $w\leq v_l$.*

Proof. The proof is by induction on l. Let $l=1$. There are at most k words w, $w \leq v_1$, of length n such that possibly $\mathrm{diff}(G_1,G_2,w) \neq 0$. Denote these words by $w_1,\ldots,w_{k'}$, $k' \leq k$. Condition (ii) yields $\mathrm{diff}(\mathrm{I}(\mathrm{K}^s(G_1)) \odot G, \mathrm{I}(\mathrm{K}^s(G_2)) \odot G, v_1) = 0$ for $0 \leq s \leq k-1$, i.e.,

$$o(w_1)^s \mathrm{diff}(G_1,G_2,w_1) + \ldots + o(w_{k'})^s \mathrm{diff}(G_1,G_2,w_{k'}) = 0$$

for $0 \leq s \leq k'-1$. By the nonsingularity of the Vandermonde matrix we obtain

$$\mathrm{diff}(G_1,G_2,w_1) = \ldots = \mathrm{diff}(G_1,G_2,w_{k'}) = 0.$$

Hence $\mathrm{diff}(G_1,G_2,w) = 0$ for all words $w \leq v_1$ of length n.

Let now $l>1$. By definition let $u_0 = a_1^n$, $u_1 = v_2, \ldots, u_{l-1} = v_l$. Then we obtain (by the case $l=1$)

$$\{w \mid \mathrm{diff}(G_1,G_2,w) \neq 0,\ u_0 \leq w \leq u_1\} = \{w \mid \mathrm{diff}(G_1,G_2,w) \neq 0,\ v_1 \leq w \leq v_2\}.$$

Hence, for all $0 \leq r \leq l-2$, $|\{w \mid \mathrm{diff}(G_1,G_2,w) \neq 0,\ u_r \leq w \leq u_{r+1}\}| \leq k$. We obtain now by our induction hypothesis $\mathrm{diff}(G_1,G_2,w) = 0$ for $w \leq u_{l-1} = v_l$. □

Our next theorem is similar to Theorem 3.2.

Theorem 3.4. *Let $k,l>0$. Let G be an unambiguous regular grammar. Assume that G_1 and G_2 are context-free grammars such that, for all $n \geq 1$, the following condition is satisfied: There exist words $v_0 = a_1^n < v_1 < \ldots \ldots < v_{l-1} < v_l = a_m^n$ of length n in $\mathrm{L}(G)$ such that*

$$|\{w \mid \mathrm{diff}(G_1,G_2,w) \neq 0,\ v_r \leq w \leq v_{r+1}\}| \leq k \qquad \text{for all } 0 \leq r \leq l-1.$$

Moreover, assume that

$$\mathrm{I}(\mathrm{K}^s(G_1)) \odot G \sim \mathrm{I}(\mathrm{K}^s(G_2)) \odot G \qquad \text{for all } 1 \leq s \leq k-1.$$

Then $G_1 \sim G_2$.

Proof. Conditions (i) and (ii) of Lemma 3.3 are satisfied by our assumptions for all $n \geq 1$. □

Before our last generalization, we need three lemmas.

Lemma 3.5. *Let $k \geq 1$, $p \geq 2k+1$. Let $l \geq 0$ and, for all $0 \leq j \leq l$, $k_j \in \mathbb{Z}$* with $0 \leq |k_j| \leq k$. *Then*

$$k_0 + k_1 p + \ldots + k_l p^l = 0 \qquad \textit{implies} \qquad k_j = 0 \ \textit{for all}\ 0 \leq j \leq l.$$

Proof. The proof is by induction on l. Since for $l=0$ Lemma 3.5 is valid, we proceed with $l>0$. The equation $k_0 + k_1 p + \ldots + k_l p^l = 0$ implies $|k_l|\, p^l \leq k\cdot(p^l-1)/(p-1)$. Hence, $|k_l|\, p^{l+1} \leq (k+|k_l|)p^l - k$. If $k_l \neq 0$, this inequality implies $p^{l+1} \leq 2kp^l$, i.e., $p \leq 2k$. This contradicts the assumption $p \geq 2k+1$. □

Lemma 3.6. *For a given context-free grammar G, for $a \in \Sigma$ and $p \geq 1$, a context-free grammar G' can be effectively constructed such that, for all w,* $\mathrm{d}(G',w) = p^{|w|_a}\mathrm{d}(G,w)$.

Proof. In the productions of G, replace each occurrence of $a \in \Sigma$ by the new nonterminals T_j and add productions $T_j \to a$, $1 \leq j \leq p$. □

In the sequel, we denote the grammar G' of Lemma 3.6 by $\mathrm{L}_p^a(G)$.

A context-free grammar G is called *d-ambiguous* iff $\mathrm{d}(G,w) \leq d$ for all $w \in \Sigma^*$.

Lemma 3.7. *Let G be an unambiguous regular grammar. Assume that G_1 and G_2 are d-ambiguous context-free grammars such that, for some $n,l \geq 1$ and $a \in \Sigma$, the following conditions are satisfied:*

(i) There exist words $v_0 = a_1^n < v_1 < \ldots < v_{l-1} < v_l$ of length n in $\mathrm{L}(G)$ *such that for all $0 \leq r \leq l-1$ and words w_1, w_2,*

$$v_r \leq w_1 < w_2 \leq v_{r+1},\ |w_1|_a = |w_2|_a \qquad \textit{imply} \qquad \mathrm{diff}(G_1,G_2,w_1) = 0 \ \textit{or}\ \mathrm{diff}(G_1,G_2,w_2) = 0.$$

(ii) $\mathrm{diff}(\mathrm{l}(\mathrm{L}_{2d+1}^a(G_1)) \odot G, \mathrm{l}(\mathrm{L}_{2d+1}^a(G_2)) \odot G, w) = 0$ *for all $w \in \Sigma^n$.*

Then $\mathrm{diff}(G_1,G_2,w) = 0$ *for all $w \in \Sigma^n$, such that $w \leq v_l$.*

Proof. The proof is by induction on l. Let $l=1$. By condition (ii), we obtain

$$\mathrm{diff}(\mathrm{I}(\mathrm{L}^{a}_{2d+1}(G_1)),\mathrm{I}(\mathrm{L}^{a}_{2d+1}(G_2)),v_1) = \sum_{v_0 \le w \le v_1} (2d+1)^{|w|_a} \mathrm{diff}(G_1,G_2,w) = 0.$$

By condition (i), there exists for each exponent j at most one word w, $v_0 \le w \le v_1$, such that $\mathrm{diff}(G_1,G_2,w) \neq 0$ and $|w|_a = j$. Hence, Lemma 3.5 implies $\mathrm{diff}(G_1,G_2,w) = 0$ for all words w, $v_0 \le w \le v_1$.

Let now $l>1$. By definition, let $u_0 = a_1^n$, $u_1 = v_2, \ldots, u_{l-1} = v_l$. Then we obtain (by the case $l=1$) that condition (i) is valid also for $u_0, u_1, \ldots, u_{l-1}$. Application of our induction hypothesis yields $\mathrm{diff}(G_1,G_2,w) = 0$ for all words w, $u_0 = v_0 \le w \le u_{l-1} = v_l$. □

Theorem 3.8. *Let G be an unambiguous regular grammar. Assume that G_1 and G_2 are d-ambiguous context-free grammars such that, for all $n \ge 1$, some $l \ge 1$ and some $a \in \Sigma$, the following conditions are satisfied:*

(i) There exist words $v_0 = a_1^n < v_1 < \ldots < v_{l-1} < v_l = a_m^n$ of length n in $\mathrm{L}(G)$ such that, for all $0 \le r \le l-1$ and words w_1, w_2,

$$v_r \le w_1 < w_2 \le v_{r+1},\ |w_1|_a = |w_2|_a \quad \textit{imply} \quad \mathrm{diff}(G_1,G_2,w_1) = 0 \ \textit{or}\ \mathrm{diff}(G_1,G_2,w_2) = 0.$$

(ii) $\mathrm{I}(\mathrm{L}^{a}_{2d+1}(G_1)) \odot G \sim \mathrm{I}(\mathrm{L}^{a}_{2d+1}(G_2)) \odot G$.

Then $G_1 \sim G_2$.

Proof. Conditions (i) and (ii) of Lemma 3.7 are satisfied by our assumptions for all $n \ge 1$. □

4. Applications

We now apply the results of Section 3. A class $\mathfrak{G}$ of context-free grammars has a *decidable multiplicity equivalence problem* iff, whenever G_1 and G_2 are in $\mathfrak{G}$, then $G_1 \sim G_2$ is decidable.

The classes of context-free grammars generating acommutative, bounded and slender languages have a decidable multiplicity equivalence problem (by Theorem 1.2; by Corollary 12 of Raz [9]; by Corollary 2.10; respectively). By help of these classes we define new classes of context-free grammars that also have a decidable multiplicity equivalence problem.

We consider languages of the form

$$L_1 = \bigcup_{\alpha \in L} \alpha a_m L_\alpha,$$

where $L \subseteq \Sigma^*$ and $L_\alpha \subseteq (\Sigma - \{a_m\})^*$, $\alpha \in L$. Let $n > 0$ be fixed, consider all words of length n in the language La_m^+ and order these words:

$$\alpha_1 a_m^{n-|\alpha_1|} < \alpha_2 a_m^{n-|\alpha_2|} < \ldots < \alpha_{l-1} a_m^{n-|\alpha_{l-1}|},$$

$\alpha_r \in L$, $1 \le r \le l-1$. Extend this notation by $\alpha_0 = a_1^n$ and $\alpha_l = \varepsilon$.

Each word $w \in L_1$ of length n has the form $w = \alpha_r a_m v$, $v \in L_{\alpha_r}$, $|v| = n - |\alpha_r| - 1$, where r is uniquely determined. Moreover, we have the inequalities

$$\alpha_{r-1} a_m^{n-|\alpha_{r-1}|} < w \le \alpha_r a_m^{n-|\alpha_r|}.$$

We now consider two conditions on all L_α, $\alpha \in L$:

(1) L_α is properly k-thin for some $k \ge 0$. (Hence, $L_\alpha = \emptyset$ is allowed; in this case $\alpha a_m L_\alpha = \emptyset$.)

(2) If $v_1, v_2 \in L_\alpha$, $|v_1| = |v_2|$, then $|v_1|_{a_1} \neq |v_2|_{a_1}$.

Assume now that condition (1) holds for all L_α. Then we obtain, for all $0 \le r \le l-1$,

$$|\{w \mid \alpha_r a_m^{n-|\alpha_r|} \le w \le \alpha_{r+1} a_m^{n-|\alpha_{r+1}|},\ w \in L_1\}| \le k+1.$$

Consider now two context-free grammars G_1 and G_2 that generate languages of the considered form

$$\mathrm{L}(G_1) = \bigcup_{\alpha \in L} \alpha a_m L_\alpha^1, \qquad \mathrm{L}(G_2) = \bigcup_{\alpha \in L} \alpha a_m L_\alpha^2,$$

such that condition (1) holds for L_α^1 and L_α^2. Let G be an unambiguous regular grammar generating the language La_m^+. Then we obtain by Theorem 3.2 that $\mathrm{I}^s(G_1) \odot G \sim \mathrm{I}^s(G_2) \odot G$, $1 \le s \le k+1$, implies $G_1 \sim G_2$. By Theorem 3.4 we obtain that $\mathrm{I}(\mathrm{K}^s(G_1)) \odot G \sim \mathrm{I}(\mathrm{K}^s(G_2)) \odot G$, $0 \le s \le k$, implies $G_1 \sim G_2$. Either of these implications yields the next theorem.

Theorem 4.1. *Consider the class $\mathfrak{L}$ of context-free languages of the form*

$$\bigcup_{\alpha \in L} \alpha a_m L_\alpha,$$

where La_m^+ is a regular acommutative, bounded or slender language and, for all $\alpha \in L$, L_α is a properly k-thin language, $k \ge 0$, not containing a_m. Then the class of context-free grammars generating languages in $\mathfrak{L}$ has a decidable multiplicity equivalence problem. □

For the proof of our next theorem, we assume that condition (2) holds for all L_α. Consider two d-ambiguous context-free grammars G_1, G_2 that generate languages

$$\mathrm{L}(G_1) = \bigcup_{\alpha \in L} \alpha a_m L_\alpha^1, \qquad \mathrm{L}(G_2) = \bigcup_{\alpha \in L} \alpha a_m L_\alpha^2,$$

such that condition (2) holds for L_α^1 and L_α^2. Let G be an unambiguous regular grammar generating the language La_m^+. Then we obtain by Theorem 3.8 that $\mathrm{I}(\mathrm{L}_{2d+1}^{a_1}(G_1)) \odot G \sim \mathrm{I}(\mathrm{L}_{2d+1}^{a_1}(G_2)) \odot G$ implies $G_1 \sim G_2$.

A definition is needed before our last theorem. A context-free grammar G is of *bounded ambiguity* iff there exists a $d \ge 1$ such that G is d-ambiguous.

Theorem 4.2. *Consider the class $\mathfrak{L}$ of context-free languages of the form*

$$\bigcup_{\alpha \in L} \alpha a_m L_\alpha,$$

where La_m^+ is a regular acommutative, bounded or slender language and, for all $\alpha \in L$, L_α is a language not containing a_m and satisfying condition (2). Then the class of context-free grammars of bounded ambiguity generating languages in $\mathfrak{L}$ has a decidable multiplicity equivalence problem. □

References

[1] M. Andrasiu, J. Dassow, Gh. Paun, A. Salomaa: *Language-theoretic problems arising from Richelieu cryptosystems.* Theoretical Computer Science **116** (1993), 339-357.

[2] S. Ginsburg: *The Mathematical Theory of Context-Free Languages.* McGraw-Hill, 1966.

[3] T. Harju, J. Karhumäki: *The equivalence problem of multitape finite automata.* Theoretical Computer Science **78** (1991), 347-355.

[4] W. Kuich, A. Salomaa: *Semirings, Automata, Languages.* Springer, 1986.

[5] Gh. Paun, A. Salomaa: *Thin and slender languages.* To appear in: Discrete Applied Mathematics.

[6] Gh. Paun, A. Salomaa: *Closure properties of slender languages.* Theoretical Computer Science **120** (1993), 293-301.

[7] Gh. Paun, A. Salomaa: *Decision problems concerning the thinness of* DOL *languages.* Bulletin of the EATCS **46** (1992), 171-181.

[8] D. Raz: *Deciding multiplicity equivalence of certain context-free languages.* Manuscript (1993), first version.

[9] D. Raz: *Deciding multiplicity equivalence of certain context-free languages.* Manuscript (1993), second version, to appear in Proceedings of the Conference on Developments in Language Theory (G. Rozenberg, A. Salomaa eds.), World Scientific Publ. Co., 1994.

[10] A. Salomaa: *Formal Languages.* Academic Press, 1973.