Unambiguous Turn Position and Rational Trace Languages

Massimiliano Goldwurm[1] and Klaus Wich[2]

[1] Dip. Scienze dell'Informazione, Università degli Studi di Milano, Via Comelico 39/41, 20135 Milano
goldwurm@dsi.unimi.it

[2] Institut für formale Methoden der Informatik, Universität Stuttgart, Universitätstr. 38, 70569 Stuttgart
Wich@informatik.uni-stuttgart.de

Technical report - March 2005

Abstract We show the existence of rational trace languages defined over direct products of free monoids that have inherent ambiguity of the order of $\log n$ and $n^{1/2}$. This result is obtained by studying the relationship between trace languages and linear context-free grammars that satisfy a special unambiguity condition on the position of the last step of derivation.

Keywords: Automata and Formal Languages, Trace Monoids, Inherent Ambiguity of rational trace languages, Linear Context-free Languages.

1 Introduction

In this paper we study the inherent ambiguity of rational trace languages. Given a trace monoid M, defined over an independence alphabet (Σ, I), the ambiguity function of a regular (string) language $L \subseteq \Sigma^*$ is defined as the map associating each $n \in \mathbb{N}$ with the maximum number of elements in L representing a trace $t \in M$ of length smaller or equal to n. This function defines the ambiguity of L as representative set of strings for the corresponding (rational) trace language $T = \{t \in M \mid t \cap L \neq \emptyset\}$. A notion of inherent ambiguity of a rational trace language $T \subseteq M$ can also be given by considering the smallest ambiguity function of a regular string language that represents T. Rational trace languages of bounded inherent ambiguity have been studied in [4, 5] (see also [3]). A rational trace language T is inherently k-ambiguous, where k is a positive integer, if T is represented by a regular string language whose ambiguity function is bounded by k, but it is not represented by any regular string language with ambiguity function bounded by k-1. A rational trace language is unambiguous if it is inherently 1-ambiguous. It is known that if the independence relation is transitive then all rational trace languages are unambiguous [5, 12]. On the contrary, if the independence relation is not transitive, then for any positive integer k there are rational trace languages that are inherently k-ambiguous [4].

It is also known that there exist rational trace languages of unbounded inherent ambiguity, meaning that the ambiguity function of every representative regular string language grows to infinity as the length of words increases [4]. In the present work, we evaluate the order of growth of the (unbounded) inherent ambiguity of some rational trace languages. We show that there exist two rational trace languages, defined over the direct product of free monoids, whose inherent ambiguities are of the order of $\log n$ and $n^{1/2}$, respectively. This result is obtained by studying the relationship between such rational trace languages and the class (denoted by $\mathcal{G}_\ell$) of linear context-free grammars that satisfy a special unambiguity condition on the position of the last step of derivation. We say that $\mathcal{G}_\ell$ is the class of grammars *with unambiguous turn position* because they are related to one-turn pushdown automata satisfying the following additional property: All accepting computations on a given input turn the pushdown (i.e. enter the popping phase) while the input tape head is reading the same cell. We prove that for any language L generated by a grammar in $\mathcal{G}_\ell$, there exists a rational trace language T over a direct product $A^* \times B^*$ of two free monoids, whose inherent ambiguity is given by the inherent ambiguity of L (in the traditional sense) restricted to the grammars in $\mathcal{G}_\ell$. This property allows us to apply the results presented in [14] concerning the existence of sublinear inherent ambiguity functions of linear context-free languages.

These results are also related to the analysis of the maximum coefficients of rational formal series defined over various types of monoids [7, 6]. The general goal of that line of research is to find out how the well-known asymptotic behaviour of the coefficients of rational series in noncommutative variables changes when a partial or total commutation of variables is allowed. Our results in the present work can be included in that general frame since the ambiguity function of any regular grammar, generating a set of representatives of a given (rational) trace language, is given by the maximum coefficients of a rational formal series in partially commutative variables.

2 Basic notions

To fix notation we recall some basic notions concerning formal languages and grammars. For a finite alphabet Σ we denote by Σ^* the corresponding free monoid and by ε the empty word. For any subset $\Gamma \subseteq \Sigma$ and every $x \in \Sigma^*$, $\pi_\Gamma(x)$ denotes the projection of x over Γ, i.e. the string obtained from x by erasing all occurrences of letters not included in Γ. The length of x is denoted by $|x|$, and $|x|_a := |\pi_{\{a\}}(x)|$ is the number of occurrences of the symbol $a \in \Sigma$ in x. The reversal of x is denoted by x^R (so, for instance, $\{x \in \Sigma^* \mid x = x^R\}$ is the set of palindromes over Σ). Finally, for every $n \in \mathbb{N}$, we define $\Sigma^{\leq n} = \{x \in \Sigma^* \mid |x| \leq n\}$.

A context-free (c.f. for short) grammar is a 4-tuple $G = (N, \Sigma, P, S)$, where N and Σ are finite disjoint alphabets representing, respectively, the set of nonterminals and the set of terminal symbols, $S \in N$ is the initial symbol and $P \subseteq N \times (\Sigma \cup N)^*$ a finite set of productions. Any production (A, α) can be also represented in the form $A \to \alpha$. The productions (A, x), where $A \in N$ and $x \in \Sigma^*$, are called *terminal* productions. The one-step derivation $\Longrightarrow_G$ is defined as a binary relation over $(\Sigma \cup N)^*$ by setting $\beta A \gamma \Longrightarrow_G \beta\alpha\gamma$ for every $(A, \alpha) \in P$ and every $\beta, \gamma \in (\Sigma \cup N)^*$. The relation $\Longrightarrow_G^*$ is the reflexive and transitive closure of $\Longrightarrow_G$ and the language $L(G)$ generated by G is the set $\{x \in \Sigma^* \mid S \Longrightarrow_G^* x\}$. A language L is context-free if it is generated by a c.f. grammar. The grammar $G = (N, \Sigma, P, S)$ is said to be *linear* if every nonterminal production in P is of the form (A, xBy) where $A, B \in N$ and $x, y \in \Sigma^*$. The language generated by such a grammar is called linear c.f. language.

A derivation of a word $\beta \in (\Sigma \cup N)^*$ from a symbol $A \in N$ is a sequence of elements $\alpha_1, \alpha_2, \ldots, \alpha_n \in (\Sigma \cup N)^*$ such that $A = \alpha_1$, $\beta = \alpha_n$, $\alpha_i \Longrightarrow_G \alpha_{i+1}$ for $i = 1, 2, \ldots, n-1$; the derivation is *leftmost* if each α_{i+1} is obtained by applying a production to the leftmost nonterminal symbol in α_i. Hence a leftmost derivation from A to β can be represented by a sequence of productions $\tau \in P^*$, and we denote it by $A \overset{\tau}{\Longrightarrow}_G \beta$. Note that each derivation of a linear c.f. grammar is leftmost.

3 Ambiguity functions of grammars and languages

A c.f. grammar $G = (N, \Sigma, P, S)$ is called *cycle-free* if for every $A \in N$ and every $\tau \in P^*$, $A \overset{\tau}{\Longrightarrow}_G A$ implies $\tau = \varepsilon$. We also say that G is reduced if every $A \in N$ appears in a derivation of some word $x \in L(G)$ from S. Also, G is said to be *proper* if it is cycle-free and reduced. It is well-known that every c.f. grammar can be transformed into a proper c.f. grammar generating the same language. If G is a proper c.f. grammar then, for every word $x \in L(G)$, there exists a finite number of leftmost derivations $S \overset{\tau}{\Longrightarrow}_G x$; we denote such a number by $d_G(x)$. Moreover, for every $n \in \mathbb{N}$, let $\hat{d}_G(n)$ be the maximum value $d_G(x)$ for $x \in \Sigma^{\leq n}$. Note that the function $\hat{d}_G : \mathbb{N} \longrightarrow \mathbb{N}$ is monotone nondecreasing. We call $\hat{d}_G$ the *ambiguity function* of G. The grammar G is said to be unambiguous if $\hat{d}_G(n) \leq 1$ for every $n \in \mathbb{N}$. In the following we only deal with proper c.f. grammars G, and hence the corresponding functions d_G and $\hat{d}_G$ are well-defined.

A c.f. language L is unambiguous if it is generated by an unambiguous c.f. grammar. Otherwise L is said to be inherently ambiguous; in this case we say that L is inherently k-ambiguous, for some integer $k \geq 2$, if L can be generated by some c.f. grammar whose ambiguity function is bounded by k, but no c.f. grammar with ambiguity function bounded by $k-1$ can generate L. This notion can be extended to nonconstant functions. However, before stating this definition formally we observe that, for any inherently k-ambiguous c.f. language with $k \geq 2$, a shortest word with at least k leftmost derivations in a c.f. grammar generating L can be arbitrarily long (its length depending on the grammar). For c.f. languages L with an infinite inherent ambiguity it is often possible to specify the length of a shortest word with a certain ambiguity up to a constant factor depending on the c.f. grammar chosen to generate L. On the contrary, for any fixed length n there is a c.f. grammar G generating L such that all words of length at most n are generated by G unambiguously. These remarks suggest the use of the following notation (introduced in [18]) to define the inherent ambiguity functions of c.f. languages.

Let $f : N \longrightarrow \mathbb{R}_+$ be a monotone nondecreasing function. Then we define

$$
\begin{aligned}
O^T(f) = \{g : N \longrightarrow \mathbb{R}_+ \mid g \text{ is monotone nondecreasing},\\
\exists c, n_0 \in \mathbb{N} : \forall n \geq n_0 \;\; g(n) \leq f(cn)\},\\
\Omega^T(f) = \{g : N \longrightarrow \mathbb{R}_+ \mid g \text{ is monotone nondecreasing}, \; f \in O^T(g)\},\\
\Theta^T(f) = O^T(f) \cap \Omega^T(f)
\end{aligned}
$$

The remarks given below follow from the previous definitions and allow a comparisons with the standard set $\Theta(f)$ representing functions of the same asymptotic order of growth as $f(n)$.

1. For every pair of distinct $k, j \in \mathbb{N}\backslash\{0\}$, $\Theta^T(k) \neq \Theta^T(j)$ while $\Theta(k) = \Theta(j)$;
2. For every pair of distinct $a, b > 0$, $\Theta^T(a \log n) \neq \Theta^T(b \log n)$ while $\Theta(a \log n) = \Theta(b \log n)$;
3. For every $a > 0$, $\Theta^T(n^a) = \Theta(n^a)$;
4. For every distinct $a, b > 1$, $\Theta^T(a^n) = \Theta^T(b^n)$ while $\Theta(a^n) \neq \Theta(b^n)$.

Note that for sublinear functions $f(n)$ the class $\Theta^T(f)$ is a refinement of $\Theta(f)$. The two classes coincide for polynomial $f(n)$'s, while $\Theta(f)$ is a refinement of $\Theta^T(f)$ for exponential functions $f(n)$.

Definition 1 *Let $\mathcal{G}$ be a family of c.f. grammars and let $f : \mathbb{N} \longrightarrow \mathbb{R}_+$ be a monotone nondecreasing function. We say that a c.f. language L is inherently f-ambiguous* with respect to $\mathcal{G}$ *if:*

1. *$L = L(G)$ for some c.f. grammar $G \in \mathcal{G}$ such that $\hat{d}_G \in \Theta^T(f)$, and*
2. *for every c.f. grammar $G' \in \mathcal{G}$ such that $L = L(G')$ we have $\hat{d}_{G'} \in \Omega^T(f)$, i.e. there exists $c \in \mathbb{N}$ such that $f(n) \leq \hat{d}_{G'}(cn)$ for all positive $n \in \mathbb{N}$ large enough.*

We simply say that L is inherently f-ambiguous if it is inherently f-ambiguous with respect to the class of context-free grammars.

The ambiguity function of c.f. grammars and languages has recently been studied in [14, 15, 16]. In particular it is known that the ambiguity function of any c.f. grammar either is polynomially bounded or it has an exponential growth (Gap theorem). A similar gap does not exist between constant ambiguity functions and polynomially bounded ambiguity functions. In [15] linear c.f. grammars are shown which have ambiguity functions of the order $\Theta^T(\log n)$ or $\Theta^T(\sqrt{n})$. Finally, in [16] it is shown that if f is the ambiguity function of some c.f. grammar then there exists some c.f. language that is inherently f-ambiguous.

4 Context-free languages with one unambiguous turn

In this work we are interested in the c.f. grammars corresponding to those one-turn pushdown automata that satisfy the following additional condition: All accepting computations over an arbitrary input x turn the pushdown after reading the same prefix of x. In order to introduce them formally, consider a c.f. grammar $G = (N, \Sigma, P, S)$ such that $P \subseteq N \times (\Sigma^* N \Sigma^* \cup \{\varepsilon\})$. Note that any linear c.f. language can be generated by a grammar of this form. Given $x \in L(G)$, we say $n \in \mathbb{N}$ is a turn position of x according to G if $S \Longrightarrow_G^* uAv \Longrightarrow_G x$, for some $A \in N$, $u, v \in \Sigma^*$, where $x = uv$ and u is the prefix of x of length n.

Definition 1. *We say that a linear c.f. grammar $G = (N, \Sigma, P, S)$ has an* unambiguous turn position *if $P \subseteq N \times (\Sigma^* N \Sigma^* \cup \{\varepsilon\})$ and every $x \in L(G)$ has a unique turn position according to G. The languages generated by these grammars are called* linear c.f. languages with unambiguous turn position.

For instance it is easy to see that the language $\{a^i b^j c^k \mid i = j \vee i = k\}$ can be generated by a grammar of that type, while the same property does not hold for $\{a^i b^j c^k \mid i = j \vee j = k\}$.

These grammars are related to one-turn pushdown automata. We recall that a pushdown automaton is one-turn if in any computation there is at most one turn from pushing to popping. Thus any computation here works in two phases: In the first one the pushdown cannot decrease while in the second phase it cannot grow. These machines are a special case of so-called reversal bounded pushdown automata [1, 11], where only a finite number of turns from pushing to popping or vice versa are allowed (they also correspond to nonterminal bounded c.f. languages [10, Section 5.7]).

One can prove that a language L is a linear c.f. language with unambiguous turn position if and only if L is accepted by a one-turn pushdown automaton satisfying the following additional condition: every string $x \in L$ admits a unique prefix u such that every accepting computation on input x terminates its pushing phase after reading u. The inherent ambiguity of these languages has been studied in [14] where two such languages are shown that are, respectively, inherently $\log n$-ambiguous and inherently $\sqrt{n}$-ambiguous. To recall this result in detail we need some preliminary definitions.

Given a special symbol $\#$ not included in the finite alphabet Σ, consider the language $E = \{x \in (\Sigma \cup \{\#\})^* \mid |x|_\# \text{ is even}\}$. We define the function $spiral : E \longrightarrow E$ by induction, setting $spiral(u) = u$ for all $u \in \Sigma^*$, and $spiral(u\#v\#x) = (u\#spiral(x)\#v)$ for every $u, v \in \Sigma^*$ and every $x \in E$. Further, given a relation $R \subseteq \Sigma^* \times \Sigma^*$ and a language $\mathcal{F} \subseteq \Sigma^*$, define

$$\mathcal{L}(R) = \{u\#v^R \mid (u, v) \in R\}\ , \quad \mathcal{S}(R) = \{u^R\#v \mid (u, v) \in R\} \ \text{ and}$$
$$L(R, \mathcal{F}) = (\mathcal{L}(R)\#)^* \mathcal{F} (\#\mathcal{S}(R))^*$$

The relation R is said to be *simple* if $\mathcal{L}(R)$ is generated by an unambiguous c.f. grammar $(N, \Sigma \cup \{\#\}, P, S)$ such that $P \subseteq N \times (\Sigma^* N \Sigma^* \cup \{\#\})$. The following proposition can be proved from the definitions by a suitable construction (for a similar analysis see [14, Sec.3]).

Proposition 1 *If $R \subseteq \Sigma^* \times \Sigma^*$ is a simple relation and $\mathcal{F} \subseteq \Sigma^*$ is a regular language then $spiral(L(R, \mathcal{F}))$ is a linear c.f. language with unambiguous turn position.*

In the present work we are especially interested in two linear c.f. languages with unambiguous turn position, denoted by L_{ln} and L_{sqrt}, respectively. These languages are defined by

$$L_{ln} = spiral(L(R_1, \mathcal{F})) \quad \text{and} \quad L_{sqrt} = spiral(L(R_2, \mathcal{F}))$$

where $R_1 = \{(a^i, a^{2i}) \mid i \in \mathbb{N}, i \geq 1\}$, $R_2 = \{(a^i, a^{i+1}) \mid i \in \mathbb{N}, i \geq 1\}$, $\mathcal{F} = \{a\}^*$. The following properties are again proved in [14, Sect.3].

Proposition 2 *The language $L_{ln} \subseteq \{a, \#\}^*$ is inherently $\log n$-ambiguous and can be generated by a linear c.f. grammar with unambiguous turn position G such that $\hat{d}_G(n) \in \Theta^T(\log n)$. Analogously, $L_{sqrt} \subseteq \{a, \#\}^*$ is inherently $\sqrt{n}$-ambiguous and can be generated by a linear c.f. grammar with unambiguous turn position G' such that $\hat{d}_{G'}(n) \in \Theta^T(\sqrt{n})$.*

5 Rational trace languages

In this section we recall the basic notions concerning trace monoids and the corresponding languages (see for instance [8, 9] for more details).

Given a finite alphabet Σ and an irreflexive and symmetric relation $I \subseteq \Sigma \times \Sigma$, let $\equiv_I$ be the reflexive and transitive closure of the relation $\sim_I$ defined by

$$xaby \sim_I xbay \qquad \forall x, y \in \Sigma^*,\ \forall (a, b) \in I.$$

The relation $\equiv_I$ is a congruence over Σ^*, i.e. an equivalence relation preserving concatenation between words. For every $x \in \Sigma^*$ the equivalence class $[x] = \{y \in \Sigma^* \mid y \equiv_I x\}$ is called trace, the quotient monoid $\Sigma^*/\equiv_I$ is called trace monoid and usually denoted by $M(\Sigma, I)$. Clearly, for every $t \in M(\Sigma, I)$, all strings in t have the same length and we denote it by $|t|$. The pair (Σ, I) is called independence alphabet and it is usually represented by an undirected graph where Σ is the set of nodes and I the set of edges. If Σ is the union of two nonempty disjoint sets A, B and $I = (A \times B) \cup (B \times A)$ then $M(\Sigma, I)$ is isomorphic to the direct product $A^* \times B^*$; in this case, any trace $[x]$ is represented by the pair $(\pi_A(x), \pi_B(x))$ and we simply write $M(\Sigma, I) = A^* \times B^*$.

For every trace monoid $M(\Sigma, I)$ the subsets $T \subseteq M(\Sigma, I)$ are called trace languages and, for every $L \subseteq \Sigma^*$, we define $[L] = \{[x] \in M(\Sigma, I) \mid x \in L\}$ as the trace language represented by L. A trace language is called *rational* if it is represented by a regular language. The class of rational trace languages over $M(\Sigma, I)$ is denoted by $\mathrm{Rat}(M(\Sigma, I))$. This class has been widely studied in the literature and it coincides with the smallest family of trace languages including the finite sets in $M(\Sigma, I)$ and closed under the operation of union, product and Kleene closure (over the trace monoid). We recall that the trace languages recognizable by finite automata define a subclass of $\mathrm{Rat}(M(\Sigma, I))$ (the two classes coincide if and only if the relation I is empty).

Here we are interested in a natural notion of ambiguity for a trace language which depends on the number of representative strings of its elements. Formally, given a language $L \subseteq \Sigma^*$, let $am_L : M(\Sigma, I) \longrightarrow \mathbb{N}$ be defined by:

$$am_L(t) = |(t \cap L)|,$$

for all $t \in M(\Sigma, I)$. In other words $am_L(t)$ is the number of representative strings of t in L. For every $n \in \mathbb{N}$ we also define

$$\widehat{am}_L(n) = \max\{am_L(t) \mid |t| \leq n\}.$$

Following [3], for every positive $k \in \mathbb{N}$ we denote by $\mathrm{Rat}_k(M(\Sigma, I))$ the family of all trace languages T represented by some regular language $L \subseteq \Sigma^*$ such that $\widehat{am}_L(n) \leq k$ for all $n \in \mathbb{N}$. Clearly we have

$$\mathrm{Rat}_1(M(\Sigma, I)) \subseteq \mathrm{Rat}_2(M(\Sigma, I)) \subseteq \ldots \subseteq \bigcup_{k=1}^{\infty} \mathrm{Rat}_k(M(\Sigma, I)) \subseteq \mathrm{Rat}(M(\Sigma, I))$$

It was proved in [4] that for some trace monoids all these inclusions are strict. Some of these separation results can also be proved by studying the corresponding generating functions [3]. The elements of $\mathrm{Rat}_1(M(\Sigma, I))$ are also called *unambiguous rational* trace languages. This class coincides with the smallest family of trace languages including the finite sets of $M(\Sigma, I)$ and closed with respect to unambiguous rational operations; it is also known that $\mathrm{Rat}_1(M(\Sigma, I)) = \mathrm{Rat}(M(\Sigma, I))$ if and only if I is transitive [5, 12].

Using the notation given in Section 3 we can define the inherent ambiguity of rational trace languages. Let $f : \mathbb{N} \longrightarrow \mathbb{N}$ be a monotone nondecreasing function and let $T \subseteq M(\Sigma, I))$ be a rational trace language. We say that T is *inherently f-ambiguous* if:

1. $T = [L]$ for some regular language $L \subseteq \Sigma^*$ such that $\widehat{am}_L = \Theta^T(f)$, and
2. for every regular language L' such that $T = [L']$ we have $\widehat{am}_{L'} = \Omega^T(f)$, i.e. there exists $c \in \mathbb{N}$ such that $f(n) \leq \widehat{am}_{L'}(cn)$ for all positive $n \in \mathbb{N}$ large enough.

We recall that in [4] an inherently k-ambiguous rational trace language $T_k \subseteq \{a\}^* \times \{b, c\}^*$ is exhibited for every positive $k \in \mathbb{N}$.

6 Bipartite grammars

Given two finite disjoint alphabets Σ, Γ and a finite set N of nonterminal symbols, let $\mathcal{P}_\ell(N, \Sigma, \Gamma)$ be the set of linear productions given by

$$\mathcal{P}_\ell(N, \Sigma, \Gamma) = N \times (\Sigma N \cup N\Gamma \cup N \cup \{\varepsilon\}).$$

We denote by $\mathcal{G}_\ell(\Sigma,\Gamma)$ the corresponding set of grammars,

$$\mathcal{G}_\ell(\Sigma,\Gamma) = \{(N, \Sigma\cup\Gamma, P, S) \mid P \subseteq \mathcal{P}_\ell(N,\Sigma,\Gamma)\}.$$

Note that every grammar in $\mathcal{G}_\ell(\Sigma,\Gamma)$ is linear context-free with unique turn position and generates a language included in $\Sigma^*\Gamma^*$.

We also define the set of regular productions $\mathcal{P}_r(N,\Sigma,\Gamma)$, and the corresponding family of grammar $\mathcal{G}_r(\Sigma,\Gamma)$, by

$$\mathcal{P}_r(N,\Sigma,\Gamma) = N \times ((\Sigma\cup\Gamma)N \cup N \cup \{\varepsilon\}),$$
$$\mathcal{G}_r(\Sigma,\Gamma) = \{(N, \Sigma\cup\Gamma, P, S) \mid P \subseteq \mathcal{P}_r(N,\Sigma,\Gamma)\}.$$

Clearly the class of languages generated by grammars in $\mathcal{G}_r(\Sigma,\Gamma)$ is the family of the regular languages over the alphabet $\Sigma\cup\Gamma$.

Now we define a natural correspondence between $\mathcal{G}_r(\Sigma,\Gamma)$ and $\mathcal{G}_\ell(\Sigma,\Gamma)$. For every $(A,\alpha) \in \mathcal{P}_r(N,\Sigma,\Gamma)$, let $\mu(A,\alpha) \in \mathcal{P}_\ell(N,\Sigma,\Gamma)$ be given by

$$\mu(A,\alpha) = \begin{cases} (A,\alpha^R) & \text{if } \alpha\in\Gamma N \\ (A,\alpha) & \text{otherwise.} \end{cases} \tag{1}$$

Function μ extends to grammars by setting $\mu : \mathcal{G}_r(\Sigma,\Gamma) \longrightarrow \mathcal{G}_\ell(\Sigma,\Gamma)$ so that, for every $G \in \mathcal{G}_r(\Sigma,\Gamma)$ where $G = (N,\Sigma\cup\Gamma,P,S)$

$$\mu(G) = (N, \Sigma\cup\Gamma, \mu(P), S).$$

Such a function is a bijection and hence $\mu^{-1}(G)$ is well-defined for every $G \in \mathcal{G}_\ell(\Sigma,\Gamma)$. Moreover, function μ defined in Equation (1) extends to a monoid isomorphism $\mu : (\mathcal{P}_r(N,\Sigma,\Gamma))^* \longrightarrow (\mathcal{P}_\ell(N,\Sigma,\Gamma))^*$. Thus, for every grammar $G = (N,\Sigma\cup\Gamma,P,S) \in \mathcal{G}_r(\Sigma,\Gamma)$, if $S \overset{\tau}{\Longrightarrow}_G w$ for some $\tau \in P^*$ and some $w \in (\Sigma\cup\Gamma)^*$, then $S \overset{\mu(\tau)}{\Longrightarrow}_{\mu(G)} z$ for some $z \in \Sigma^*\Gamma^*$.

The previous functions give rise to analogous correspondences among languages. Let $\varphi : (\Sigma\cup\Gamma)^* \longrightarrow \Sigma^*\Gamma^*$ be the map such that

$$\varphi(w) = \pi_\Sigma(w)\cdot(\pi_\Gamma(w))^R$$

for every $w \in (\Sigma\cup\Gamma)^*$. This function is useful to deal with the representative strings of traces in the monoid $\mathcal{M} = \Sigma^*\times\Gamma^*$. The following properties are easily proved which refer to that monoid:

Remark 1. For every $u, v \in (\Sigma\cup\Gamma)^*$, $[u] = [v]$ if and only if $\varphi(u) = \varphi(v)$;

Remark 2. The extension $\varphi : \mathcal{M} \longrightarrow \mathcal{M}$ defined by $\varphi([w]) = [\varphi(w)]$, is an involution, i.e. $\varphi(\varphi(t)) = t$ for every $t \in \mathcal{M}$.

The first property above implies the following

Proposition 3 *Let G_1, G_2 be two grammars in $\mathcal{G}_r(\Sigma,\Gamma)$. Then $[L(G_1)] = [L(G_2)]$ if and only if $\varphi(L(G_1)) = \varphi(L(G_2))$.*

We now present the main property of functions μ and φ, which allows us to define a correspondence between derivations in regular and linear c.f. grammars.

Proposition 4 *For every $G \in \mathcal{G}_r(\Sigma, \Gamma)$, $G = (N, \Sigma \cup \Gamma, P, S)$, every $\tau \in P^*$ and every $w \in (\Sigma \cup \Gamma)^*$, we have*

$$(\exists u \in [w] \text{ such that } S \overset{\tau}{\Longrightarrow}_G u) \quad \text{if and only if} \quad (S \overset{\mu(\tau)}{\Longrightarrow}_{\mu(G)} \varphi(w))$$

Proof. Let w be an arbitrary word in $(\Sigma \cup \Gamma)^*$. Reasoning by induction on the length of derivation one can prove the following statements:

1. If $S \overset{\tau}{\Longrightarrow}_G uA$ holds for any $A \in N$, $\tau \in P^*$ and $u \in (\Sigma \cup \Gamma)^*$, then $S \overset{\mu(\tau)}{\Longrightarrow}_{\mu(G)} xAy$, where $x = \pi_\Sigma(u)$ and $y = (\pi_\Gamma(u))^R$. Moreover if $u \in [w]$ then $\varphi(u) = xy = \varphi(w)$;
2. If $S \overset{\tau'}{\Longrightarrow}_{\mu(G)} xAy$ for any $A \in N$, $\tau' \in \mu(P)^*$, $x \in \Sigma^*$ and $y \in \Gamma^*$ such that $xy = \varphi(w)$, then $\mu^{-1}(\tau')$ defines a derivation $S \overset{\mu^{-1}(\tau')}{\Longrightarrow}_G uA$ for some sentence uA, where $u \in (\Sigma \cup \Gamma)^*$ is obtained by a shuffle of x and y^R. Note that, for every shuffle u of x and y^R, we have $u \in [xy^R]$ and $\varphi(u) = xy$.

Then the result follows by observing that in both grammars G and $\mu(G)$ any derivation ends with an empty production. $\square$

The previous result has two useful consequences.

Corollary 5 *For every $G \in \mathcal{G}_r(\Sigma, \Gamma)$ and every $w \in (\Sigma \cup \Gamma)^*$*

$$|[w] \cap L(G)| \leq \sum_{v \in [w]} d_G(v) = d_{\mu(G)}(\varphi(w)) \tag{2}$$

Corollary 6 *For every $G \in \mathcal{G}_r(\Sigma, \Gamma)$ we have $L(\mu(G)) = \varphi(L(G))$.*

Therefore φ maps regular languages over the alphabet $\Sigma \cup \Gamma$ into linear c.f. languages with one turn position. The correspondence is not bijective. Moreover, in Corollary 5 the inequality of Equation (2) becomes an identity whenever the grammar G is unambiguous. This proves the following

Proposition 7 *For every unambiguous $G \in \mathcal{G}_r(\Sigma, \Gamma)$ and every $w \in (\Sigma \cup \Gamma)^*$ we have $am_{L(G)}([w]) = d_{\mu(G)}(\varphi(w))$.*

The previous properties allow to prove the relationship between the inherent ambiguity of rational sets in $\Sigma^* \times \Gamma^*$ and the inherent (with respect to $\mathcal{G}_\ell(\Sigma, \Gamma)$) ambiguity of linear c.f. languages with unambiguous turn position. The following theorem states the correspondence in the simpler direction.

Theorem 8 *Given a pair of disjoint alphabets Σ, Γ and a monotone nondecreasing function $f : \mathbb{N} \longrightarrow \mathbb{R}_+$, let T be an inherently f-ambiguous rational trace language over the monoid $\Sigma^* \times \Gamma^*$. Then there exists a linear c.f. language $L \subseteq \Sigma^* \Gamma^*$ that is inherently f-ambiguous with respect to the set $\mathcal{G}_\ell(\Sigma, \Gamma)$.*

Proof. Let $R \subseteq (\Sigma \cup \Gamma)^*$ be a regular language such that $[R] = T$ and assume $\widehat{am}_R \in \Theta^T(f)$. Consider an unambiguous grammar $G \in \mathcal{G}_r(\Sigma, \Gamma)$ generating R. By Corollary 6, $\mu(G)$ generates the language $L = \varphi(R)$ and, by Proposition 7, $\widehat{am}_R = \hat{d}_{\mu(G)}$. This means that L is generated by a grammar in $\mathcal{G}_\ell(\Sigma, \Gamma)$ whose ambiguity function is in $\Theta^T(f)$.

On the other hand, let $G' \in \mathcal{G}_\ell(\Sigma, \Gamma)$ be a grammar generating $L = \varphi(R)$ and consider the grammar $F \in \mathcal{G}_r(\Sigma, \Gamma)$ given by $F = \mu^{-1}(G')$. By Proposition 4 we get $[L(F)] = T$, which implies $\widehat{am}_{L(F)} \in \Omega^T(f)$. As a consequence also $\hat{d}_{G'} \in \Omega^T(f)$ because, by Corollary 5, $\widehat{am}_{L(F)}(n) \leq \hat{d}_{G'}(n)$ for every $n \in \mathbb{N}$. □

6.1 Example

As a further application of the previous properties we describe a trace language of inherent ambiguity 2 that is slightly simpler than an analogous example given in [4].

Let $\mathcal{M}$ be the trace monoid given by $\mathcal{M} = \{a\}^* \times \{b, c\}^*$. Let L be the language defined by the regular expression $c^*(ab)^* \cup (ac)^* b^*$ and set $T = [L]$. We want to show that T is inherently 2-ambiguous. To this end first observe that $am_L(t) \leq 2$ for all $t \in T$. Now, consider a regular language $R \subseteq \{a, b, c\}^*$ such that $[R] = T$ and let $G \in \mathcal{G}_r(\{a\}, \{b, c\})$ be an unambiguous grammar generating R. We have to prove that $am_R([x]) \geq 2$ for some $x \in R$.

By Corollary 6 we know that $\mu(G)$ generates the language $\varphi(R)$, the map $\varphi : \{a, b, c\}^* \longrightarrow a^*\{b, c\}^*$ being defined by

$$\varphi(w) = \pi_a(w) \cdot (\pi_{\{b,c\}}(w))^R \ , \qquad \forall w \in \{a, b, c\}^*$$

Since R represents T and

$$T = \{[w] \mid \pi_{\{b,c\}}(w) \in c^* b^* \ \wedge \ (|w|_a = |w|_b \ \vee \ |w|_a = |w|_c)\},$$

for every $w \in R$ we have $\varphi(w) = \pi_a(w) \cdot \pi_b(w) \cdot \pi_c(w)$, and hence

$$\varphi(R) = \{a^i b^j c^k \mid i = j \ \vee \ i = k\}$$

It is known that this is is a linear inherently 2-ambiguous c.f. language [10]; therefore there exists $u \in \varphi(R)$ such that $d_{\mu(G)}(u) \geq 2$. Thus, choosing x in R so that $\varphi(x) = u$, by Proposition 7 we get $am_R([x]) = d_{\mu(G)}(u) \geq 2$.

7 Main result

Here we present our main contribution stating a sort of reverse version of Theorem 8.

Theorem 9 *Let Σ be a finite alphabet and let $f : \mathbb{N} \longrightarrow \mathbb{R}_+$ be a monotone nondecreasing function. Assume $L \subseteq \Sigma^*$ is an inherently f-ambiguous c.f. language generated by a linear c.f. grammar G with unambiguous turn position such that $\hat{d}_G \in \Theta^T(f)$. Then there exists a rational trace language $T \subseteq \Sigma^* \times \overline{\Sigma}^*$ that is inherently f-ambiguous, where $\overline{\Sigma}$ is an isomorphic copy of Σ.*

Proof. Without loss of generality we may assume that G is a 4-tuple $G = (N, \Sigma, P, S)$ where $P \subseteq N \times (\Sigma N \cup N\Sigma \cup \{\varepsilon\})$. We also define $\overline{\Sigma} = \{\overline{a} \mid a \in \Sigma\}$. For every $u \in \Sigma^*$, let $\overline{u} \in \overline{\Sigma}^*$ be the word obtained from u by replacing every occurrence of a by $\overline{a}$, for each $a \in \Sigma$. We define the grammar $\overline{G}$ by $G = (N, \Sigma \cup \overline{\Sigma}, \overline{P}, S)$ where

$$\overline{P} = (P \cap N \times (\Sigma N \cup \{\varepsilon\})) \ \cup \ \{A \to B\overline{a} \mid A \to Ba \in P\}.$$

Note that $\overline{G} \in \mathcal{G}_\ell(\Sigma, \overline{\Sigma})$. Since G has an unambiguous turn position, for every $x \in L$ there exists a unique pair of words $u, v \in \Sigma^*$ such that $x = uv$ and $u\overline{v}$ is generated by $\overline{G}$; moreover

$$d_G(x) = d_{\overline{G}}(u\overline{v})$$

and $L(\overline{G}) = \{u\overline{v} \in \Sigma^*\overline{\Sigma}^* \mid uv \in L\}$. As a consequence $\hat{d}_{\overline{G}} = \hat{d}_G$ proving that also the ambiguity function of $\overline{G}$ is in $\Theta^T(f)$. Actually, the correspondence between G and $\overline{G}$ extends to a bijective map between the set of all linear c.f. grammars with unambiguous turn position generating L, and the family of all grammars in $\mathcal{G}_\ell(\Sigma, \overline{\Sigma})$ that generate $L(\overline{G})$. Since such a map keeps the number of derivations of all words we deduce that $L(\overline{G})$ is inherently f-ambiguous with respect to $\mathcal{G}_\ell(\Sigma, \overline{\Sigma})$.

Now consider the grammar $F = \mu^{-1}(\overline{G})$. Since $F \in \mathcal{G}_r(\Sigma, \overline{\Sigma})$, the language $T = [L(F)]$ is a rational trace language over the monoid $\mathcal{M} = \Sigma^* \times \overline{\Sigma}^*$. By Corollary 5, for every $w \in (\Sigma \cup \overline{\Sigma})^*$ we have

$$am_{L(F)}([w]) = |[w] \cap L(F)| \leq d_{\overline{G}}(\varphi(w))$$

and hence, since $|w| = |\varphi(w)|$, for every integer $n \geq 1$ we get

$$\widehat{am}_{L(F)}(n) \leq f(n),$$

proving that the inherent ambiguity of T is of the order $O^T(f)$.

In order to prove that T is at least f-ambiguous, consider an unambiguous grammar $G_r \in \mathcal{G}_r(\Sigma, \overline{\Sigma})$ such that $[L(G_r)] = T$. By Proposition 7, for every $w \in (\Sigma \cup \overline{\Sigma})^*$ we have $am_{L(G_r)}([w]) = d_{\mu(G_r)}(\varphi(w))$ and hence for every $n \in \mathbb{N}$

$$\widehat{am}_{L(G_r)}(n) = \hat{d}_{\mu(G_r)}(n).$$

Moreover, by Proposition 3 we have $\varphi(L(G_r)) = \varphi(L(F))$ implying that $\mu(G_r)$ and $\overline{G}$ generate the same language. Since $L(\overline{G})$ is inherently f-ambiguous with respect to $\mathcal{G}_\ell(\Sigma, \overline{\Sigma})$, we get $\hat{d}_{\mu(G_r)} \in \Omega^T(f)$ and hence also $\widehat{am}_{L(G_r)} \in \Omega^T(f)$, which concludes the proof. □

By applying Theorem 9 and Proposition 2 we obtain the following

Corollary 10 *There exist two rational languages T_1 and T_2 over the trace monoid $\{a, \#\}^* \times \{\overline{a}, \overline{\#}\}^*$ that are, respectively, inherently $\log n$-ambiguous and inherently $\sqrt{n}$-ambiguous.*

We conclude by recalling that the class of ambiguity functions of c.f. grammars (and c.f. languages) is wider than the family of functions considered in this work. It also includes sublogarithmic functions, and even divergent functions growing as slowly as any computable total (nondecreasing divergent) function [17]. However, the corresponding grammars obtained so far are not linear and hence our approach cannot be applied to determine trace languages of analogous inherent ambiguity. Thus, a natural open problem arises whether there exist rational trace languages of sublogarithmic inherent ambiguity.

References

[1] B. Baker and R. Book. Reversal bounded multipushdown machines. *Journal of Computer and System Science*, 8:315–332, 1974.

[2] J. Berstel and C. Reutenauer. *Rational series and their languages*, Springer-Verlag, New York - Heidelberg - Berlin, 1988.

[3] A. Bertoni, M. Goldwurm, G. Mauri and N. Sabadini. Counting techniques for inclusion, equivalence and membership problems. In *The book of traces*, V. Diekert and G. Rozenberg editors, World Scientific, 1995, 131–163.

[4] A. Bertoni, G. Mauri and N. Sabadini. A hierarchy of regular trace languages and some combinatorial applications. In *Proceedings of the 2nd World Conference on Mathematics at the Service of Man*, A. Ballester, D. Cardus and E. Trillas editors, Universidad Politecnica de Las Palmas, 1982, 146–153.

[5] A. Bertoni, G. Mauri and N. Sabadini. Unambiguous regular trace languages. In *Proceedings of the Coll. on Algebra, Combinatorics and Logic in Computer Science*, Colloquia Mathematica Soc. J. Bolyai, vol. n.42, North Holland 1985, 113–123.

[6] C. Choffrut. Rational relations as rational series. In *Theory is forever (Salomaa Festschrift)*, Lecture Notes in Comput. Sci., vol. n.3113, Springer-Verlag, 2004, 29–34.

[7] C. Choffrut, M. Goldwurm, and V. Lonati. On the maximum coefficient of rational formal series in commuting variables. In *Proceedings of the 8th DLT*, Lecture Notes in Comput. Sci., vol. n.3340, Springer-Verlag, 2004, 114–126.

[8] V. Diekert and Y. Metivier. Partial commutation and traces. In *Handbook of formal languages: beyond words*, G. Rozenberg and A. Salomaa editors, Springer 1997, 457–534.

[9] V. Diekert and G. Rozenberg (editors). *The book of traces*, World Scientific, 1995.

[10] M. Harrison. *Introduction to Formal Language Theory.* Addison-Wesley, 1978.

[11] O. Ibarra. Reversal-bounded multicounter machines and their decision problems. *Journal of the ACM*, 25(1): 116–133, 1978.

[12] J. Sakarovitch. On regular trace languages. *Theoret. Comput. Sci.* 52:59–75, 1987.

[13] A. Salomaa and M. Soittola. *Automata-Theoretic Aspects of Formal Power Series*, Springer-Verlag, 1978.

[14] K. Wich. Sublinear ambiguity. In *Proceedings of the 25th MFCS*, M. Nielsen and B. Rovan editors. Lecture Notes in Comput. Sci., vol. n.1893, Springer-Verlag, 2000, 690–698.

[15] K. Wich. Characterization of context-free languages with polynomially bounded ambiguity. In *Proceedings of the 26th MFCS*, J. Sgall, A. Pultr and P. Kolman editors. Lecture Notes in Comput. Sci., vol. n.2136, Springer-Verlag, 2001, 703–714.

[16] K. Wich. Universal inherence of cycle-free context-free ambiguity functions. In *Proceedings of the 29th ICALP*, P. Widmayer et al. editors. Lecture Notes in Comput. Sci., vol. n.2380, Springer-Verlag, 2002, 669–680.
[17] K. Wich. Sublogarithmic ambiguity. In *Proceedings of the 29th MFCS*, J. Fiala et al. editors. Lecture Notes in Comput. Sci., vol. n.3153, Springer-Verlag, 2004, 794–806.
[18] K. Wich. Ambiguity functions of context-free grammars and languages. Phd Thesis, Fakultät 5 Informatik, Elektrotechnik und Informationstechnik, Universität Stuttgart, preliminary version, August 2004.